



UNIVERSIDADE ESTADUAL DE CAMPINAS  
FACULDADE DE ENGENHARIA AGRÍCOLA

**DANIEL DOS SANTOS COSTA**

**DESENVOLVIMENTO DE PROTÓTIPO OPTOELETRÔNICO  
PARA DETERMINAÇÃO NÃO DESTRUTIVA DOS  
ATRIBUTOS DE QUALIDADE E ESTÁGIO DE  
AMADURECIMENTO EM UVAS VINÍFERAS**

CAMPINAS

2018

**DANIEL DOS SANTOS COSTA**

**DESENVOLVIMENTO DE PROTÓTIPO OPTOELETRÔNICO  
PARA DETERMINAÇÃO NÃO DESTRUTIVA DOS  
ATRIBUTOS DE QUALIDADE E ESTÁGIO DE  
AMADURECIMENTO EM UVAS VINÍFERAS**

Tese apresentada à Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Agrícola, na Área de Máquinas Agrícolas.

Orientadora: Profa. Dra. Barbara Janet Teruel Mederos

Coorientador: Prof. Dr. Rodrigo Pereira Ramos

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO DANIEL DOS SANTOS COSTA E ORIENTADA PELA PROFA. DRA. BARBARA JANET TERUEL MEDEROS.

CAMPINAS

2018

**Agência(s) de fomento e nº(s) de processo(s):** Não se aplica.

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca da Área de Engenharia e Arquitetura  
Rose Meire da Silva - CRB 8/5974

C823d Costa, Daniel dos Santos, 1986-  
Desenvolvimento de protótipo optoeletrônico para determinação não destrutiva dos atributos de qualidade e estágio de amadurecimento em uvas viníferas / Daniel dos Santos Costa. – Campinas, SP : [s.n.], 2018.

Orientador: Barbara Janet Teruel Mederos.  
Coorientador: Rodrigo Pereira Ramos.  
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Agrícola.

1. Espectroscopia de refletância. 2. Aprendizado de máquina. 3. Optoeletrônica. 4. Processamento de sinais. 5. Vitis vinifera. I. Teruel Mederos, Barbara Janet, 1966-. II. Ramos, Rodrigo Pereira. III. Universidade Estadual de Campinas. Faculdade de Engenharia Agrícola. IV. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Development of optoelectronic prototype for non-destructive determination of quality attributes and maturation stage in vinifera grape

**Palavras-chave em inglês:**

Reflectance spectroscopy

Machine learning

Optoelectronics

Signal processing

Vitis vinifera

**Área de concentração:** Máquinas Agrícolas

**Titulação:** Doutor em Engenharia Agrícola

**Banca examinadora:**

Barbara Janet Teruel Mederos [Orientador]

Inácio Maria dal Fabbro

Ricardo Yassushi Inamasu

Anderson Gomide Costa

Juliana Aparecida Fracarolli

**Data de defesa:** 14-12-2018

**Programa de Pós-Graduação:** Engenharia Agrícola

Este exemplar corresponde à redação final da **Tese de Doutorado** defendida por **Daniel dos Santos Costa**, aprovada pela Comissão Julgadora em 14 de dezembro de 2018, na Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas.

FEAAGRI

---

Profa. Dra. Barbara Janet Teruel Mederos - Presidente e Orientador(a)

---

Prof. Dr. Inácio Maria Dal Fabbro - Membro Titular

---

Dr. Ricardo Yassushi Inamasu - Membro Titular

Faculdade de

---

Engenharia Agrícola

Unicamp

---

Profa. Dra. Juliana Aparecida Fracarolli - Membro Titular

**A Ata da defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica do discente.**

## DEDICATÓRIA

A esposa da minha juventude.

## AGRADECIMENTOS

A meu Deus, cujo nome é Jeová, por conceder-me sabedoria, perspicácia e força continuamente;

Aos meus pais, Laurizete e Aderval, por valorizar o conhecimento adquirido, e que me ajudaram de maneiras distintas a concluir esse trabalho;

À minha esposa, pela paciência, companheirismo e amizade, mesmo quando afastado por motivo de força maior;

À professora DSc. Barbara Janet Teruel Mederos pela orientação e ensinamento, além de acreditar no potencial do trabalho proposto;

Ao professor DSc. Rodrigo Pereira Ramos pela orientação, ensinamento, parceria, paciência e amizade que foram desenvolvidos nos últimos anos;

À Universidade Estadual de Campinas (UNICAMP) por conceder a oportunidade de realizar tal trabalho;

À Faculdade de Engenharia Agrícola (FEAGRI) pelo conhecimento e experiência adquiridos, que está conceituada faculdade tem na história da Engenharia Agrícola no Brasil;

À Universidade Federal do Vale do São Francisco, nos nomes dos professores Acácio Figueiredo Neto, José Aliçandro Bezerra da Silva e Icaro Cardoso Maia, pela concessão dos espaços e equipamentos para realização desse trabalho;

À Fazenda Santa Maria por conceder-me realizar os trabalhos de tese e colher as melhores uvas viníferas do Vale do Submédio São Francisco;

Aos colegas de curso Nelson, Nicolas, German, Ivan, Luciene, Perdigão, Danilo e outros que por falta de papel não poderão ser mencionados, pelo grande apoio nas atividades desenvolvidas ao longo dos anos;

Aos amigos e colegas do Colegiado Acadêmico de Engenharia Agrícola, em especial as professoras Jardênia Feitosa e Thais Azevedo;

Aos meus queridos alunos e parceiros do Laboratório de Energia na Agricultura (LENA-UNIVASF), onde não citarei nomes com o intuito de não esquecer alguém;

E a todos aqueles que de alguma forma contribuíram para a realização deste trabalho.

Se  $A$  significa sucesso na vida, então  $A$  é igual a  $X$  mais  $Y$  mais  $Z$ , onde  $X$  é trabalho,  $Y$  é lazer e  $Z$  é manter a boca fechada. – *Albert Einstein*

## RESUMO

Os consumidores estão cada vez mais preocupados com a qualidade dos alimentos, exigindo continuamente produtos de excelência, impactando diretamente na aceitabilidade e comercialização. Assim sendo, o constante monitoramento dos produtos alimentícios é importante para assegurar a boa qualidade. As tecnologias ópticas apresentam-se promissoras para a análise não destrutiva de alimentos frescos, e especificamente na agroindústria Vitivinícola, justificado no alto valor agregado dos vinhos. Portanto, o objetivo do trabalho foi desenvolver um protótipo de dispositivo optoeletrônico para realizar a mensuração não destrutiva dos atributos de qualidade (sólidos solúveis totais, antocianinas totais e flavonoides amarelos) nos frutos de videiras viníferas em diferentes estádios de amadurecimento. As amostras de uvas viníferas de duas variedades (Syrah e Cabernet Sauvignon) foram provenientes da Vitivinícola Santa Maria, localizada no município de Lagoa Grande – PE. Foram colhidas 72 bagas semanalmente, de cada variedade, antes da fase de ‘*véraison*’ até o ponto de colheita. As amostras foram submetidas às medições de referência precedidas pelas medições de refletância. A aquisição dos espectros de refletância foi feita com o uso do espectrorradiômetro FieldSpec3 (*Analytical Spectral Devices, Boulder, Colorado, USA*). Identificada a resposta espectral, que corresponde aos comprimentos de onda que estimulam os atributos de qualidade, desenvolveu-se o protótipo do dispositivo optoeletrônico. Esse foi constituído por fontes de excitação (LEDs) com comprimentos de onda específicos, subsidiado pelas informações do espectrorradiômetro. Um fotodetector de InGaAs coleta a resposta espectral emitida da baga de uva por efeito da fonte de excitação. Os sinais de saída do detector foram condicionados com o uso de circuitos elétricos, e posteriormente processados e armazenados em memória não volátil. A avaliação de desempenho do protótipo foi realizada no processo de calibração e validação cruzada. Análises químicas em laboratório dos atributos de qualidade foram realizadas como método de referência. Os dados obtidos pelos métodos ópticos e de referência foram submetidos às análises quimiométricas para obtenção dos modelos preditivos que são embarcados no dispositivo optoeletrônico para determinação dos atributos de qualidade em tempo real. Foram identificados seis comprimentos de ondas sensíveis para os atributos de qualidade, sólidos solúveis totais, antocianinas totais e flavonoides amarelos, sendo três assinaturas espectrais no visível (570, 660 e 690 nm) e três no infravermelho próximo (970, 1200 e 1300 nm). O instrumento optoeletrônico para mensuração dos atributos de qualidade em uvas viníferas, apresentou-se promissor como ferramenta para determinação não destrutiva de atributos de qualidade.

**Palavras-chave:** Espectroscopia de refletância, Aprendizado de máquina, Optoeletrônica, Processamento de sinais, *Vitis vinifera* L.

## ABSTRACT

Consumers are increasingly concerned about food quality, continually demanding products of excellence, directly impacting the acceptability and marketing of products. Therefore, the constant monitoring of food products is important to ensure the good quality of the final product. Optical technologies are promising for the non-destructive analysis of fresh food, and specifically in the Vitivinícola agroindustry, justified by the high added value of wines. Therefore, the objective of the project was to develop a prototype of an optoelectronic device to perform the non - destructive measurement of the quality attributes (total soluble solids, total anthocyanins and yellow flavonoids) in the vines of different stages of maturation. Grape samples from two varieties (Syrah and Cabernet Sauvignon) were obtained from Viña Santa Maria, located in the city of Lagoa Grande - PE. 72 berries were collected weekly, from each variety, before the '*véraison*' phase to the point of harvest. The samples were subjected to reference measurements preceded by reflectance measurements. The acquisition of reflectance spectra was performed using the FieldSpec3 spectroradiometer (Analytical Spectral Devices, Boulder, Colorado, USA). Identifying the spectral response, which corresponds to the wavelengths that stimulate the quality attributes, the prototype of the optoelectronic device was developed. This was constituted by excitation sources (LEDs) with specific wavelengths, subsidized by the information of the spectroradiometer. An InGaAs photodetector collects the spectral response emitted from the grape berry by effect of the excitation source. The output signals from the detector were conditioned using electrical circuits, and then processed and stored in non-volatile memory. The performance evaluation of the prototype was performed in the calibration and cross validation process. Laboratory chemical analyzes of the quality attributes were performed as reference method. The data obtained by the optical and reference methods were submitted to the chemometric analysis to obtain the predictive models that are loaded in the optoelectronic device to determine the attributes of quality in real time. Six sensitive wavelengths were identified for quality attributes, total soluble solids, total anthocyanins and yellow flavonoids, with three visible signatures (570, 660 and 690 nm) and three near-infrared (970, 1200 and 1300 nm) spectral signatures. The optoelectronic instrument for measuring quality attributes in wine grapes presented a promising tool for the non-destructive determination of quality attributes.

**Keyword:** Reflectance spectroscopy, Machine learning, Optoelectronics, Signal processing, *Vitis vinifera* L.

## LISTA DE ILUSTRAÇÕES

Figura 2.1-1. Diagrama esquemático demonstrando o padrão de desenvolvimento dos frutos de uvas tintas (adaptado de KENNEDY, 2002).....	22
Figura 2.4-1. Fluxograma da dinâmica de desenvolvimento do trabalho de tese nos anos de 2017 e 2018.....	27
Figura 3.1-1. Amostragem das bagas de uva nas vinhas (a) e amostragem das bagas de uvas nos cachos (b) das variedades Syrah e Cabernet Sauvignon. ....	33
Figura 3.1-2. Diagrama do sistema de aquisição dos espectros de refletância VIS-NIR, sendo este constituído por: computador (1); espectrorradiômetro (2); câmara escura (3); e fonte de luz (4). ....	34
Figura 3.1-3. Diagrama do processo de desenvolvimento dos modelos matemáticos de regressão e classificação.....	36
Figura 3.1-4. Espectros brutos de absorbância de diferentes estádios de amadurecimento (verde, <i>véraison</i> e madura) para três amostras de bagas de uva.....	42
Figura 3.1-5. Gráficos de carregamento de pesos das variáveis independentes para SST com pré-tratamento SNV (1, 7, 5, 1) MF (a), antocianinas totais com pré-tratamento SNV (1, 21, 27, 1) MF (b) e flavonoides amarelos com pré-tratamento SNV (1, 7, 27, 1) MF (c). ....	47
Figura 3.1-6. Dados previstos pela espectroscopia VIS-NIR <i>versus</i> dados de referência para o conjunto de calibração e validação externa para SST (a), antocianinas totais (b) e flavonoides amarelos (c). ....	49
Figura 3.1-7. Gráficos de pontuação da PCA com os pré-tratamentos SNV (1, 7, 5, 1) MF (a), SNV (1, 7, 13, 1) MF (b) e SNV (1,7, 27, 1) MF (c). ....	51
Figura 3.1-8. Gráficos de coeficientes de correlação (a) e carregamento de pesos (b) das variáveis independentes no modelo PLS-DA. ....	54
Figura 3.2-1. Alisamento utilizando a Transformada de Fourier para remoção de ruídos de alta frequência realizada no espectro inteiro. Sinal espectral bruto (a). Representação do sinal bruto no domínio da frequência (b). Sinal após a alisamento (c).....	67
Figura 3.2-2. Gráfico de <i>box-plot</i> da média e desvio padrão dos estádios de amadurecimento (verde, <i>véraison</i> e madura) para os atributos de qualidade SST (a), antocianinas totais (b) e flavonoides amarelos (c). ....	75
Figura 3.2-3. Gráficos de absorbância versus comprimento de onda: (a) Dados brutos; (b) limiarização adaptativa brusca com <i>db</i> 4, 3 níveis de decomposição, <i>Savitzky-Golay</i> com	

polinômio de segunda ordem, 7 segmentos com primeira derivada e SNV; (c) limiarização global suave em duas faixas do espectro com *db* 4, 4 níveis de decomposição, *Savitzky-Golay* com polinômio de segunda ordem, 21 segmentos com primeira derivada e SNV; (d) filtro da mediana com 27 segmentos, *Savitzky-Golay* com polinômio de segunda ordem, 21 segmentos com primeira derivada e SNV; (e) limiarização global suave com *sym* 3, 10 níveis de decomposição, *Savitzky-Golay* com polinômio de segunda ordem, 21 segmentos com primeira derivada e SNV; (f) filtro da mediana com 17 segmentos, *Savitzky-Golay* com polinômio de segunda ordem, 21 segmentos com primeira derivada e SNV; (g) limiarização global suave com *coif* 3, 10 níveis de decomposição, *Savitzky-Golay* com polinômio de segunda ordem, 21 segmentos com primeira derivada e SNV. .... 77

Figura 3.2-4. Valores de referência e previstos para os melhores modelos, sendo: (a) e (b) modelos PLSR e SVM-R para determinação de SST, respectivamente; (c) e (d) modelos PLSR e SVM-R para determinação de antocianinas totais, respectivamente; (e) e (f) modelos PLSR e SVM-R para determinação de flavonoides amarelos, respectivamente.

..... 80

Figura 3.3-1. Diagrama em bloco dos principais componentes do dispositivo OPTOFruit..... 91

Figura 3.3-2. Diagrama do sistema de aquisição dos sinais do dispositivo OPTOFruit. .... 93

Figura 3.3-3. Evolução da concentração de sólidos solúveis totais (SST), antocianinas totais (Anth) e flavonoides amarelos (Flav) durante as semanas ao longo do amadurecimento. 95

Figura 3.3-4. Gráficos de dispersão: MLR para SST (a); *Random Forest* para SST (b); MLR para antocianinas totais (c); *Random Forest* para antocianinas totais (d); MLR para flavonoides amarelos (e); *Random Forest* para flavonoides amarelos (f). .... 96

## LISTA DE TABELAS

Tabela 2.4-1 – Características das três principais categorias de dispositivos ópticos.....	25
Tabela 3.1-1. Estatística descritiva das variedades Syrah, Cabernet Sauvignon e ambas as variedades juntas (Geral).....	41
Tabela 3.1-2. Modelos matemáticos de regressão da variedade Syrah, Cabernet Sauvignon e ambas as variedades (Syrah e Cabernet Sauvignon) usando o espectro completo (450 nm a 1800 nm). ....	43
Tabela 3.1-3. Modelos matemáticos de regressão com seleção de variáveis Jack-Knife o conjunto de ambas variedades (Syrah e Cabernet Sauvignon). ....	46
Tabela 3.1-4. Precisão, sensibilidade, seletividade e falsos positivos da classificação dos estádios de amadurecimento usando espectros entre 450 e 1800 nm por modelos LDA e PLS-DA.....	53
Tabela 3.2-1. Estatística descritiva das bagas de uvas das variedades Syrah e Cabernet Sauvignon.....	75
Tabela 3.2-2. Métricas de desempenho dos modelos de regressão PLSR e SVM-R para as variedades Syrah e Cabernet Sauvignon, com o sinal espectral bruto e processado. ....	78
Tabela 3.2-3. Métricas de desempenho dos modelos de classificação PLS-DA e SVM para as variedades Syrah e Cabernet Sauvignon, com o sinal espectral bruto e processado. ....	81
Tabela 3.3-1. Descritores estatísticos das análises físico-químicas das bagas de uva da variedade Cabernet Sauvignon. ....	95

## LISTA DE ABREVIATURAS E SIGLAS

A/D - *Analog/Digital*

ANOVA - *Análise de Variância*

AOAC - *Association of Official Agricultural Chemists*

EEPROM - *Electrically-Erasable Programmable Read-Only Memory*

FIR - *Finite Impulse Response*

IEEE - *Institute of Electrical and Electronic Engineers*

I/O - *Input/Output*

InGaAs - *Índio-Gálio-Arsênio*

IR - *Infrared*

LDA - *Linear Discriminant Analysis*

LED - *Light Emitting Diode*

LPF - *Low-Pass Filtering*

MLR - *Multiple Linear Regression*

MPLS - *Modified Partial Least Squares*

MSC - *Multiplicative Scatter Correction*

NIPALS - *Non-linear Iterative Partial Least Squares*

NIR - *Near Infrared*

PC - *Principal Component*

PCA - *Principal Components Analysis*

PCR - *Principal Components Regression*

pH - *Potencial Hidrogeniônico*

PLS - *Partial Least Squares*

PLS-DA - *Partial Least Squares Discriminant Analysis*

PLSR - *Partial Least Squares Regression*

R<sup>2</sup> - *Coefficiente de determinação*

RAM - *Random Access Memory*

RER - *Range Error Ratio*

RMSEC - *Root Mean Squared Error of Calibration*

RMSEP - *Root Mean Squared Error of Prediction*

RMSECV - *Root Mean Squared Error of Cross-Validation*

RPD - *Residual Predictive Deviation*

SD - *standard deviation*

SEC - *Standard Error of Calibration*

SEP - *Standard Error of Prediction*

SECV - *Standard Error of Cross-Validation*

SPI - *Serial Peripheral Interface*

SNV - *Standard Normal Variate*

SST - *Sólidos Solúveis Totais*

SVM - *Support Vector Machine*

USART - *Universal Synchronous Asynchronous Receiver Transmitter*

UV - *Ultraviolet*

VIS - *Visible*

## SUMÁRIO

<b>1. INTRODUÇÃO GERAL.....</b>	<b>17</b>
1.1. OBJETIVO GERAL .....	19
1.1.1. Objetivos Específicos .....	19
1.2. HIPÓTESE.....	19
<b>2. REVISÃO BIBLIOGRÁFICA.....</b>	<b>21</b>
2.1. FISIOLOGIA E AMADURECIMENTO DA UVA.....	21
2.2. ASPECTOS DE QUALIDADE.....	22
2.3. ESPECTROSCOPIA DE REFLETÂNCIA.....	23
2.4. INSTRUMENTOS ÓPTICOS .....	25
<b>3. SEÇÕES DE ARTIGOS .....</b>	<b>27</b>
3.1. DESENVOLVIMENTO DE MODELOS PREDITIVOS PARA OS ATRIBUTOS DE QUALIDADE E ESTÁDIO DE AMADURECIMENTO EM UVAS VINÍFERAS USANDO A ESPECTROSCOPIA DE REFLETÂNCIA VIS-NIR.....	29
3.2. TÉCNICAS DE PROCESSAMENTO DIGITAL DE SINAIS E APRENDIZAGEM DE MÁQUINA COM ESPECTROSCOPIA DE REFLETÂNCIA VIS-NIR EM UVAS VINÍFERAS.....	61
3.3. INSTRUMENTO OPTOELETRÔNICO VIS-NIR PARA MENSURAÇÃO DE ATRIBUTOS DE QUALIDADE EM UVAS VINÍFERAS DA VARIEDADE CABERNET SAUVIGNON.....	88
<b>4. SÍNTESE E DISCUSSÕES GERAIS .....</b>	<b>102</b>
<b>5. CONCLUSÕES GERAIS .....</b>	<b>105</b>
5.1. SUGESTÕES PARA CONTINUIDADE DO TRABALHO .....	105
<b>6. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>106</b>

## 1. INTRODUÇÃO GERAL

O incremento na produção agrícola mundial observado nos últimos anos e a disseminação de informação contribuem para que os consumidores estejam mais preocupados com a qualidade dos alimentos, exigindo continuamente produtos de excelência, impactando diretamente na aceitabilidade e comercialização dos produtos agrícolas (ALANDER *et al.*, 2013).

O constante monitoramento da qualidade dos produtos alimentícios é importante nas diferentes fases da cadeia de produção e comercialização para assegurar um produto final de excelência (RUIZ-ALTISENT *et al.*, 2010), sendo também uma demanda do setor Vitivinícola.

O monitoramento do amadurecimento e atributos de qualidade, frequentemente, envolve técnicas analíticas que são demoradas, destrutivas e, às vezes, pode exigir equipamentos sofisticados. Além disso, as análises destrutivas podem ser executadas somente em um número limitado de amostras, permitindo assim conhecer meramente uma média do amadurecimento do vinhedo, não possibilitando monitorar o desenvolvimento desde os estádios iniciais até o ponto de colheita de um mesmo fruto por repetidas avaliações (KOLB *et al.*, 2006).

Nos últimos anos, as pesquisas se concentraram sobre o desenvolvimento de técnicas não destrutivas adequadas para aumentar o número de amostras que podem ser analisadas, acompanhando o desenvolvimento fisiológico do fruto e, conseqüentemente, obtendo informações em tempo real. O monitoramento da qualidade dos frutos durante o amadurecimento torna-se um problema cada vez mais importante, assim como prever o comportamento dessa qualidade durante o seu desenvolvimento sem danificá-los (LECHAUDEL *et al.*, 2010).

Os métodos não destrutivos e não invasivos proporcionam, no âmbito do controle de qualidade, pouca ou nenhuma preparação da amostra, versatilidade e flexibilidade no processo de medição e, conseqüentemente, podem ser incorporados à linha de processamento permitindo a tomada de decisão em tempo real e redução de custos (FLORES *et al.*, 2009). As aplicações bem-sucedidas destes métodos requerem a combinação de sensores eficazes com sofisticados modelos matemáticos e algoritmos computacionais para estabelecer as relações entre as propriedades físico-químicas selecionadas e os atributos de qualidade do produto (RUIZ-ALTISENT *et al.*, 2010).

As técnicas de ensaios não destrutivas para avaliação da qualidade ganharam popularidade recentemente. Alguns desses esforços usando tomografia e raios-x (ABDULLAH

*et al.*, 2004), sensores fluorescentes (CEROVIC *et al.*, 2008; HAZIR *et al.*, 2012b), *biospeckle laser* (ANSARI e NIRALA, 2013; ZDUNEK *et al.*, 2014; COSTA *et al.*, 2017), visão artificial com intensidade de cor (RAZALI *et al.*, 2008; COSTA *et al.*, 2018) e dados espectrais de refletância (JUNKWON *et al.*, 2009; SAEED *et al.*, 2012) têm sido relatados para o monitoramento de diversos processos, desde a classificação até o controle de qualidade de frutas e vegetais.

Contudo, algumas técnicas são onerosas, em virtude dos equipamentos necessários para as análises. Por conseguinte, estão sendo desenvolvidos sistemas ópticos simplificados. Muitos equipamentos ópticos encontram-se no mercado com diferentes princípios de funcionamento, dentre os quais pode-se citar a espectroscopia do ultravioleta, visível e infravermelho próximo (MARTINS *et al.*, 2010), a fluorescência (CEROVIC *et al.*, 2008; CEROVIC *et al.*, 2009; GHOZLEN *et al.*, 2010; TUCCIO *et al.*, 2011) e a refletância em conjunto com fluorescência (GRAÇA, 2014), que têm sido aplicados na vitivinicultura. Ademais, em virtude desses equipamentos serem importados, torna-se difícil o seu uso em ambiente brasileiro em decorrência da falta de suporte técnico e manutenção.

Os métodos químicos analíticos, que são os convencionalmente utilizados para análise de atributos de qualidade em uvas, são invasivos, destrutivos, demorados e necessitam de equipamentos e insumos, na maioria das vezes sofisticados. As amostras necessitam ser retiradas do vinhedo e submetidas a procedimentos laboratoriais, que demandam tempo e equipamentos sofisticados para análise, além de consumirem reagentes, que ocasionalmente são caros. Isso limita o número de amostras a serem analisadas, não representando a população do vinhedo, além de haver um intervalo de tempo entre a amostragem e a obtenção dos resultados, o que pode dificultar a tomada de decisão sobre tratamentos culturais ou outra eventual necessidade. Também, tais métodos desperdiçam amostras que são produtos perdidos, e conseqüentemente representam perdas e prejuízos financeiros.

A determinação dos atributos de qualidade por métodos não destrutivos, *in loco*, auxilia a identificação das variabilidades que existem no vinhedo, proporcionando agrupar uvas com características similares, no que tange aos atributos de qualidade, para obtenção de vinhos de alto valor agregado, e conseqüentemente produtos de excelência. Além disso, economiza tempo, tendo em vista que as amostras não necessitam ser submetidas a procedimentos laboratoriais. E por fim, economiza recursos como reagentes, promovendo a otimização do processo produtivo.

Considerando a importância e necessidade da medição dos atributos de qualidade de uvas viníferas, proporcionando informações para seleção pós-colheita, controle de padrões de

qualidade e tomada de decisão, no menor tempo possível, o desenvolvimento de um dispositivo eletrônico para realizar testes não destrutivos, pode ser uma contribuição para o estado da arte no setor Vitivinícola.

Diante do exposto, existe uma demanda por equipamentos nacionais para análise de atributos de qualidade em uvas viníferas que operem de maneira não invasiva de forma a tornar a colheita dos frutos mais eficiente.

## **1.1. OBJETIVO GERAL**

Desenvolver um protótipo de dispositivo optoeletrônico para realizar a mensuração não destrutiva dos atributos de qualidade (sólidos solúveis totais, concentração de antocianinas totais e flavonoides amarelos) nos frutos de videiras viníferas em diferentes estádios de amadurecimento.

### **1.1.1. Objetivos Específicos**

- Desenvolver modelos preditivos para quantificação dos atributos de qualidade em uvas viníferas com uso de espectroscopia;
- Identificar os melhores pré-tratamentos para os sinais espectrais relacionados a espectroscopia de refletância no visível e infravermelho próximo (VIS-NIR);
- Identificar as respostas espectrais que caracterizam os atributos de qualidade sólidos solúveis totais, concentração de antocianinas totais e flavonoides amarelos em uvas viníferas;
- Projetar um protótipo de dispositivo optoeletrônico com aquisição de dados em tempo real, com base nas respostas espectrais, para estimativa rápida dos atributos de qualidade em uvas viníferas da variedade Cabernet Sauvignon;
- Analisar o desempenho do dispositivo optoeletrônico na predição dos atributos de qualidade de uvas viníferas.

## **1.2. HIPÓTESE**

Um dispositivo optoeletrônico, como ferramenta de análise não destrutiva, proporcionará a medição dos atributos de qualidade (sólidos solúveis totais, concentração de

antocianinas totais e flavonoides amarelos) e identificação dos diferentes estádios de amadurecimento em uvas viníferas com precisão de no mínimo 85%.

## 2. REVISÃO BIBLIOGRÁFICA

### 2.1. FISIOLOGIA E AMADURECIMENTO DA UVA

A videira (*Vitis vinifera* L.) destaca-se como uma das frutíferas mais cultivadas no mundo e seu fruto possui dois nichos de mercado, podendo ser consumido *in natura* ou ainda usado na produção de vinhos e sucos. De acordo com a *International Organisation of Vine and Wine* (OIV), em 2015, aproximadamente 52,8% da produção mundial de uvas foi destinada à produção de vinho e sucos.

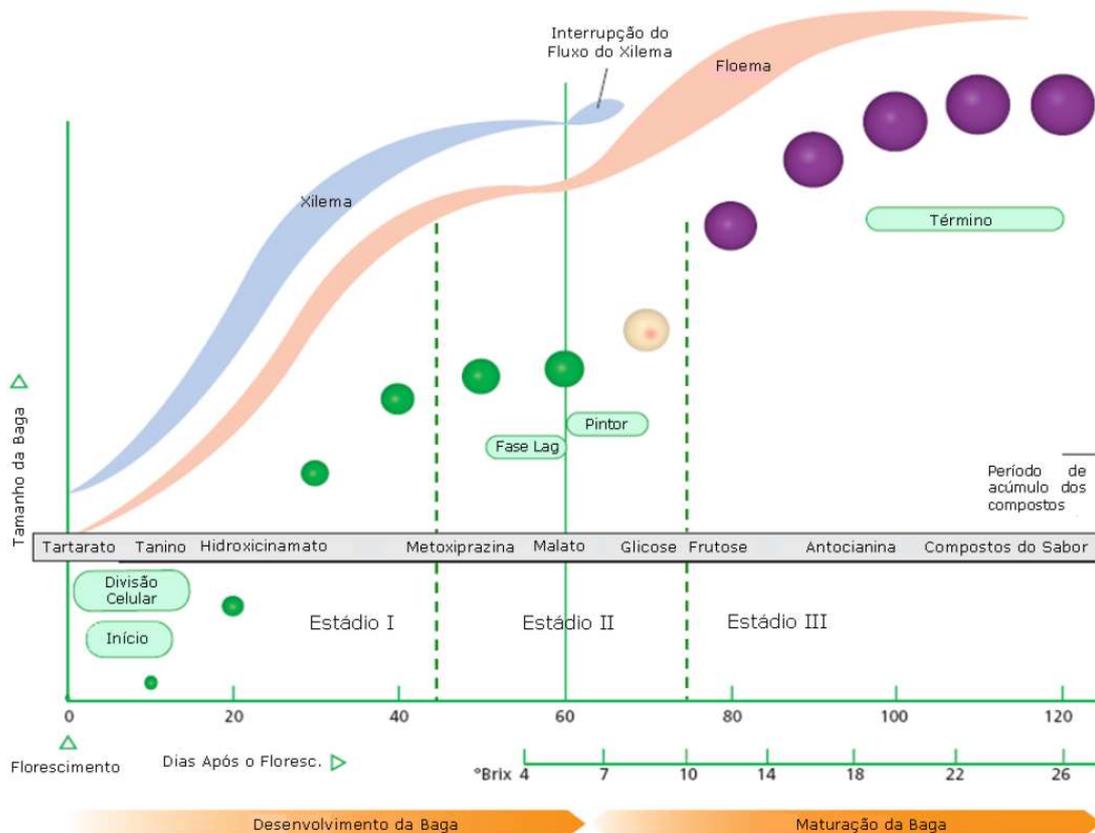
Os fatores mais importantes que conferem o sabor dos frutos são caracterizados pelos açúcares e ácidos. Como a uva é um fruto não climatérico, deve ser colhida somente após ter atingido estágio de amadurecimento que lhe confira aparência, sabor e textura adequados ao consumo (SOARES e LEÃO, 2009).

O desenvolvimento e o amadurecimento da uva compõem um processo contínuo, não havendo uma fase específica em que o fruto pode ser considerado inteiramente maduro (SOUZA, 2009). O padrão de desenvolvimento dos frutos segue uma curva dupla-sigmóide (Figura 2.1-1) dividida em três estádios. O estágio I é caracterizado pela intensa divisão e expansão celular, o estágio II é caracterizado pelo reduzido crescimento durante o amadurecimento da semente e o estágio III é caracterizado pelo acúmulo de açúcares, redução da acidez, expansão das células da polpa e acúmulo de antocianinas (KENNEDY, 2002). Com o amadurecimento, o fruto torna-se mais atraente e saboroso, por causa do equilíbrio entre os açúcares e ácidos.

Durante o amadurecimento, ocorrem mudanças na aparência física dos frutos, do engajo, e na composição química dos frutos. O ponto de colheita da uva é determinado pelo teor de sólidos solúveis totais (SST), associado aos atributos aparência, coloração, textura e sabor. A colheita de uvas viníferas é realizada quando estas apresentam SST igual ou superior a 20°Brix para um adequado volume alcoólico no vinho (SOARES e LEÃO, 2009). Contudo, essa variável não é suficiente para determinação da época de colheita. O amadurecimento ótimo da uva envolve diferentes eventos dentre eles o amadurecimento fisiológico (biossíntese evolucionária na baga), o amadurecimento tecnológico (acúmulo de açúcar e ácidos) e o amadurecimento fenólico (acúmulo qualitativo e quantitativo de taninos, pigmentos e compostos ligados ao sabor e ao aroma) (MANDELLI *et al.*, 2003).

O amadurecimento dos frutos é geralmente acompanhado de uma mudança na cor do epicarpo, devido a uma modificação da concentração de pigmentos nos tecidos superficiais (AGATI *et al.*, 2005). Tal alteração de coloração, resultante de reações bioquímicas no interior dos frutos, pode ser relacionada à maturidade (CHOONG *et al.*, 2006). Por exemplo, em uvas de variedade tinta os frutos ainda verdes têm uma maior proporção de clorofila, que diminui gradualmente com a maturidade, aumentando progressivamente a concentração de antocianinas e flavonoides (AGATI *et al.*, 2007; AGATI *et al.*, 2008; HAGEN *et al.*, 2006).

Figura 2.1-1. Diagrama esquemático demonstrando o padrão de desenvolvimento dos frutos de uvas tintas (adaptado de KENNEDY, 2002).



Fonte: Souza (2009)

## 2.2. ASPECTOS DE QUALIDADE

O termo qualidade implica no grau de excelência de um produto ou na sua adequação para um determinado uso. A qualidade é uma construção humana compreendendo muitas propriedades ou características. Assim, a qualidade dos produtos abrange propriedades sensoriais (aparência, textura, sabor e aroma), valores nutritivos, constituintes químicos,

propriedades mecânicas, propriedades funcionais e defeitos. A preocupação sobre a qualidade dos alimentos está crescendo constantemente. Os consumidores estão mais exigentes no que tange aos aspectos de qualidade dos produtos alimentícios, e estão dispostos a adquirir produtos com elevado padrão de excelência, independentemente do custo (ABBOTT, 1999).

No caso de frutas, a qualidade da fruta está intimamente relacionada à identificação do amadurecimento na colheita, sendo esta uma questão crucial. Assim, o monitoramento do amadurecimento da uva no campo pode fornecer dados valiosos para controlar a qualidade do vinho. O amadurecimento é um processo complexo, ocorrendo alterações bioquímicas no nível do epicarpo e mesocarpo. Para monitorar essas alterações, rotineiramente, são realizadas mensurações dos parâmetros de amadurecimento (por exemplo, teor de sólidos solúveis, acidez, fenóis, antocianinas, firmeza, etc.), essenciais para frutas e vegetais (GIOVENZANA *et al.*, 2015). Historicamente, o principal parâmetro que é medido para esta finalidade é o teor de açúcar, que está diretamente relacionado com o grau alcoólico do vinho a ser produzido (LARRAÍN *et al.*, 2008).

Para se atingir os excelentes padrões de qualidade, um constante monitoramento dos produtos alimentícios é de extrema importância. Como resultado, a indústria de alimentos necessita suceder métodos tradicionais de medição por métodos rápidos, não destrutivos e não invasivos (KEMPS *et al.*, 2010), e isso não é diferente na vitivinicultura. Tradicionalmente, os atributos de qualidade são mensurados com técnicas e instrumentos destrutivos e invasivos. Essas metodologias são procedimentos lentos, que requerem insumos e reagentes, onerando as análises quanto maior o número de amostras a serem medidas.

### **2.3. ESPECTROSCOPIA DE REFLETÂNCIA**

A espectroscopia de refletância apresenta-se promissora para a análise não destrutiva de produtos alimentícios e é ideal para as exigências da indústria agro-alimentar. Essas tecnologias utilizam a faixa espectral do ultra-violeta (UV – 1 a 380 nm), do visível (VIS – 380 a 780 nm) e infravermelho (IR – 780 a 2500 nm), irradiando o objeto da análise, e medindo-se a radiação refletida ou transmitida (CEN e HE, 2007). Quando a radiação atinge o produto, suas características espectrais mudam através de processos de absorção e espalhamento. Essas mudanças dependem da composição química da amostra, assim como suas propriedades de espalhamento que estão relacionadas com a microestrutura. As informações relacionadas às propriedades físico-químicas, ou seja, às características de qualidade do produto, encontram-se no espectro obtido (NICOLAI *et al.*, 2007).

A espectroscopia de refletância fornece informações estruturais complexas relacionadas com o comportamento vibracional das ligações químicas. A radiação excita as ligações moleculares de O-H, C-H e N-H, sendo que estas estão sujeitas às mudanças de energia vibracional. Essas ligações moleculares influenciam muitas propriedades químicas dos produtos alimentícios (ALANDER *et al.*, 2013).

O ramo alimentício tem-se beneficiado da aplicabilidade da espectroscopia de refletância, pois esta fornece informações rápidas e não destrutivas de aspectos de qualidade do produto. Assim, pesquisas envolvendo aplicações com uso da tecnologia de espectroscopia de refletância na predição de atributos de qualidade foram desenvolvidas em uma grande variedade de legumes e frutas, como tomates (SHAO *et al.*, 2007), peras (PAZ *et al.*, 2009) e maçãs (BOBELYN *et al.*, 2010).

Recentemente, a investigação com uso da espectroscopia de refletância centrou-se em produtos com alto valor agregado, como por exemplo uvas para processamento destinadas a vinhos e sucos. Herrera *et al.* (2003), utilizando a técnica da espectroscopia de refletância para desenvolver modelos de predição que permitam estimar os valores de SST de diferentes variedades de uva, encontraram resultados promissores na mensuração dessa variável. Estudo semelhante foi realizado por Fernández-Navales *et al.* (2009a), que mensuraram o conteúdo de açúcares redutores durante o amadurecimento, encontrando resultados satisfatórios de calibração e validação. Ademais, Gonzáles-Caballero *et al.* (2010), além de mensurarem o SST e conteúdo de açúcares redutores, mediram também pH, acidez titulável, conteúdos de ácidos tartárico e málico, e mesmo assim, os melhores modelos de predição foram para as variáveis relacionadas à doçura da uva. Por fim, Kemps *et al.* (2010) incluíram também a mensuração de densidade e compostos fenólicos, como antocianinas e polifenóis, não obtendo sucesso unicamente na predição de polifenóis.

Uma particularidade da tecnologia de espectroscopia de refletância deve ser considerada para o desenvolvimento dos modelos de predição e, conseqüentemente, mensuração dos atributos de qualidade em frutas. Algumas regiões espectrais podem não conter informações úteis sobre as variações químicas nas amostras. Assim, o desenvolvimento de um modelo de calibração envolve uma decisão sobre quais comprimentos de onda podem ser usados para estabelecer um modelo ideal. A existência de ruído pode agravar a capacidade preditiva do modelo, sendo necessária a eliminação dessas variáveis irrelevantes na modelagem de calibração para melhorar os resultados em termos de robustez e precisão (WU *et al.*, 2010). As evidências experimentais sugerem que os comprimentos de onda específicos, em vez do espectro completo, podem melhorar o desempenho dos modelos preditivos (FERNÁNDEZ-

NOVALES *et al.*, 2009b; GONZÁLES-CABALLERO *et al.*, 2010; WU *et al.*, 2010). Regiões específicas podem gerar modelos mais estáveis com boa interpretabilidade. Assim, é importante selecionar variáveis que contenham informações úteis, reproduzindo menores erros de previsão.

## 2.4. INSTRUMENTOS ÓPTICOS

Sistemas ópticos têm sido utilizados para detecção de qualidade de frutos em várias culturas hortícolas (RUIZ-ALTISENT *et al.*, 2010). Para este fim, três tipos principais de dispositivos ópticos podem ser identificados: (1) instrumentos laboratoriais para aplicações em centros de pesquisa ou em laboratórios da indústria; (2) dispositivos de triagem e classificação destinados especificamente para as indústrias de frutas e vegetais, por exemplo, em *packhouse*; (3) dispositivos portáteis para uso diretamente no campo (CIVELLI *et al.*, 2015). A Tabela 2.4-1 mostra as principais diferenças entre os três tipos de dispositivos ópticos.

Tabela 2.4-1 – Características das três principais categorias de dispositivos ópticos.

	Área de aplicação	Flexibilidade de uso	Aplicabilidade	Medição, precisão e reprodutibilidade	Custo
Dispositivos de laboratório	Pesquisa científica/Indústria	Adaptável a matrizes diferentes	Sistema fixo	Ótimo	Médio/Alto
Dispositivos de triagem e classificação	Indústria	Categorias específicas de produtos	Sistema fixo	Razoável	Médio/Alto
Dispositivos portáteis	Campo	Dedicado para produtos individuais	Portátil/de mão	Razoável	Médio

Fonte: CIVELLI *et al.* (2015)

As técnicas ópticas comumente utilizadas incluem espectroscopia de refletância no infravermelho próximo, espectroscopia de fluorescência, imagens multiespectrais e hiperespectrais. Contudo, algumas técnicas são onerosas, em virtude dos equipamentos necessários para as análises. Por conseguinte, estão sendo desenvolvidos sistemas ópticos simplificados que têm como base a diferença de absorvância, em comprimentos de onda específicos, para monitorar a qualidade de, por exemplo, maçã (BETEMPS *et al.*, 2012), palma (SAEED *et al.*, 2012; HAZIR *et al.*, 2012a) e alface-da-terra (BEGHI *et al.*, 2014).

Os métodos ópticos são particularmente adequados para a viticultura de precisão que tenta responder ao problema da heterogeneidade espacial e temporal das culturas, especialmente na maturidade da fruta. A colheita seletiva tem sido proposta como uma solução para este problema com base na gestão zonal do vinhedo (CEROVIC *et al.*, 2008). Isto posto, existe um

esforço para estimar com precisão a cinética do amadurecimento da uva, com o intuito de prever a melhor época para a colheita.

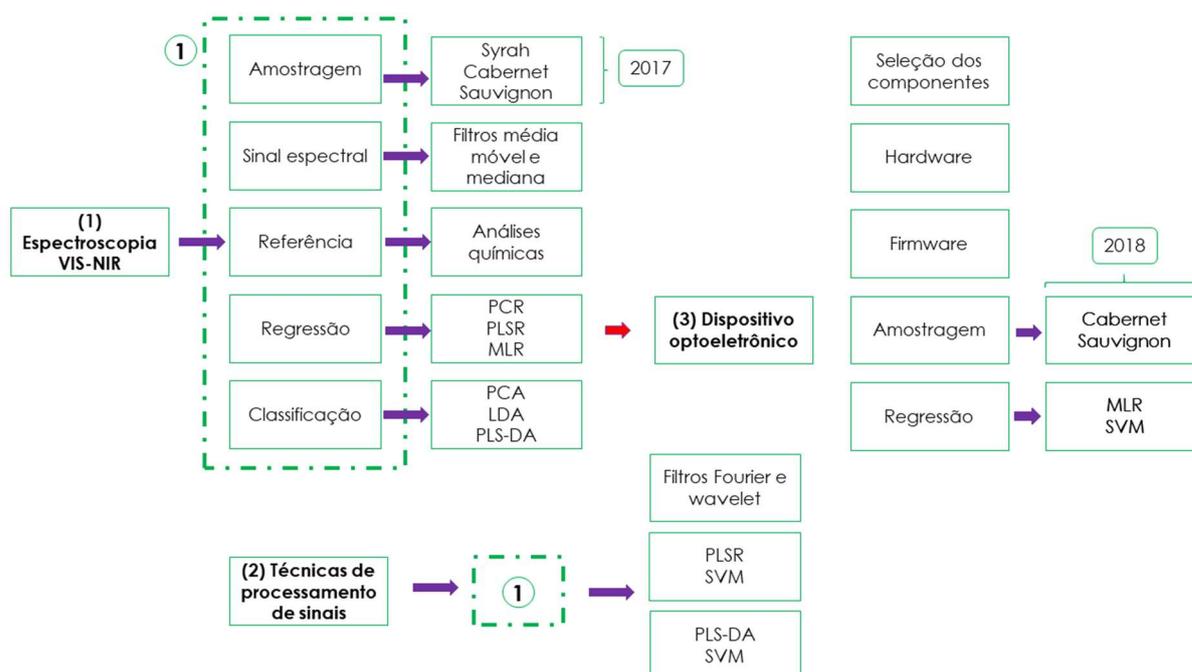
Larraín *et al.* (2008), desenvolveram um instrumento não invasivo para mensurar três parâmetros de amadurecimento (SST, pH e concentração de antocianinas) em uvas viníferas de cinco variedades, encontrando resultados satisfatórios para SST ( $R^2$  entre 0,874 e 0,930) e pH ( $R^2$  entre 0,560 e 0,798), não apresentando o mesmo desempenho para concentração de antocianinas ( $R^2$  entre 0,398 e 0,681). Ghazlen *et al.* (2010), usaram sensores ópticos de fluorescência para monitoramento de antocianinas na detecção para maturidade de cachos de uva, sob condições de laboratório e de campo, técnica que se mostrou promissora. Já Giovenzana *et al.* (2015), projetaram e avaliaram um dispositivo óptico que opera em quatro comprimentos de onda (630, 690, 750 e 850 nm) para uma rápida estimativa do teor de sólidos solúveis (SST) e acidez titulável (TA) em uvas no campo, onde o instrumento desenvolvido mostrou-se preciso, apresentando desempenho satisfatório para SST ( $R^2 = 0,85$ ) e TA ( $R^2 = 0,87$ ).

### 3. SEÇÕES DE ARTIGOS

O trabalho de tese foi dividido em três artigos. O primeiro artigo propôs desenvolver modelos preditivos para os atributos de qualidade e estágio de amadurecimento em uvas viníferas usando a espectroscopia de refletância VIS-NIR. Também, nesse artigo, objetivou-se identificar bandas e assinaturas espectrais em relação aos atributos de qualidade SST, antocianinas totais e flavonoides amarelos. O segundo artigo visou avaliar o desempenho de técnicas de processamento digital de sinais e aprendizagem de máquina na espectroscopia VIS-NIR para a determinação dos atributos de qualidade e discriminação dos estádios de amadurecimento em uvas viníferas. Nesse mesmo artigo, técnicas avançadas de processamento (transformada de Fourier e transformada *wavelet*) associadas com aprendizagem de máquina (máquinas de vetores de suporte) são implementadas. Por fim, o terceiro artigo, pretendeu-se desenvolver e avaliar um instrumento optoeletrônico, com base nas assinaturas espectrais previamente identificadas, para mensuração não destrutiva de sólidos solúveis totais (SST), antocianinas totais e flavonoides amarelos em uvas viníferas da variedade Cabernet Sauvignon.

A Figura 2.4-1 apresenta um fluxograma da dinâmica de desenvolvimento do trabalho de tese. O trabalho foi desenvolvido utilizando uvas viníferas da variedade Syrah e Cabernet Sauvignon das safras 2017 e 2018.

Figura 2.4-1. Fluxograma da dinâmica de desenvolvimento do trabalho de tese nos anos de 2017 e 2018.



Os artigos a seguir possuem os seguintes títulos:

3.1. Desenvolvimento de modelos preditivos para os atributos de qualidade e estágio de amadurecimento em uvas viníferas usando a espectroscopia de refletância VIS-NIR;

3.2. Técnicas de processamento digital de sinais e aprendizagem de máquina com espectroscopia de refletância VIS-NIR em uvas viníferas;

3.3. Instrumento optoeletrônico VIS-NIR para mensuração de atributos de qualidade em uvas viníferas da variedade Cabernet Sauvignon.

### 3.1. DESENVOLVIMENTO DE MODELOS PREDITIVOS PARA OS ATRIBUTOS DE QUALIDADE E ESTÁDIO DE AMADURECIMENTO EM UVAS VINÍFERAS USANDO A ESPECTROSCOPIA DE REFLETÂNCIA VIS-NIR

**RESUMO:** O mercado vitivinícola vem exigindo continuamente produtos de excelência, impactando diretamente na aceitabilidade e comercialização dos produtos. Assim sendo, o constante monitoramento das vinhas, quanto aos atributos de qualidade e estágio de amadurecimento, é importante para assegurar vinhos de boa qualidade, justificado pelo alto valor agregado desses produtos. Logo, as tecnologias ópticas apresentam-se promissoras para a análise não destrutiva de uvas viníferas tendo em vista a redução de insumos e tempo de análise. Isto posto, o objetivo deste estudo foi desenvolver modelos preditivos para os atributos de qualidade e estágio de amadurecimento em uvas viníferas usando a espectroscopia de refletância VIS-NIR. Foram coletadas um total de 432 bagas da variedade Syrah e 576 bagas da variedade Cabernet Sauvignon e feita a aquisição dos espectros de refletância com o espectrorradiômetro FieldSpec® 3 utilizando faixa espectral útil de 450 a 1800 nm. Posteriormente, foi realizada a determinação dos sólidos solúveis totais, antocianinas totais e flavonoides amarelos com os métodos de referência. Antes do desenvolvimento dos modelos, os dados espectrais foram submetidos a uma combinação de diferentes pré-tratamentos, dentre estes alisamentos, derivadas e correções. Os modelos preditivos de regressão utilizados foram a Regressão por Componentes Principais (PCR) e a Regressão por Quadrados Mínimos Parciais (PLSR) com o espectro completo (450 – 1800 nm) e espectro selecionado pelo método de Jack-Knife. Também foram desenvolvidos modelos preditivos utilizando a Regressão Linear Múltipla (MLR) com as assinaturas espectrais dos atributos de qualidade. Os estádios de amadurecimento foram discriminados utilizando os classificadores de Análise de Componentes Principais – Análise Discriminante Linear (PCA-LDA) e a Análise Discriminante dos Quadrados Mínimos Parciais (PLS-DA). A construção dos modelos de regressão PCR, PLSR e MLR proporcionou a predição de sólidos solúveis totais e conteúdo de antocianinas totais com precisão ( $R^2 \geq 0,90$ ), além do conteúdo de flavonoides com certo grau de rigor ( $R^2 \geq 0,70$ ). Além disso, possibilitou a discriminação dos diferentes estádios de amadurecimento em videira, com taxas de acerto de 93,15% usando o PLS-DA. Portanto, a espectroscopia de refletância VIS-NIR é uma ferramenta poderosa para avaliação não destrutiva dos atributos de qualidade e estágio de amadurecimento em uvas intactas da variedade Syrah e Cabernet Sauvignon.

**PALAVRAS-CHAVES:** compostos fenólicos; métodos não destrutivos; quimiometria; sólidos solúveis totais; Vale do São Francisco; vitivinicultura;

## INTRODUÇÃO

A região do Vale do São Francisco (VSF), localizada nos estados de Pernambuco e Bahia, é responsável por quase 23% da produção brasileira de vinhos finos, sendo a maior produtora da área tropical semiárida (RAMOS *et al.*, 2015). Os vinhos produzidos no VSF apresentam características favoráveis para o consumidor devido à presença de componentes bioativos e alta atividade antioxidante (PADILHA *et al.*, 2017), sendo esses relacionados à redução dos riscos de doenças crônico-degenerativas e cardiovasculares.

Uma das razões para os vinhos desta região agregarem tais características e serem considerados de excelente qualidade se deve às condições climáticas, que são distintas das de outras áreas produtoras de vinho no Brasil e no mundo, uma vez que o VSF possui temperaturas médias de 26 °C, alta intensidade da luz solar (3.000 h/ano) e baixa precipitação anual (em torno de 500 mm) (TEIXEIRA *et al.*, 2013).

O clima local, aliado à ausência de inverno e à água para irrigação, permitem obter duas safras no mesmo ano. Isso também influencia a maturidade das uvas, retardando ou acelerando o ciclo vegetativo, sendo necessárias avaliações a fim de decidir o tempo ideal para a colheita (FERRER-GALLEGO *et al.*, 2012). Esse monitoramento do amadurecimento e dos atributos de qualidade envolve, frequentemente, métodos químicos analíticos, que são invasivos, destrutivos e demorados. Além disso, tais métodos demandam reagentes que são onerosos e necessitam ser descartados na natureza de forma adequada.

Desse modo, a determinação dos atributos de qualidade por métodos não destrutivos, *in loco*, auxilia a identificação das heterogeneidades que existem no vinhedo, proporcionando agrupar uvas com características similares, no que tange aos atributos de qualidade, para obtenção de vinhos de alto valor agregado, e conseqüentemente produtos de excelência. Também, reduz o tempo de análise, tendo em vista que as amostras não necessitam ser submetidas a procedimentos laboratoriais, promovendo a otimização do processo produtivo. As aplicações bem-sucedidas destes métodos requerem a combinação de sensores eficazes com sofisticados modelos matemáticos e algoritmos computacionais para estabelecer as relações entre as propriedades físico-químicas selecionadas e os atributos de qualidade do produto (RUIZ-ALTISENT *et al.*, 2010).

O uso de métodos não destrutivos e não invasivos para monitoramento da maturidade em uvas viníferas surge da necessidade de acompanhar a evolução temporal de certas variáveis antes da colheita (HERRERA *et al.*, 2003). O ramo alimentício tem se beneficiado da aplicabilidade da espectroscopia de refletância, pois esta fornece informações rápidas dos aspectos de qualidade interna do produto. Assim, pesquisas envolvendo aplicações com uso da tecnologia de espectroscopia de refletância na predição de atributos de qualidade foram desenvolvidas em uma grande variedade de legumes e frutas, como tomate (SIRISOMBOON *et al.*, 2012), amora (LEIVA-VALENZUELA *et al.*, 2013) e manga (RUNGPICHAYAPICHET *et al.*, 2016).

Nos últimos anos, a investigação com uso da espectroscopia centrou-se em produtos com alto valor agregado, a exemplo de uvas destinadas a vinhos e sucos (FERNÁNDEZ-NOVALES *et al.*, 2009b; FERRER-GALLEGO *et al.*, 2011; HERRERA *et al.*, 2003; MARTÍNEZ-SANDOVAL *et al.*, 2016; NOGALES-BUENO *et al.*, 2014). Entretanto, tais investigações voltam-se para regiões limitadas do espectro eletromagnético, na grande maioria para o infravermelho próximo (NIR). Assim, uma investigação com uma ampla faixa espectral necessita ser realizada, pois algumas regiões espectrais podem conter informações úteis sobre as variações químicas nas amostras (GONZÁLEZ-CABALLERO *et al.*, 2010; WU *et al.*, 2010).

Portanto, o principal objetivo deste estudo foi desenvolver modelos preditivos para os atributos de qualidade e estágio de amadurecimento em uvas viníferas usando a espectroscopia de refletância VIS-NIR. Os objetivos específicos foram: (1) desenvolver modelos preditivos robustos para os atributos sólidos solúveis totais (SST), antocianinas totais e flavonoides amarelos com a região espectral entre o visível e o infravermelho; (2) identificar bandas e assinaturas espectrais em relação aos atributos de qualidade analisados; (3) desenvolver modelos de classificação para discriminar os diferentes estádios de amadurecimento.

## **MATERIAIS E MÉTODOS**

### **Amostragem das bagas de uvas e processamento**

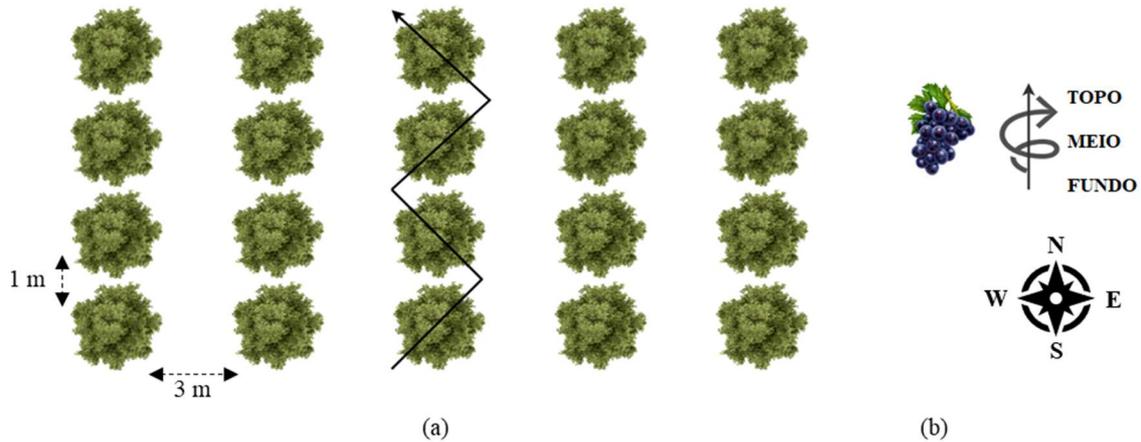
As amostras de uva (*Vitis vinifera* L.), das variedades Syrah e Cabernet Sauvignon, foram coletadas de um vinhedo localizado na região do Vale do Submédio São Francisco (Lagoa Grande, Pernambuco, Brasil: -9,05363; -40,19868). A região é caracterizada por possuir

clima do tipo BSh, segundo a classificação de Köppen (ALVARES *et al.*, 2013), correspondendo a uma região de clima árido. As vinhas de Syrah e Cabernet Sauvignon possuem espaçamento entre linhas de 3 m, espaçamento entre plantas de 1 m, linhas com comprimentos de 160 m, com área de 3,0 hectares e 2,2 hectares, respectivamente, sendo irrigadas por irrigação localizada.

Foram coletadas um total de 432 bagas da variedade Syrah e 576 bagas da variedade Cabernet Sauvignon. A coleta das bagas da variedade Syrah iniciou-se no dia 18 de maio de 2017, 77 dias após a poda, enquanto a coleta da variedade Cabernet Sauvignon iniciou-se no dia 06 de julho de 2017, 73 dias após a poda. Houve diferentes números de amostras para cada variedade devido à precocidade da Syrah nas condições ambientais locais. Seis datas de coletas foram realizadas para Syrah e oito datas de coletas foram realizadas para Cabernet Sauvignon.

Foram colhidas 72 bagas semanalmente entre os meses de maio e agosto de 2017, abrangendo diversos estádios, desde antes do *véraison* até o estágio de colheita. Para obter amostras representativas, foi adotado o método de amostragem estratificado (ARAÚJO, 2007). Uma fileira do meio da vinha foi escolhida e, dela, selecionou-se seis plantas do centro, sendo dois cachos marcados por planta (Figura 3.1-1a). Os cachos selecionados foram divididos em três posições: topo, meio e fundo (Figura 3.1-1b). De cada posição foram retiradas duas bagas, perfazendo um total de seis bagas por cacho. Essas amostras foram tomadas como representativas do cacho. Ressalta-se que é impraticável realizar a medição dos atributos SST, antocianinas totais e flavonoides amarelos no cacho inteiro. As bagas individuais foram acondicionadas em embalagem impermeável e armazenadas em gelo para transporte, imediato processamento e aquisição dos espectros. Antes de cada medição, foi necessário aguardar que as amostras se estabilizassem à temperatura ambiente de 25 °C.

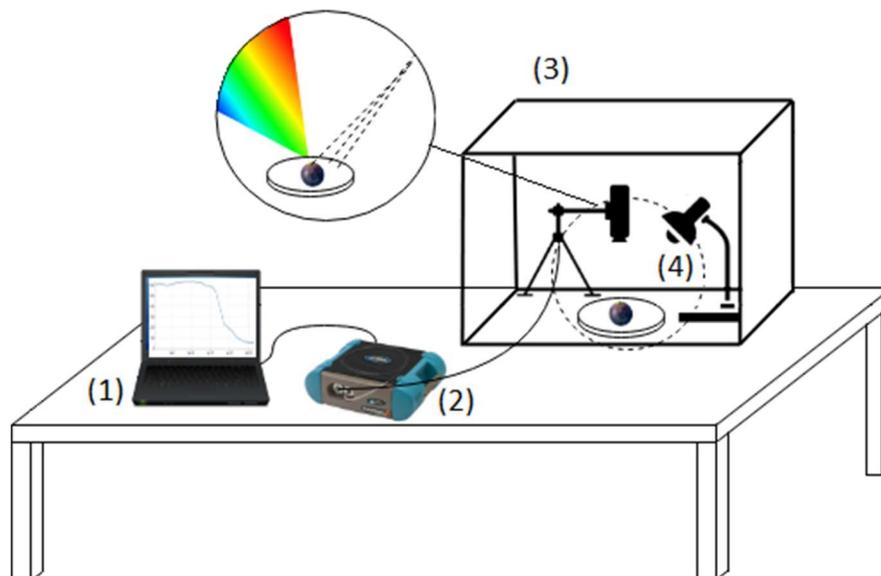
Figura 3.1-1. Amostragem das bagas de uva nas vinhas (a) e amostragem das bagas de uvas nos cachos (b) das variedades Syrah e Cabernet Sauvignon.



### Aquisição dos espectros de refletância

O sistema de aquisição dos espectros de refletância foi constituído dos seguintes componentes: espectrorradiômetro FieldSpec® 3 (*Analytical Spectral Devices*, Boulder, Colorado, USA), com campo de visão do sensor óptico de 8°, operando na faixa de comprimento de onda de 350 nm a 2500 nm, com resolução de 3 a 10 nm e tempo por varredura de 100 ms; fonte de luz de quartzo-tungstênio-halogênio de 70 W; câmara escura com dimensões de 100 x 50 x 50 cm; e computador com o *software* RS3 (*Analytical Spectral Devices*, Boulder, Colorado, USA). O sistema foi ligado 60 minutos antes da aquisição dos espectros, visando aumentar a qualidade e homogeneidade dos dados espectrais. Ademais, a fibra óptica foi posicionada perpendicularmente ao plano de medição com altura de 7,0 cm, com as bagas sendo localizadas no centro do referido plano (Figura 3.1-2).

Figura 3.1-2. Diagrama do sistema de aquisição dos espectros de refletância VIS-NIR, sendo este constituído por: computador (1); espectrorradiômetro (2); câmara escura (3); e fonte de luz (4).



Uma placa cerâmica Spectralon® (*Labsphere Inc.*, North Sutton, NH, EUA) foi usada como referência branca, com a qual operações de calibração (aquisição de referência do branco padrão) foram realizadas antes da tomada dos dados. Cada espectro foi constituído pela média de trinta varreduras feitas pelo sensor, para cada uma das bagas. Os valores de refletância foram transformados em absorbância  $\log \frac{1}{R}$  com o *software* ViewSpec Pro (*Analytic Spectral Devices*, Boulder, Colorado, EUA). As regiões compreendidas entre as faixas de 350 nm a 449 nm e 1801 nm a 2500 nm foram removidas da análise por apresentarem ruído aleatório excessivo.

### **Determinação dos atributos de qualidade das amostras de uva**

Após as medições de refletância, as bagas individuais foram submetidas às medições de referência, a saber, sólidos solúveis totais (SST), antocianinas totais (Anth) e flavonoides amarelos (Flav).

A determinação dos sólidos solúveis totais foi realizada a partir da leitura direta em refratômetro digital (HI 96804, *Hanna Instruments*, USA), com faixa de medição entre 0 e 85 °Brix e precisão de  $\pm 0,2$  °Brix, utilizando-se duas gotas da amostra triturada e homogeneizada, com os resultados expressos em °Brix.

Os teores de antocianinas totais e flavonoides amarelos da casca foram determinados segundo Francis (1982), sendo este método de pH único não havendo diferença com o método AOAC (FAVARO, 2008). Pesou-se 0,5 g de casca e polpa, em seguida, adicionou-se 25 mL da solução extratora de etanol (95%) acidificado com HCl (1,5 N), na proporção 85:15. As amostras foram maceradas por um minuto e, posteriormente, transferidas para recipientes ao abrigo da luz ficando em repouso e armazenados na geladeira. Após 24 horas, o sobrenadante era recolhido para leitura em espectrofotômetro (Spectronic BioMate 5 UV-Vis, *Thermo Electron*, UK) a 535 nm, para antocianinas totais, e 374 nm, para flavonoides amarelos. Os resultados foram expressos em  $\text{mg} \cdot 100 \text{ g}^{-1}$ , por meio das equações (1) e (2):

$$\text{Anth} = \text{abs} * \text{fator de diluição} / E1\% \quad (1)$$

$$\text{Flav} = \text{abs} * \text{fator de diluição} / E2\% \quad (2)$$

em que o fator de diluição é 5000, o coeficiente de extinção para antocianinas totais (E1%) é 98,2 e o coeficiente de extinção para flavonoides amarelos (E2%) é 76,6.

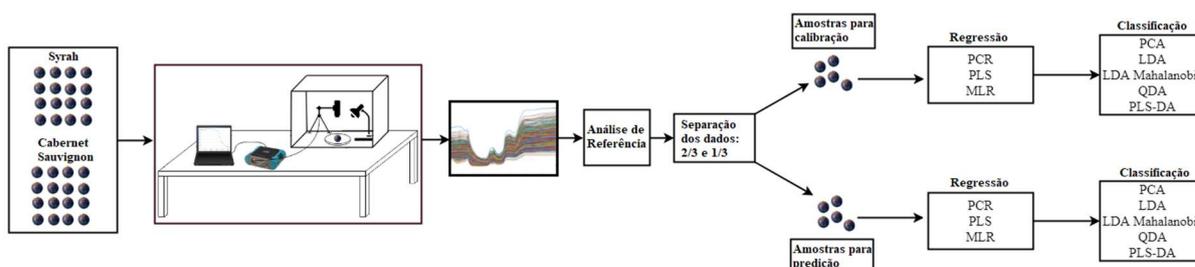
### **Desenvolvimento dos modelos de regressão multivariados**

A análise quimiométrica foi realizada utilizando o *software* Unscrambler X 10.4 (CAMO ASA, Oslo, Noruega). Antes do desenvolvimento dos modelos de calibração, os dados espectrais foram submetidos a uma combinação de diferentes pré-tratamentos matemáticos. Utilizou-se duas técnicas de alisamento, separadamente, sendo elas o filtro da mediana (MF) e o filtro da média móvel (MAF), com diferentes tamanhos de segmentos (5, 9, 13, 17, 21 e 27 pontos). Sobre os dados suavizados, utilizou-se a primeira derivada e a segunda derivada, calculadas pelo método de *Savitzky-Golay* (SAVITZKY e GOLAY, 1964; GORRY, 1990), com polinômio de segunda ordem e dois diferentes segmentos (7 e 21 pontos). Por fim, sobre os dados suavizados e com as derivadas, utilizou-se a correção multiplicativa de espalhamento (MSC) (ISAKSSON e NAES, 1988) e a padronização normal de sinal (SNV) (BARNES *et al.*, 1989). Os pré-tratamentos foram denotados como (a, b, c, d), onde o primeiro dígito representa o grau da derivada, o segundo representa o número de segmentos sobre o qual a derivada é calculada, o terceiro representa o número de segmentos na suavização, e o quarto representa o segundo alisamento (SHENK e WESTERHAUS, 1995).

O número de amostras segue a recomendação de Ferreira (2015), onde o conjunto de calibração deve ser no mínimo igual a  $6(A+1)$ , sendo  $A$  o número de fatores (componentes principais ou variáveis latentes) do modelo. Os conjuntos de dados foram separados na proporção 2/3-1/3, ou seja, 2/3 dos dados foram usados para o conjunto de calibração e 1/3 para

o conjunto de predição. Foram desenvolvidos modelos matemáticos de regressão para a variedade Syrah e Cabernet Sauvignon, separadamente, e o conjunto de ambas as variedades (Figura 3.1-3).

Figura 3.1-3. Diagrama do processo de desenvolvimento dos modelos matemáticos de regressão e classificação.



Os modelos de regressão, baseados nos atributos de qualidade e nos espectros, foram desenvolvidos usando regressão pelo método das componentes principais (PCR) e regressão pelo método dos quadrados mínimos parciais (PLSR). Para construir os modelos, utilizaram-se os dados espectrais brutos ou pré-tratados e os atributos de qualidade (SST, antocianinas totais e flavonoides amarelos), como variáveis de entrada independentes (X) e dependentes (Y), respectivamente, (WOLD *et al.*, 1984) das uvas individuais no conjunto de calibração e predição. Os dados espectrais foram centrados na média antes da construção dos modelos. A validação cruzada foi realizada com todas as amostras do conjunto de calibração e o número ótimo de fatores definido através do menor valor de erro quadrático médio de validação cruzada (RMSECV). O algoritmo utilizado para a construção dos modelos foi o NIPALS (GELADI e KOWALSKI, 1986) com 100 interações. Um total de 1728 modelos de regressão dos atributos de qualidade (SST, antocianinas e flavonoides) e conjunto de dados (Syrah, Cabernet Sauvignon e Syrah combinada com Cabernet Sauvignon) foram desenvolvidos.

A partir da identificação dos melhores pré-tratamentos aplicados sobre os dados espectrais nos modelos PCR e PLSR utilizando parâmetros estatísticos, foi realizada uma seleção de variáveis dos comprimentos de onda através do teste de incerteza com o método Jack-Knife (MARTENS e MARTENS, 2000) e teste-t de Student com nível de significância de 5%. Foram reconstruídos os melhores modelos com as variáveis selecionadas (comprimentos de onda) utilizando PCR e PLSR no conjunto de dados composto por ambas as variedades (Syrah combinada com Cabernet Sauvignon).

Neste estudo, a regressão linear múltipla (MLR) foi utilizada, uma vez ser mais simples essa modelagem. Contudo, em virtude dos problemas de colinearidades, foi necessário fazer uma seleção de variáveis para usar como entrada para a MLR. Assim, os vinte comprimentos de onda com os maiores coeficientes de correlação e carregamento da variável independente dos modelos PCR e PLSR dos espectros selecionados pelo método Jack-Knife foram selecionados. Os vinte comprimentos de onda foram definidos como variáveis espectrais independentes, utilizando os valores de absorvância dos respectivos comprimentos de onda selecionados e posteriormente submetidos a ANOVA e ao teste de significância a 5% dos coeficientes da regressão e validação externa.

A detecção de amostras anômalas foi realizada usando critérios baseados nos parâmetros estatísticos de alavancagem e resíduos t de *Student*. As amostras anômalas foram retiradas quando os limites de alavancagem e resíduos de *Student* foram ultrapassados simultaneamente.

Após o desenvolvimento dos modelos de regressão, foi utilizado um conjunto de dados que não foram empregados durante o processo de calibração, ou seja, foi utilizada a validação externa ou predição sobre os modelos desenvolvidos com a finalidade de determinar a capacidade preditiva.

Os desempenhos dos modelos resultantes dos diferentes algoritmos de pré-tratamentos foram comparados usando diferentes parâmetros estatísticos, sendo: coeficiente de determinação ( $R^2$ ); erro quadrático médio de calibração (RMSEC); erro padrão de calibração (SEC); erro quadrático médio de validação cruzada (RMSECV); erro padrão de validação cruzada (SECV); erro quadrático médio de predição (RMSEP); erro padrão de predição (SEP); e viés (Bias). Esses parâmetros são definidos pelas equações (3) à (10):

$$R^2(\text{Pearson}) = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}) * (y_i - \bar{y})^2}{(n - 1) * \sigma_r * \sigma_p} \quad (3)$$

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

$$\text{SEC} = \sqrt{\frac{n * \text{RMSEC}^2}{n - 1}} \quad (5)$$

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_{i*})^2}{n}} \quad (6)$$

$$\text{SECV} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i - \text{bias})^2}{n - 1}} \quad (7)$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{m}} \quad (8)$$

$$SEP = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i - bias)^2}{m - 1}} \quad (9)$$

$$bias = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{m} \quad (10)$$

em que:  $\hat{y}_i$  é o valor previsto pelo modelo na calibração;  $\hat{y}_{i*}$  é o valor previsto pelo modelo na validação cruzada;  $y_i$  é o valor de referência;  $\bar{y}$  é a média dos valores de referência;  $\bar{\hat{y}}$  é a média dos valores previstos;  $n$  é o número de amostras de calibração ou validação;  $m$  é o número de amostras de predição;  $\sigma_r$  é o desvio padrão dos valores de referência;  $\sigma_p$  é o desvio padrão dos valores previstos; e  $S_r^2$  é a variância dos valores de referência.

Para a escolha dos melhores modelos foram utilizados o coeficiente de determinação de Pearson ( $R^2$ ), o erro padrão de calibração (SEC), e o erro padrão de validação cruzada (SECV) (GÓMEZ *et al.*, 2006). A avaliação do SECV deve ser realizada em associação com o desvio padrão (SD) e o intervalo do atributo de qualidade estudado. Assim, são definidos o desvio preditivo residual (RPD) e a relação de erro de intervalo (RER) como sendo nas equações (11) e (12):

$$RPD = \frac{SD}{SECV} \quad (11)$$

$$RER = \frac{Alcance}{SECV} \quad (12)$$

Por fim, para conjuntos de calibração que compõem cem ou mais amostras e conjuntos de predição constituídos por nove ou mais amostras, os seguintes limites de controle foram assumidos (SHENK *et al.*, 1989) como nas equações (13) e (14):

$$\text{Limite controle SEP} = 1,30 * RMSEC \text{ (erro quadrático médio da calibração)} \quad (13)$$

$$\text{Limite controle bias} = \pm 0,60 * RMSEC \text{ (erro quadrático médio da calibração)} \quad (14)$$

## Desenvolvimento dos modelos de classificação

A análise de componentes principais (PCA) foi utilizada para explorar o possível agrupamento de amostras e fornecer informações sobre a estrutura latente da matriz espectral. A matriz de dados foi constituída dos espectros de bagas de uvas viníferas da variedade Syrah e Cabernet Sauvignon, em diferentes estádios de amadurecimento.

Identificados os agrupamentos entre os diferentes estádios de amadurecimento pela PCA, modelos discriminantes supervisionados foram construídos para classificar as uvas por fase de amadurecimento ao longo do desenvolvimento dos frutos, utilizando a análise discriminante linear (LDA) e a análise discriminante pelo método de quadrados mínimos parciais (PLS-DA).

Um conjunto de treinamento e outro de teste com 672 e 336 amostras, respectivamente, compostos por espectros de bagas em três estádios de amadurecimento (verde, *véraison* e madura), classificadas segundo Coombe (1995), foram utilizados para desenvolvimento dos modelos matemáticos de classificação. A classificação colorimétrica foi realizada para cada baga. A matriz de dados foi constituída por valores de absorbância, sendo esta variável independente, e por classes definidas de amadurecimento, como variável dependente.

O método LDA foi utilizado com diferentes técnicas para o cálculo das distâncias (linear, quadrática e Mahalanobis) entre amostras, e suas conseqüentes similaridades a uma determinada classe. Tendo em vista que se utilizou todo o espectro adquirido (450 a 1800 nm), sendo o número de amostras inferior ao número de variáveis (1008 amostras vs. 1350 variáveis), foi necessário executar uma PCA com seis componentes principais e os escores desta foram utilizados como variáveis de entrada na LDA, eliminando o problema de colinearidade.

O método PLS-DA foi utilizado com a mesma abordagem do método de regressão PLS. Contudo a variável dependente utilizada não assumiu valores contínuos, como ocorre na análise quantitativa, mas valores discretos (-1; 0 e 1), representando cada classe.

É comum definir-se um limiar de 0,5 para o PLS-DA. Entretanto, essa técnica pode resultar em vários erros de classificação, devido à sua natureza generalista. Um limiar mais refinado foi obtido através da teoria Bayesiana, buscando minimizar o risco de ocorrência de erros de classificação (ZHANG, 2000). Dadas as funções de probabilidade condicional  $p(y | A)$  e  $p(y | B)$  das classes A e B, respectivamente, onde  $y$  é o valor estimado da amostra pelo modelo PLS, tem-se que o limiar é o valor de  $y$  para o qual a equação (15) é válida:

$$p(y|A) * P(A) = p(y|B) * P(B), \quad (15)$$

Tal igualdade pode ser resolvida analiticamente através da equação (16):

$$(s_B^2 - s_A^2)^2 y^2 + (2\bar{y}_B s_A^2 - 2\bar{y}_A s_B^2) y + \bar{y}_A^2 s_B^2 - \bar{y}_B^2 s_A^2 - 2s_B^2 s_A^2 \ln\left(\frac{P(A) s_B}{P(B) s_A}\right) = 0 \quad (16)$$

em que:  $A$  são os valores previstos para as amostras pertencentes à classe A;  $B$  são os valores previstos para as amostras pertencentes à classe B;  $s_A$  é desvio padrão dos valores em A;  $s_B$  é desvio padrão dos valores em B;  $\bar{y}_A$  é média dos valores em A;  $\bar{y}_B$  é média dos valores em B;  $P(A)$  é a probabilidade da classe A; e  $P(B)$  é a probabilidade da classe B.

Os desempenhos dos modelos de classificação supervisionada foram avaliados usando a matriz de confusão para um problema de três classes. Os valores de precisão (*Prec*), sensibilidade (*Sen*), seletividade (*Sel*) e taxa de falsos positivos (TFP) para cada classe foram obtidos a partir das matrizes de confusão de cada modelo, conforme equações (17) à (20):

$$Prec = \frac{VP + VN}{VP + FN + FP + VN} * 100 \quad (17)$$

$$Sel = \frac{VN}{FP + VN} * 100 \quad (18)$$

$$Sen = \frac{VP}{VP + FN} * 100 \quad (19)$$

$$TFP = \frac{FP}{FP + VN} * 100 \quad (20)$$

em que  $VP$ ,  $VN$ ,  $FP$  e  $FN$  são os valores verdadeiro positivo, verdadeiro negativo, falso positivo e falso negativo, respectivamente, para cada uma das três classes individuais. As medidas gerais de desempenho, ou seja, precisão, sensibilidade, seletividade e taxa de falsos positivos, foram calculadas com base nas medidas de desempenho das três classes.

## RESULTADOS E DISCUSSÕES

### Análise química das uvas

A Tabela 3.1-1 mostra os principais descritores estatísticos para os atributos de qualidade que caracterizam os conjuntos de calibração e predição, isto é, o número de amostras, a média, os valores de mínimo e máximo, o intervalo, o desvio padrão e a variância. Pode-se notar uma alta variabilidade dos atributos de qualidade mensurados, especialmente para antocianinas totais e flavonoides amarelos, devido aos diferentes estádios de amadurecimento da uva, proporcionando assim o desenvolvimento de modelos matemáticos versáteis de predição.

Tabela 3.1-1. Estatística descritiva das variedades Syrah, Cabernet Sauvignon e ambas as variedades juntas (Geral).

Parâmetro	Amostras	Média	Min	Max	Alcance	SD	Variância	Amostras	Média	Min	Max	Alcance	SD	Variância
Calibração								Predição						
Syrah								Cabernet Sauvignon						
SST	288	13,6	4,3	22,9	18,6	5,4	28,6	144	14,1	4,6	22,2	17,6	4,9	24,2
Anth	288	82,1	1,4	194,1	192,8	71,6	5133,2	144	87,9	1,4	193,7	192,3	71,0	5042,4
Flav	288	29,0	8,5	88,4	79,9	17,4	301,9	144	31,6	9,5	83,6	74,2	17,6	311,5
Cabernet Sauvignon								Geral						
SST	384	13,3	3,9	26,2	22,3	6,6	43,2	192	14,5	4,1	24,9	20,8	6,7	44,6
Anth	384	105,4	0,8	197,9	197,0	83,3	6946,6	192	120,6	1,3	197,5	196,2	84,4	7129,7
Flav	384	41,8	9,7	153,3	143,6	26,9	723,0	192	59,9	12,0	137,9	125,8	33,5	1121,4
SST	672	13,5	3,9	26,2	22,3	6,1	36,8	336	14,3	4,1	25,9	21,8	6,0	36,1
Anth	672	95,4	0,8	197,9	197,0	79,3	6293,9	336	106,6	1,3	197,5	196,2	80,5	6481,7
Flav	672	36,3	8,5	153,3	144,8	24,1	580,3	336	47,8	9,5	137,9	128,4	31,2	971,0

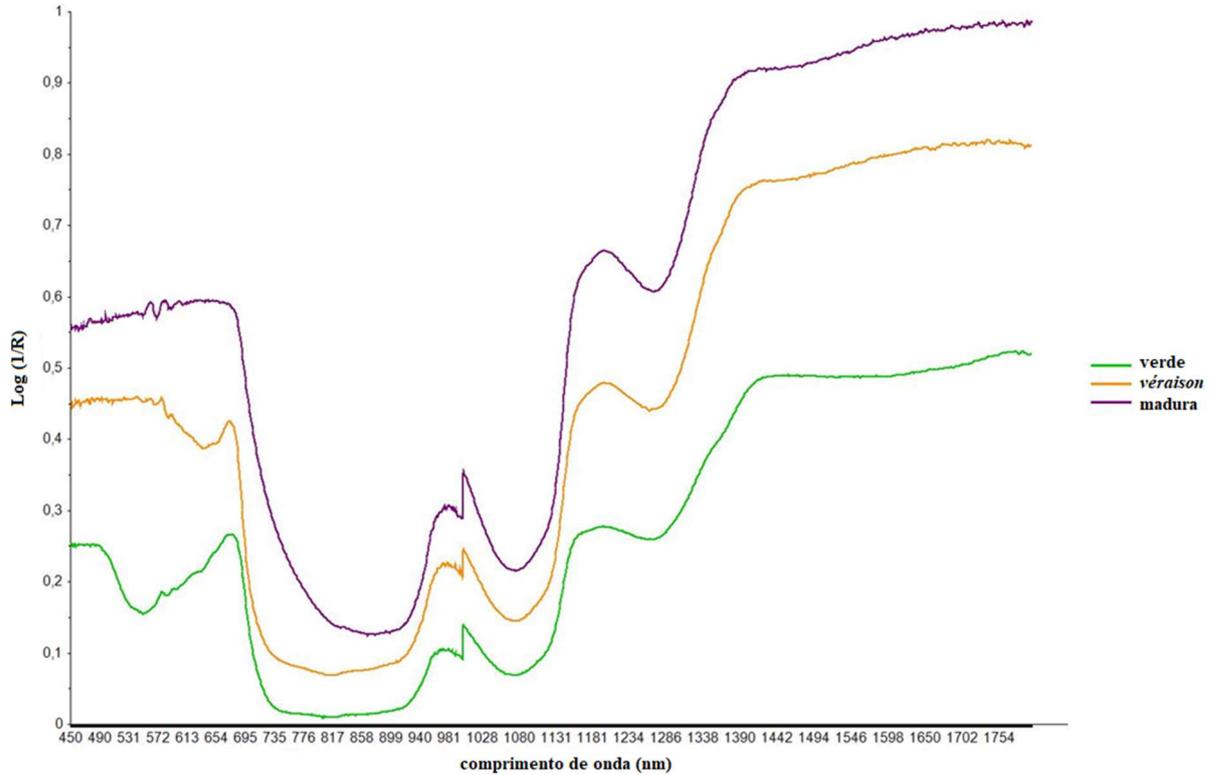
Legenda: Min – valor mínimo; Max – valor máximo; SD – desvio padrão

### Análise exploratória do espectro

Os espectros brutos de absorvância  $\log \frac{1}{R}$  dos três estádios de amadurecimento (verde, *véraison* e madura) em bagas de uva viníferas são mostrados na Figura 3.1-4. Observa-se uma diferenciação dos espectros em torno do comprimento de 550 nm, estando este compreendido na região do visível. O amadurecimento das bagas de uvas tintas é geralmente acompanhada de uma mudança na cor do epicarpo, devido a uma modificação da concentração de pigmentos nos tecidos superficiais (AGATI *et al.*, 2005). Normalmente, o desenvolvimento da coloração verde para roxa é modulada pelo acúmulo de antocianinas, juntamente com a degradação das clorofilas (AGATI *et al.*, 2008; HAGEN *et al.*, 2006).

Além disso, o padrão de desenvolvimento dos frutos segue uma curva dupla-sigmóide dividida em três estádios. O estágio I é caracterizado pela intensa divisão e expansão celular; o estágio II é caracterizado pelo reduzido crescimento durante o amadurecimento da semente; e o estágio III é caracterizado pelo acúmulo de açúcares, redução da acidez, expansão das células da polpa e acúmulo de antocianinas (KENNEDY, 2002). Como grande parte das substâncias relacionadas à coloração dos frutos (clorofila e antocianinas) encontra-se na epiderme (AGATI *et al.*, 2007), ocorre uma interação, a nível molecular, destas com a energia radiante nos diferentes comprimentos de onda.

Figura 3.1-4. Espectros brutos de absorbância de diferentes estádios de amadurecimento (verde, *véraison* e madura) para três amostras de bagas de uva.



### Modelos PCR e PLSR com espectro completo

Os modelos matemáticos foram construídos com os métodos de PCR e PLSR para as variedades Syrah e Cabernet Sauvignon, separadamente, e para a combinação do conjunto de dados de ambas as variedades (modelo geral), após a aplicação das técnicas de pré-tratamento. A Tabela 3.1-2, mostra os melhores modelos PCR e PLSR para os atributos de qualidade analisados.

Tabela 3.1-2. Modelos matemáticos de regressão da variedade Syrah, Cabernet Sauvignon e ambas as variedades (Syrah e Cabernet Sauvignon) usando o espectro completo (450 nm a 1800 nm).

Parâmetro	Modelo	Pré-tratamento	Fator	R <sup>2</sup> <sub>cal</sub>	RMSEC	SEC	R <sup>2</sup> <sub>cv</sub>	RMSECV	SECV	R <sup>2</sup> <sub>pred</sub>	RMSEP	SEP	Bias <sub>pred</sub>
<b>Syrah</b>													
SST	PLSR	MSC (1, 7, 5, 1) MF	8	0,97	0,89	0,89	0,94	1,33	1,33	0,95	1,15	1,13	0,25
	PCR	SNV (1, 7, 5, 1) MAF	2	0,84	2,11	2,11	0,84	2,13	2,14	0,92	1,41	1,41	-0,03
Anth	PLSR	SNV (1, 7, 9, 1) MF	3	0,94	17,54	17,57	0,93	18,35	18,39	0,94	16,90	16,74	2,71
	PCR	SNV (1, 7, 13, 1) MAF	5	0,93	18,38	18,41	0,93	18,63	18,66	0,94	16,79	16,71	2,10
Flav	PLSR	(1, 21, 27, 1) MAF	7	0,76	8,54	8,56	0,73	8,99	9,01	0,75	8,73	8,76	-0,01
	PCR	(1, 21, 27, 1) MAF	11	0,75	8,72	8,73	0,73	9,08	9,09	0,75	8,82	8,85	0,29
<b>Cabernet Sauvignon</b>													
SST	PLSR	SNV (1, 21, 27, 1) MAF	6	0,97	1,08	1,09	0,97	1,13	1,13	0,96	1,39	1,40	-0,03
	PCR	MSC (1, 21, 17, 1) MAF	2	0,90	2,08	2,08	0,90	2,09	2,09	0,94	1,63	1,63	0,11
Anth	PLSR	SNV (1, 21, 27, 1) MF	3	0,97	15,11	15,13	0,97	15,46	15,48	0,98	13,56	13,57	0,85
	PCR	SNV (1, 21, 13, 1) MAF	4	0,96	15,65	15,67	0,96	15,86	15,88	0,97	13,66	13,63	1,27
Flav	PLSR	SNV (1, 21, 27, 1) MF	2	0,73	13,88	13,90	0,73	14,00	14,02	0,74	21,21	17,53	-12,01
	PCR	SNV (1, 21, 27, 1) MF	2	0,73	13,93	13,94	0,73	14,01	14,03	0,74	21,23	17,54	-12,03
<b>Geral</b>													
SST	PLSR	SNV (1, 7, 5, 1) MF	7	0,96	1,21	1,21	0,95	1,37	1,37	0,95	1,38	1,39	0,02
	PCR	MSC (1, 21, 27, 1) MF	2	0,87	2,22	2,22	0,86	2,23	2,23	0,92	1,66	1,66	-0,06
Anth	PLSR	SNV (1, 21, 27, 1) MF	3	0,95	18,04	18,06	0,95	18,28	18,29	0,96	16,74	16,60	2,36
	PCR	SNV (1, 7, 13, 1) MF	3	0,94	20,06	20,08	0,94	20,19	20,21	0,95	18,46	18,31	2,51
Flav	PCR	SNV (1, 7, 27, 1) MF	3	0,72	12,84	12,85	0,71	12,92	12,93	0,73	18,55	17,08	-7,30
	PLSR	(1, 21, 17, 1) MF	5	0,70	13,23	13,24	0,69	13,48	13,49	0,67	20,07	18,61	-7,58

Legenda: R<sup>2</sup><sub>cal</sub> - coeficiente de determinação da calibração; RMSEC - erro quadrático médio de calibração; SEC - erro padrão de calibração; R<sup>2</sup><sub>cv</sub> - coeficiente de determinação da validação cruzada; RMSECV - erro quadrático médio de validação cruzada; SECV - erro padrão de validação cruzada; R<sup>2</sup><sub>pred</sub> - coeficiente de determinação da predição; RMSEP - erro quadrático médio de predição; SEP - erro padrão de predição; Bias<sub>pred</sub> - viés da predição.

Para o conjunto de dados da Syrah, os melhores pré-tratamentos foram conseguidos com o modelo PLSR associado a MSC (1, 7, 5, 1) MF, SNV (1, 7, 9, 1) MF e (1, 21, 27, 1) MAF, sendo os valores de RPD 4,06, 3,89 e 1,93; os valores de RER 13,98, 10,49 e 8,87; os limites de controle SEP 1,16 °Brix, 22,80 mg.100 g<sup>-1</sup> e 11,11 mg.100 g<sup>-1</sup>; e os limites de controle de viés 0,53 °Brix, 10,52 mg.100 g<sup>-1</sup> e 5,13 mg.100 g<sup>-1</sup> para as variáveis respostas SST, antocianinas totais e flavonoides amarelos, respectivamente.

Para o conjunto de dados da Cabernet Sauvignon, os melhores pré-tratamentos foram conseguidos com o modelo PLSR associado a SNV (1, 21, 27, 1) MAF, SNV (1, 21, 27, 1) MF e SNV (1, 21, 27, 1) MF, sendo os valores de RPD 5,83, 5,38 e 1,92; os valores de RER 19,68,

12,72 e 10,24; os limites de controle SEP 1,41 °Brix, 19,65 mg.100 g<sup>-1</sup> e 18,05 mg.100 g<sup>-1</sup>; e os limites de controle de viés 0,65 °Brix, 9,07 mg.100 g<sup>-1</sup> e 8,33 mg.100 g<sup>-1</sup> para as variáveis respostas SST, antocianinas totais e flavonoides amarelos, respectivamente.

Para o conjunto de dados do modelo geral (Syrah e Cabernet Sauvignon), os melhores pré-tratamentos foram conseguidos com o modelo PLSR associado ao SNV (1, 7, 5, 1) MF, SNV (1, 21, 27, 1) MF e PCR associado ao SNV (1, 7, 27, 1) MF, sendo os valores de RPD 4,44, 4,34 e 1,86; os valores de RER 16,24, 10,77 e 11,20; os limites de controle SEP 1,57 °Brix, 23,45 mg.100 g<sup>-1</sup> e 16,70 mg.100 g<sup>-1</sup>; e os limites de controle de viés 0,73 °Brix, 10,83 mg.100 g<sup>-1</sup> e 7,71 mg.100 g<sup>-1</sup> para as variáveis respostas SST, antocianinas totais e flavonoides amarelos, respectivamente.

Shenk e Westerhaus (1996) sugerem que um valor de R<sup>2</sup> superior a 0,9 indica excelente ajuste do modelo, fato observado nos modelos PLSR para as variáveis respostas SST e antocianinas totais nas etapas de calibração, validação cruzada e predição.

Ademais, Williams e Sobering (1996) recomendam que a razão RPD seja próxima de três e que a relação RER seja igual ou superior a dez. Nesse sentido, no que se refere aos valores de RPD, estes não foram satisfatórios para o atributo flavonoides amarelos, considerando todos os conjuntos de dados estudados e, no que diz respeito aos valores de RER, esse foi ligeiramente inferior (RER = 8,87) para o mesmo atributo na variedade Syrah. As aplicabilidades dos modelos dependem dos valores de RPD: quando estes estão compreendidos entre 3,1 e 4,9 indicam que o modelo é adequado para a triagem; entre 5,0 e 6,4, adequado para o controle de qualidade; entre 6,5 e 8,0, adequado para o controle do processo; e maiores que 8,0, adequados para qualquer aplicação (WILLIAMS, 2001). Os modelos desenvolvidos nesse estudo para os atributos SST e antocianinas totais se enquadram nos critérios de triagem e controle de qualidade, a depender do conjunto de dados analisado.

Além disso, de acordo com o procedimento proposto por Shenk *et al.* (1989), que trata da avaliação do desempenho dos modelos desenvolvidos, o SEP obtido foi inferior ao limite de confiança das variáveis respostas, exceto para flavonoides amarelos no modelo geral (SEP = 17,08 mg.100 g<sup>-1</sup>), onde essa métrica excedeu ligeiramente o limite máximo recomendado (16,70 mg.100 g<sup>-1</sup>). Ao mesmo tempo, o viés foi inferior ao limite de confiança das variáveis respostas, exceto novamente para flavonoides amarelos para o modelo com a variedade Cabernet Sauvignon, onde essa métrica (viés = -12,01 mg.100 g<sup>-1</sup>) excedeu o limite máximo recomendado ( $\pm 8,33$  mg.100 g<sup>-1</sup>).

A capacidade preditiva dos modelos para SST em cada variedade, separadamente, e com o conjunto de ambas (Syrah e Cabernet Sauvignon) foram melhores do que os relatados por

Herrera *et al.* (2003) onde, para esse atributo na variedade Cabernet Sauvignon utilizando a espectroscopia na faixa espectral entre 650 nm a 1100 nm, obtiveram valores de  $R^2_{\text{pred}}$  e RMSEP de 0,896 e 1,25 °Brix. Em um estudo mais recente, Nogales-Bueno *et al.* (2014), utilizando imagens hiperespectrais numa faixa mais ampla (950 nm a 1650 nm), obtiveram resultados semelhantes com valores de  $R^2$  e SEP de 0,99 e 1,37 °Brix, respectivamente, para uvas da variedade Tempranillo e Syrah.

Normalmente, os estudos de desenvolvimento dos modelos de regressão para uvas tende para predição dos compostos fenólicos, em virtude da sua importância como agente antioxidante (FERRER-GALLEGO *et al.*, 2011; KEMPS *et al.*, 2010; MARTÍNEZ-SANDOVAL *et al.*, 2016; NOGALES-BUENO *et al.*, 2015). Para antocianinas totais, os resultados encontrados nesse estudo são melhores do que os relatados por Nogales-Bueno *et al.* (2015) que desenvolveram modelos de predição com uso de imagens hiperespectrais (900 nm a 1700 nm) e conseguiram valor de  $R^2_{\text{cal}}$  igual a 0,79, e por Martínez-Sandoval *et al.* (2016) predizendo antocianinas totais e antocianinas não aciladas, que encontraram valores de  $R^2_{\text{pred}}$  0,76 e 0,75, respectivamente.

Contudo, Kemps *et al.* (2010) mesmo utilizando espectroscopia VIS-NIR para predição de antocianinas a pH 1,0, encontraram valores de  $R^2_{\text{pred}}$  de 0,66 e 0,86 para Cabernet Sauvignon e Syrah, respectivamente. Nota-se que essas pesquisas fazem uso na grande maioria da tecnologia NIR, entretanto, utilizando uma faixa espectral mais ampla (VIS-NIR) em associação com a aplicação de diferente pré-tratamentos matemáticos, como utilizada nesse estudo, consegue-se resultados mais satisfatórios, principalmente para antocianinas tendo em vista que este é um composto pigmentante.

Portanto, os modelos preditivos desenvolvidos para os atributos sólidos solúveis totais e antocianinas totais, utilizando a região espectral de 450 nm a 1800 nm (visível e infravermelho), apresentam precisão para o monitoramento das vinhas Syrah e Cabernet Sauvignon.

### **Modelos PCR e PLSR com seleção de variáveis**

Na discussão acima, nenhuma consideração foi dada à contribuição dos comprimentos de onda para os modelos desenvolvidos. Por isso, foi realizada uma seleção de variáveis com a combinação do conjunto de dados de ambas as variedades (Syrah e Cabernet Sauvignon) utilizando o método Jack-Knife. Novamente, reconstruiu-se os modelos pelos métodos de PCR e PLSR e os resultados são mostrados na Tabela 3.1-3.

Tabela 3.1-3. Modelos matemáticos de regressão com seleção de variáveis Jack-Knife o conjunto de ambas variedades (Syrah e Cabernet Sauvignon).

Parâmetro	Modelo	Pré-tratamento	Fator	$R^2_{cal}$	RMSEC	SEC	$R^2_{cv}$	RMSECV	SECV	$R^2_{pred}$	RMSEP	SEP	Bias <sub>pred</sub>
SST	PLSR	SNV (1, 7, 5, 1) MF	4	0,95	1,40	1,40	0,94	1,44	1,44	0,94	1,47	1,48	0,01
	PCR	MSC (1, 21, 27, 1) MF	2	0,87	2,22	2,22	0,86	2,23	2,23	0,92	1,66	1,66	-0,06
Anth	PLSR	SNV (1, 21, 27, 1) MF	3	0,95	18,00	18,01	0,95	18,21	18,23	0,96	16,64	16,50	2,35
	PCR	SNV (1, 7, 13, 1) MF	3	0,94	20,07	20,09	0,94	20,20	20,21	0,95	18,43	18,30	2,46
Flav	PCR	SNV (1, 7, 27, 1) MF	2	0,71	12,91	12,92	0,71	12,97	12,98	0,73	18,64	17,12	-7,43
	PLSR	(1, 21, 17, 1) MF	4	0,69	13,37	13,38	0,69	13,53	13,55	0,66	20,22	18,86	-7,37

Legenda:  $R^2_{cal}$  - coeficiente de determinação da calibração; RMSEC - erro quadrático médio de calibração; SEC - erro padrão de calibração;  $R^2_{cv}$  - coeficiente de determinação da validação cruzada; RMSECV - erro quadrático médio de validação cruzada; SECV - erro padrão de validação cruzada;  $R^2_{pred}$  - coeficiente de determinação da predição; RMSEP - erro quadrático médio de predição; SEP - erro padrão de predição; Bias<sub>pred</sub> - vies da predição.

Os melhores modelos reconstruídos, com seus respectivos pré-tratamentos matemáticos, foram os mesmos que aqueles obtidos com o espectro completo (Tabela 3.1-2). De modo geral, a complexidade e precisão dos modelos foi reduzida, fato evidenciado quando se compara o número de fatores, coeficientes de determinação e erros padrões dos modelos por espectro completo com modelos por seleção de variáveis pelo método Jack-Knife.

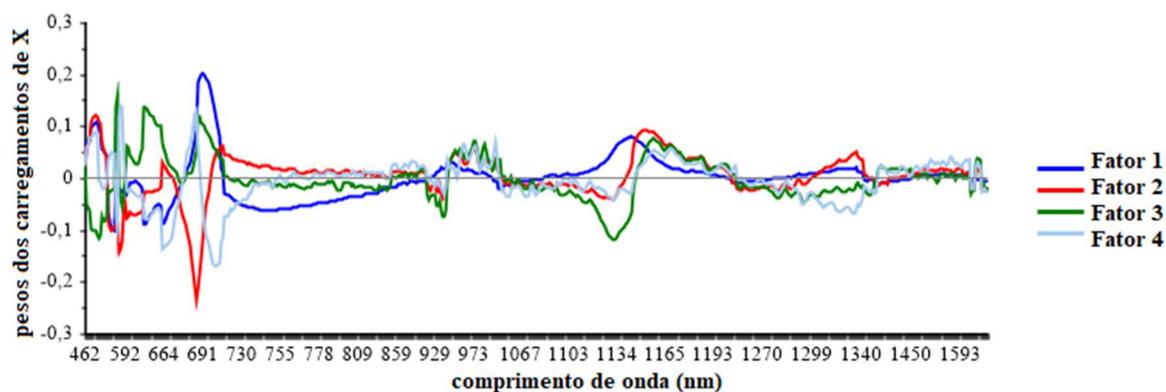
Isso está de acordo com Wu *et al.* 2010, os quais observaram que determinados comprimentos de onda têm pouca ou nenhuma contribuição sobre os modelos desenvolvidos, não modelando essas variáveis sem informações, e conseqüentemente, diminuindo o número de fatores dos modelos. Além disso, os mesmos autores realizaram seleção de variáveis e observaram que modelos com comprimentos de onda selecionados foram piores do que com o espectro completo, isso por que o algoritmo de seleção foi aplicado sobre um espectro com baixa relação sinal/ruído, o que também é observado nesse estudo.

Contudo, o modelo PLSR com pré-tratamento SNV (1, 21, 27, 1) MF para a variável resposta antocianinas totais apresentou métricas que descrevem o erro (RMSE e SE) ligeiramente menores com o espectro selecionado do que com o espectro completo. Resultado semelhante foi obtido por Cao *et al.* (2010), para predição de SST com modelos desenvolvidos através de comprimentos de onda selecionados. Outrossim, os modelos com variáveis selecionadas apresentam entradas reduzidas, minimizando o tempo de calibração dos modelos (CAO *et al.*, 2010; WU *et al.*, 2010).

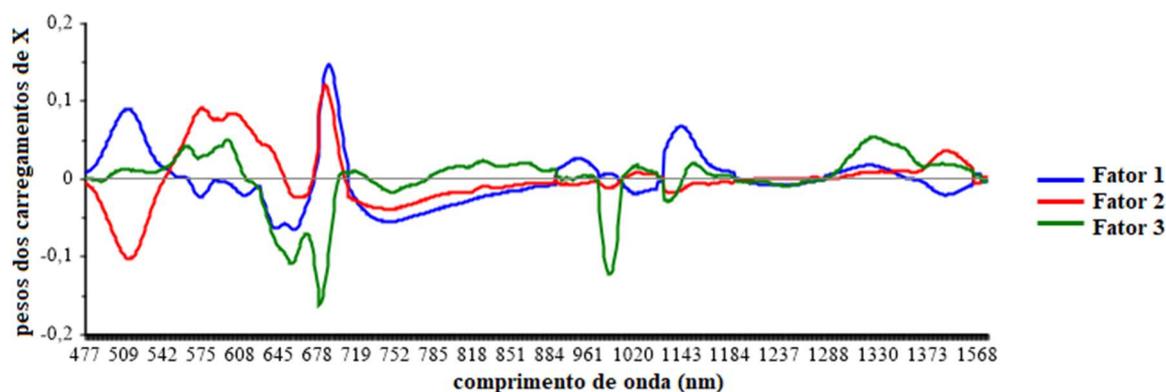
## Modelos MLR com seleção de variáveis

Os melhores pré-tratamentos do conjunto de dados composto por ambas as variedades (Syrah e Cabernet Sauvignon) foram selecionados com base no desempenho dos modelos PCR e PLSR com o método Jack-Kfine. Contudo, em virtude dos problemas de colinearidade, foi necessário fazer uma seleção de variáveis para usar como entrada para a MLR. Sendo assim, uma análise dos gráficos de carregamento das variáveis independentes nos atributos SST, antocianinas totais e flavonoides amarelos foi realizada e evidenciou uma possível associação entre os comprimentos de onda sensíveis e os atributos de qualidade, conforme mostrado na Figura 3.1-5.

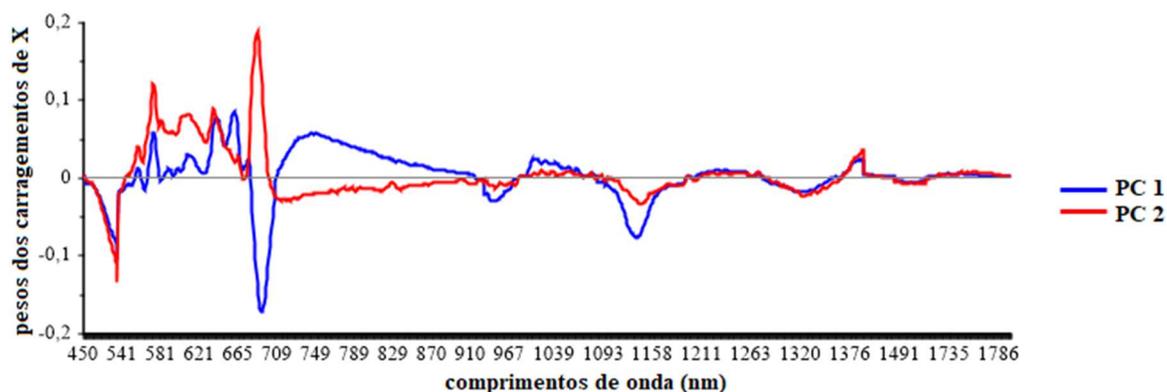
Figura 3.1-5. Gráficos de carregamento de pesos das variáveis independentes para SST com pré-tratamento SNV (1, 7, 5, 1) MF (a), antocianinas totais com pré-tratamento SNV (1, 21, 27, 1) MF (b) e flavonoides amarelos com pré-tratamento SNV (1, 7, 27, 1) MF (c).



(a)



(b)



(c)

Os comprimentos de onda com maiores correlações selecionados foram: 507, 581, 587, 644, 664, 684, 692, 700, 945, 973, 980, 986, 1132, 1154, 1161, 1280, 1324, 1412, 1581 e 1745 nm para SST; 515, 576, 599, 625, 657, 684, 689, 692, 714, 890, 955, 1001, 1023, 1129, 1130, 1144, 1154, 1327, 1369 e 1389 nm para antocianinas totais; e 508, 558, 566, 574, 615, 627, 642, 661, 671, 686, 692, 710, 748, 951, 1019, 1143, 1149, 1324, 1388 e 1541 nm para flavonoides amarelos. Uma vez que os comprimentos de onda individuais podem ser particularmente importantes para os atributos de qualidade das uvas viníferas, todos eles foram selecionados nas curvas de carregamento, independentemente de serem ou não as assinaturas espectrais reais e seus desempenhos foram avaliados.

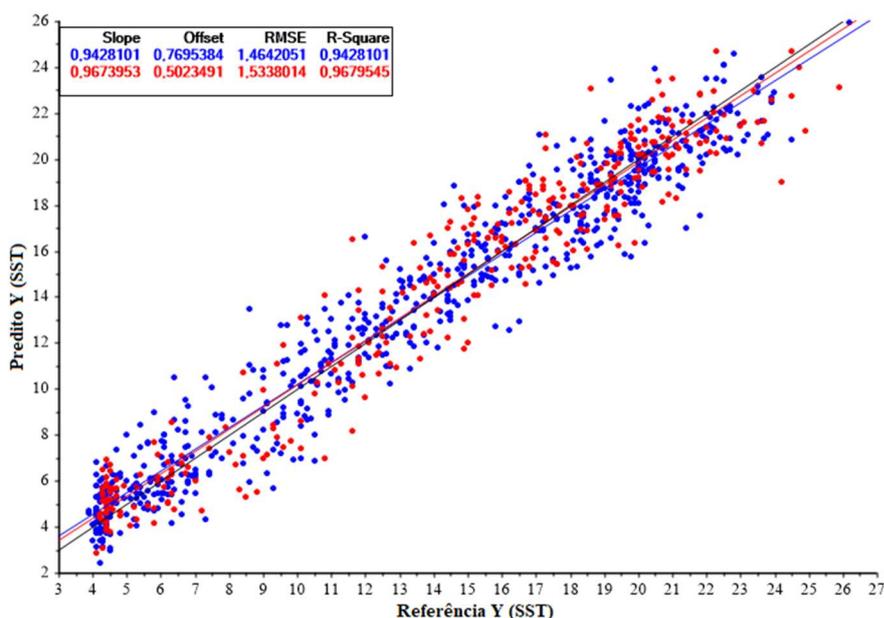
Os modelos MLR submetidos a ANOVA e teste de significância a 5% selecionaram as assinaturas espectrais: 507, 587, 692, 700, 973, 980, 986, 1132, 1154, 1161, 1280, 1324 e 1581 nm para SST; 625, 657, 890, 955, 1154, 1327 e 1389 nm para antocianinas totais; e 671, 692 e 748 nm para flavonoides amarelos.

No presente estudo, as assinaturas espectrais identificadas na região do visível (380 a 780 nm), estão relacionadas à presença de pigmentos verdes (clorofila) e vermelhos (antocianinas e flavonoides), durante o amadurecimento das bagas de uvas (AGATI *et al.*, 2008, 2005; CAO *et al.*, 2010; CEN e HE, 2007; KEMPS *et al.*, 2010). Por sua vez, as assinaturas identificadas na região do infravermelho (780 a 2500 nm) estão relacionadas às bandas de absorção por: (1) açúcares (em torno de 980 nm), devido ao segundo sobretom de O-H; (2) água (em torno de 973, 1324 e 1581 nm), devido ao terceiro sobretom de O-H; (3) e antocianinas (em torno de 1154 nm) devido ao segundo sobretom C-H que está relacionado aos compostos coloridos (CAO *et al.*, 2010; CEN e HE, 2007; CHEN *et al.*, 2015; GONZÁLEZ-CABALLERO *et al.*, 2010; WILLIAMS, 2001).

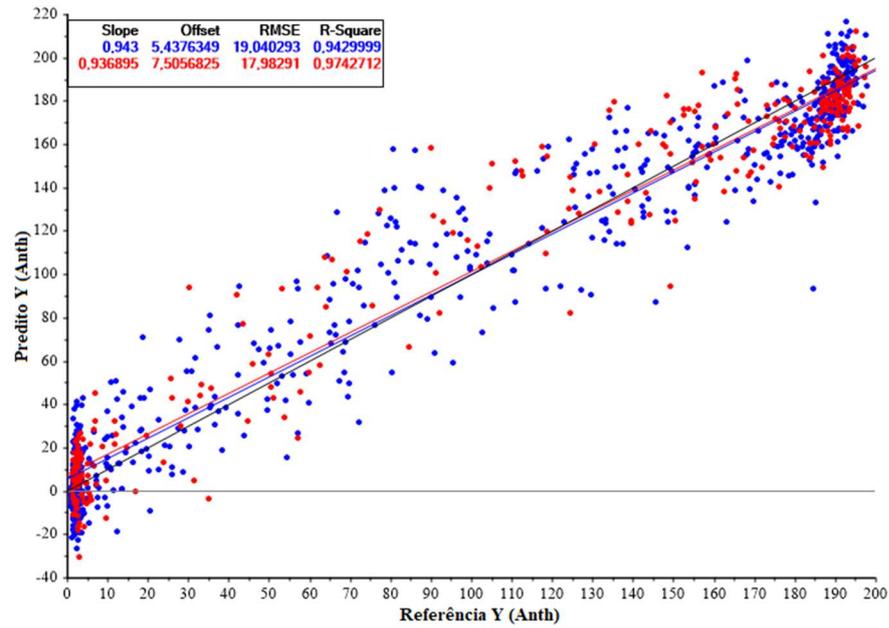
Os resultados de calibração e predição dos modelos MLR construídos com as assinaturas espectrais são mostrados na Figura 3.1-6. Os modelos MLR apresentaram alta correlação ( $R^2 > 0,90$ ), com exceção do qual prediz o atributo flavonoides amarelos ( $R^2 = 0,70$ ). Para esses modelos, foram obtidos: SECV de 1,45 °Brix, SEP de 1,54 °Brix, RPD de 4,21 e RER de 15,38, para SST; SECV de 18,94 mg.100 g<sup>-1</sup>, SEP de 17,99 mg.100 g<sup>-1</sup>, RPD de 4,19 e RER de 10,40, para antocianinas totais; e SECV de 12,55 mg.100 g<sup>-1</sup>, SEP de 22,52 mg.100 g<sup>-1</sup>, RPD de 1,92 e RER de 11,54, para flavonoides amarelos. Além disso, sobre o monitoramento da qualidade dos modelos desenvolvidos, o SEP foi inferior ao limite de confiança das variáveis respostas, exceto para flavonoides amarelos (SEP = 22,52 mg.100 g<sup>-1</sup>), onde essa métrica excede o limite máximo recomendado (16,35 mg.100 g<sup>-1</sup>). Para o viés, esse foi inferior ao limite de confiança das variáveis respostas.

Observa-se que os modelos com variáveis selecionadas apresentaram desempenho ligeiramente inferior quando se compara com o espectro completo (Tabela 3.1-2) e seleção de variáveis pelo método Jack Knife (Tabela 3.1-3) nos modelos PCR e PLSR para os atributos SST, antocianinas totais e flavonoides amarelos, sendo algo esperado de acordo com Fernández-Navales *et al.* (2009b) e Wu *et al.* (2010). Contudo, essa inferioridade não compromete a atividade de previsão dos atributos de qualidade.

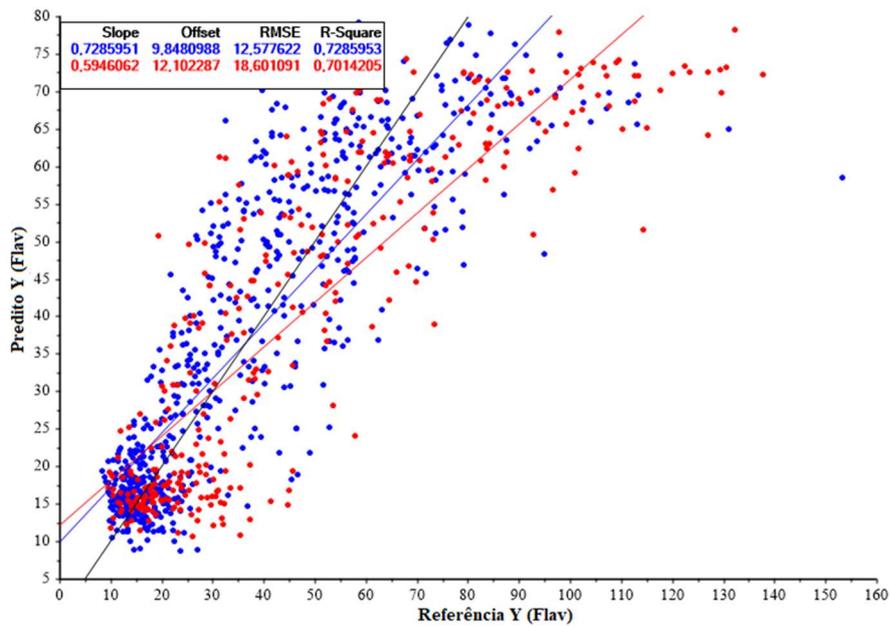
Figura 3.1-6. Dados previstos pela espectroscopia VIS-NIR *versus* dados de referência para o conjunto de calibração e validação externa para SST (a), antocianinas totais (b) e flavonoides amarelos (c).



(a)



(b)



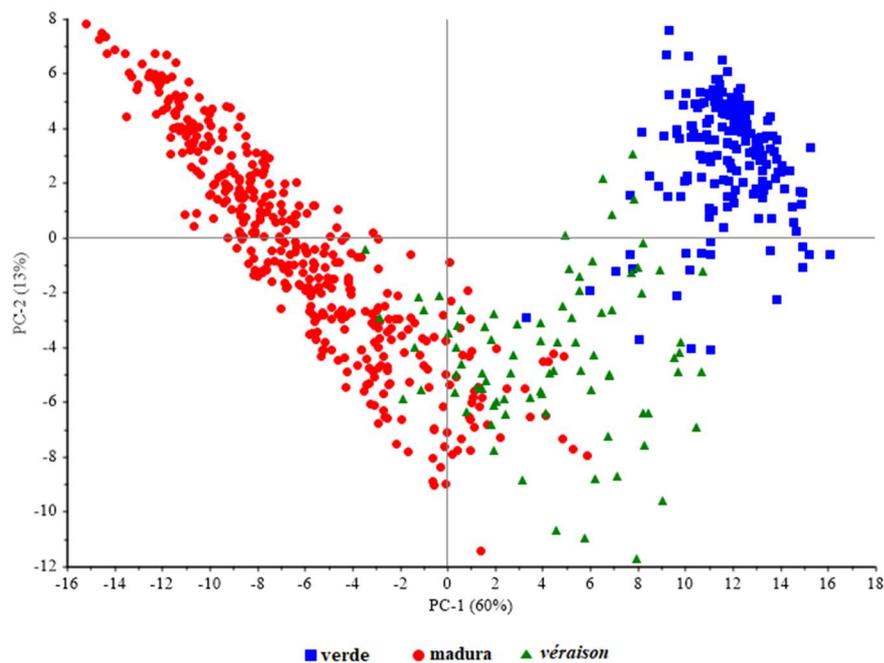
(c)

Assim, as assinaturas espectrais identificadas ao longo do espectro visível e infravermelho podem ser utilizadas para desenvolver sensores e instrumentos optoeletrônicos para determinação não destrutiva e previsão de sólidos solúveis totais, antocianinas totais e flavonoides amarelos destinados à indústria vitivinícola.

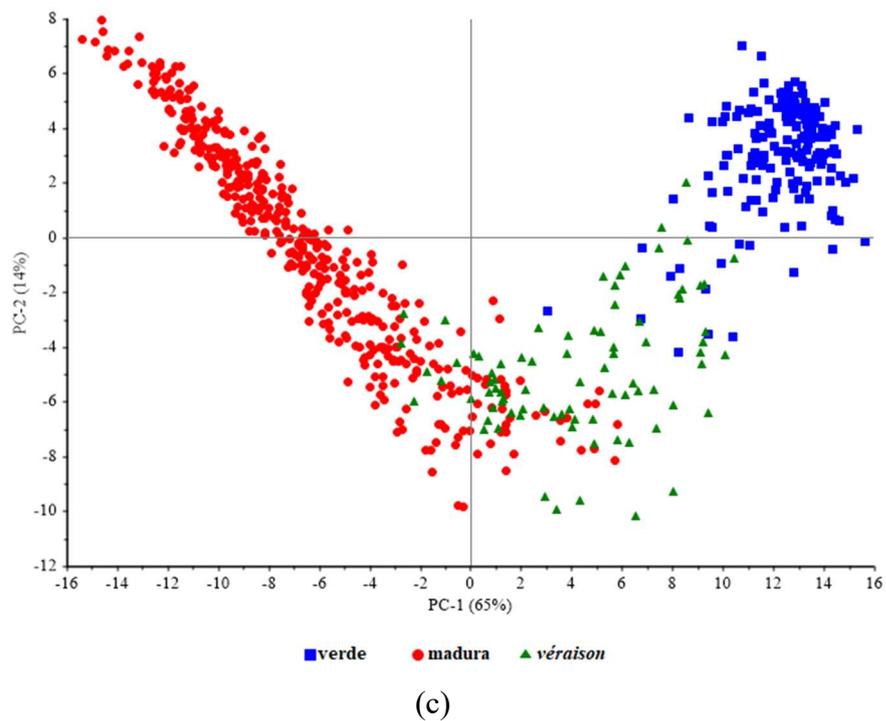
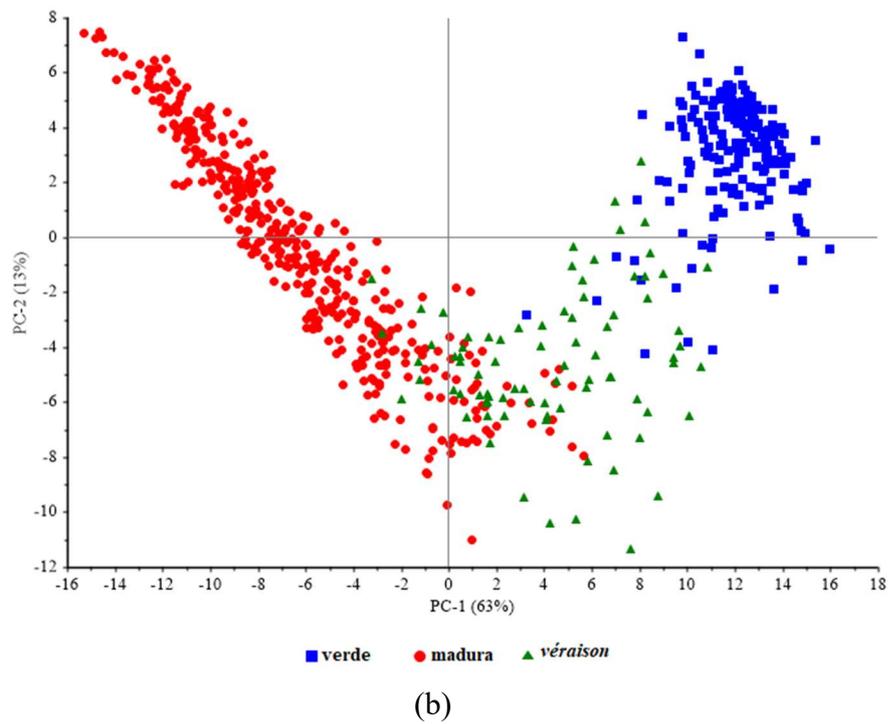
### Análise de componentes principais (PCA)

A Figura 3.1-7, mostra os gráficos de pontuação das amostras de uva no espaço definido pela primeira e segunda componente principal. Os pré-tratamentos que se destacaram na PCA foram: SNV (1, 7, 5, 1) MF que descreve 60% e 13% da variabilidade espectral na primeira e segunda componente principal, respectivamente, e variância total de 88% com seis componentes principais; SNV (1, 7, 13, 1) MF que descreve 63% e 13% da variabilidade espectral na primeira e segunda componente principal, respectivamente, e variância total de 91% com seis componentes principais; e SNV (1,7, 27, 1) MF que descreve 65% e 14% da variabilidade espectral na primeira e segunda componente principal, respectivamente e variância total de 90% com seis componentes principais.

Figura 3.1-7. Gráficos de pontuação da PCA com os pré-tratamentos SNV (1, 7, 5, 1) MF (a), SNV (1, 7, 13, 1) MF (b) e SNV (1,7, 27, 1) MF (c).



(a)



Nos gráficos da Figura 3.1-7, as diferenças entre os três estádios de amadurecimento (verde, *véraison* e madura) são evidentes ao longo da primeira componente principal e mostra agrupamentos de amostras ao longo do amadurecimento das bagas. Existe uma sobreposição de amostras no estágio de amadurecimento *véraison* em relação aos estádios verde e madura. Isso ocorre devido à mudança de coloração não ocorrer de forma abrupta ao longo do

desenvolvimento da baga da uva. A discriminação dos estádios de amadurecimento em uva foi relatada por Musingarabwi *et al.* (2016), onde conseguiu-se diferenciar cinco estádios de amadurecimento (verde, *pré-véraison*, *véraison*, *pós-véraison* e madura) e por Nogales-Bueno *et al.* (2014), onde foi possível diferenciar os primeiros dias de amadurecimento do resto das amostras. Contudo, Martínez-Sandoval *et al.* (2016), não conseguiram evidenciar uma diferenciação entre os estádios de amadurecimento.

### Métodos supervisionados de classificação (LDA e PLS-DA)

A capacidade de discriminação dos estádios de amadurecimento com uso de espectroscopia VIS-NIR e métodos de classificação supervisionada é mostrada na Tabela 3.1-4. Os modelos classificaram corretamente os três estádios de amadurecimento das bagas de uvas, sendo o LDA com distância calculada pelo método linear e o PLS-DA os que apresentaram maiores taxas de acerto, na ordem de 92,86% e 93,15%, respectivamente. Melhores resultados foram conseguidos nesse estudo em comparação ao estudo de González-Caballero *et al.* (2012), que empregaram a espectroscopia (1600 a 2400 nm) associada com PLS-DA como método de classificação para distinguir o amadurecimento dos cachos de uva da variedade Cabernet Sauvignon com precisão de 88%, permitindo uma colheita seletiva para produção de vinhos tintos jovens, vintage e suaves. Além disso, resultados semelhantes foram alcançados por Fernández-Navales *et al.* (2009a), utilizando a espectroscopia (700 a 1060 nm) e o PLS-DA para discriminar variedades de uvas brancas e tintas, os autores conseguiram taxas de acerto de 91,3% e 100%, respectivamente.

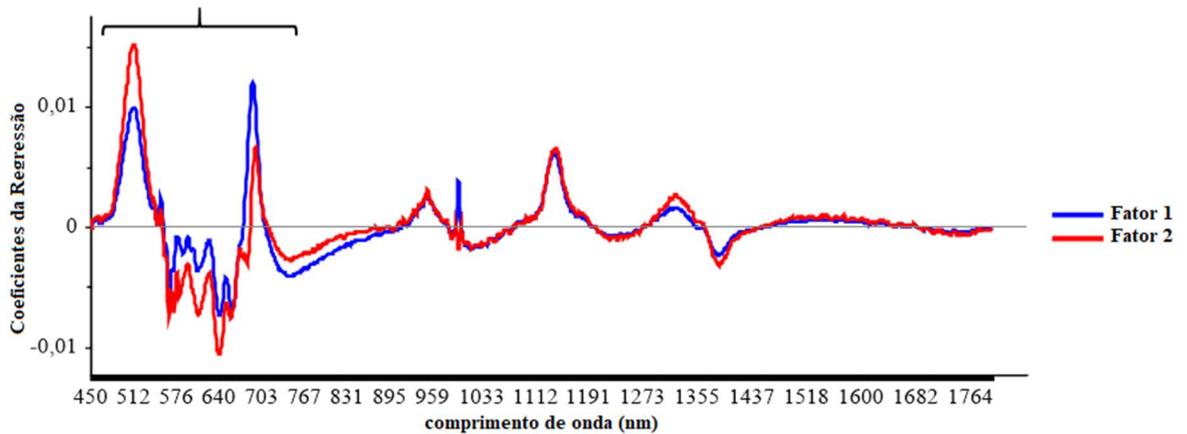
Tabela 3.1-4. Precisão, sensibilidade, seletividade e falsos positivos da classificação dos estádios de amadurecimento usando espectros entre 450 e 1800 nm por modelos LDA e PLS-DA.

Modelo	Pré-tratamento	Precisão (%)	Sensibilidade (%)	Seletividade (%)	Falsos Positivos (%)
LDA_linear	SNV (1, 7, 5, 1) MF	92,86	85,59	95,46	4,54
LDA_quadratic	SNV (1, 7, 27, 1) MF	92,26	86,54	95,32	4,68
LDA_mahalanobis	SNV (1, 7, 13, 1) MF	92,26	88,03	95,55	4,45
PLS-DA	SNV (1, 7, 13, 1) MF	93,15	83,51	95,17	4,83

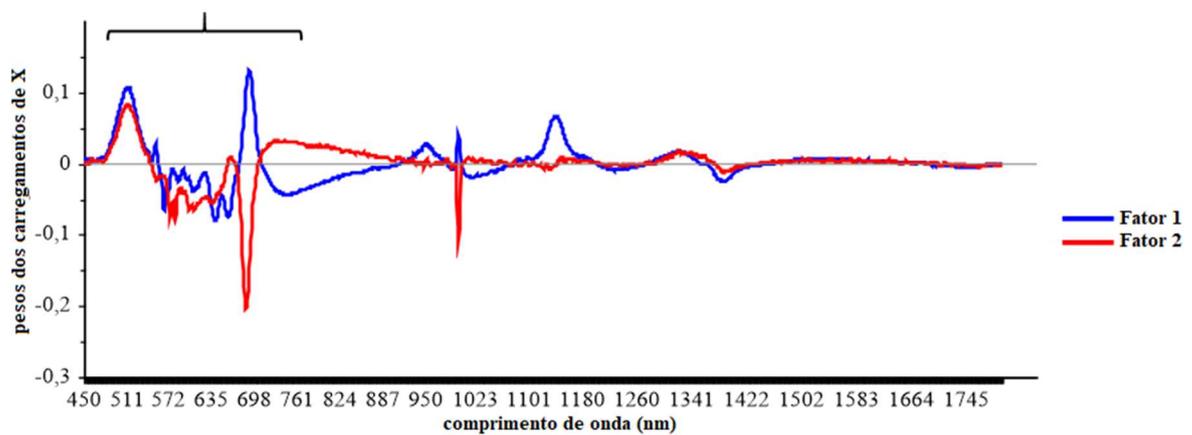
A classificação dos estádios de amadurecimento está associada à mudança de coloração em virtude da degradação e acúmulo de substâncias pigmentantes (clorofila e antocianinas). Tal

fato pode ser verificado nos gráficos de coeficiente da regressão e carregamento da variável independente (comprimentos de onda) ilustrado na Figura 3.1-8, onde as maiores contribuições para o modelo PLS-DA são de variáveis que se encontram na região do visível (450 a 780 nm).

Figura 3.1-8. Gráficos de coeficientes de correlação (a) e carregamento de pesos (b) das variáveis independentes no modelo PLS-DA.



(a)



(b)

Por fim, os modelos de classificação possibilitaram discriminar os diferentes estádios de amadurecimento em uvas viníferas, com uso da espectroscopia VIS-NIR. A faixa espectral que corresponde ao visível teve maior contribuição, considerando a mudança de cor que as bagas sofrem no processo de amadurecimento.

## CONCLUSÕES

A espectroscopia de refletância VIS-NIR foi uma ferramenta poderosa para avaliação não destrutiva de uvas intactas da variedade Syrah e Cabernet Sauvignon utilizando PCR e PLSR para predição de sólidos solúveis totais e antocianinas totais ( $R^2 \geq 0,90$ ).

As assinaturas espectrais dos atributos de qualidade sólidos solúveis totais, antocianinas totais e flavonoides amarelos, associada aos modelos de MLR, se mostraram adequadas, considerando o  $R^2$  de 0,97, 0,97 e 0,70, respectivamente.

O desenvolvimento dos modelos de classificação possibilitou a discriminação dos diferentes estádios de amadurecimento em videira da variedade Syrah e Cabernet Sauvignon com precisão de 93,15% com o PLS-DA, tendo a faixa espectral do visível maior contribuição.

## REFERÊNCIAS

AGATI, G.; TRAVERSI, M.L.; CEROVIC, Z.G. Chlorophyll fluorescence imaging for the noninvasive assessment of anthocyanins in whole grape (*Vitis Vinifera* L.) bunches. **Photochem. Photobiol.**, 84, 1431–1434, 2008.

AGATI, G.; MEYER, S.; MATTEINI, P.; CEROVIC, Z.G. Assessment of anthocyanins in grape (*Vitis vinifera* L) berries using a noninvasive chlorophyll fluorescence method. **J. Agric. Food Chem.**, 55, 1053–1061, 2007.

AGATI, G.; PINELLI, P.; EBNER, S.C.; ROMANI, A.; CARTELAT, A.; CEROVIC, Z.G. Nondestructive evaluation of anthocyanins in olive (*Olea europaea*) fruits by in situ chlorophyll fluorescence spectroscopy. **J. Agric. Food Chem.**, 53, 1354–1363, 2005.

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. M.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorol. Z.**, 22, 711–728, 2013.

ARAÚJO, E. G. **Métodos de amostragem e tamanho de amostra para avaliar o estado de maturação da uva Niágara Rosada**. 63p. Dissertação (Mestrado em Agronomia) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2017.

BARNES, R. J.; DHANOA, M. S.; LEISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, **Appl. Spectrosc.**, 43, 772–777, 1989.

CAO, F.; WU, D.; HE, Y. Soluble solids content and pH prediction and varieties discrimination of grapes based on visible-near infrared spectroscopy. **Comput. Electron. Agric.**, 71, 2010.

CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. **Trends Food Sci. Technol.**, 18, 72–83, 2007.

CHEN, S.; ZHANG, F.; NING, J.; LIU, X.; ZHANG, Z.; YANG, S. Predicting the anthocyanin content of wine grapes by NIR hyperspectral imaging. **Food Chem.**, 172, 788–793, 2015.

COOMBE, B. G. Adoption of a system for identifying grapevine growth stages. **Australian Journal of Grape and Wine Research**, 1, 100–110, 1995.

FAVARO, M. M. A. **Extração, estabilidade e quantificação de antocianinas de frutas típicas brasileiras para aplicação como corante**. 102p. Dissertação (Mestrado em Química) – Instituto de Química, Universidade Estadual de Campinas, Campinas, 2008.

FERNÁNDEZ-NOVALES, J.; LÓPEZ, M.I.; SÁNCHEZ, M.T.; GARCÍA-MESA, J.A.; GONZÁLEZ-CABALLERO, V. Assessment of quality parameters in grapes during ripening using a miniature fiber-optic near-infrared spectrometer. **Int. J. Food Sci. Nutr.**, 60, 265–277, 2009a.

FERNÁNDEZ-NOVALES, J.; LÓPEZ, M.I.; SÁNCHEZ, M.T.; MORALES, J.; GONZÁLEZ-CABALLERO, V. Shortwave-near infrared spectroscopy for determination of reducing sugar content during grape ripening, winemaking, and aging of white and red wines. **Food Res. Int.**, 42, 285–291, 2009b.

FERREIRA, M. M. C. **Quimiometria – conceitos, métodos e aplicações**. SP, Brazil: Editora da Unicamp, 2015.

FERRER-GALLEGO, R.; HERNÁNDEZ-HIERRO, J.M.; RIVAS-GONZALO, J.C.; ESCRIBANO-BAILÓN, M.T. Determination of phenolic compounds of grape skins during ripening by NIR spectroscopy. **LWT - Food Sci. Technol.**, 44, 847–853, 2011.

FERRER-GALLEGO, R.; HERNÁNDEZ-HIERRO, J.M.; RIVAS-GONZALO, J.C.; ESCRIBANO-BAILÓN, M.T. Influence of climatic conditions on the phenolic composition of *Vitis vinifera* L. cv. Graciano. **Anal. Chim. Acta**, 732, 73–77, 2012.

FRANCIS, F. J. Analysis of Anthocyanins. In: MARKAIS, P. **Anthocyanins as food colors**. New York: Academic Press. p. 181-207, 1982.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: A tutorial. **Anal. Chim. Acta**, 185, 1–17, 1986.

GÓMEZ, A.H.; HE, Y.; PEREIRA, A.G. Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. **J. Food Eng.**, 77, 313–319, 2006.

GONZÁLEZ-CABALLERO, V.; SÁNCHEZ, M.T.; FERNÁNDEZ-NOVALES, J.; LÓPEZ, M.I.; PÉREZ-MARÍN, D. On-Vine Monitoring of Grape Ripening Using Near-Infrared Spectroscopy. **Food Anal. Methods**, 5, 1377–1385, 2012.

GONZÁLEZ-CABALLERO, V.; SÁNCHEZ, M.T.; LÓPEZ, M.I.; PÉREZ-MARÍN, D. First steps towards the development of a non-destructive technique for the quality control of wine grapes during on-vine ripening and on arrival at the winery. **J. Food Eng.**, 101, 158–165, 2010.

GORRY, P. A. General least-square smoothing and differentiation by the convolution (Savitzky- Golay) method. **Anal. Chem.**, 62, 570–573, 1990.

HAGEN, S.F.; SOLHAUG, K.A.; BENGTSSON, G.B.; BERGE, G.I.A.; BILGER, W. Chlorophyll fluorescence as a tool for non-destructive estimation of anthocyanins and total flavonoids in apples. **Postharvest Biol. Technol.**, 41, 156–163, 2006.

HERRERA, J.; GUESALAGA, A.; AGOSIN, E. Shortwave near infrared spectroscopy for non-destructive determination of maturity of wine grapes. **Meas. Sci. Technol.**, 14, 689–697, 2003.

ISAKSSON, T.; NAES, T. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy, **Appl. Spectrosc.**, 42, 1273–1284, 1988.

KEMPS, B.; LEON, L.; BEST, S.; DE BAERDEMAEKER, J.; DE KETELAERE, B. Assessment of the quality parameters in grapes using VIS/NIR spectroscopy. **Biosyst. Eng.**, 105, 507–513, 2010.

KENNEDY, B.Y.J. Understanding grape berry development. **Pract. Winer. Vineyard**, 1–5, 2002.

LEIVA-VALENZUELA, G.A.; LU, R.; AGUILERA, J.M. Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging. **J. Food Eng.**, 115, 91–98, 2013.

MARTENS, H.; MARTENS, M. Modified jack-knife estimation of parameter uncertainty in bilinear modeling by partial least squares regression (PLSR). **Food Qual. Preference**, 11, 5–16, 2000.

MARTÍNEZ-SANDOVAL, J.R.; NOGALES-BUENO, J.; RODRÍGUEZ-PULIDO, F.J.; HERNÁNDEZ-HIERRO, J.M.; SEGOVIA-QUINTERO, M.A.; MARTÍNEZ-ROSAS, M.E.; HEREDIA, F.J. Screening of anthocyanins in single red grapes using a non-destructive method based on the near infrared hyperspectral technology and chemometrics. **J. Sci. Food Agric.**, 96, 1643–1647, 2016.

MUSINGARABWI, D.M.; NIEUWOUDT, H.H.; YOUNG, P.R.; EYÉGHÈ-BICKONG, H.A.; VIVIER, M.A. A rapid qualitative and quantitative evaluation of grape berries at various stages of development using Fourier-transform infrared spectroscopy and multivariate data analysis. **Food Chem.**, 190, 253–262, 2016.

NOGALES-BUENO, J.; BACA-BOCANEGRA, B.; RODRÍGUEZ-PULIDO, F.J.; HEREDIA, F.J.; HERNÁNDEZ-HIERRO, J.M. Use of near infrared hyperspectral tools for the screening of extractable polyphenols in red grape skins. **Food Chem.**, 172, 559–564, 2015.

NOGALES-BUENO, J.; HERNÁNDEZ-HIERRO, J.M.; RODRÍGUEZ-PULIDO, F.J.; HEREDIA, F.J. Determination of technological maturity of grapes and total phenolic compounds of grape skins in red and white cultivars during ripening by near infrared hyperspectral image: A preliminary approach. **Food Chem.**, 152, 586–591, 2014.

PADILHA, C.V. DA S.; CAMARÃO TELLES BIASOTO, A.; CORRÊA, L.C.; DOS SANTOS LIMA, M.; PEREIRA, G. E. Phenolic compounds profile and antioxidant activity of commercial tropical red wines (*Vitis vinifera* L.) from São Francisco Valley, Brazil. **J. Food Biochem.**, 41, 1–9, 2017.

RAMOS, V.; RAMALHO, P.; VIVAS, C.; SOUSA, A. Strategy and innovation in order to be competitive in a globalized world: an evolutionary analysis of the emerging wine sector pole of Vale do São Francisco (Brazil). **Rev. Ciênc. Empres.**, UNIPAR, Umuarama, 16, 2, 157–167, 2015.

RUIZ-ALTISENTA, M.; RUIZ-GARCIA, L.; MOREDA, G. P.; RENFU, LU.; HERNANDEZ-SANCHEZ, N.; CORREA, E. C.; DIEZMA, B.; NICOLAI, B.; GARCÍA-

RAMOS, J. Sensors for product characterization and quality of specialty crops-A review. **Comput. Electron. Agric.**, 74, 176-194, 2010.

RUNGPICHAYAPICHET, P.; MAHAYOTHEE, B.; NAGLE, M.; KHUWIJITJARU, P.; MÜLLER, J. Robust NIRS models for non-destructive prediction of postharvest fruit ripeness and quality in mango. **Postharvest Biol. Technol.**, 111, 31–40, 2016.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures, **Anal. Chem.**, 36, 1627–1639, 1964.

SHENK, J. S.; WESTERHAUS, M. O.; ABRAMS, S. M. Protocol for NIR calibrations: Monitoring analysis results and recalibration. In G. C. Martens, J. S. Shenk, & F. E. Barton, II (Eds.), **Near infrared spectroscopy (NIRS): Analysis of forage quality**. Agriculture Handbook, No. 643 (pp. 104–110). Washington, DC: USDA-ARS, US Government Printing Office, 1989.

SHENK, J. S.; WESTERHAUS, M. O. **Routine operation, calibration, development and network system management manual**, NIRSystems. MD, USA: Silver Spring, 1995.

SHENK, J. S.; WESTERHAUS, M. O. Calibration the ISI way. In A. M. C. Davies & P. C. Williams (Eds.), **Near infrared spectroscopy: The future waves** (pp. 198–202). Chichester: NIR Publications, 1996.

SIRISOMBOON, P.; TANAKA, M.; KOJIMA, T.; WILLIAMS, P. Nondestructive estimation of maturity and textural properties on tomato “Momotaro” by near infrared spectroscopy. **J. Food Eng.**, 112, 218–226, 2012.

TEIXEIRA, A. H. C.; SCHERER-WARREN, M.; HERNANDEZ, F. B. T.; ANDRADE, R. G.; LEIVAS, J. F. Large-scale water productivity assessments with MODIS images in a changing semi-arid environment: A Brazilian case study. **Remote Sensing**, 5, 5783–5804, 2013.

WILLIAMS, P. C. Implementation of near-infrared technology. In P. Williams & K. Norris (Eds.), **Near-infrared technology in the agricultural and food industries** (2nd ed., pp. 145–169). St. Paul, MN: American Association of Cereal Chemists, 2001.

WILLIAMS, P. C.; SOBERING, D. C. How do we do it: A brief summary of the methods we use in developing near infrared calibrations. In A. M. C. Davies & P. Williams (Eds.), **Near infrared spectroscopy: The future waves** (pp. 185–188). Chichester: NIR Publications, 1996.

WOLD, S.; RUHE, A.; WOLD, H. The collinearity problem in linear regression. The partial least square (PLS) approach to generalized inverses. **SIAM Journal on Scientific and Statistical Computing**, 5(3), 735–743, 1984.

WU, D.; HE, Y.; NIE, P.; CAO, F.; BAO, Y. Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice. **Anal. Chim. Acta**, 659, 229–237, 2010.

ZHANG, G. P. Neural networks for classification: a survey. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)** 30, 4, 451–462, 2000.

### 3.2. TÉCNICAS DE PROCESSAMENTO DIGITAL DE SINAIS E APRENDIZAGEM DE MÁQUINA COM ESPECTROSCOPIA DE REFLETÂNCIA VIS-NIR EM UVAS VINÍFERAS

**RESUMO:** A espectroscopia VIS-NIR tem se destacado na determinação não destrutiva de atributos de qualidade em frutas. Os sinais espectrais fornecem informações físico-químicas de diferentes compostos da amostra. Contudo, tais sinais apresentam não apenas as informações analíticas pertencentes à amostra, mas também informações de fundo e ruído, tornando-se crucial eliminá-las ou atenuar sua influência no sinal espectral. Isto posto, o objetivo principal deste estudo foi avaliar o desempenho de técnicas de processamento digital de sinais e aprendizagem de máquina na espectroscopia VIS-NIR para a determinação dos atributos de qualidade e discriminação dos estádios de amadurecimento em uvas viníferas. Foram coletadas um total de 1008 amostras da variedade Syrah e Cabernet Sauvignon e feita a aquisição dos espectros de refletância utilizando a faixa espectral de 450 a 1800 nm. Os sinais espectrais foram pré-processados utilizando diferentes técnicas de alisamentos (filtros da mediana e média móvel, transformadas de Fourier e *wavelet*), além das diferenciações e normalizações. Foi realizada a determinação dos atributos de qualidade sólidos solúveis totais (SST), antocianinas totais e flavonoides amarelos com os métodos de referência. As predições dos atributos de qualidade foram realizadas através da regressão dos quadrados mínimos parciais (PLSR) e máquinas de vetores de suporte (SVM). Os estádios de amadurecimento foram discriminados utilizando os classificadores análise de discriminante por quadrados mínimos parciais (PLS-DA) e SVM. A regressão SVM associada ao alisamento com transformada *wavelet* apresentou os melhores desempenhos para predição de SST ( $R^2 = 0,96$  e RMSEP = 1,21 °Brix) antocianinas totais ( $R^2 = 0,97$  e RMSEP = 13,52 mg.100 g<sup>-1</sup>) e flavonoides amarelos ( $R^2 = 0,78$  e RMSEP = 17,18 mg.100 g<sup>-1</sup>). A classificação SVM associada a transformada *wavelet* proporcionou uma precisão de 0,95 e taxa de falso positivo de 0,04. Portanto, os modelos de regressão e classificação utilizando a técnica de SVM associada a transformada *wavelet* na espectroscopia VIS-NIR apresentou o maior desempenho na determinação dos atributos de qualidade (SST, antocianinas totais e flavonoides amarelos) e discriminação dos estádios de amadurecimento.

**PALAVRAS-CHAVES:** Cabernet Sauvignon, *machine learning*, Syrah, transformada *wavelet*.

## INTRODUÇÃO

A espectroscopia VIS-NIR tem se destacado na determinação dos atributos de qualidade de alimentos por ser uma técnica óptica não destrutiva, proporcionando a caracterização de diversos compostos em uma mesma amostra de forma rápida (HUANG *et al.*, 2008). Esta técnica tem sido empregada na avaliação dos atributos de qualidade em frutas, como exemplo, em pera (LI *et al.*, 2013), oliva (GARCÍA MARTIN, 2015), tangerinas (ANTONUCCI *et al.*, 2011) dentre outras.

A espectroscopia de refletância VIS-NIR utiliza a faixa espectral do visível (VIS – 380 a 780 nm) e infravermelho (IR – 780 a 2500 nm), irradiando o objeto da análise, e medindo-se a radiação refletida (CEN e HE, 2007; NICOLAÏ *et al.*, 2007b), fornecendo informações do comportamento vibracional das ligações químicas de grupos funcionais C–H, N–H, O–H e S–H (BLANCO e VILLARROYA, 2002).

O espectro VIS-NIR é dominado por picos de bandas de absorção e, portanto, é essencial o uso da quimiometria para extrair informações químicas e físicas dos dados espectrais (ARENDSE *et al.*, 2018). Para isso, os métodos dos Quadrados Mínimos Parciais (PLS) e das Máquinas de Vetor de Suporte (SVM) são utilizados para modelar o relacionamento do sinal espectral com os atributos químicos e físicos de interesse, sendo o primeiro baseado na aprendizagem estatística e o segundo na aprendizagem de máquina (PÉREZ-ENCISO *et al.*, 2003; LU *et al.*, 2011).

Contudo, os espectrômetros podem variar na resposta fotométrica, e conseqüentemente, na calibração do equipamento. Isto ocorre devido a mudanças ao longo do tempo na reposta instrumental (envelhecimento de fonte e luz, substituição de algumas peças, etc.), mudanças nas condições ambientais (temperatura, umidade e radiação incidente) e mudanças nas condições de medição (caminho óptico e ângulo) (FEARN, 2001; RANDOLPH, 2006).

Assim, os dados espectrais contêm não apenas as informações analíticas pertencentes à amostra, mas também informações de fundo e ruído, tornando-se crucial eliminá-las ou atenuar sua influência no sinal espectral (MOGHIMI *et al.*, 2010). A remoção dessas informações indesejáveis aumenta a relação sinal/ruído do espectro melhorando o desempenho dos modelos preditivos (NICOLAÏ *et al.*, 2007a). Também, favorece a transferência de calibração, padronizando espectros de equipamentos diferentes, pois elimina a necessidade de realização de novas medições destrutivas (referência) e não destrutivas (espectrais), otimizando o processo de calibração para outros espectrômetros (GREENSILL *et al.*, 2001).

A remoção das informações de fundo e ruído e, por conseguinte, a padronização dos sinais espectrais, é realizada por meio de normalizações e alisamentos dos espectros. As normalizações frequentemente utilizadas são a padronização normal de sinal (SNV) (BARNES *et al.*, 1989) e a correção multiplicativa de espalhamento (MSC) (ISAKSSON e NAES, 1988). Além disso, são aplicados alisamentos, sendo os mais comumente utilizados os filtros de média móvel e mediana, como também o filtro *Savitzky-Golay* (SAVITZKY e GOLAY, 1964).

Diversos trabalhos relatam a aplicação da padronização dos sinais espectrais VIS-NIR convencionais (filtros média móvel, mediana e *Savitzky-Golay*) e modelagem PLS com as propriedades físico-químicas para fins de análise de qualidade de frutas, principalmente em uvas (DAMBERGS *et al.*, 2003; HERRERA *et al.*, 2003; KEMPS *et al.*, 2010; GALLEGO *et al.*, 2011; BAIANO *et al.*, 2012). Não obstante, existem outros métodos de padronização com eficiente remoção dos ruídos espectrais e melhora no desempenho dos modelos de predição, dentre eles destacando-se a transformada *wavelet* (GREENSILL *et al.*, 2001; ZHU *et al.*, 2007). Nicolai *et al.* (2007a) sugere a combinação, convenientemente, da transformada *wavelet* com as SVM para a predição do conteúdo de açúcares em maçãs utilizando a espectroscopia NIR.

O objetivo deste estudo foi avaliar o desempenho de técnicas de processamento digital de sinais e aprendizagem de máquina na espectroscopia VIS-NIR para a determinação dos atributos de qualidade e discriminação dos estádios de amadurecimento em uvas viníferas da variedade Syrah e Cabernet Sauvignon. Diversas técnicas de pré-processamento foram avaliadas e modelos de predição com base em atributos de qualidade foram aplicados, comparando-se os sinais com e sem a aplicação dos pré-processamentos. Além disso, foram aplicados modelos de classificação de dados para avaliação dos estádios de maturação dos frutos.

## **MATERIAIS E MÉTODOS**

### **Amostragem das bagas e aquisição dos espectros**

As amostras de uva (*Vitis vinifera* L.), das variedades Syrah e Cabernet Sauvignon, foram coletadas de um vinhedo localizado na região do Vale do Submédio São Francisco (Lagoa Grande, Pernambuco, Brasil: -9,05363; -40,19868). Foram coletadas um total de 432 bagas da variedade Syrah e 576 bagas da variedade Cabernet Sauvignon. Houve diferentes números de amostras para cada variedade devido à precocidade da Syrah nas condições ambientais locais.

Foram colhidas 72 bagas semanalmente entre os meses de maio e agosto de 2017, abrangendo diversos estádio, desde antes do *véraison* até o estádio de colheita. Para obter amostras representativas, foi adotado o método de amostragem estratificado (ARAÚJO, 2007). As bagas individuais foram acondicionadas em embalagem impermeável e armazenadas em gelo para transporte, imediato processamento e aquisição dos espectros. Antes de cada medição, foi necessário aguardar que as amostras se estabilizassem à temperatura ambiente de 25 °C.

O sistema de aquisição dos espectros de refletância foi constituído pelos seguintes componentes: espectrorradiômetro FieldSpec® 3 (*Analytical Spectral Devices*, Boulder, Colorado, USA), com sensor de fibra óptica com campo de visão de 8°, operando na faixa de comprimento de onda de 350 nm a 2500 nm, com resolução de 3 a 10 nm e tempo por varredura de 100 ms; fonte de luz de quartzo-tungstênio-halogênio de 70 W; câmara escura com dimensões de 100 x 50 x 50 cm; e computador com o *software* RS3 (*Analytical Spectral Devices*, Boulder, Colorado, USA). Ademais, a fibra óptica foi posicionada perpendicularmente ao plano de medição com altura de 7,0 cm, com as bagas sendo localizadas no centro do referido plano.

Uma placa cerâmica Spectralon® (*Labsphere Inc.*, North Sutton, NH, EUA) foi usada como referência branca, com a qual operações de calibração (aquisição de referência do branco padrão) foram realizadas antes da tomada dos dados. Cada espectro foi constituído pela média de trinta varreduras feitas pelo sensor, para cada uma das bagas. Os valores de refletância foram transformados em absorbância  $\log \frac{1}{R}$  com o *software* ViewSpec Pro (*Analytical Spectral Devices*, Boulder, Colorado, EUA). As regiões compreendidas entre as faixas de 350 nm a 449 nm e 1801 nm a 2500 nm foram removidas da análise por apresentarem ruído aleatório excessivo.

### **Determinação dos atributos de qualidade das amostras de uva**

Após as medições de refletância, as bagas individuais foram submetidas às medições de referência, a saber sólidos solúveis totais (SST), antocianinas totais (Anth) e flavonoides amarelos (Flav).

A determinação dos sólidos solúveis totais foi realizada a partir da leitura direta em refratômetro digital (HI 96804, *Hanna Instruments*, USA), com faixa de medição entre 0 e 85 °Brix e precisão de  $\pm 0,2$  °Brix, utilizando-se duas gotas da amostra triturada e homogeneizada, com os resultados expressos em °Brix.

Os teores de antocianinas totais e flavonoides amarelos da casca foram determinados segundo Francis (1982), sendo este método de pH único não havendo diferença com o método AOAC (FAVARO, 2008). Pesou-se 0,5 g de casca e polpa, em seguida, adicionou-se 25 mL da solução extratora de etanol (95%) acidificado com HCl (1,5 N), na proporção 85:15. As amostras foram maceradas por um minuto e, posteriormente, transferidas para recipientes ao abrigo da luz, ficando em repouso e armazenados na geladeira. Após 24 horas, o sobrenadante era recolhido para leitura em espectrofotômetro (Spectronic BioMate 5 UV-Vis, *Thermo Electron*, UK) a 535 nm, para antocianinas totais, e 374 nm, para flavonoides amarelos. Os resultados foram expressos em  $\text{mg} \cdot 100 \text{ g}^{-1}$ , por meio das equações (1) e (2):

$$\text{Anth} = \text{abs} * \text{fator de diluição} / E1\% \quad (1)$$

$$\text{Flav} = \text{abs} * \text{fator de diluição} / E2\% \quad (2)$$

em que o fator de diluição é 5000, o coeficiente de extinção para antocianinas totais (E1%) é 98,2 e o coeficiente de extinção para flavonoides amarelos (E2%) é 76,6.

### **Pré-processamento dos dados espectrais**

O pré-processamento dos dados espectrais e o desenvolvimento dos modelos de regressão e classificação foram realizados no *software* MATLAB (*MathWorks Inc.*, Massachusetts, EUA) com o auxílio das ferramentas Wavelet Toolbox, Signal Processing Toolbox e PLS\_Toolbox (*Eigenvector Research Inc.*, Washington, EUA). Antes do desenvolvimento dos modelos matemáticos, os dados espectrais foram submetidos a uma combinação de diferentes pré-processamentos. Utilizou-se quatro técnicas de alisamento, separadamente, sendo eles os filtros da mediana (MF) e média móvel (MAF), com diferentes tamanhos de segmentos (5, 9, 13, 17, 21 e 27 pontos), além das transformadas de Fourier, com filtragem passa-baixa (LPF), e transformada *wavelet*, com três funções-mães distintas (Daubechies - *db*, Symlets - *sym* e Coiflets - *coif*) para remoção de ruídos. Sobre os dados alisados, utilizou-se a primeira derivada ( $D^1$ ) e a segunda derivada ( $D^2$ ), calculadas pelo método de *Savitzky-Golay* (SAVITZKY e GOLAY, 1964; GORRY, 1990), com polinômio de segunda ordem e dois diferentes tamanhos de segmento (7 e 21 pontos). Por fim, sobre os dados alisados e de posse das derivadas, utilizou-se a correção multiplicativa de espalhamento (MSC) (ISAKSSON e NAES, 1988) e a padronização normal de sinal (SNV) (BARNES *et al.*, 1989).

Observou-se que, com o alisamento, houve a deterioração de faixas do sinal que não continham ruídos. Assim, optou-se por aplicar o alisamento pelas transformadas de Fourier e

*wavelet* não somente no espectro inteiro, mas também em faixas do sinal que continham os ruídos. Logo, as faixas selecionadas foram 450-600 nm e 1400-1800 nm.

#### Alisamento com transformada de Fourier

O alisamento com transformada de Fourier foi realizado utilizando a ferramenta Signal Processing Toolbox. A transformada discreta de Fourier do sinal espectral foi usando como entrada uma matriz  $X$  de comprimentos de onda por absorbância. Assim, através da transformada, foi possível determinar uma frequência de corte adequada ao sinal no domínio da frequência. A transformada foi calculada por meio da equação (3):

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-i\omega n} \quad (3)$$

em que  $x(n)$  é o sinal discreto de entrada (amostras dos valores de absorbância para diferentes comprimentos de onda) e  $\omega$  é a frequência angular discreta.

Foi utilizada a classe de filtros de Resposta ao Impulso Finita (FIR, do inglês, *Finite Impulse Response*), cuja função de transferência é mostrada na equação (4):

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{i=0}^M b_i z^{-i} \quad (4)$$

em que  $M$  é a ordem do filtro,  $b$  são os coeficientes do numerador da função de transferência e  $X(z) = 1$ , indicando que o filtro contém apenas zeros.

A função de transferência do filtro foi calculada através dos parâmetros  $n$ , que é a ordem do filtro (10, 16 e 20), do tipo de filtro, sendo usado um passa-baixas, e  $Wn$ , que é a frequência discreta normalizada entre 0 e 1, calculada por meio da equação (5):

$$Wn = \frac{2f_{corte}}{F_s} \quad (5)$$

em que  $f_{corte}$  é a frequência de corte desejada (200 e 1000 Hz, neste trabalho) e  $F_s$  é a frequência de amostragem, definida como o inverso do tempo de varredura do espectrorradiômetro.

Por fim, o alisamento foi realizado através da equação de diferenças no domínio espectral, definida pelos parâmetros  $b$ ,  $a$  e  $x$ , em que  $b$  e  $a$  correspondem aos coeficientes do numerador e denominador do filtro, respectivamente. O parâmetro  $x$  corresponde aos dados de entrada, sendo estes as absorbâncias das amostras em cada comprimento de onda. A saída  $y$  alisada foi calculada através da equação (6):

$$y(k) = \sum_{i=0}^M b_i x(k-i) - \sum_{j=1}^N a_j y(k-j) \quad (6)$$

em que  $x(n)$  é o sinal de entrada,  $k$  indica a amostra corrente,  $M$  e  $N$  representam as ordens no numerador e denominador, respectivamente, da função de transferência do filtro, e  $b$  e  $a$  representam os coeficientes no numerador e denominador, respectivamente, da função de transferência do filtro, no domínio  $z$ .

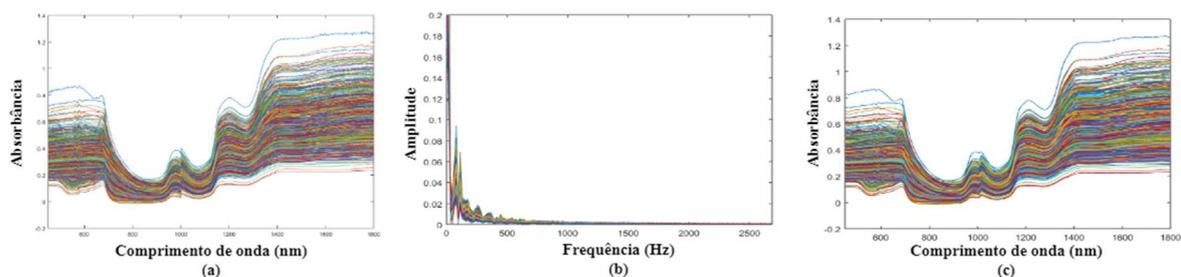
Entretanto, devido ao efeito de cauda, tornou-se necessário expandir temporalmente o sinal, através da repetição dos  $n$  primeiros pontos, em que  $n$  é a ordem do filtro. Após a filtragem, os  $n$  pontos adicionados foram removidos e a transformada de Fourier inversa foi calculada, de forma a obter o sinal filtrado no domínio espectral, por meio da equação (7):

$$X(j) = \frac{1}{n} \sum_{k=1}^n Y(k) Q_n^{-(j-1)(k-1)} \quad (7)$$

onde  $Q_n = e^{\frac{-2\pi i}{n}}$ ,  $n$  é o comprimento do sinal,  $Y(k)$  é a transformada do sinal e  $j$  é uma amostra.

A Figura 3.2-1 mostra o sinal espectral do conjunto de dados, sua representação no domínio da frequência, e a saída alisada por um filtro passa-baixa de ordem 20 e frequência de corte de 1 kHz.

Figura 3.2-1. Alisamento utilizando a Transformada de Fourier para remoção de ruídos de alta frequência realizada no espectro inteiro. Sinal espectral bruto (a). Representação do sinal bruto no domínio da frequência (b). Sinal após a alisamento (c).



#### Alisamento com transformada *wavelet*

O alisamento com transformada *wavelet* foi realizado utilizando a Wavelet Toolbox, de forma a implementar a remoção dos ruídos de alta frequência pela limiarização dos coeficientes do sinal. Os limiares foram obtidos pelos métodos global e adaptativo, sendo estes aplicados sobre o sinal espectral por limiarização suave ou brusca. As funções-mãe escolhidas para a

decomposição do sinal foram Daubechies (*db*) e Symlets (*sym*), por já apresentarem bons resultados em estudos anteriores (LIU *et al.*, 2015; SHAO *et al.*, 2007), e a Coiflets (*coif*).

Inicialmente, o sinal foi decomposto usando o algoritmo à *Trous* (Mallat, 1998), através de filtragens passa-baixas e passa-altas simultâneas, resultando nos coeficientes de aproximação e detalhes, respectivamente, conforme equações (8) e (9):

$$y_a(n) = \sum_{k=-\infty}^{\infty} x(k)g(n-k) \quad (8)$$

$$y_d(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k) \quad (9)$$

onde  $x(n)$  é o sinal de entrada,  $g(n)$  é a resposta ao impulso do filtro passa-baixa, e  $h(n)$  é a resposta ao impulso do filtro passa-altas.

As saídas foram então subamostradas  $m$  vezes, em que  $m$  é a quantidade de níveis de decomposição, sendo que, para cada subamostragem, a saída foi filtrada por novos filtros passa-baixas e passa-altas. Os coeficientes dos filtros foram dados pelos coeficientes *wavelet* calculados a partir da equação (10):

$$X_{a,b} = \int_{-\infty}^{\infty} x(k) \frac{1}{\sqrt{2^j}} \Psi\left(\frac{k-b2^a}{2^a}\right) dk \quad (10)$$

onde  $a$  e  $b$  são parâmetros de escala e deslocamento, respectivamente,  $\Psi(t)$  é a função-mãe, e  $x(k)$  é o sinal de entrada.

A decomposição do sinal foi obtida através dos parâmetros *thrtype*,  $X$ , *wname*, *THR*, *SORH*, *KEEPAPP*, em que  $X$  corresponde à matriz de absorvância por comprimentos de onda, *wname* é o tipo de função-mãe e  $N$  é o número de níveis de decomposição. Os parâmetros *thrtype*, *THR*, *SORH* e *KEEPAPP* referem-se ao alisamento do sinal e remoção de ruído do mesmo, onde calcula-se o limiar a ser aplicado nos coeficientes *wavelet*.

O limiar global, definido por Donoho e Johnstone (1994), foi obtido através da equação (11), sendo este aplicado sobre o sinal em todos os níveis de decomposição, variando de um a doze níveis:

$$\lambda = \sigma\sqrt{2 \ln N} \quad (11)$$

em que  $N$  corresponde ao comprimento do sinal e  $\sigma$  à variância do ruído.

Empregou-se os parâmetros *INI*, *IN2*,  $X$  para a obtenção do limiar global, onde *INI* corresponde ao tipo de transformada *wavelet* (normal ou pacote), *IN2* é o tipo de operação (alisamento ou compressão) e  $X$  é a matriz de absorvância por comprimento de onda.

Na abordagem de limiar adaptativo, definida por Birgé-Massart (1997), um limiar foi determinado e aplicado para cada nível  $i$ . Para tal, os  $n_i$  maiores coeficientes *wavelet* foram selecionados, sendo esta quantidade determinada através da equação (12):

$$n_i = \frac{M}{(j + 2 - i)^\alpha} \quad (12)$$

em que  $M$  é a quantidade de coeficientes do primeiro nível,  $j$  é o número de níveis, e  $\alpha$  é um parâmetro empírico.

Dados os  $n_i$  maiores valores, o limiar obtido foi pela equação (13):

$$\lambda = |c(t)| \quad (13)$$

em que  $|c(t)|$  é o vetor de valores absolutos dos coeficientes *wavelet* ordenados de forma decrescente e  $t$  assume um valor a minimizar a equação (14):

$$crit(t) = - \sum_{i=0}^N c(i) + 2\sigma^2 t(\alpha + \log(n/t)) \quad (14)$$

em que  $N$  é o número de coeficientes *wavelet* e  $\sigma$  é a variância do ruído.

O limiar adaptativo foi obtido através dos parâmetros  $C$ ,  $L$ ,  $\alpha$ ,  $M$ , em que  $C$  é o vetor de decomposição do sinal,  $L$  é um vetor de número de coeficientes por nível,  $\alpha$  e  $M$  correspondem a valores padrão do limiar adaptativo, sendo iguais a 3 e ao primeiro valor de  $L$ , respectivamente.

Após a obtenção dos limiares pelos métodos global e adaptativo, emprega-se a limiarização suave e brusca para tratamento dos coeficientes através das equações (15) e (16):

$$Y = u(X)(|X| - \lambda) \quad (15)$$

$$Y = \begin{cases} X, & \text{se } X \geq \lambda \\ 0, & \text{se } X \leq \lambda \end{cases} \quad (16)$$

em que  $X$  corresponde a um determinado coeficiente *wavelet*,  $Y$  é o novo valor deste coeficiente, e  $u(X)$  corresponde à função degrau aplicada a um coeficiente  $X$ .

As limiarizações suave ou brusca foram aplicadas após a decomposição do sinal e obtenção dos limiares (global ou adaptativo). Para tal, os parâmetros *thrtype* e *THR* corresponderam ao tipo de limiar e seu(s) valor(es), respectivamente, *SORH* indica o tipo de limiarização e *KEEPAPP* determina a permanência dos coeficientes de aproximação.

Após a decomposição (filtros passa-baixas e passa-altas), obtenção dos limiares (métodos global e adaptativo) e limiarização (suave ou brusca), os doze sinais alisados resultantes foram reconstruídos por superamostragem com o emprego de filtros inversos, conforme as equações (17) e (18):

$$g' = (g'_1, g'_2 \dots g'_N) = (g_N, g_{N-1} \dots g_1) \quad (17)$$

$$h' = (h'_1, h'_2 \dots h'_N) = (g_1, -g_2 \dots (-1)^{n+1}g_n \dots (-1)^{N+1}g_N) \quad (18)$$

em que  $g'_1, g'_2 \dots g'_N$  e  $h'_1, h'_2 \dots h'_N$  são os coeficientes dos filtros passa-baixas e passa-altas inversos, respectivamente, e  $g_N, g_{N-1} \dots g_1$  são os coeficientes do filtro passa-baixas.

### Desenvolvimento dos modelos de regressão

Os modelos de regressão foram construídos para a predição de SST, antocianinas totais e flavonoides amarelos do conjunto de dados contendo as variedades Syrah e Cabernet Sauvignon. O número de amostras segue a recomendação de Ferreira (2015), onde o conjunto de calibração deve ser no mínimo igual a  $6(A+1)$ , sendo  $A$  o número de fatores do modelo. Os conjuntos de dados foram separados na proporção 2/3-1/3, ou seja, 2/3 dos dados foram usados para o conjunto de calibração e 1/3 para o conjunto de predição, resultando em 672 amostras para calibração e 336 amostras para predição.

A construção dos modelos foi realizada por meio da regressão dos quadrados mínimos parciais (PLSR) e regressão de máquina de vetor de suporte (SVM-R), em que a matriz de dados independentes foi dada pela absorbância de cada amostra nos diferentes comprimentos de onda, enquanto a matriz de dados dependentes consistiu nos valores de referência dos atributos de qualidade para cada amostra.

Para a construção do modelo PLSR, foi necessário definir o número de variáveis latentes por validação cruzada, de forma que o modelo construído apresentasse um baixo erro quadrático médio de validação cruzada (RMSECV) e alta variância explicada. Assim, foi escolhida uma quantidade tal que, caso fosse adicionada mais uma variável, não houvesse a diminuição no valor de RMSECV e aumento da variância em pelo menos 2%. (WISE *et al.*, 2006). O algoritmo PLS empregado para a regressão foi o NIPALS (GELADI e KOWALSKI, 1986), com 100 iterações e 20 variáveis latentes.

O tipo de SVM-R utilizada foi a *epsilon*-SVM-R, logo, foi preciso definir os parâmetros *epsilon* e custo ( $C$ ). Empregou-se a técnica *grid search* para obtenção dos valores ótimos de *epsilon* e  $C$ . Assim, foi escolhida a faixa de valores entre  $10^{-2}$  a  $10^2$  para *epsilon* e  $10^{-3}$  a  $10^4$  para  $C$ . Buscou-se um modelo que possuísse baixa complexidade e menor custo computacional na construção da função *epsilon-intensive loss*, através da equação (19):

$$\vec{w}x_i + b - y_i \leq \varepsilon \quad (19)$$

em que  $\vec{w}$  é o vetor normal à função,  $x_i$  é a amostra corrente,  $b$  é o viés, e  $y_i$  é o valor de referência para a amostra corrente.

Para cada combinação possível de *epsilon* e  $C$ , foram construídas funções *epsilon-intensive loss* em que erros situados a uma distância máxima igual à *epsilon* foram ignorados.

A determinação da melhor função foi realizada através da minimização da equação (20):

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (20)$$

sujeito a:

$$y_i - (wx_i - b) \leq \varepsilon + \xi_i \quad (21)$$

$$(wx_i) + b - y_i \leq \varepsilon + \xi_i^* \quad (22)$$

$$\xi_i, \xi_i^* \geq 0, i = 1 \dots N \quad (23)$$

em que  $\|w\|$  corresponde à norma do vetor  $w$ ,  $\xi_i$  e  $\xi_i^*$  são variáveis *slack* adicionadas pelo modelo,  $b$  é o viés e  $N$  é o número de amostras.

Os valores ótimos de *epsilon* e  $C$  foram então determinados por validação cruzada 10-*fold*, de forma que fosse selecionado o custo para o qual o RMSECV fosse mínimo. A função kernel empregada foi a linear, não havendo necessidade de mapear os dados para um espaço dimensional maior.

As métricas consideradas para avaliação dos modelos na calibração, validação cruzada e predição foram o coeficiente de determinação ( $R^2$ ), a raiz quadrada do erro médio quadrático (RMSE), o desvio preditivo residual (RPD) e os limites de controle SEP e bias, calculadas por meio das equações (24) a (28):

$$R^2 = \left( \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}) * (y_i - \bar{y})}{(n - 1) * \sigma_r * \sigma_p} \right)^2 \quad (24)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (25)$$

$$RPD = \frac{SD}{SECV} \quad (26)$$

$$\text{Limite controle SEP} = 1,30 * RMSEC \text{ (erro quadrático médio da calibração)} \quad (27)$$

$$\text{Limite controle bias} = \pm 0,60 * RMSEC \text{ (erro quadrático médio da calibração)} \quad (28)$$

onde  $n$  é o número de amostras para calibração, validação cruzada ou predição;  $\hat{y}_i$  é o valor previsto pelo modelo na calibração, validação cruzada ou predição;  $y_i$  é o valor de referência das amostras de validação cruzada ou predição;  $\bar{y}$  é a média dos valores de referência;  $\bar{\hat{y}}$  é a média dos valores previstos;  $\sigma_r$  é o desvio padrão dos valores de referência;  $\sigma_p$  é o desvio padrão dos valores previstos; SD é o desvio padrão; SECV é o erro padrão de validação cruzada.

### Desenvolvimento dos modelos de classificação

A classificação foi realizada para o conjunto de dados contendo as variedades Syrah e Cabernet Sauvignon, visando a discriminação das amostras quanto ao estágio de amadurecimento. A matriz de dados independentes consistiu nos valores de absorvância de cada amostra nos diferentes comprimentos de onda, submetidos aos pré-processamentos que proporcionaram os melhores modelos de regressão. A matriz de dados dependentes consistiu nas classes de amadurecimento (verde, *veráison* e madura), classificadas segundo Coombe (1995). As técnicas de classificação empregadas foram a máquinas de vetores de suporte (SVM-C) e a análise de discriminante por quadrados mínimos parciais (PLS-DA).

A PLS-DA foi utilizada com a mesma abordagem do método de regressão PLS. Entretanto as variáveis dependentes de entrada assumiram valores discretos (1, 0, -1), representando cada classe (estádio de amadurecimento). Contudo, como as saídas obtidas pelo modelo PLS não consistiam em números discretos como as classes especificadas, tornou-se necessário definir uma técnica de discriminação das classes com base nos valores contínuos obtidos. Adotou-se a teoria Bayesiana, visando minimizar o risco de ocorrência de erros de classificação (ZHANG, 2000). Dadas as funções de probabilidade  $p(y|M)$ ,  $p(y|V)$  e  $p(y|V')$  das classes verde, *veráison* e madura, respectivamente, onde  $y$  é o valor estimado da amostra pelo modelo PLS, obteve-se dois limiares tais que as probabilidades nas equações (29) e (30) fossem iguais.

$$p(y|M)P(M) = p(y|V)P(V) \quad (29)$$

$$p(y|V)P(V) = p(y|V')P(V') \quad (30)$$

As funções de probabilidade foram calculadas através das equações (31), (32) e (33) a seguir:

$$P(M) = \frac{n_M}{n_M + n_V + n_{V'}} \quad (31)$$

$$P(V) = \frac{n_V}{n_M + n_V + n_{V'}} \quad (32)$$

$$P(V') = \frac{n_{V'}}{n_M + n_V + n_{V'}} \quad (33)$$

em que  $n_M$  é a quantidade de amostras da classe M;  $n_V$  é a quantidade de amostras da classe V; e  $n_{V'}$  é a quantidade de amostras da classe V'. Os limiares foram então obtidos analiticamente através das equações (34) e (35):

$$(s_V^2 - s_M^2)^2 y^2 + (2m_V s_M^2 - 2m_M s_V^2) y + m_M^2 s_V^2 - m_V^2 s_M^2 - 2s_V^2 s_M^2 \ln\left(\frac{P(M) s_V}{P(V) s_M}\right) = 0 \quad (34)$$

$$(s_{V'}^2 - s_V^2)^2 y^2 + (2m_{V'} s_V^2 - 2m_V s_{V'}^2) y + m_V^2 s_{V'}^2 - m_{V'}^2 s_V^2 - 2s_{V'}^2 s_V^2 \ln\left(\frac{P(V) s_{V'}}{P(V') s_V}\right) = 0 \quad (35)$$

onde  $s_M$  é o desvio padrão dos valores previstos para amostras da classe M;  $s_V$  é o desvio padrão dos valores previstos para amostras da classe V;  $s_{V'}$  é o desvio padrão dos valores previstos para amostras da classe V';  $m_M$  é a média dos valores previstos para amostras da classe M;  $m_V$  é a média dos valores previstos para amostras da classe V e  $m_{V'}$  é a média dos valores previstos para amostras da classe V'.

A SVM-C foi utilizada com a mesma abordagem do método SVM-R, entretanto, o método utilizado para a classificação foi o C-SVM, onde é necessário definir somente o custo (C). Assim, para cada valor possível de C, foram construídos diferentes hiperplanos com respectivas quantidades de vetores de suporte, de forma que fosse encontrada a margem que melhor separasse as classes, definida pela equação (36):

$$\vec{w}\vec{x} - b = 0 \quad (36)$$

em que  $\vec{w}$  é o vetor normal ao hiperplano e  $\vec{x}$  é o conjunto de amostras. Assim, para que a margem escolhida separasse os dados com a maior distância, realizou-se uma otimização através das equações (37):

$$\min_{w,b} \frac{\|w\|}{2} + C \sum_{i=1}^N \xi_i \quad (37)$$

sujeita a:

$$y_i(w^T x_i + b) \geq 1 - \xi_i; i = 1 \dots N \quad (38)$$

em que  $N$  é o número de amostras;  $\frac{\|w\|}{2}$  é o inverso da distância euclidiana entre as classes;  $\xi_i$  são variáveis *slack* adicionadas pelo modelo,  $y_i$  é a classe obtida;  $w^T$  é o vetor  $w$  transposto e  $b$  é o viés.

Ademais, o valor ótimo de  $C$  foi determinado por validação cruzada *10-fold*, de tal forma que fosse obtida a menor quantidade de erros de discriminação.

O desempenho dos modelos de classificação supervisionada foi avaliado usando a matriz de confusão para um problema de três classes. Os valores de precisão, especificidade, sensibilidade e taxa de falsos positivos para cada classe foram obtidos a partir das matrizes de confusão de cada modelo, conforme equações (39) a (42):

$$\text{Precisão} = \frac{PV + NV}{n} \quad (39)$$

$$\text{Especificidade} = \frac{NV}{NV + PF} \quad (40)$$

$$\text{Sensibilidade} = \frac{PV}{PV + NF} \quad (41)$$

$$\text{Taxa de falsos positivos} = 1 - \text{Especificidade} \quad (42)$$

em que  $n$  é o tamanho da amostra,  $PV$ ,  $NV$ ,  $PF$  e  $NF$  são os valores positivos verdadeiros, negativos verdadeiros, positivos falsos e negativos falsos, respectivamente, para cada uma das três classes individuais. As medidas gerais de desempenho, ou seja, precisão, sensibilidade, seletividade e taxa de falsos positivos, foram calculadas com base nas medidas de desempenho das três classes.

## RESULTADOS E DISCUSSÃO

### Análise química das uvas

A Tabela 3.2-1 mostra os principais descritores estatísticos para os atributos de qualidade que caracterizam os conjuntos de calibração e predição, isto é, o número de amostras, a média, os valores de mínimo e máximo, o intervalo, o desvio padrão e a variância. Pode-se notar uma alta variabilidade dos atributos de qualidade mensurados, especialmente para antocianinas totais e flavonoides amarelos, devido aos diferentes estádios de amadurecimento da uva, proporcionando assim o desenvolvimento de modelos preditivos versáteis.

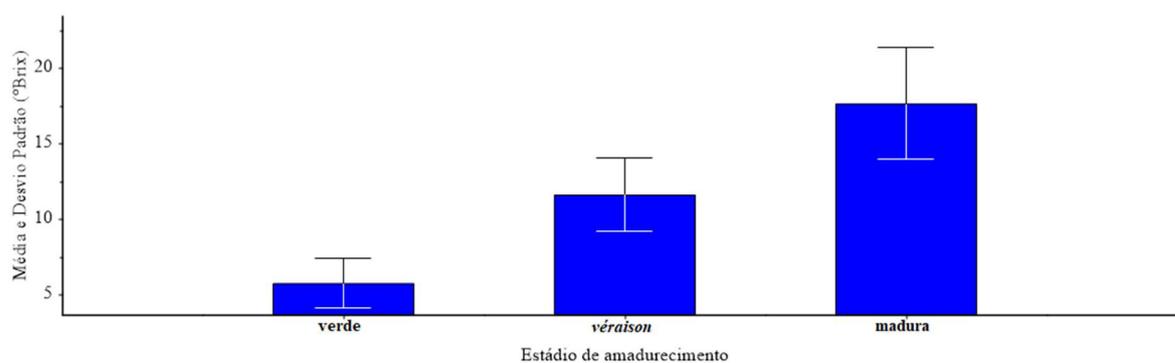
Tabela 3.2-1. Estatística descritiva das bagas de uvas das variedades Syrah e Cabernet Sauvignon.

Parâmetro	Amostras	Média	Min	Max	Alcance	SD	Variância	Syrah e Cabernet Sauvignon						
								Calibração				Predição		
SST	672	13,5	3,9	25,9	22	6,3	36,8	336	14,3	4,1	26,2	21,8	6,0	36,1
Anth	672	95,4	0,8	197,9	197,1	79,3	6293,9	336	106,6	1,3	197,5	196,2	80,5	6481,7
Flav	672	36,3	8,5	153,3	144,8	24,1	580,3	336	47,8	9,5	137,9	128,4	31,2	971,0

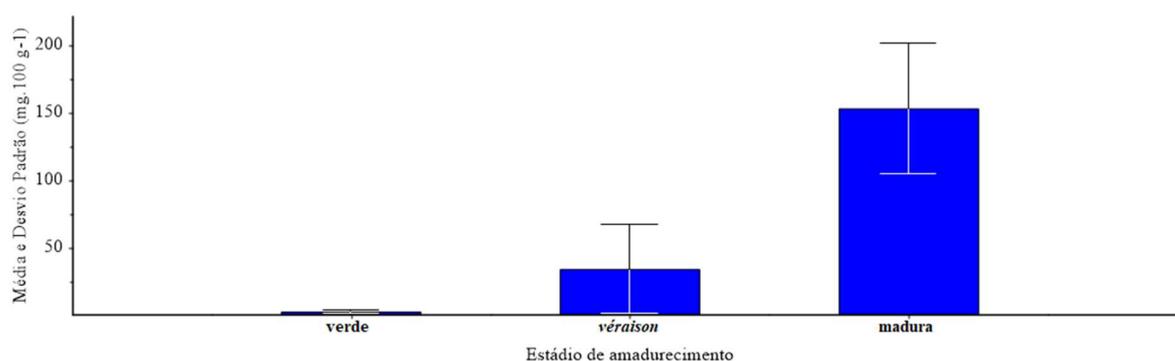
Min: valor mínimo; Max: valor máximo; SD: desvio padrão.

A Figura 3.2-2 mostra os gráficos de *box-plot* da média e desvio padrão para os atributos SST, antocianinas totais e flavonoides amarelos para cada estágio de amadurecimento. Ocorre um incremento dos atributos de qualidade ao longo do amadurecimento com acúmulo de açúcares (SST) e pigmentos (antocianinas totais e flavonoides amarelos) para as uvas da variedade Syrah e Cabernet Sauvignon.

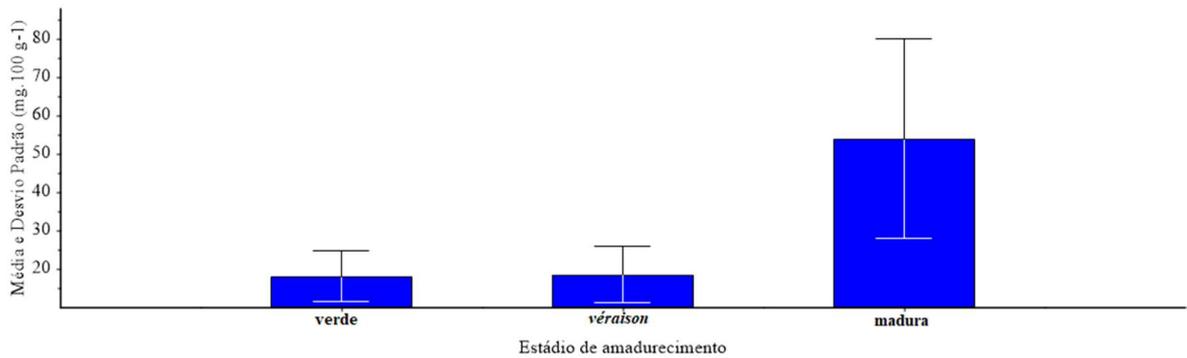
Figura 3.2-2. Gráfico de *box-plot* da média e desvio padrão dos estádios de amadurecimento (verde, *véraison* e madura) para os atributos de qualidade SST (a), antocianinas totais (b) e flavonoides amarelos (c).



(a)



(b)



(c)

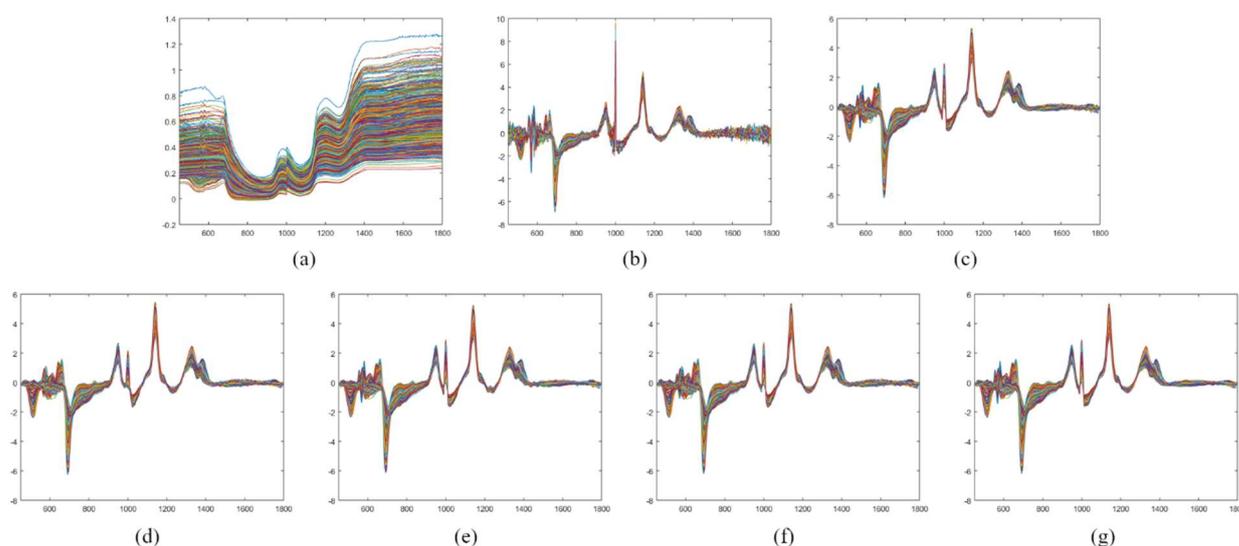
### Pré-processamento dos dados espectrais

A Figura 3.2-3, mostra os dados espectrais brutos, assim como seis sinais resultantes dos melhores pré-processamentos encontrados nos modelos de regressão PLS e SVM-R.

Com a primeira etapa de alisamento, foram obtidos sinais suavizados e os ruídos de alta frequência foram minimizados, isso em decorrência da transformada *wavelet* que possui a capacidade de tratar pontos bem localizados no sinal, diferentemente das demais filtragens onde pontos do sinal são eliminados ou tratados indiscriminadamente (ERGEN, 2012).

Além disso, nota-se que no sinal bruto o deslocamento na linha de base é muito mais evidente que a inclinação da mesma. Assim, a primeira derivada se sobressaiu como técnica de diferenciação com a remoção dos termos constantes do sinal espectral. Após o deslocamento do sinal para a origem do eixo das ordenadas, foi necessário aplicar a normalização para a remoção dos efeitos multiplicativos que foram introduzidos devido à sensibilidade do instrumento óptico empregado. A SNV se sobressaiu na normalização, de forma que no sinal resultante as amostras não mais estavam deslocadas entre si, mas concentradas em uma faixa ao redor da origem.

Figura 3.2-3. Gráficos de absorvância versus comprimento de onda: (a) Dados brutos; (b) limiarização adaptativa brusca com  $db$  4, 3 níveis de decomposição, *Savitzky-Golay* com polinômio de segunda ordem, 7 segmentos com primeira derivada e SNV; (c) limiarização global suave em duas faixas do espectro com  $db$  4, 4 níveis de decomposição, *Savitzky-Golay* com polinômio de segunda ordem, 21 segmentos com primeira derivada e SNV; (d) filtro da mediana com 27 segmentos, *Savitzky-Golay* com polinômio de segunda ordem, 21 segmentos com primeira derivada e SNV; (e) limiarização global suave com  $sym$  3, 10 níveis de decomposição, *Savitzky-Golay* com polinômio de segunda ordem, 21 segmentos com primeira derivada e SNV; (f) filtro da mediana com 17 segmentos, *Savitzky-Golay* com polinômio de segunda ordem, 21 segmentos com primeira derivada e SNV; (g) limiarização global suave com  $coif$  3, 10 níveis de decomposição, *Savitzky-Golay* com polinômio de segunda ordem, 21 segmentos com primeira derivada e SNV.



Com a utilização dos sinais pré-processados, notou-se a melhora das métricas e redução da complexidade alcançada nos modelos de regressão e classificação. Sendo assim, comprovou-se que o pré-processamento dos dados espectrais é uma etapa crucial e necessária antes da calibração dos modelos.

### Desempenho dos modelos preditivos para os atributos de qualidade

Os modelos matemáticos de regressão foram construídos com os métodos de PLSR e SVM-R para as variedades Syrah e Cabernet Sauvignon, consideradas conjuntamente no conjunto de dados, após a aplicação das técnicas de pré-tratamento. A Tabela 3.2-2 mostra os

melhores modelos PLSR e SVM-R para os atributos de qualidade analisados, assim como também os modelos obtidos sem pré-tratamento do sinal espectral.

Tabela 3.2-2. Métricas de desempenho dos modelos de regressão PLSR e SVM-R para as variedades Syrah e Cabernet Sauvignon, com o sinal espectral bruto e processado.

Parâmetro	Modelo	Pré-tratamento	LV/SV	Calibração		Validação		Predição	
				R <sup>2</sup> <sub>cal</sub>	RMSEC	R <sup>2</sup> <sub>cv</sub>	RMSECV	R <sup>2</sup> <sub>pred</sub>	RMSEP
SST	PLSR	Nenhum	9	0,90	1,91	0,89	1,97	0,88	2,11
	SVM-R	Nenhum	631	0,89	1,99	0,89	2,05	0,90	1,94
	PLSR	HSA-db4-l3/2x7-D <sup>1</sup> /SNV	7	0,96	1,24	0,95	1,37	0,94	1,43
	SVM-R	R2-SG-db4-l4/2x21-D <sup>1</sup> /SNV	205	0,98	0,86	0,97	1,03	0,96	1,21
Anth	PLSR	Nenhum	7	0,91	24,01	0,90	24,67	0,91	24,75
	SVM-R	Nenhum	458	0,92	22,65	0,91	23,72	0,90	24,90
	PLSR	MF(27)/2x21-D <sup>1</sup> /SNV	3	0,95	18,04	0,95	18,28	0,96	16,76
	SVM-R	SG-sym3-l10/2x21-D <sup>1</sup> /SNV	614	0,98	11,81	0,97	13,12	0,97	13,52
Flav	PLSR	Nenhum	4	0,68	13,67	0,67	13,80	0,65	20,53
	SVM-R	Nenhum	224	0,75	12,21	0,71	12,92	0,70	19,81
	PLSR	MF(17)/2x21-D <sup>1</sup> /SNV	10	0,80	10,88	0,76	11,89	0,73	18,50
	SVM-R	SG-coif3-l10/2x21-D <sup>1</sup> /SNV	591	0,80	10,92	0,77	11,71	0,78	17,18

MF(17) or MF(27): filtro mediana com segmentos de tamanho 17 ou 27; 2x7 ou 2x21: sinal ajustado com polinômio de segunda ordem e segmentos de tamanho 7 ou 21; R2: alisamento aplicado a duas bandas espectrais; HSA ou SG: limiares adaptativo brusco ou global suave; D<sup>1</sup>: primeira derivada; db4, sym3 ou coif3: função-mãe *wavelet*; l3, l4 ou l10: número dos níveis de decomposição.

Legenda: LV/SV – variáveis latentes ou vetores de suporte; R<sup>2</sup><sub>cal</sub> - coeficiente de determinação da calibração; RMSEC - erro quadrático médio de calibração; R<sup>2</sup><sub>cv</sub> - coeficiente de determinação da validação cruzada; RMSECV - erro quadrático médio de validação cruzada; R<sup>2</sup><sub>pred</sub> - coeficiente de determinação da predição; RMSEP - erro quadrático médio de predição.

Para a determinação de SST, o melhor modelo preditivo foi conseguido com a SVM-R associada à transformada *wavelet* (*db*) alisando duas faixas, obtendo os valores do R<sup>2</sup><sub>pred</sub> e RMSEP de 0,96 e 1,21 °Brix, respectivamente. Em contrapartida, o modelo PLSR com o sinal bruto apresentou R<sup>2</sup><sub>pred</sub> de 0,88 e RMSEP de 2,11 °Brix. Algo interessante a observar foi a redução do número de variáveis latentes e vetores de suporte após o pré-tratamento, tanto para a PLSR como para a SVM (Tabela 3.2-2).

Para a determinação de antocianinas totais, o melhor modelo preditivo foi conseguido com a SVM-R associada à transformada *wavelet (sym)*, sendo os valores de  $R^2_{\text{pred}}$  e RMSEP de 0,97 e 13,52 mg.100 g<sup>-1</sup>, respectivamente. Entretanto, o modelo a SMV-R com o sinal bruto apresentou  $R^2_{\text{pred}}$  igual a 0,90 e RMSEP igual a 24,90 mg.100 g<sup>-1</sup>.

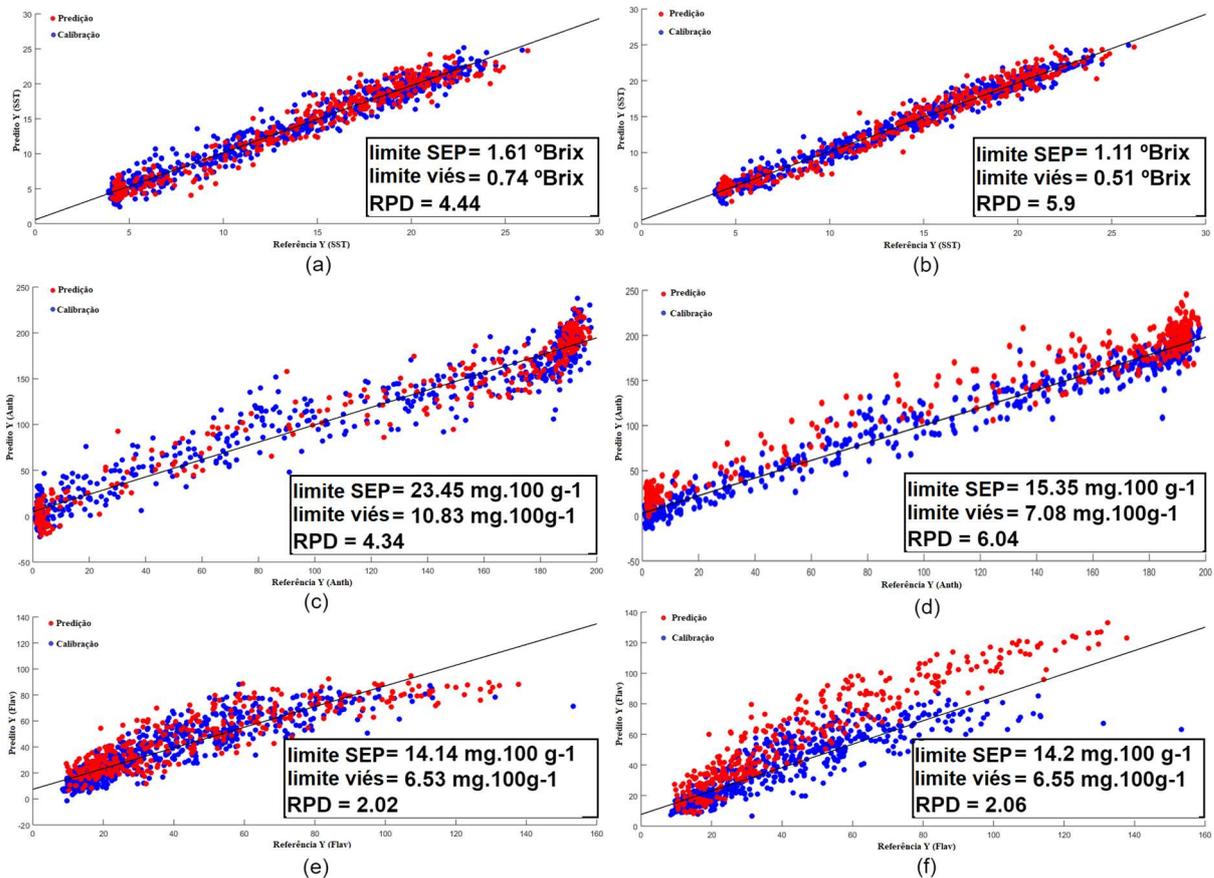
Para a determinação dos flavonoides amarelos, o melhor modelo preditivo foi conseguido com a SVM-R associada à transformada *wavelet (coif)*, sendo os valores de  $R^2_{\text{pred}}$  e RMSEP de 0,78 e 17,18 mg.100 g<sup>-1</sup>, respectivamente. Todavia, o modelo PLSR com o sinal bruto apresentou  $R^2_{\text{pred}}$  de 0,65 e RMSEP de 20,53 mg.100 g<sup>-1</sup>.

A capacidade preditiva do modelo SVM-R associada à transformada *wavelet (db)* para SST foi melhor do que o relatado por Herrera *et al.* (2003) onde, para esse atributo, utilizando a espectroscopia de transmitância na faixa espectral entre 650 nm a 1100 nm e variedade Cabernet Sauvignon, os autores obtiveram valores de  $R^2_{\text{pred}}$  e RMSEP de 0,94 e 1,06 °Brix, respectivamente. Além disso, Xiao *et al.* (2018), utilizando a espectroscopia VIS-NIR (400-1000 nm), encontraram valores de  $R^2_{\text{pred}}$  de 0,93 e RMSEP de 1,30 °Brix, apresentando desempenho inferior ao demonstrado nesse estudo. Os resultados estão de acordo com Nicolaï *et al.* (2007a), onde os melhores desempenhos foram obtidos com a transformada *wavelet db* para a predição do conteúdo de açúcares em maçãs.

Assim como para a determinação de SST, para o conteúdo de antocianinas totais, a capacidade preditiva do modelo SVM-R associada à transformada *wavelet (sym)* foi melhor do que o encontrado por Kemps *et al.* (2010), que obteve coeficiente de correlação na validação de 0,93 e 0,84 na medição com pH = 1,0 e pH = 3,2, respectivamente. Contudo, para o conteúdo de flavonoides amarelos os modelos não foram satisfatórios (Tabela 3.2-2).

A Figura 3.2-4, mostra os valores de referência e previstos dos atributos SST, antocianinas totais e flavonoides amarelos, com a linha de regressão ajustada, para os melhores modelos PLSR e SVM-R.

Figura 3.2-4. Valores de referência e previstos para os melhores modelos, sendo: (a) e (b) modelos PLSR e SVM-R para determinação de SST, respectivamente; (c) e (d) modelos PLSR e SVM-R para determinação de antocianinas totais, respectivamente; (e) e (f) modelos PLSR e SVM-R para determinação de flavonoides amarelos, respectivamente.



Os valores de RPD dos modelos para o SST foram de 4,44 e 5,90, para antocianinas totais foram de 4,34 e 6,04, e para flavonoides amarelos foram de 2,02 e 2,06, para PLSR e SVM, nessa ordem. Williams e Sobering (1996) recomendam que a razão RPD seja maior que 3,00. Nesse sentido, no que se refere aos valores de RPD, estes não foram satisfatórios para o atributo flavonoides amarelos.

A aplicabilidade dos modelos depende dos valores de RPD: quando estes estão compreendidos entre 3,1 e 4,9, indicam que o modelo é adequado para a triagem; entre 5,0 e 6,4, é adequado para o controle de qualidade; entre 6,5 e 8,0, é adequado para o controle do processo; e maiores que 8,0, é adequado para qualquer aplicação (WILLIAMS, 2001). Os modelos desenvolvidos nesse estudo para os atributos SST e antocianinas totais se enquadram nos critérios de triagem (modelos PLSR) e controle de qualidade (modelos SVM-R), a depender do tipo de modelo.

### Determinação qualitativa dos estádios de amadurecimento

A discriminação dos estádios de amadurecimento foi realizada com os métodos de classificação PLS-DA e SVM, cujos resultados encontram-se Tabela 3.2-3. Os melhores modelos obtidos, tanto PLS-DA e SVM, foram aqueles associados à transformada *wavelet*, sendo o modelo SVM ligeiramente superior (precisão = 0,95; taxa de falso positivos = 0,04). A remoção de informações redundantes e irrelevantes do sinal espectral contribui para o aumento do desempenho dos modelos (PLS-DA e SVM) (RINNAN *et al.*, 2009).

Tabela 3.2-3. Métricas de desempenho dos modelos de classificação PLS-DA e SVM para as variedades Syrah e Cabernet Sauvignon, com o sinal espectral bruto e processado.

Modelo	Pré-tratamento	LV/SV	Sensibilidade	Especificidade	Precisão	Taxa de falso positivo
PLS-DA	Nenhum	6	0,83	0,94	0,89	0,06
	HSA-db4-13/2x7-D <sup>1</sup> /SNV	2	0,87	0,95	0,92	0,05
	MF(27)/2x21-D <sup>1</sup> /SNV	2	0,86	0,95	0,93	0,05
	MF(17)/2x21-D <sup>1</sup> /SNV	2	0,87	0,96	0,93	0,04
	R2-SG-db4-14/2x21-D <sup>1</sup> /SNV	2	0,88	0,96	0,93	0,04
	SG-sym3-110/2x21-D <sup>1</sup> /SNV	2	0,87	0,96	0,93	0,04
	SG-coif3-110/2x21-D <sup>1</sup> /SNV	2	0,86	0,96	0,93	0,04
SVM	Nenhum	106	0,82	0,95	0,93	0,05
	HSA-db4-13/SG1st2x7/SNV	130	0,86	0,96	0,95	0,04
	MF(27)/2x21-D <sup>1</sup> /SNV	112	0,83	0,95	0,94	0,05
	MF(17)/2x21-D <sup>1</sup> /SNV	111	0,83	0,95	0,94	0,05
	R2-SG-db4-14/2x21-D <sup>1</sup> /SNV	104	0,84	0,95	0,93	0,05
	SG-sym3-110/2x21-D <sup>1</sup> /SNV	112	0,84	0,95	0,94	0,05
	SG-coif3-110/2x21-D <sup>1</sup> /SNV	100	0,85	0,95	0,94	0,05

---

MF(17) or MF(27): filtro mediana com segmentos de tamanho 17 ou 27; 2x7 ou 2x21: sinal ajustado com polinômio de segunda ordem e segmentos de tamanho 7 ou 21; R2: alisamento aplicado a duas bandas espectrais; HSA ou SG: limiães adaptativo brusco ou global suave; D<sup>1</sup>: primeira derivada; db4, sym3 ou coif3: função-mãe *wavelet*; l3, l4 or l10: número dos níveis de decomposição.

---

Além disso, houve uma grande diminuição da complexidade dos modelos, principalmente para PLS-DA, que empregou apenas 2 variáveis latentes com os sinais espectrais padronizados, enquanto que o sinal espectral bruto precisou de 6 variáveis latentes para explicar a covariância entre a matriz de absorbância e o vetor de classes. Por outro lado, o número de vetores de suporte nos modelos SVM não sofreu influência das padronizações de sinal, contudo o custo computacional foi reduzido. Para o modelo SVM com sinal bruto, o custo empregado foi de 100, enquanto que para os demais modelos com a padronização dos sinais espectrais o maior custo foi de 1.

Os melhores desempenhos dos modelos, tanto PLSR e SVM-R, associados à transformada *wavelet*, deve-se à ênfase que tal procedimento coloca nas regiões espectrais relacionadas à propriedade de interesse, semelhante à seleção de comprimento de onda, auxiliando no desenvolvimento de modelos de classificação precisos (GREENSILL *et al.*, 2001). Isso está de acordo com Xiao *et al.* (2018) que, empregando a seleção de variáveis no espectro VIS-NIR associado ao modelo LS-SVM obteve precisão de 0,90 e 1,00 na discriminação dos estádios de amadurecimento de uvas das variedades Manicure Finger e Ugni Blanc, respectivamente. Ademais, Yang *et al.* (2012), utilizando o espectro UV-VIS-IR associado à PCA-SVM para discriminação de sementes de uva, alcançaram precisão de 0,95, e Lv *et al.* (2012) adotando a mesma abordagem (PCA-SVM) com o espectro VIS-NIR visando a discriminação dos estádios de amadurecimento de uvas obtiveram precisão de 0,96.

## CONCLUSÕES

Os modelos de predição para sólidos solúveis totais, antocianinas totais e flavonoides amarelos se mostraram adequados, considerando  $R^2_{\text{pred}}$  de 0,96, 0,97 e 0,78, respectivamente, para regressão SVM associada a transformada *wavelet*.

O desenvolvimento dos modelos de classificação possibilitou a discriminação dos diferentes estádios de amadurecimento em uvas da variedade Syrah e Cabernet Sauvignon com precisão de 95% para o modelo SVM associada a transformada *wavelet*.

A transformada *wavelet* apresentou maior desempenho como técnica de pré-processamento do sinal espectral, tanto nos modelos de predição quanto de classificação.

## REFERÊNCIAS

ANTONUCCI, F.; PALLOTTINO, F.; PAGLIA, G.; PALMA, A.; D'AQUINO, S.; MENESATTI, P. Non-destructive estimation of mandarin maturity status through portable VIS-NIR spectrophotometer. **Food and Bioprocess Technology**, v.4, n.5, 809-813, 2011.

ARAÚJO, E. G. **Métodos de amostragem e tamanho de amostra para avaliar o estado de maturação da uva Niágara Rosada**. 63p. Dissertação (Mestrado em Agronomia) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2017.

ARENDSE, E.; FAWOLE, O. A.; MAGWAZA, L. S.; OPARA, U. L. Non-destructive prediction of internal and external quality attributes of fruit with thick rind: A review. **Journal of Food Engineering**, 217, 11–23, 2018.

BAIANO, A.; TERRACONE, C.; PERI, G.; ROMANIELLO, R. Application of hyperspectral imaging for prediction of physico-chemical and sensory characteristics of table grapes. **Computers and Electronics in Agriculture**, 87, 142-151, 2012.

BARNES, R. J.; DHANOA, M. S.; LEISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, **Appl. Spectrosc.**, 43, 772–777, 1989.

BIRGÉ, L.; PASCAL M. **From model selection to adaptive estimation. Festschrift for Lucien le cam**. Springer, New York, NY, 55-87, 1997.

BLANCO, M.; VILLARROYA, I. N. I. R. NIR spectroscopy: a rapid-response analytical tool. **TrAC Trends in Analytical Chemistry**, v.21, n.4, 240-250, 2002.

CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. **Trends in Food Science & Technology**, v.18, n.2, 72-83, 2007.

COOMBE, B. G. Adoption of a system for identifying grapevine growth stages. **Australian Journal of Grape and Wine Research**, 1, 100–110, 1995.

DAMBERGS, R.G.; COZZOLINO, D.; CYNKAR, W.; KAMBOURIS, A.; FRANCIS, I.; HØJ, P.B.; GISHEN, M. The use of near infrared spectroscopy for grape quality measurement. **Australian Grapegrower Winemaker**, 473(a), 69–76, 2003.

DONOHO, D. L.; JOHNSTONE, I. M. Threshold selection for wavelet shrinkage of noisy data. Engineering in Medicine and Biology Society, 1994. *Engineering Advances: New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE*. Vol. 1. IEEE, 1994.

ERGEN, B. Signal and image denoising using wavelet transform. In: **Advances in Wavelet Theory and Their Applications in Engineering, Physics and Technology**. InTech, 2012.

FAVARO, M. M. A. **Extração, estabilidade e quantificação de antocianinas de frutas típicas brasileiras para aplicação como corante**. 102p. Dissertação (Mestrado em Química) – Instituto de Química, Universidade Estadual de Campinas, Campinas, 2008.

FEARN, T. Standardisation and Calibration Transfer for near Infrared Instruments: A Review. **Journal of Near Infrared Spectroscopy**, v.9, n.4, 229–244, 2001.

FERREIRA, M. M. C. **Quimiometria – conceitos, métodos e aplicações**. SP, Brazil: Editora da Unicamp, 2015.

FERRER-GALLEGO, R.; HERNÁNDEZ-HIERRO, J. M.; RIVAS-GONZALO, J. C.; ESCRIBANO-BAILÓN, M. T. Determination of phenolic compounds of grape skins during ripening by NIR spectroscopy. **LWT-Food Science and Technology**, 44.4, 847-853, 2011.

FRANCIS, F. J. **Analysis of anthocyanins. Anthocyanins as food colors**. Academic Press, New York, NY, 181-207, 1982.

GARCÍA MARTÍN, J. F. Optical path length and wavelength selection using Vis/NIR spectroscopy for olive oil's free acidity determination. **International Journal of Food Science & Technology**, v. 50, n.6, 1461-1467, 2015.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: A tutorial. **Anal. Chim. Acta**, 185, 1–17, 1986.

GORRY, P. A. General least-square smoothing and differentiation by the convolution (Savitzky- Golay) method. **Anal. Chem.**, 62, 570–573, 1990.

GREENSILL, C. V.; WOLFS, P. J.; SPIEGELMAN, C. H.; WALSH, K. B. Calibration Transfer between PDA-Based NIR Spectrometers in the NIR Assessment of Melon Soluble Solids Content. **Applied Spectroscopy**, v.55, n.5, 647–653, 2001.

HERRERA, J.; GUESALAGA, A.; AGOSIN, E. Shortwave – near infrared spectroscopy for non-destructive determination of maturity of wine grapes. **Measurement Science and Technology**, 14, 689–697, 2003

HUANG, H.; YU, H.; XU, H.; YING, Y. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. **Journal of Food Engineering**, v.87, n.3, 303-313, 2008.

ISAKSSON, T.; NAES, T. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. **Appl. Spectrosc.**, 42, 1273–1284, 1988.

KEMPS, B.; LEON, L.; BEST, S.; DE BAERDEMAEKER, J.; DE KETELAERE, B. Assessment of the quality parameters in grapes using VIS-NIR spectroscopy. **Biosystems Engineering**, v.105, n.4, 507-513, 2010.

LI, J.; HUANG, W.; ZHAO, C.; ZHANG, B. A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy. **Journal of Food Engineering**, v.116, n.2, 324-332, 2013.

LIU, C.; YANG, S. X.; LIE D. Determination of internal qualities of Newhall navel oranges based on NIR spectroscopy using machine learning. **Journal of Food Engineering**, 161, 16-23, 2015.

LU, H.; ZHENG, H.; HU, Y.; LOU, H.; KONG, X. Bruise detection on red bayberry (*Myrica rubra* Sieb. & Zucc.) using fractal analysis and support vector machine. **Journal of Food Engineering**, v.104, n.1, 149-153, 2011.

LV, G.; YANG, H.; XU, N.; MOUAZEN, A. M. Identification of less-ripen, ripen, and over-ripen grapes during harvest time based on visible and near-infrared (Vis-NIR) spectroscopy. **Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on.** IEEE, 2012.

MALLAT, S. **A Wavelet Tour of Signal Processing**, 2a. ed. Academic Press, 1998.

MOGHIMI, A.; AGHKHANI, M. H.; SAZGARNIA, A.; SARMAD, M. Vis-NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. **Biosystems engineering**, v.106, n.3, 295-302, 2010.

NICOLAÏ, B. M.; THERON, K. I.; LAMMERTYN, J. Kernel PLS regression on wavelet transformed NIR spectra for prediction of sugar content of apple. **Chemometrics and Intelligent Laboratory Systems**, v.85, n.2, 243–252, 2007a.

NICOLAÏ, B. M.; BEULLENS, K.; BOBELYN, E.; PEIRS, A.; SAEYS, W.; THERON, K. I.; LAMMERTYN, J. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. **Postharvest Biology and Technology**, v.46, n.2, 99–118, 2007b.

PÉREZ-ENCISO, M.; TENENHAUS, M. Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. **Human Genetics**, v.112, n.5-6, 581-592, 2003.

RANDOLPH, T. W. Scale-based normalization of spectral data. **Cancer Biomarkers**, 2.3-4, 135-144, 2006.

RINNAN, Å.; VAN DEN BERG, F.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC Trends in Analytical Chemistry**, v.28, n.10, 1201-1222, 2009.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures, **Anal. Chem.**, 36, 1627–1639, 1964.

SHAO, Y.; YONG H.; YIDAN B. A new approach to predict acidity of bayberry juice by using vis/near infrared spectroscopy. **International Journal of Food Properties**, v.10, n.3, 631-638, 2007.

XIAO, H.; LI, A.; LI, M.; SUN, Y.; TU, K.; WANG, S.; PAN, L. Quality assessment and discrimination of intact white and red grapes from *Vitis vinifera* L. at five ripening stages by visible and near-infrared spectroscopy. **Scientia Horticulturae**, 233, 99-107, 2018.

WILLIAMS, P. C. Implementation of near-infrared technology. In P. Williams & K. Norris (Eds.), **Near-infrared technology in the agricultural and food industries** (2nd ed., pp. 145–169). St. Paul, MN: American Association of Cereal Chemists, 2001.

WILLIAMS, P. C.; SOBERING, D. C. How do we do it: A brief summary of the methods we use in developing near infrared calibrations. In A. M. C. Davies & P. Williams (Eds.), **Near infrared spectroscopy: The future waves** (pp. 185–188). Chichester: NIR Publications, 1996.

WISE, B. M.; GALLAGHER, N. B.; BRO, R.; SHAVER, J. M.; WINDIG, W.; KOCH, R. S. Chemometrics tutorial for PLS\_Toolbox and Solo. **Eigenvector Research, Inc**, 3905, 102-159, 2006.

YANG, H. Q.; WEI Q. L.; WEN J. W. Nondestructive discrimination of grape seed varieties using UV-VIS-NIR spectroscopy and chemometrics. **Applied Mechanics and Materials**. v.236. Trans Tech Publications, 2012.

ZHANG, G. P. Neural networks for classification: a survey. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, v.30, n.4, 451-462, 2000.

ZHU, D.; JI, B.; MENG, C.; SHI, B.; TU, Z.; QING, Z. Study of wavelet denoising in apple's charge-coupled device near-infrared spectroscopy. **Journal of agricultural and food chemistry**, v.55, n.14, 5423-5428, 2007.

### 3.3. INSTRUMENTO OPTOELETRÔNICO VIS-NIR PARA MENSURAÇÃO DE ATRIBUTOS DE QUALIDADE EM UVAS VINÍFERAS DA VARIEDADE CABERNET SAUVIGNON

**RESUMO:** A determinação da época de colheita é essencial na vinha. Para isso, atributos de qualidade como sólidos solúveis totais e compostos fenólicos são monitorados constantemente ao longo do amadurecimento. Portanto, objetivou-se com esse trabalho desenvolver e avaliar um instrumento optoeletrônico para mensuração não destrutiva de sólidos solúveis totais (SST), antocianinas totais e flavonoides amarelos em uvas viníferas. Foram colhidas 360 bagas da variedade Cabernet Sauvignon da região do Vale do Submédio São Francisco em diferentes estádios de amadurecimento. Um dispositivo optoeletrônico, denominado OPTOFruit, foi desenvolvido com a finalidade de determinar, de forma não destrutiva, atributos de qualidade das uvas, a saber, sólidos solúveis totais, antocianinas totais e flavonoides amarelos. Esse foi constituído por: (1) diodos emissores de luz (LEDs); (2) fotodetectores de InGaAs (Índio-Gálio-Arsênio); (3) condicionadores de sinal; (4) unidade de processamento; e (5) transceptor de radiofrequência. Análises laboratoriais foram realizadas nas bagas para aquisição dos atributos de qualidade (SST, antocianinas totais e flavonoides amarelos) posteriormente às análises ópticas. Foram desenvolvidos modelos matemáticos usando a Regressão Linear Múltipla (MLR) e a Regressão *Random Forest*. A validação cruzada foi realizada com todas as amostras do conjunto de calibração. O número de árvores usado para a construção dos modelos *Random Forest* foi de 500. O desempenho do modelo *Random Forest* foi superior em relação ao MLR nas métricas  $R^2$  e MSE para os atributos SST ( $R^2 = 0,432$  vs.  $R^2 = 0,896$ ; MSE = 22,892 vs. MSE = 4,796), antocianinas totais ( $R^2 = 0,391$  vs.  $R^2 = 0,878$ ; MSE = 3657,167 vs. MSE = 891,295) e flavonoides amarelos ( $R^2 = 0,239$  vs.  $R^2 = 0,866$ ; MSE = 190,386 vs. MSE = 46,864). Portanto, o instrumento optoeletrônico associado aos modelos *Random Forest* para mensuração de sólidos solúveis totais (SST), antocianinas totais e flavonoides amarelos em uvas viníferas da variedade Cabernet Sauvignon, apresentou-se promissor como ferramenta para análise não destrutiva.

**PALAVRAS-CHAVES:** dispositivo optoeletrônico; *Random Forest*; regressão linear múltipla, métodos não destrutivos

## INTRODUÇÃO

A qualidade da uva está intimamente relacionada à identificação do amadurecimento na colheita, sendo então uma questão crucial. O monitoramento do amadurecimento em uva no campo pode fornecer dados valiosos para controlar a qualidade do vinho. O amadurecimento é um processo complexo, ocorrendo alterações bioquímicas no nível do epicarpo e mesocarpo. Para monitorar essas alterações, rotineiramente, são realizadas mensurações dos parâmetros de amadurecimento (por exemplo, teor de sólidos solúveis, acidez, fenóis, antocianinas, firmeza, etc.), essencial para frutas e vegetais (GIOVENZANA *et al.*, 2015). Historicamente, o principal parâmetro que é medido para esta finalidade é o teor de açúcar, que está diretamente relacionado com o grau alcoólico do vinho a ser produzido (LARRAÍN *et al.*, 2008).

Tradicionalmente, os parâmetros de qualidade são mensurados com técnicas e instrumentos destrutivos e invasivos. Essas metodologias são procedimentos lentos, que requerem insumos e reagentes. Também, não há uma representatividade de amostras a serem medidas, pois o número de amostras a serem analisadas é diretamente proporcional ao custo, onerando o processo de monitoramento por conta dos insumos (reagentes, equipamentos e descarte adequado).

A alteração de coloração, resultantes de reações bioquímicas no interior das frutas, pode ser relacionada a maturidade (CHOONG *et al.*, 2006). Por exemplo, frutos verdes têm uma maior proporção de clorofila, que diminui gradualmente com a maturidade (AGATI *et al.*, 2008). Estas alterações e mudanças bioquímicas podem ser observadas utilizando dispositivos ópticos.

Os sistemas ópticos têm sido utilizados para o monitoramento da qualidade de frutos em várias culturas hortícolas (RUIZ-ALTISENT *et al.*, 2010). Contudo, algumas dessas técnicas são dispendiosas, em virtude dos equipamentos necessários para as análises (CIVELLI *et al.*, 2015). Por conseguinte, estão sendo desenvolvidos sistemas ópticos simplificados. Estes têm como base a diferença de absorvância, em comprimentos de onda específicos, para monitorar a qualidade de, por exemplo, maçã (BETEMPS *et al.*, 2012), palma (SAEED *et al.*, 2012; HAZIR *et al.*, 2012) e alface-da-terra (BEGHI *et al.*, 2014).

Muitos equipamentos ópticos encontram-se disponíveis comercialmente com diferentes princípios de funcionamento, dentre eles a espectroscopia do ultravioleta, visível e infravermelho próximo curto (MARTINS *et al.*, 2010), fluorescência (CEROVIC *et al.*, 2008; CEROVIC *et al.*, 2009; GHOZLEN *et al.*, 2010; TUCCIO *et al.*, 2011) e reflectância em conjunto com fluorescência (GRAÇA, 2014).

Os métodos ópticos são particularmente adequados para a viticultura de precisão que tenta responder ao problema da heterogeneidade espacial e temporal da cultura, especialmente na maturidade da fruta. A colheita seletiva tem sido proposta como uma solução para este problema com base na gestão zonal do vinhedo (CEROVIC *et al.*, 2008). Isto posto, existe um esforço para estimar com precisão a cinética de amadurecimento da uva, com o intuito de prever a melhor época para a colheita.

A busca por métodos não destrutivos que proporcionem uma rápida tomada de decisão é desejável. Portanto, o objetivo desse estudo foi desenvolver e avaliar um instrumento optoeletrônico para mensuração não destrutiva de sólidos solúveis totais (SST), antocianinas totais e flavonoides amarelos em uvas viníferas da variedade Cabernet Sauvignon.

## **MATERIAIS E MÉTODOS**

### **Amostragem**

As amostras de uva (*Vitis vinifera* L.), da variedade Cabernet Sauvignon, foram coletadas de um vinhedo localizado na região do Vale do Submédio São Francisco (Lagoa Grande, Pernambuco, Brasil: -9,05363; -40,19868). Foram coletados 30 cachos de uvas tintas em 2018, em cinco datas de amostragem (14, 21, 28 de agosto, 04 e 11 de setembro), abrangendo diversos estádios de amadurecimento, desde do estágio de *véraison* à maduro. Para obter amostras representativas, foi adotado o método de amostragem estratificado (ARAÚJO, 2007). Os cachos selecionados foram divididos em três posições: topo, meio e fundo. De cada posição foram retiradas quatro bagas, perfazendo um total de doze bagas por cacho. Um total de 360 bagas individuais foram utilizadas para leitura óptica e, posteriormente, análises químicas destrutivas.

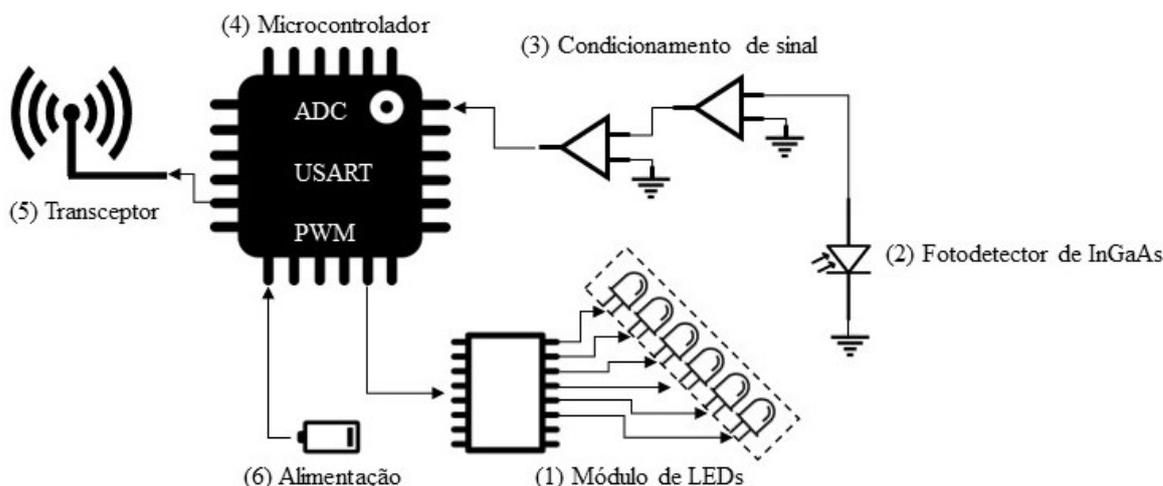
Os cachos individuais foram acondicionados em embalagem impermeável e armazenadas em gelo para transporte, imediato processamento e aquisição óptica. Antes de cada medição, foi necessário aguardar que as amostras se estabilizassem à temperatura ambiente de 25 °C.

### **Descrição do dispositivo optoeletrônico**

Um dispositivo optoeletrônico, denominado OPTOFruit, foi desenvolvido com a finalidade de determinar, de forma não destrutiva, atributos de qualidade das uvas, a saber,

sólidos solúveis totais, antocianinas totais e flavonoides amarelos. Esse foi constituído por: (1) diodos emissores de luz (LEDs); (2) fotodetectores de InGaAs (Índio-Gálio-Arsênio); (3) condicionadores de sinal; (4) unidade de processamento; e (5) transceptor de radiofrequência (Figura 3.3-1).

Figura 3.3-1. Diagrama em bloco dos principais componentes do dispositivo OPTOFruit.



A fonte de excitação é constituída por doze LEDs (*Roithner Lasertechnik*, Viena, Áustria) na faixa espectral do visível (570, 660 e 690 nm) e infravermelho (970, 1200 e 1300 nm). Esses LEDs são acionados por um multiplexador 74HC237 (*Nexperia*, Nimega, Holanda), sendo controlado por três linhas de dados e multiplexando para oito saídas.

A refletância da fonte de excitação foi adquirida pelo fotodetector PT521 (*Roithner Lasertechnik*, Viena, Áustria), tendo esse um fotodiodo de InGaAs uma resposta na faixa espectral de 500 a 1650 nm. A resposta desse fotodiodo é em corrente elétrica.

O sinal de corrente elétrica, proporcional à refletância que excita o fotodetector, foi amplificado pelo circuito amplificador Multiboard (*Roithner Lasertechnik*, Viena, Áustria). Este circuito foi configurado para o funcionamento em dois estádios de amplificação, sendo o primeiro de transimpedância, para converter o sinal de corrente elétrica em tensão elétrica, com ganho ajustável de  $10^4$  a  $10^6$  V/A, e o segundo uma configuração inversora, com pré-ganho de 10 V/V e ganho ajustável de 0,1 a 10 V/V. Esses ganhos foram ajustados em ambiente escuro com uma placa cerâmica Spectralon® (*Labsphere Inc.*, North Sutton, NH, EUA) sendo usada como referência branca, com a qual operações de calibração (aquisição de referência do branco padrão) foram realizadas antes da tomada dos dados, acionando os LEDs em diferentes comprimentos de onda sequencialmente.

A unidade de processamento do sinal foi constituída por um microcontrolador ATmega328P (*Microchip Technology Inc.*, Arizona EUA), com 32 kbytes de memória Flash, 2 kbytes de memória RAM, 1 kbytes de memória EEPROM, 23 portas I/O, 6 conversores A/D (*Analog/Digital*) de 10 bits, 2 temporizadores e periféricos de comunicação SPI (*Serial Peripheral Interface*) e USART (*Universal Synchronous Asynchronous Receiver Transmitter*).

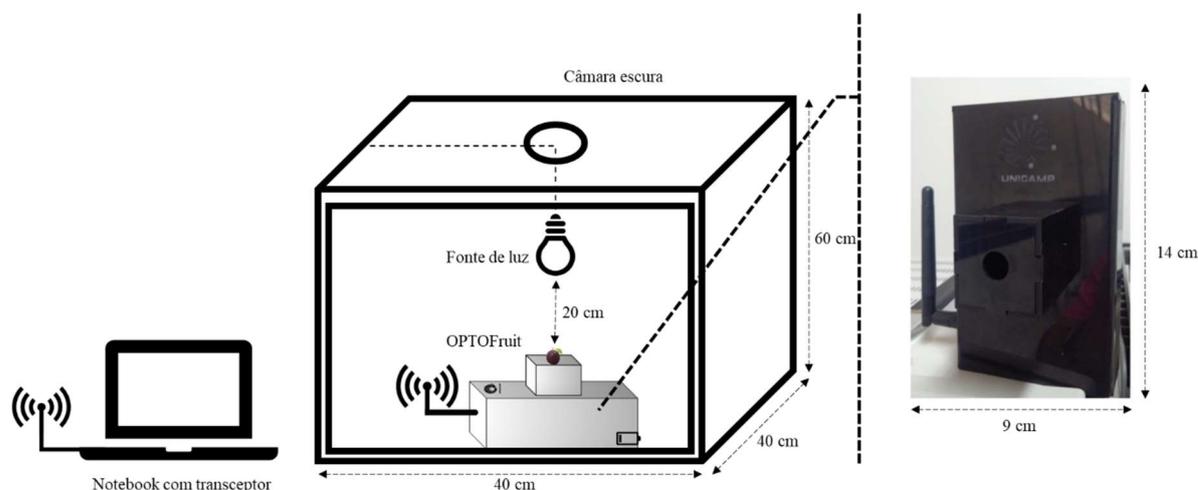
Foi utilizado um *notebook* para o armazenamento dos dados, e a comunicação sem fio entre o OPTOFruit e o notebook foi realizada através do transceptor de radiofrequência Xbee Series 2 (*Digi International Inc.*, Minnesota, EUA), que opera na banda de 2,4 GHz baseado no padrão IEEE 802.15.4, alcançando taxa de transmissão de 250 kbps.

A alimentação do dispositivo é fornecida por uma fonte simétrica de 12 V para o circuito de amplificação e uma fonte assimétrica, de mesma tensão, para o restante do circuito. Esse módulo é composto por reguladores de tensão que proporcionam uma queda para 5 V aos circuitos integrados (multiplexador, microcontrolador e transceptor).

### **Aquisição dos dados experimentais**

O dispositivo OPTOFruit foi acondicionado em uma câmara escura com dimensões de 60 x 40 x 40 cm com fonte de luz de quartzo-tungstênio-halogênio de 120 W e temperatura de cor de 2800 K, com a finalidade de sensibilizar o fotodetector com maior intensidade (Figura 3.3-2). O aparato foi ligado 30 minutos antes das leituras, visando aumentar a qualidade e homogeneidade dos dados. O dispositivo apresenta uma coluna com altura de 4 cm e um orifício no centro de 1 cm onde coloca-se as bagas para as leituras. Cada sinal foi constituído pela média de três varreduras feitas pelo sensor, para cada uma das bagas, e este foi transmitido para um computador pessoal para posterior processamento e análise.

Figura 3.3-2. Diagrama do sistema de aquisição dos sinais do dispositivo OPTOFruit.



Após as medições não destrutivas com o dispositivo OPTOFruit, as bagas individuais foram submetidas às medições de referência, a saber sólidos solúveis totais (SST), antocianinas totais (Anth) e flavonoides amarelos (Flav).

A determinação dos sólidos solúveis totais foi realizada a partir da leitura direta em refratômetro digital (HI 96804, Hanna Instruments, USA), com faixa de medição entre 0 e 85 °Brix e precisão de  $\pm 0,2$  °Brix, utilizando-se duas gotas da amostra triturada e homogeneizada, com os resultados expressos em °Brix.

Os teores de antocianinas totais e flavonoides amarelos da casca foram determinados segundo Francis (1982), sendo este método de pH único não havendo diferença com o método AOAC (FAVARO, 2008). Pesou-se 0,5 g de casca e polpa, em seguida, adicionou-se 25 mL da solução extratora de etanol (95%) acidificado com HCl (1,5 N), na proporção 85:15. As amostras foram maceradas por um minuto e, posteriormente, transferidas para recipientes ao abrigo da luz ficando em repouso e armazenados na geladeira. Após 24 horas, o sobrenadante era recolhido para leitura em espectrofotômetro (Spectronic BioMate 5 UV-Vis, Thermo Electron, UK) a 535 nm, para antocianinas totais, e 374 nm, para flavonoides amarelos. Os resultados foram expressos em  $\text{mg} \cdot 100 \text{ g}^{-1}$ , por meio das seguintes expressões:

$$\text{Anth} = \text{abs} * \text{fator de diluição} / E1\% \quad (1)$$

$$\text{Flav} = \text{abs} * \text{fator de diluição} / E2\% \quad (2)$$

em que o fator de diluição é 5000, o coeficiente de extinção para antocianinas totais (E1%) é 98,2 e o coeficiente de extinção para flavonoides amarelos (E2%) é 76,6.

## Processamento e análise dos dados experimentais

O desenvolvimento dos modelos de regressão foi realizado utilizando o pacote computacional R (*R Foundation*, Viena, Áustria). Para construção dos modelos matemáticos, utilizaram-se os dados advindos do OPTOFruit e os atributos de qualidade (SST, antocianinas totais e flavonoides amarelos), como variáveis de entrada independentes (X) e dependentes (Y), respectivamente, (WOLD *et al.*, 1984) das uvas individuais no conjunto de calibração e validação cruzada.

Os modelos de regressão foram desenvolvidos usando a regressão linear múltipla (MLR) e a regressão *Random Forest*. A validação cruzada foi realizada com todas as amostras do conjunto de calibração. O número de árvores usado para a construção dos modelos *Random Forest* foi 500.

O desempenho da regressão linear múltipla e *Random Forest* foi avaliado utilizando três parâmetros estatísticos, sendo eles o coeficiente de correlação ( $r$ ), o coeficiente de determinação ( $R^2$ ), e o erro quadrático médio (MSE), definidos nas Equações (3), (4) e (5):

$$r(\text{Pearson}) = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}) * (y_i - \bar{y})}{(n - 1) * \sigma_r * \sigma_p} \quad (3)$$

$$MSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

$$R^2 (r_{\text{square}}) = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}) * (y_i - \bar{y})}{(n - 1) * \sigma_r * \sigma_p}} \quad (5)$$

em que  $\hat{y}_i$  é o valor previsto pelo modelo;  $y_i$  é o valor de referência;  $\bar{y}$  é a média dos valores de referência;  $\bar{\hat{y}}$  é a média dos valores previstos;  $n$  é o número de amostras;  $\sigma_r$  é o desvio padrão dos valores de referência; e  $\sigma_p$  é o desvio padrão dos valores previstos.

## RESULTADOS E DISCUSSÕES

Os descritores estatísticos obtidos a partir das análises físico-químicas dos atributos SST, antocianinas totais e flavonoides amarelos estão listados na Tabela 3.3-1. Pode-se observar uma grande variação dos valores obtidos em decorrência da mudança dos estádios de amadurecimento, o que proporciona a obtenção de modelos preditivos versáteis.

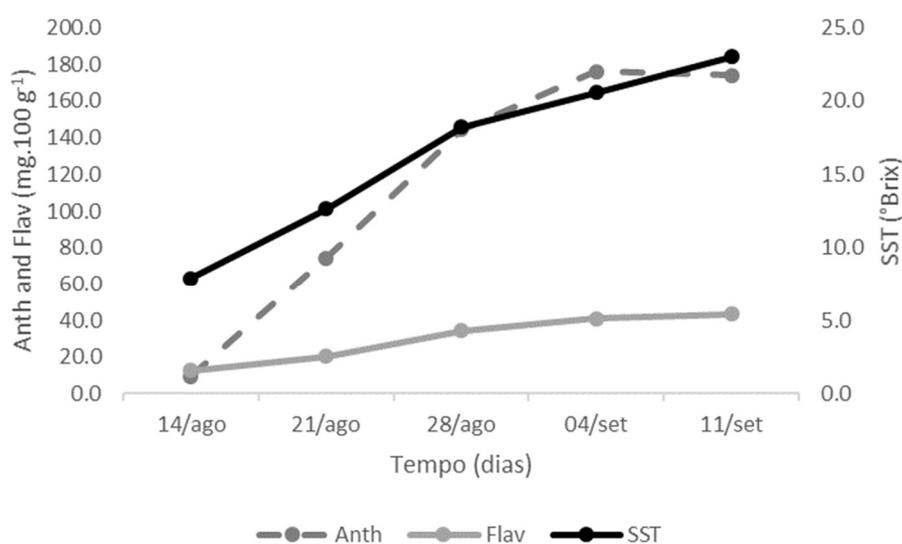
Tabela 3.3-1. Descritores estatísticos das análises físico-químicas das bagas de uva da variedade Cabernet Sauvignon.

Parâmetro	Amostras	Média	Mínimo	Máximo	Alcance	Desvio Padrão	Variância
SST	360	16,5	2,9	26,7	23,8	6,4	40,6
Anth	360	115,7	0,0	195,5	195,6	77,6	6021,1
Flav	360	30,3	5,7	69,5	63,8	15,8	250,9

A evolução dos atributos de qualidade SST, antocianinas totais e flavonoides amarelos durante o amadurecimento estão mostrados na Figura 3.3-3. Ocorre uma mudança significativa dos atributos de qualidade ao longo do processo de amadurecimento das uvas da variedade Cabernet Sauvignon. Considerando o atributo SST para a tomada de decisão sobre a colheita, as vinhas obtiveram a partir do 04 de setembro valores acima de 20 °Brix. Contudo, as mesmas foram colhidas com SST em torno de 23 °Brix.

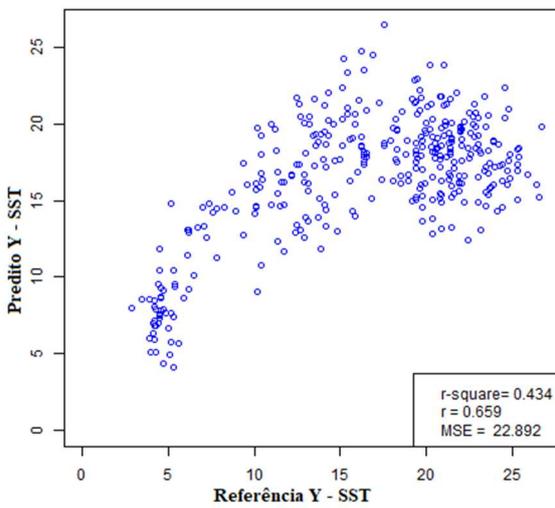
Ao longo do amadurecimento das uvas observa-se um aumento na concentração de flavonoides amarelos (12,8 mg.100 g<sup>-1</sup> para 43,3 mg.100 g<sup>-1</sup>) e, principalmente, de antocianinas totais (9,6 mg.100 g<sup>-1</sup> para 174,0 mg.100 g<sup>-1</sup>) (Figura 3.3-3). Isto se deve à degradação das clorofilas no tecido superficial e acúmulo de compostos pigmentantes, como as antocianinas totais e flavonoides amarelos (AGATI *et al.*, 2007, 2008).

Figura 3.3-3. Evolução da concentração de sólidos solúveis totais (SST), antocianinas totais (Anth) e flavonoides amarelos (Flav) durante as semanas ao longo do amadurecimento.

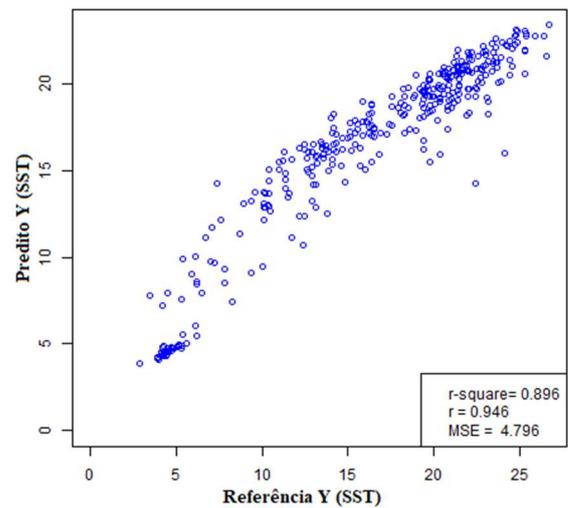


Na Figura 3.3-4 são apresentados os gráficos de dispersão, juntamente com as métricas MSE, coeficiente de correlação ( $r$ ) e coeficiente de determinação ( $R^2$ ), para os atributos de qualidade sólidos solúveis totais, antocianinas totais e flavonoides amarelos com os modelos MLR e *Random Forest*.

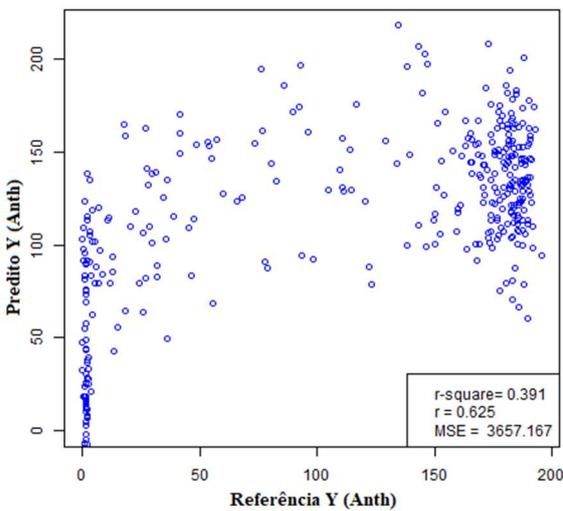
Figura 3.3-4. Gráficos de dispersão: MLR para SST (a); *Random Forest* para SST (b); MLR para antocianinas totais (c); *Random Forest* para antocianinas totais (d); MLR para flavonoides amarelos (e); *Random Forest* para flavonoides amarelos (f).



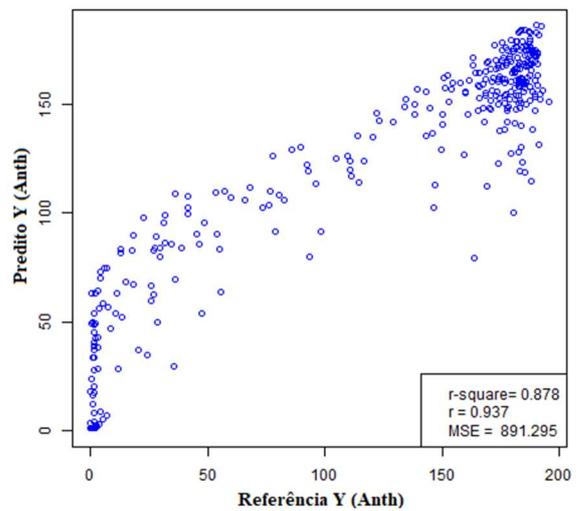
(a)



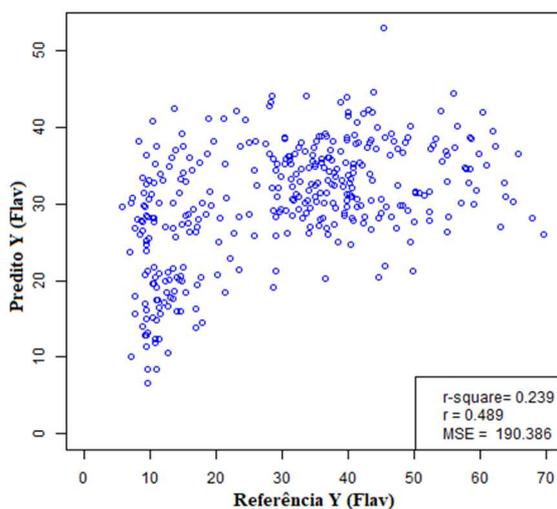
(b)



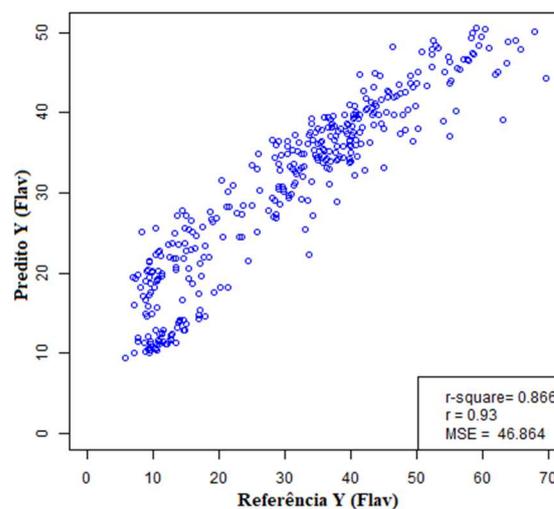
(c)



(d)



(e)



(f)

O desempenho do modelo *Random Forest* foi superior em relação ao MLR nas métricas  $R^2$  e MSE para os atributos SST ( $R^2 = 0,432$  vs.  $R^2 = 0,896$ ; MSE = 22,892 vs. MSE = 4,796), antocianinas totais ( $R^2 = 0,391$  vs.  $R^2 = 0,878$ ; MSE = 3657,167 vs. MSE = 891,295) e flavonoides amarelos ( $R^2 = 0,239$  vs.  $R^2 = 0,866$ ; MSE = 190,386 vs. MSE = 46,864).

Os modelos baseados em aprendizagem de máquina, como as redes neurais artificiais (RNA) e o *Random Forest*, apresentam desempenho superior aos modelos baseados na aprendizagem estatística. Zheng *et al.* (2011) utilizando a análise de imagens com fractais, obtiveram melhores resultados com a RNA ( $R^2 = 0,939$ ) do que com a PLSR ( $R^2 = 0,918$ ) para predição de antocianinas. Ademais, os mesmos autores, obtiveram também desempenho superior com a RNA ( $R^2 = 0,655$ ) do que com a PLSR ( $R^2 = 0,641$ ) para a predição de flavonoides amarelos. Pereira *et al.* (2018), por sua vez, usaram o *Random Forest* para predição de amadurecimento de mamão *papaya*, e obtiveram acurácia de 94,3% na classificação com validação cruzada.

A capacidade preditiva do OPTOFruit para o atributo SST, utilizando o modelo *Random Forest*, foi semelhante ao relatado por Larraín *et al.* (2008), que desenvolveram um instrumento não invasivo para cinco variedades de uvas viníferas obtendo valores de  $R^2$  entre 0,874 à 0,930 para SST, e entre 0,398 à 0,681 para conteúdo de antocianinas, e Giovenzana *et al.* (2015), que desenvolveram um dispositivo óptico que opera com quatro comprimentos de onda (630, 690, 750 e 850 nm) obtendo valores de  $R^2$  de 0,85 e 0,87 para predição do SST e acidez titulável, respectivamente. O desempenho do OPTOFruit foi superior ao dispositivo relatado por Larraín

*et al.* (2008) para a determinação de conteúdo pigmentantes (conteúdo de antocianinas e flavonoides amarelos).

Portanto, o OPTOFruit associado com o modelo *Random Forest* apresenta-se promissor na determinação do SST, antocianinas totais e flavonoides amarelos em uvas viníferas da variedade Cabernet Sauvignon.

## CONCLUSÕES

O desempenho do modelo *Random Forest* foi superior em relação ao MLR nas métricas coeficientes de correlação, determinação e erro quadrático médio para todos os atributos de qualidade (SST, antocianinas totais e flavonoides amarelos).

O instrumento optoeletrônico, associado com o *Random Forest*, para determinação de sólidos solúveis totais, antocianinas totais e flavonoides amarelos, apresentou desempenho adequado, considerando  $R^2$  (0,90, 0,88 e 0,87, respectivamente).

## REFERÊNCIAS

AGATI, G.; MEYER, S.; MATTEINI, P.; CEROVIC, Z. G. Assessment of anthocyanins in grape (*Vitis vinifera* L) berries using a noninvasive chlorophyll fluorescence method. **Journal of Agricultural and Food Chemistry**, v.55, n.4, 1053-1061, 2007.

AGATI, G.; TRAVERSI, M. L.; CEROVIC, Z. G. Chlorophyll fluorescence imaging for the noninvasive assessment of anthocyanins in whole grape (*Vitis vinifera* L.) bunches. **Photochemistry and photobiology**, v.84, 1431-1434, 2008.

ARAÚJO, E. G. **Métodos de amostragem e tamanho de amostra para avaliar o estado de maturação da uva Niágara Rosada**. 63p. Dissertação (Mestrado em Agronomia) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2017.

BEGHI, R.; GIOVENZANA, V.; CIVELLI, R.; MALEGORI, C.; BURATTI, S.; GUIDETTI, R. Setting-up of a simplified handheld optical device for decay detection in fresh-cut *Valerianella locusta* L. **Journal of Food Engineering**, v.127, 10-15, 2014.

BETEMPS, D. L.; FACHINELLO, J. C.; GALARÇA, S. P.; PORTELA, N. M.; REMORINI, D.; MASSAI, R.; AGATI, G. Non-destructive evaluation of ripening and quality traits in

apples using a multiparametric fluorescence sensor. **Journal of the Science of Food and Agriculture**, v.92, n.9, 1855-1864, 2012

CEROVIC, Z. G.; MOISE, N.; AGATI, G.; LATOUCHE, G.; BEN GHOZLEN, N.; MEYER, S. New portable optical sensors for the assessment of winegrape phenolic maturity based on berry fluorescence. **Journal of Food Composition and Analysis**, v.21, 650-654, 2008.

CEROVIC, Z. G.; GOUTOULY, J-P.; HILBERT, G.; DESTRAK-IRVINE, A.; MARTINON, V.; MOISE, N. **Mapping winegrape quality attributes using portable fluorescence-based sensors**. In: FRUTIC 09, PROGAP INIA, 301-310, 2009.

CHOONG, T. S. Y.; ABBAS, S.; SHARIFF, A. R.; HALIM, R.; ISMAIL, M. H. S.; YUNUS, R.; ALI, S.; AHMADUN, F.-R. Digital Image Processing of Palm Oil Fruits. **International Journal of Food Engineering**, v.2, n.2, 1556-1560, 2006.

CIVELLI, R.; GIOVENZANA, V.; BEGHI, R.; NALDI, E.; GUIDETTI, R.; OBERTI, R. A Simplified, Light Emitting Diode (LED) Based, Modular System to be Used for the Rapid Evaluation of Fruit and Vegetable Quality: Development and Validation on Dye Solutions. **Sensors**, v.15, n.9, 22705–22723, 2015.

FAVARO, M. M. A. **Extração, estabilidade e quantificação de antocianinas de frutas típicas brasileiras para aplicação como corante**. 102p. Dissertação (Mestrado em Química) – Instituto de Química, Universidade Estadual de Campinas, Campinas, 2008.

FRANCIS, F. J., 1982. **Analysis of Anthocyanins**. In: MARKAIS, P. Anthocyanins as food colors. New York: Academic Press. p. 181-207.

GIOVENZANA, V.; CIVELLI, R.; BEGHI, R.; OBERTI, R.; GUIDETTI, R. Testing of a simplified LED based vis/NIR system for rapid ripeness evaluation of white grape (*Vitis vinifera* L.) for Franciacorta wine. **Talanta**, v.144, 584-591, 2015.

GHOZLEN, N. B.; CEROVIC, Z. G.; GERMAIN, C.; TOUTAIN, S.; LATOUCHE, G. Non-destructive optical monitoring of grape maturation by proximal sensing. **Sensors**, v.10, 10040-10068, 2010.

GHOZLEN, N.; MOISE, N.; LATOUCHE, G.; MARTINON, V.; MERCIER, L.; BESANÇON, E.; CEROVIC, Z. Assessment of Grapevine Maturity Using a New Portable Sensor: Non-Destructive Quantification of Anthocyanins. **J. Int. Sci. Vigne Vin.**, v.44, 1-8, 2010.

GRAÇA, J. N. D. **Validação de um dispositivo não-invasivo e não-destrutivo para avaliação do estado de maturação das uvas**. 84p. Dissertação (Mestrado em Viticultura e Enologia) - Instituto Superior de Agronomia, Faculdade de Ciências, Universidade do Porto, Lisboa. 2014.

HAZIR, M. H. M.; SHARIFF, A. R. M.; AMIRUDDIN, M. D. Determination of oil palm fresh fruit bunch ripeness-Based on flavonoids and anthocyanin content. **Industrial Crops and Products**, v.36, n.1, 466-475, 2012.

LARRAIN, M.; GUESALAGA, A. R.; AGOSIN, E. A multipurpose portable instrument for determining ripeness in wine grapes using NIR spectroscopy. **IEEE Transactions on Instrumentation and Measurement**, v.57, n.2, 294-302, 2008.

MARTINS, R.C.; LOPES, V. V.; FERREIRA, A. C. S. **Facing the Complexity of Grapes Quality Management and Delivering an Hightroughput Device: VinePAT**. International IVIF – Congress, 2010.

SANTOS PEREIRA, L. F.; BARBON, S.; VALOUS, N. A.; BARBIN, D. F. Predicting the ripening of papaya fruit with digital imaging and Random Forests. **Computers and Electronics in Agriculture**, v.145, n.230, 76-82, 2018.

RUIZ-ALTISENTA, M.; RUIZ-GARCIA, L.; MOREDA, G. P.; RENFU, LU.; HERNANDEZ-SANCHEZ, N.; CORREA, E. C.; DIEZMA, B.; NICOLAI, B.; GARCÍA-RAMOS, J. Sensors for product characterization and quality of specialty crops-A review. **Computers and Electronics in Agriculture**, v.74, n.2, 176-194, 2010.

SAEED, O. M. B.; SANKARAN, S.; SHARIFF, A. R. M.; SHAFRI, H. Z. M.; EHSANI, R.; ALFATNI, M. S.; HAZIR, M. H. M. Classification of oil palm fresh fruit bunches based on their maturity using portable four-band sensor system. **Computers and Electronics in Agriculture**, v.82, 55-60, 2012.

TUCCIO, L.; REMORINI, D.; PINELLI, P.; FIERINI, E.; TONUTTI, P.; SCALABRELLI, G.; AGATI, G. Rapid and non-destructive method to assess in the vineyard grape berry anthocyanins under different seasonal and water conditions. **Australian Journal of Grape and Wine Research**, v.17, n.2, 181-189, 2011.

WOLD, S.; RUHE, A.; WOLD, H.; DUNN, III, W. J. The collinearity problem in linear regression. The partial least square (PLS) approach to generalized inverses. **SIAM Journal on Scientific and Statistical Computing**, v.5, n.3, 735-743, 1984.

ZHENG, H.; JIANG, L.; LOU, H.; HU, Y.; KONG, X.; LU, H. Application of artificial neural network (ANN) and partial least-squares regression (PLSR) to predict the changes of anthocyanins, ascorbic acid, total phenols, flavonoids, and antioxidant activity during storage of red bayberry juice based on fractal analysis and red, green, and blue (RGB) intensity values. **Journal of Agricultural and Food Chemistry**, v.59, n.2, 592-600, 2010.

#### 4. SÍNTESE E DISCUSSÕES GERAIS

No artigo 1, mostrou-se que a espectroscopia VIS-NIR (450 a 1800 nm) associada a modelos preditivos PCR e PLSR é uma excelente ferramenta para a quantificação não destrutiva de sólidos solúveis totais, antocianinas totais e flavonoides amarelos em uvas viníferas da variedade Syrah e Cabernet Sauvignon.

Para as uvas da variedade Syrah, os melhores modelos foram obtidos com a PLSR com desempenho na predição avaliando o  $R^2_{\text{pred}}$  de 0,95, 0,94 e 0,75 e RMSEP de 1,15 °Brix, 16,90 mg.100 g<sup>-1</sup> e 8,73 mg.100 g<sup>-1</sup>, para SST, antocianinas totais e flavonoides amarelos, respectivamente. Para as uvas da variedade Cabernet Sauvignon, os melhores modelos foram conseguidos com a PLSR com desempenho na predição avaliando o  $R^2_{\text{pred}}$  de 0,96, 0,98 e 0,74 e RMSEP de 1,39 °Brix, 13,56 mg.100 g<sup>-1</sup> e 21,21 mg.100 g<sup>-1</sup>, para SST, antocianinas totais e flavonoides amarelos, respectivamente.

O excelente desempenho dos modelos para os atributos SST e antocianinas totais deve-se ao pré-tratamento dos dados espectrais, sendo esta uma etapa fundamental no processo de desenvolvimento dos modelos preditivos, pois Kempes *et al.* (2010) mesmo utilizando a espectroscopia VIS-NIR (320 a 1660 nm) para predição de antocianinas a pH 1,0 e pH 3,2 e concentração de açúcares obtiveram modelos PLSR com desempenhos inferiores comparado aos resultados apresentados no presente trabalho.

Entretanto, modelos mais simplificados e generalistas são mais desejáveis. Assim, foi realizada uma seleção de variáveis com o intuito de eliminar determinados comprimentos de onda que têm pouca ou nenhuma informação espectral relevante para o atributo resposta de interesse. Mostrou-se que apesar da diminuição do número de variáveis, eliminando o sinal espectral pouco informativo, não houve uma diminuição do desempenho dos modelos preditivos para SST e antocianinas totais (Tabela 3.1-3) e além disso os modelos ficaram mais simplificados. Isso está de acordo com Wu *et al.* (2010), os quais observaram que determinados comprimentos de onda têm pouca ou nenhuma contribuição sobre os modelos desenvolvidos, não modelando essas variáveis sem informações e diminuindo o número de fatores dos modelos, e conseqüentemente, tornando o modelo mais simplificado.

Além disso, foram identificadas as assinaturas espectrais dos atributos de qualidade de interesse, sendo elas: 507, 587, 692, 700, 973, 980, 986, 1132, 1154, 1161, 1280, 1324 e 1581 nm para SST; 625, 657, 890, 955, 1154, 1327 e 1389 nm para antocianinas totais; e 671, 692 e 748 nm para flavonoides amarelos. No presente estudo, as assinaturas espectrais identificadas

na região do visível (380 a 780 nm) estão relacionadas à presença de pigmentos verdes (clorofila) e vermelhos (antocianinas e flavonoides), durante o amadurecimento das bagas de uvas (AGATI *et al.*, 2008, 2005; CAO *et al.*, 2010; CEN e HE, 2007; KEMPS *et al.*, 2010). Por sua vez, as assinaturas identificadas na região do infravermelho (780 a 2500 nm) estão relacionadas às bandas de absorção por: (1) açúcares (em torno de 980 nm), devido ao segundo sobretom de O-H; (2) água (em torno de 973, 1324 e 1581 nm), devido ao terceiro sobretom de O-H; (3) e antocianinas (em torno de 1154 nm) devido ao segundo sobretom C-H que está relacionado aos compostos coloridos (CAO *et al.*, 2010; CEN e HE, 2007; CHEN *et al.*, 2015; GONZÁLEZ-CABALLERO *et al.*, 2010; WILLIAMS, 2001).

Além da quantificação dos atributos de qualidade, a espectroscopia mostrou-se capaz de discriminar os estádios de amadurecimento considerados (verde, *véraison* e madura), com precisão de 93,15 % e taxa de erro de 4,83 % associado ao modelo PLS-DA. Fernández-Novales *et al.* (2009a), utilizando a espectroscopia (700 a 1060 nm) e o PLS-DA para discriminar variedades de uvas brancas e tintas conseguiram taxas de acerto de 91,3% e 100%, respectivamente. No presente estudo, a faixa do espectro visível (450 a 780 nm) apresentou maior contribuição em virtude da modulação de coloração ao longo do amadurecimento das bagas.

Tendo em vista a importância dos pré-tratamentos nos sinais espectrais, no artigo 2 foi exposta a aplicação de técnicas de processamentos de sinais, como a transformada *wavelet* e a transformada de Fourier, associadas a classificadores SVM, tanto para a quantificação dos atributos de qualidade, quanto para a qualificação dos estádios de amadurecimento. Os melhores modelos foram conseguidos, na maioria, com a transformada *wavelet* associada ao SVM.

Comparando-se os resultados apresentados no artigo 1 (Tabela 3.1-2 e 3.1-3) com os resultados apresentados no artigo 2 (Tabela 3.2-2) quanto à predição de SST, antocianinas totais e flavonoides amarelos, observa-se uma pequena melhora no desempenho dos modelos construídos. Isso pode estar relacionado ao uso da transformada *wavelet* associada com a SVM, pois a *wavelet* remove informações redundantes e irrelevantes do sinal espectral contribuindo para o aumento do desempenho dos modelos (RINNAN *et al.*, 2009). Isso está de acordo com Nicolai *et al.* (2007a) que sugeriram a combinação de ambas as técnicas (transformada *wavelet* com SVM) para a predição do conteúdo de açúcares em maçãs utilizando a espectroscopia NIR. Em contrapartida o custo computacional para aplicação das técnicas de processamento de sinais, como a transformada *wavelet*, associada ao SVM, é oneroso.

Identificados os comprimentos de onda que correspondem à assinatura espectral, como exposto no artigo 1, foi desenvolvido um dispositivo optoeletrônico com seis comprimentos de ondas (570, 660, 690, 970, 1200 e 1300 nm) para a determinação não destrutiva dos atributos de qualidade em uvas da variedade Cabernet Sauvignon. A calibração e validação cruzada da matriz de dados foi realizada com os modelos *Random Forest* que apresentaram desempenho superior ao MLR para todos os atributos de qualidade. Os valores de  $R^2$  para SST nos modelos MLR e *Random Forest* foram de 0,434 e 0,896, respectivamente. Os valores de  $R^2$  para antocianinas totais nos modelos MLR e *Random Forest* foram de 0,391 e 0,878, respectivamente. Também, os valores de  $R^2$  para flavonoides amarelos nos modelos MLR e *Random Forest* foram de 0,239 e 0,866, respectivamente. Ademais, os valores de MSE para SST, antocianinas totais e flavonoides amarelos nos modelos MLR foram de 22,892 °Brix, 3657,167 mg.100 g<sup>-1</sup> e 190,386 mg.100 g<sup>-1</sup>, respectivamente.

A capacidade preditiva do OPTOFruit para o atributo SST, utilizando o modelo *Random Forest*, foi semelhante ao relatado por Larraín *et al.* (2008), que desenvolveram um instrumento não invasivo para cinco variedades de uvas viníferas obtendo valores de  $R^2$  entre 0,874 à 0,930 para SST, e entre 0,398 à 0,681 para conteúdo de antocianinas, e Giovenzana *et al.* (2015), que desenvolveram um dispositivo óptico que opera com quatro comprimentos de onda (630, 690, 750 e 850 nm) obtendo valores de  $R^2$  de 0,85 e 0,87 para predição do SST e acidez titulável, respectivamente. O desempenho do OPTOFruit foi superior ao dispositivo relatado por Larraín *et al.* (2008) para a determinação de conteúdo pigmentantes (conteúdo de antocianinas e flavonoides amarelos).

Portanto, o OPTOFruit associado com o modelo *Random Forest* apresenta-se promissor na determinação do SST, antocianinas totais e flavonoides amarelos em uvas viníferas da variedade Cabernet Sauvignon.

## 5. CONCLUSÕES GERAIS

Os objetivos do trabalho foram todos atingidos e a hipótese se mostrou verdadeira, para todas as condições estabelecidas. Portanto, em face aos resultados as conclusões foram:

- I. Apresenta-se um instrumento optoeletrônico (OPTOFruit), baseado em seis assinaturas espectrais (570, 660, 690, 970, 1200 e 1300 nm), com precisão comprovada nos testes de validação, para as variáveis:
  - Sólidos solúveis totais:  $R^2 = 0,896$  e  $MSE = 4,796$  °Brix;
  - Antocianinas totais:  $R^2 = 0,878$  e  $MSE = 891,295$  mg.100 g<sup>-1</sup>;
  - Flavonoides amarelos:  $R^2 = 0,866$  e  $MSE = 46,864$  mg.100 g<sup>-1</sup>.
- II. O instrumento pode ser utilizado, com o desempenho obtido, para medições em uvas viníferas da variedade Cabernet Sauvignon em diferentes estádios de maturação. Para outras variedades de uvas é recomendada uma calibração preliminar.
- III. O instrumento OPTOFruit executa a aquisição de dados em tempo real e de forma não destrutiva, podendo auxiliar na tomada de decisão quanto ao momento da colheita das uvas, diminuindo recursos e tempo de análise.

### 5.1. SUGESTÕES PARA CONTINUIDADE DO TRABALHO

Sugere-se que trabalhos futuros avaliem tal dispositivo utilizando LEDs com maior emissão de potência luminosa, com o intuito de minimizar o efeito da luz natural e, consequentemente usá-lo nas condições de campo não retirando as bagas do cacho. Além disso, incorporar outro fotodetector com maior sensibilidade na faixa do visível para registrar às nuances da mudança de coloração nas bagas ao longo do processo de amadurecimento das uvas viníferas tintas.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

ABBOTT, J.A. Quality measurement of fruits and vegetables. **Postharvest Biology and Technology**, v.15, n.3, 207–225, 1999.

ABDULLAH, M. Z.; GUAN, L. C.; LIM, K. C.; KARIM, A. A. The applications of computer vision system and tomographic radar imaging for assessing physical properties of food. **Journal of Food Engineering**, v.61, n.1, 125-135, 2004.

AGATI, G.; PINELLI, P.; CORTÉS EBNER, S.; ROMANI, A.; CARTELAT, A.; CEROVIC, Z. G. Nondestructive evaluation of anthocyanins in olive (*Olea europaea*) fruits by in situ chlorophyll fluorescence spectroscopy. **Journal of Agricultural and Food Chemistry**, v.53, n.5, 1354-1363, 2005.

AGATI, G.; MEYER, S.; MATTEINI, P.; CEROVIC, Z. G. Assessment of anthocyanins in grape (*Vitis vinifera* L) berries using a noninvasive chlorophyll fluorescence method. **Journal of Agricultural and Food Chemistry**, v.55, n.4, 1053-1061, 2007.

AGATI, G.; TRAVERSI, M. L.; CEROVIC, Z. G. Chlorophyll fluorescence imaging for the noninvasive assessment of anthocyanins in whole grape (*Vitis vinifera* L.) bunches. **Photochemistry and photobiology**, v.84, 1431-1434, 2008.

ALANDER, J. T.; BOCHKO, V.; MARTINKAUPPI, B.; SARANWONG, S.; MANTERE, T. A Review of Optical Nondestructive Visual and Near-Infrared Methods for Food Quality and Safety. **International Journal of Spectroscopy**, v.2013, 1-36, 2013.

ANSARI, M. D. Z.; NIRALA, A. K. Biospeckle activity measurement of Indian fruits using the methods of cross-correlation and inertia moments. **Optik**, v.124, n.15, 2180-2186, 2013.

BEGHI, R.; GIOVENZANA, V.; CIVELLI, R.; MALEGORI, C.; BURATTI, S.; GUIDETTI, R. Setting-up of a simplified handheld optical device for decay detection in fresh-cut *Valerianella locusta* L.. **Journal of Food Engineering**, v.127, 10-15, 2014.

BETEMPS, D. L.; FACHINELLO, J. C.; GALARÇA, S. P.; PORTELA, N. M.; REMORINI, D.; MASSAI, R.; AGATI, G. Non-destructive evaluation of ripening and quality traits in apples using a multiparametric fluorescence sensor. **Journal of the Science of Food and Agriculture**, v.92, n.9, 1855-1864, 2012

BOBELYN, E.; SERBAN, A.-S.; NICU, M.; LAMMERTYN, J.; NICOLAI, B. M.; SAEYS, W. Postharvest quality of apple predicted by NIR-spectroscopy: Study of the effect of

biological variability on spectra and model performance. **Postharvest Biology and Technology**, v.55, n.3, 133-143, 2010.

CAO, F.; WU, D.; HE, Y. Soluble solids content and pH prediction and varieties discrimination of grapes based on visible-near infrared spectroscopy. **Comput. Electron. Agric.**, v.71, 15-18, 2010.

CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. **Trends in Food Science & Technology**, v.18, n.2, 72-83, 2007.

CEROVIC, Z. G.; MOISE, N.; AGATI, G.; LATOUCHE, G.; BEN GHOZLEN, N.; MEYER, S. New portable optical sensors for the assessment of winegrape phenolic maturity based on berry fluorescence. **Journal of Food Composition and Analysis**, v.21, 650-654, 2008.

CEROVIC, Z. G.; GOUTOULY, J-P.; HILBERT, G.; DESTRAK-IRVINE, A.; MARTINON, V.; MOISE, N. **Mapping winegrape quality attributes using portable fluorescence-based sensors**. In: FRUTIC 09, PROGAP INIA, 301-310, 2009.

CHEN, S.; ZHANG, F.; NING, J.; LIU, X.; ZHANG, Z.; YANG, S. Predicting the anthocyanin content of wine grapes by NIR hyperspectral imaging. **Food Chem.**, v.172, 788–793, 2015.

CHOONG, T. S. Y.; ABBAS, S.; SHARIFF, A. R.; HALIM, R.; ISMAIL, M. H. S.; YUNUS, R.; ALI, S.; AHMADUN, F.-R. Digital Image Processing of Palm Oil Fruits. **International Journal of Food Engineering**, v.2, n.2, 1556-1560, 2006.

CIVELLI, R.; GIOVENZANA, V.; BEGHI, R.; NALDI, E.; GUIDETTI, R.; OBERTI, R. A Simplified, Light Emitting Diode (LED) Based, Modular System to be Used for the Rapid Evaluation of Fruit and Vegetable Quality: Development and Validation on Dye Solutions. **Sensors**, v.15, n.9, 22705–22723, 2015.

COSTA, A. G. **Identificação da maturação de frutos da macaúba por meio de características óticas**. 143p. Tese (Doutorado Engenharia Agrícola) – Departamento de Engenharia Agrícola, Universidade Federal de Viçosa, Viçosa, 2015.

COSTA, A. G.; PINTO, F. A. C.; MOTOIKE, S. Y.; BRAGA JÚNIOR, R. A.; GRACIA, L. M. N. Determination of macaw fruit harvest period by biospeckle laser technique. **African Journal of Agricultural Research**, v.12, 674-683, 2017.

COSTA, A. G.; PINTO, F. A. C.; MOTOIKE, S. Y.; BRAGA JÚNIOR, R. A.; GRACIA, L. M. N. Classification of macaw palm fruits from colorimetric properties for determining the harvest moment. **Engenharia Agrícola**, v.38, 634-641, 2018.

FERNÁNDEZ-NOVALES, J.; LÓPEZ, M.I.; SÁNCHEZ, M.T.; GARCÍA-MESA, J.A.; GONZÁLEZ-CABALLERO, V. Assessment of quality parameters in grapes during ripening using a miniature fiber-optic near-infrared spectrometer. **Int. J. Food Sci. Nutr.**, v.60, 265–277, 2009a

FERNÁNDEZ-NOVALES, J.; LÓPEZ, M.-I.; SÁNCHEZ, M.-T.; MORALES, J.; GONZÁLEZ-CABALLERO, V. Shortwave-near infrared spectroscopy for determination of reducing sugar content during grape ripening, winemaking, and aging of white and red wines. **Food Research International**, v.42, n.2, 285-291, 2009b.

FLORES, K.; SÁNCHEZ, M.-T.; PÉREZ-MARÍN, D.; GUERRERO, J.-E.; GARRIDO-VARO, A. Feasibility in NIRS instruments for predicting internal quality in intact tomato. **Journal of Food Engineering**, v.91, n.2, 311-318, 2009.

GIOVENZANA, V.; CIVELLI, R.; BEGHI, R.; OBERTI, R.; GUIDETTI, R. Testing of a simplified LED based vis/NIR system for rapid ripeness evaluation of white grape (*Vitis vinifera* L.) for Franciacorta wine. **Talanta**, v.144, 584-591, 2015.

GHOZLEN, N. B.; CEROVIC, Z. G.; GERMAIN, C.; TOUTAIN, S.; LATOUCHE, G. Non-destructive optical monitoring of grape maturation by proximal sensing. **Sensors**, v.10, 10040-10068, 2010.

GHOZLEN, N.; MOISE, N.; LATOUCHE, G.; MARTINON, V.; MERCIER, L.; BESANÇON, E.; CEROVIC, Z. Assessment of Grapevine Maturity Using a New Portable Sensor: Non-Destructive Quantification of Anthocyanins. **J. Int. Sci. Vigne Vin.**, v.44, 1-8, 2010.

GONZÁLEZ-CABALLERO, V.; SÁNCHEZ, M.T.; LÓPEZ, M.I.; PÉREZ-MARÍN, D. First steps towards the development of a non-destructive technique for the quality control of wine grapes during on-vine ripening and on arrival at the winery. **Journal of Food Engineering**, v.101, n.2, 158-165, 2010.

GRAÇA, J. N. D. **Validação de um dispositivo não-invasivo e não-destrutivo para avaliação do estado de maturação das uvas**. 84p. Dissertação (Mestrado em Viticultura e Enologia) - Instituto Superior de Agronomia, Faculdade de Ciências, Universidade do Porto, Lisboa. 2014.

HAGEN, S. F.; SOLHAUG, K. A.; BENGTSSON, G. B.; BORGE, G. I. A.; BILGER, W. Chlorophyll fluorescence as a tool for non-destructive estimation of anthocyanins and total flavonoids in apples. **Postharvest Biology and Technology**, v.41, n.2, 156-163, 2006.

HAZIR, M. H. M.; SHARIFF, A. R. M.; AMIRUDDIN, M. D. Determination of oil palm fresh fruit bunch ripeness-Based on flavonoids and anthocyanin content. **Industrial Crops and Products**, v.36, n.1, 466-475, 2012a.

HAZIR, M. H. M.; SHARIFF, A. R. M.; AMIRUDDIN, M. D.; RAMLI, A. R.; IQBAL SARIPAN, M. Oil palm bunch ripeness classification using fluorescence technique. **Journal of Food Engineering**, v.113, n.4, 534-540, 2012b.

HERRERA, J.; GUESALAGA, A.; AGOSIN, E. Shortwave near infrared spectroscopy for non-destructive determination of maturity of wine grapes. **Measurement Science and Technology**, v.14, n.5, 698-697, 2003.

JUNKWON, P.; TAKIGAWA, T.; OKAMOTO, H.; HASEGAWA, H.; KOIKE, M.; SAKAI, K.; SIRUNTAWINETI, J.; CHAEYCHOMSRI, W.; SANEVAS, N.; TITTINUCHANON, P.; BAHALAYODHIN, B. Potential Application of Color and Hyperspectral Images for Estimation of Weight and Ripeness of Oil Palm (*Elaeis guineensis* Jacq. var. *tenera*). **Agricultural Information Research**, v.18, n.2, 72-81, 2009.

KEMPS, B.; LEON, L.; BEST, S.; DE BAERDEMAEKER, J.; DE KETELAERE, B. Assessment of the quality parameters in grapes using VIS/NIR spectroscopy. **Biosystems Engineering**, v.105, n.4, 507-513, 2010.

KENNEDY, J. Understanding grape berry development. **Practical Winery & Vineyard**. p. 5, july/august, 2002.

KOLB, C. A.; WIRTH, E.; KAISER, W. M.; MEISTER, A.; RIEDERER, M.; PFÜNDEL, E. E. Noninvasive evaluation of the degree of ripeness in grape berries (*Vitis Vinifera* L Cv. *Bacchus* and *Silvaner*) by chlorophyll fluorescence. **Journal of Agricultural and Food Chemistry**, v.54, n.2, 299-305, 2006.

LARRAIN, M.; GUESALAGA, A. R.; AGOSIN, E. A multipurpose portable instrument for determining ripeness in wine grapes using NIR spectroscopy. **IEEE Transactions on Instrumentation and Measurement**, v.57, n.2, 294-302, 2008.

LECHAUDEL, M.; URBAN, L.; JOAS, J. Chlorophyll fluorescence, a nondestructive method to assess maturity of mango fruits (Cv. 'cogshall') without growth conditions bias. **Journal of Agricultural and Food Chemistry**, v.58, n.13, 7532-7538, 2010.

MANDELLI, F.; BERLATTO, M. A.; TONIETTO, J.; BERGAMASCHI, H. Fenologia da videira na Serra Gaúcha. **Pesquisa Agropecuária Gaúcha**, Porto Alegre, v.9, 129-144, 2003.

MARTINS, R.C.; LOPES, V. V.; FERREIRA, A. C. S. **Facing the Complexity of Grapes Quality Management and Delivering an Hightthroughput Device: VinePAT**. International IVIF – Congress, 2010.

NICOLAÏ, B. M.; BEULLENS, K.; BOBELYN, E.; PEIRS, A.; SAEYS, W.; THERON, K. I.; LAMMERTYN, J. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. **Postharvest Biology and Technology**, v.46, n.2, 99-118, 2007

NICOLAÏ, B. M.; THERON, K. I.; LAMMERTYN, J. Kernel PLS regression on wavelet transformed NIR spectra for prediction of sugar content of apple. **Chemometrics and Intelligent Laboratory Systems**, v.85, n.2, 243–252, 2007a.

OIV. International Organisation of Vine and Wine. **2017 World Vitiviniculture Situation: OIV Statistical Report on World Vitiviniculture**, 2017.

PAZ, P.; SÁNCHEZ, M.-T.; PÉREZ-MARÍN, D.; GUERRERO, J. E.; GARRIDO-VARO, A. Instantaneous quantitative and qualitative assessment of pear quality using near infrared spectroscopy. **Computers and Electronics in Agriculture**, v.69, n.1, 24-31, 2009.

RAZALI, M. H.; WAN ISMAIL, W. I.; RAMLI, A. R.; SULAIMAN, M. N. Modeling of Oil Palm Fruit Maturity for the Development of an Outdoor Vision System. **International Journal of Food Engineering**, v.4, n.3, 2008.

RINNAN, Å.; VAN DEN BERG, F.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC Trends in Analytical Chemistry**, v.28, n.10, 1201-1222, 2009.

RUIZ-ALTISENTA, M.; RUIZ-GARCIA, L.; MOREDA, G. P.; RENFU, LU.; HERNANDEZ-SANCHEZ, N.; CORREA, E. C.; DIEZMA, B.; NICOLAI, B.; GARCÍA-RAMOS, J. Sensors for product characterization and quality of specialty crops-A review. **Computers and Electronics in Agriculture**, v.74, n.2, 176-194, 2010.

SAEED, O. M. B.; SANKARAN, S.; SHARIFF, A. R. M.; SHAFRI, H. Z. M.; EHSANI, R.; ALFATNI, M. S.; HAZIR, M. H. M. Classification of oil palm fresh fruit bunches based on their maturity using portable four-band sensor system. **Computers and Electronics in Agriculture**, v.82, 55-60, 2012.

SHAO, Y.; HE, Y.; GÓMEZ, A. H.; PEREIR, A. G.; QIU, Z.; ZHANG, Y. Visible/near infrared spectrometric technique for nondestructive assessment of tomato 'Heatwave' (*Lycopersicon esculentum*) quality characteristics. **Journal of Food Engineering**, v.31, n.4, 672-678, 2007.

SOARES, J. M.; LEÃO, P. C. S. **A vitivinicultura no Semiárido brasileiro**. Petrolina: Embrapa, 2009.

SOUZA, G. M. **Avaliação não invasiva do desenvolvimento de frutos de videira pela fluorescência da clorofila a**. 80p. Dissertação (Mestrado em Produção Vegetal) - Centro de Ciências e Tecnologias Agropecuárias, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campo dos Goytacazes. 2009.

TUCCIO, L.; REMORINI, D.; PINELLI, P.; FIERINI, E.; TONUTTI, P.; SCALABRELLI, G.; AGATI, G. Rapid and non-destructive method to assess in the vineyard grape berry anthocyanins under different seasonal and water conditions. **Australian Journal of Grape and Wine Research**, v.17, n.2, 181-189, 2011.

WILLIAMS, P. C., 2001. Implementation of near-infrared technology. In P. Williams & K. Norris (Eds.), **Near-infrared technology in the agricultural and food industries** (2nd ed., pp. 145–169). St. Paul, MN: American Association of Cereal Chemists.

WU, D.; HE, Y.; NIE, P.; CAO, F.; BAO, Y. Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice. **Analytica Chimica Acta**, v.659, 229-237, 2010.

ZDUNEK, A.; ADAMIAK, A.; PIECZYWEK, P. M.; KURENDA, A. The biospeckle method for the investigation of agricultural crops: A review. **Optics and Lasers in Engineering**, v.52, n.1, 276-285, 2014.