



**UNIVERSIDADE ESTADUAL DE CAMPINAS  
FACULDADE DE ENGENHARIA DE ALIMENTOS**

**INGRID ALVES DE MORAES**

**AVALIAÇÃO E CLASSIFICAÇÃO DE CARAMBOLA USANDO  
ESPECTROMETRO PORTÁTIL NO INFRAVERMELHO PRÓXIMO  
(NIR)**

**CAMPINAS**

**2021**

INGRID ALVES DE MORAES

**AVALIAÇÃO E CLASSIFICAÇÃO DE CARAMBOLA USANDO  
ESPECTROMETRO PORTÁTIL NO INFRAVERMELHO PRÓXIMO  
(NIR)**

Dissertação apresentada à Faculdade de Engenharia de Alimentos da Universidade Estadual de Campinas como parte dos requisitos exigidos para obtenção do título de Mestra em Engenharia de Alimentos.

**Orientador:** Prof. Dr. Douglas Fernandes Barbin

Este trabalho corresponde à versão final da dissertação defendida pela aluna Ingrid Alves de Moraes, e orientado pelo Prof. Dr. Douglas Fernandes Barbin.

**CAMPINAS**

**2021**

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca da Faculdade de Engenharia de Alimentos  
Claudia Aparecida Romano - CRB 8/5816

M791a Moraes, Ingrid Alves de, 1996-  
Avaliação e classificação de carambola usando espectrômetro portátil no infravermelho próximo (NIR) / Ingrid Alves de Moraes. – Campinas, SP : [s.n.], 2021.

Orientador: Douglas Fernandes Barbin.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia de Alimentos.

1. Aprendizado de máquina. 2. Frutas – Qualidade. 3. Maturação. 4. Espectroscopia. I. Barbin, Douglas Fernandes. II. Universidade Estadual de Campinas. Faculdade de Engenharia de Alimentos. III. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** The evaluation and classification of carambola using portable near infrared spectrometer (NIR)

**Palavras-chave em inglês:**

Machine learning

Fruit - Quality

Maturation

Spectroscopy

**Área de concentração:** Engenharia de Alimentos

**Titulação:** Mestra em Engenharia de Alimentos

**Banca examinadora:**

Douglas Fernandes Barbin [Orientador]

Daniela Souza Ferreira

Raul Benito Siche Jara

**Data de defesa:** 10-08-2021

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

**Identificação e informações acadêmicas do(a) aluno(a)**

- ORCID do autor: <https://orcid.org/0000-0002-8887-423>

- Currículo Lattes do autor: <http://lattes.cnpq.br/6944338935374468>

**COMISSÃO EXAMINADORA**

**Prof. Dr. Douglas Fernandes Barbin**

ORIENTADOR - Faculdade de Engenharia de Alimentos, UNICAMP

**Prof<sup>a</sup>. Dr<sup>a</sup>. Daniela Souza Ferreira**

MEMBRO TITULAR

**Prof. Dr. Raúl Benito Siche Jara**

MEMBRO TITULAR - Universidad Nacional de Trujillo

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

## DEDICATÓRIA

*À maior dádiva divina a mim: minha família.*

*Aos meus pais Alessandro e Sabrina,  
minha irmã Giovanna, meu noivo Victor e  
meus avós Orivaldo, Marly, Pedro e Cleuza.*

## **AGRADECIMENTOS**

Agradeço, a Deus por me abençoar e me dar forças e sabedoria para seguir meu caminho.

Aos meus pais, meus pilares, Alessandro e Sabrina por todos os esforços dedicados a realização dos meus objetivos.

À minha irmã Giovanna, meu noivo Victor e meus avós por todo o apoio, ensinamentos, amor incondicional e por sempre me incentivar e acreditar em mim.

Ao meu orientador Prof. Dr. Douglas Fernandes Barbin pelos ensinamentos e oportunidades de crescimento, paciência, incentivo e compreensão durante o mestrado.

Ao meu colega de trabalho, Luis Pier, por ter tido tanta paciência e por agregar tantos conhecimentos à pesquisa.

Às minhas amigas Isabella, Gabriela, Rafaella que mesmo distantes sempre me motivaram, ouviram e me apoiaram.

Aos amigos que fiz no Mestrado Dani, Adriana, Ederson, pelas reuniões, risadas e cervejas.

A todos os colegas da FEA que, diretamente ou indiretamente colaboraram para a realização do projeto, em especial, Marciano, Marcus, Yasmin, Lucimar, Renann, Karina, Matheus e Tatiana.

Este trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) Projeto 2019/12625-1 (bolsa de mestrado) e 2015/24351-2 (Jovem Pesquisador), e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## RESUMO

*Averrhoa carambola* conhecida como fruta estrela, devido a sua forma, é consumida em altas quantidades na Ásia e Américas do Sul e Central por ser rica em vitaminas e minerais. A classificação da qualidade da carambola é conduzida por avaliadores treinados através de inspeções visuais. Trata-se de um método dispendioso e subjetivo, que pode gerar alta variabilidade nos resultados, pois a análise repetitiva de milhares de amostras, conduzidas diariamente, dificulta a padronização das classes. Assim, o objetivo deste trabalho foi utilizar espectroscopia de infravermelho próximo (NIR) como análise não destrutiva para predição e classificação dos estádios de maturação da carambola. Primeiramente, foram obtidos os espectros de absorbância num espectrômetro NIR portátil. Em seguida, foram determinadas as características de cor e composição de sólidos solúveis, vitamina C, umidade, pH e acidez titulável, por métodos analíticos de referência. Análises multivariadas como análises de componentes principais (PCA), regressão de mínimos quadrados parciais (PLSR) e a análise discriminante pelo método de mínimos quadrados parciais (PLS-DA) foram utilizadas para prever os parâmetros físicos e físico-químicos das amostras. Foram testados pré-tratamentos nos espectros, incluindo primeira e segunda derivadas, a fim de reduzir as variações causadas por fatores não relacionados à composição, e melhorar a precisão dos modelos estatísticos. A seleção de variáveis pelo método dos mínimos quadrados por intervalo (iPLS) e Algoritmo genético (GA-PLS) permitiram uma melhoria no desempenho dos modelos PLSR, atingindo ( $R^2_P$  de 0,78 e 0,74), (RMSEP de 0,2 e 0,87), (RPD de 2,01 e 2,23) e (RER de 8,02 e 10,38), para pH e teor de umidade, respectivamente, o que sugere serem modelos aceitáveis para controle de qualidade com previsões quantitativas com baixos erros. O modelo PLS-DA com seleção de variáveis pelo iPLS obteve o melhor desempenho com 84,2% de precisão para classificar a carambola conforme o estágio de maturação. O espectrômetro NIR portátil juntamente com a quimiometria pode ser uma ferramenta promissora de baixo custo para quantificar parâmetros físico-químicos e classificar a carambola a partir do estágio de maturação.

**Palavras-chave:** aprendizado de máquina; qualidade da fruta; maturação; espectroscopia

## ABSTRACT

Averrhoa carambola is known as star fruit, due to its shape. It is consumed in high quantities in Asia and South and Central America as it is a source of vitamins and minerals. The classification of carambola is carried out by trained evaluators through visual inspections. It is an expensive and subjective method, which can generate high variability in the results, since the repetitive analysis of thousands of samples, makes it difficult to standardize this task. Thus, the aim of this project was to use near-infrared spectroscopy (NIR) as a non-destructive method for predicting and classifying carambola maturity stages. First, the absorbance spectra were obtained on a portable NIR spectrometer. Then, the colour and composition characteristics of soluble solids, vitamin C, moisture, pH and titratable acidity were determined by reference analytical methods. Multivariate analyses such as principal component analysis (PCA), partial least squares regression (PLSR) partial least squares and discriminant analysis (PLS-DA) were used to predict the physical and physical-chemical parameters of the samples. Pre-treatment was tested in spectra, including first and second derivatives, to reduce variations caused by factors unrelated to composition and improve the accuracy of statistical models. The selection of variables by Iterative partial least square (iPLS) and Genetic algorithm (GA-PLS) improved the performance of the PLSR models, reaching ( $R^2_P$  of 0.78 and 0.74), (RMSEP of 0.2 and 0.87), (RPD of 2.01 and 2.23) and (RER of 8.02 and 10.38), for pH and moisture content, respectively, which suggests it to be acceptable models for quality control with quantitative predictions with low errors. The PLS-DA model with variable selection by iPLS obtained the best performance with 84.2 % accuracy to classify a carambola according to the stage of maturity. The portable NIR spectrometer together with chemometry can be a promising low-cost tool to quantify the physicochemical attributes and classify the carambola according to the maturity stage.

**Keywords:** machine learning; fruit quality; maturation; spectroscopy

## LISTA DE FIGURAS

### CAPÍTULO 3 – REVISAO BIBLIOGRÁFICA

|                                                                                     |    |
|-------------------------------------------------------------------------------------|----|
| <b>FIGURA 3.1.</b> Carambolas em diferentes estádios de maturação. ....             | 22 |
| <b>FIGURA 3.2.</b> Organização da matriz x a partir da aquisição espectral NIR..... | 25 |
| <b>FIGURA 3.3.</b> Contribuição existente no sinal medido.....                      | 26 |

### CAPÍTULO 4 – ONLINE MEASUREMENT OF CARAMBOLA (AVERRHOA CARAMBOLA L.) PHYSICOCHEMICAL PROPERTIES AND ESTIMATION OF MATURITY STAGES USING A PORTABLE NIR SPECTROMETER

|                                                                                                                                                                                                                                                                                           |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>FIGURE 4.1.</b> Maturity stages for carambola according to hue ( $h^{\circ}_{ab}$ ) value .....                                                                                                                                                                                        | 53 |
| <b>FIGURE 4.2.</b> A) Carambola in different maturity stages (MS1, MS2, MS3 AND MS4), B) mean spectra of carambola in different maturity stages, C) scores plot based on 1 <sup>st</sup> s-g derivative + smoothing s-g spectra ( 902–1637 NM), and D) loadings plot of PCA analysis..... | 56 |
| <b>FIGURE 4.3.</b> Confusion matrix of PLS-DA models for calibration (cal) set and validation (test) set using full spectra (A and B) and reduced spectra by iPLS (C and D).....                                                                                                          | 58 |
| <b>FIGURE 4.4.</b> PLSR models with best performance for total titratable acidity (TTA), total of soluble solids (TSS), pH, moisture and ascorbic acid (AA). ....                                                                                                                         | 63 |
| <b>FIGURE 4.5.</b> A) Raw spectra, B) spectra with region 913–1637 nm, C) full spectral data and D) mean spectra pre-processing with 1 <sup>st</sup> s-g derivative + smoothing s-g spectra.....                                                                                          | 79 |
| <b>figure 4.6.</b> Loadings of PLS-DA model in spectral range of 905–1640 nm.....                                                                                                                                                                                                         | 80 |
| <b>FIGURE 4.7.</b> High light dispersion.....                                                                                                                                                                                                                                             | 81 |

## LISTA DE TABELAS

### CAPÍTULO 3 – REVISAO BIBLIOGRÁFICA

|                                                                        |    |
|------------------------------------------------------------------------|----|
| <b>Tabela 3.1.</b> Trabalhos com o uso do NIR em diversas frutas. .... | 36 |
|------------------------------------------------------------------------|----|

### CAPÍTULO 4 – ONLINE MEASUREMENT OF CARAMBOLA (AVERRHOA CARAMBOLA L.) PHYSICOCHEMICAL PROPERTIES AND ESTIMATION OF MATURITY STAGES USING A PORTABLE NIR SPECTROMETER

|                                                                                                                                                                                 |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Table 4.1.</b> Statistical results of the physicochemical properties of the carambola samples in four different maturity stages. ....                                        | 53 |
| <b>Table 4.2.</b> Statistics for the calibration and prediction sets for physical and physico-chemical analysis in carambola using portable NIR and PLSR, iPLS and GA-PLS. .... | 59 |
| <b>Tabela 4.3.</b> PLS-DA models based on spectra data to discriminate four maturation level (MS1, MS2, MS3 AND MS4). ....                                                      | 72 |
| <b>Table 4.4.</b> PLSR models based on spectra data to the parameters analysed in carambola fruit. ....                                                                         | 75 |

## LISTA DE ABREVIATURAS

*AA*– Teor de ácido ascórbico

*ATT*– Teor de Acidez titulável

*GA* – Algoritmo genético

*iPLS* – Método dos mínimos quadrados por intervalo

*LV* – Variáveis latentes

*MC* – Centragem na média

*MSC* –Correção multiplicativa de espalhamento

*PCA* – Análise por componentes principais

*PLSR* – Método por mínimos quadrados parciais

*PLS-DA* – Análise discriminante de mínimos quadrados parciais

*RMSEC* – Raiz da média quadrática de calibração

*RMSECV* – Raiz da média quadrática de validação cruzada

*RMSEP* – Raiz da média quadrática de predição

*RER* – Razão da faixa de erro

*RPD* – Taxa de desvio padrão

$R^2_c$  – Coeficiente de correlação de calibração

$R^2_{cv}$  – Coeficiente de correlação de validação cruzada

$R^2_p$  – Coeficiente de correlação de predição

*S-G* – Savitzky–Golay

*SNV* – Padronização normal de sinal

*SST* – Teor de sólidos solúveis

*VSN* – Classificação de variáveis para normalização

## SUMÁRIO

|                                                                                                                                                                                        |           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>CAPÍTULO 1 – INTRODUÇÃO GERAL .....</b>                                                                                                                                             | <b>14</b> |
| <b>1.1. INTRODUÇÃO GERAL .....</b>                                                                                                                                                     | <b>15</b> |
| <b>CAPÍTULO 2 – OBJETIVOS .....</b>                                                                                                                                                    | <b>17</b> |
| <b>2.1. OBJETIVO GERAL .....</b>                                                                                                                                                       | <b>18</b> |
| <b>2.2. OBJETIVOS ESPECÍFICOS.....</b>                                                                                                                                                 | <b>18</b> |
| <b>CAPÍTULO 3 – REVISAO BIBLIOGRÁFICA.....</b>                                                                                                                                         | <b>19</b> |
| <b>3.1. CARAMBOLA (<i>AVERRHOA CARAMBOLA</i> L.). .....</b>                                                                                                                            | <b>20</b> |
| <b>3.2. PARÂMETROS DA QUALIDADE DA CARAMBOLA .....</b>                                                                                                                                 | <b>21</b> |
| <b>3.2.1. Maturação e parâmetros de classificação .....</b>                                                                                                                            | <b>21</b> |
| <b>3.3. TÉCNICAS ANALÍTICAS NÃO DESTRUTIVAS .....</b>                                                                                                                                  | <b>23</b> |
| <b>3.3.1. Espectroscopia no infravermelho próximo (NIR).....</b>                                                                                                                       | <b>23</b> |
| <b>3.4. QUIMIOMETRIA.....</b>                                                                                                                                                          | <b>25</b> |
| <b>3.4.1. Pré-tratamento .....</b>                                                                                                                                                     | <b>26</b> |
| <b>3.4.2. Análise exploratória .....</b>                                                                                                                                               | <b>28</b> |
| <b>3.4.3. Regressão.....</b>                                                                                                                                                           | <b>29</b> |
| <b>3.4.4. Discriminação .....</b>                                                                                                                                                      | <b>30</b> |
| <b>3.4.5. Avaliação da performance dos modelos .....</b>                                                                                                                               | <b>31</b> |
| 3.4.5.1. Modelos de regressão .....                                                                                                                                                    | 31        |
| 3.4.5.2. Modelos de classificação .....                                                                                                                                                | 32        |
| <b>3.4.6. Seleção de variáveis .....</b>                                                                                                                                               | <b>33</b> |
| <b>CAPÍTULO 4 - ONLINE MEASUREMENT OF CARAMBOLA (<i>AVERRHOA CARAMBOLA</i> L.) PHYSICOCHEMICAL PROPERTIES AND ESTIMATION OF MATURITY STAGES USING A PORTABLE NIR SPECTROMETER.....</b> | <b>42</b> |
| <b>4.1. INTRODUCTION .....</b>                                                                                                                                                         | <b>47</b> |

|                                                                       |           |
|-----------------------------------------------------------------------|-----------|
| <b>4.2. MATERIAL AND METHODS.....</b>                                 | <b>48</b> |
| <b>4.2.1. Carambola samples .....</b>                                 | <b>48</b> |
| <b>4.2.2. Physical and physico-chemical analyses.....</b>             | <b>49</b> |
| <b>4.2.3. Acquisition of NIR spectra .....</b>                        | <b>49</b> |
| 4.2.3.1. Spectra pre-processing.....                                  | 49        |
| 4.2.3.2. Exploratory analysis .....                                   | 50        |
| 4.2.3.3. Classification of maturity stage .....                       | 50        |
| 4.2.3.4. Prediction of physicochemical properties.....                | 51        |
| 4.2.3.5. Variable Selection.....                                      | 51        |
| <b>4.2.4. Statistical analysis .....</b>                              | <b>52</b> |
| <b>4.3. RESULTS AND DISCUSSION.....</b>                               | <b>52</b> |
| <b>4.3.1. Physicochemical properties .....</b>                        | <b>52</b> |
| <b>4.3.2. Spectra profile and exploratory analysis .....</b>          | <b>54</b> |
| <b>4.3.3. Discrimination model.....</b>                               | <b>57</b> |
| <b>4.3.4. Prediction of physico-chemical properties by PLSR .....</b> | <b>58</b> |
| <b>4.4. CONCLUSION .....</b>                                          | <b>63</b> |
| <b>4.5. REFERENCES .....</b>                                          | <b>64</b> |
| <b>CAPÍTULO 5 – DISCUSSÃO GERAL.....</b>                              | <b>82</b> |
| <b>5.1. DISCUSSÃO GERAL .....</b>                                     | <b>83</b> |
| <b>CAPÍTULO 6 – CONCLUSÃO GERAL .....</b>                             | <b>84</b> |
| <b>6.1. CONCLUSÃO GERAL .....</b>                                     | <b>85</b> |
| <b>CAPÍTULO 7 – REFERÊNCIAS .....</b>                                 | <b>86</b> |
| <b>7.1. REFERÊNCIAS .....</b>                                         | <b>87</b> |

## **CAPÍTULO 1 – INTRODUÇÃO GERAL**

---

### 1.1. INTRODUÇÃO GERAL

A presença dos frutos na alimentação tem resultado em um crescimento global na produção e impulsionado pesquisas para melhorias da produção e qualidade (FAO, 2019). O Brasil, atualmente, é o 3º maior produtor mundial de frutas (MAPA, 2021). A *Averrhoa carambola* L. também conhecida como *star fruit* ou carambola é fonte rica de minerais e vitaminas, incluindo magnésio, potássio e antioxidantes, como vitamina C e carotenoides (JIA et al., 2017; MUTHU et al., 2016). Devido ao seu valor nutricional, a carambola está presente na alimentação, ao ser consumida in natura, assim como, em saladas, bebidas de coquetéis, chás, sucos e geleias (LIM, 2012; MUTHU et al., 2016). Essa diversidade do uso tem estimulado os comerciantes a proporcionarem frutos com mais alto padrão de qualidade, a preços acessíveis e que atendam aos consumidores; que estão se tornando mais conscientes dos parâmetros físico-químicos relacionados a qualidade dos frutos (ALANDER et al., 2013; YAHAYA; OMAR, 2017).

A qualidade dos frutos está diretamente relacionada com parâmetros como estágio de maturação, sólidos solúveis, acidez titulável, vitamina C; e a medição desses parâmetros pode ser de grande auxílio a uma adequada classificação (AMIR et al., 2019; KYRIACOU; ROUPHAEL, 2018). A colheita dos frutos no estágio de maturação adequado atribui qualidade ao produto; frutos colhidos imaturos ou excessivamente maduros estão sujeitos a ficarem murchos, sem nutrientes e sabor adequado, e a sofrerem danos mecânicos (JIE; WEI, 2018; MADUWANTHI; MARAPANA, 2019). As técnicas analíticas tradicionais para determinação desses parâmetros físico-químicos são destrutivas, lentas e envolvem o uso de produtos químicos e que podem ser prejudiciais para o técnico de laboratório e o meio ambiente (ALANDER et al., 2013; YAHAYA; OMAR, 2017). Entretanto, a indústria de frutas requer técnicas analíticas rápidas, não destrutivas, e com o mínimo de reagentes, para implementação de medições on-line em linha.

A Espectroscopia no Infravermelho Próximo (NIR) é capaz de padronizar, prever e quantificar os parâmetros de qualidade (CHANDRASEKARAN et al., 2019). Recentemente, a redução no tamanho dos componentes permitiu o desenvolvimento de espectrômetros NIR de mão / portáteis a baixos custos quando comparadas a outras técnicas não destrutivas (PASQUINI, 2018). Os espectrômetros NIR portáteis possuem design ergonômico, são transportáveis para diferentes ambientes de trabalho e, quando bem calibrados, podem ser uma alternativa para dispositivos de bancada (CRUZ-TIRADO; MEDEIROS; BARBIN, 2021). Trata-se de um método de análise rápido com variada aplicação como previsão de parâmetros de qualidade do tomate (BORBA et al., 2021),

determinação da qualidade em maçã (PISSARD et al., 2021), previsões de sólidos solúveis, acidez, acidez titulável, índice de maturação, firmeza de polpa de laranja (CAYUELA; WEILAND, 2010) e a avaliação da qualidade e os estádios de maturação em melancias (QI et al., 2014) e em mangas (SARANWONG; SORNSRIVICHAI; KAWANO, 2004).

Na espectroscopia no infravermelho próximo (NIR) em frutas frescas, a interação da luz resulta em dois fenômenos principais, absorção e dispersão (Lu et al., 2020). A absorbância está relacionada aos componentes químicos da fruta, e a dispersão à microestrutura física da casca e polpa (SAEYS et al., 2019). A dispersão pode conter informações importantes durante o amadurecimento do fruto, e esse conhecimento é relevante para a predição de algumas propriedades físico-químicas, pois a estrutura do fruto está correlacionada com o estágio de maturação do mesmo (Lu et al., 2020; Saeys et al., 2019). Com isso, há necessidade de estudos aprofundados em análises quimiométricas e o seu uso em frutas frescas, testando qual a melhor abordagem ou pré-tratamento a ser utilizado (MISHRA et al., 2021b).

No entanto, até onde sabemos, não há pesquisas disponíveis na literatura sobre avaliação de propriedades físico-químicas da carambola e classificação da maturidade ou mesmo o uso de equipamentos portáteis. Portanto, esse trabalho teve como objetivo desenvolver uma metodologia baseada em um espectrômetro NIR portátil para a determinação online das propriedades físico-químicas e do estágio de maturação da carambola. Além disso, foram construídos modelos utilizando o espectro com todas as faixas espectrais e modelos com seleção de variáveis relevantes às propriedades de interesse.

## **CAPÍTULO 2 – OBJETIVOS**

---

### **2.1. OBJETIVO GERAL**

Classificar a fruta carambola quanto ao estágio de maturação e prever parâmetros físicos e físico-químicos utilizando um equipamento portátil de espectroscopia no infravermelho próximo (NIR).

### **2.2. OBJETIVOS ESPECÍFICOS**

- Definir o estágio de maturação da carambola usando o parâmetro de cor Hue.
- Medir e comparar as propriedades físico-químicas para cada estágio de maturação.
- Usar métodos supervisionados para classificar as frutas de carambola segundo seu estágio de maturação.
- Estimar os valores dos parâmetros físico-químicos, que inclui teor de sólidos solúveis, pH, vitamina C, acidez titulável e umidade, usando modelos de regressão com os espectros NIR como preditores.

## **CAPÍTULO 3 – REVISAO BIBLIOGRÁFICA**

---

### 3.1. CARAMBOLA (*AVERRHOA CARAMBOLA* L.).

Durante séculos, *Averrhoa carambola* L., pertencente à família Oxalidaceae, tem sido cultivada nos trópicos devido seu succulento fruto, bem como seu caráter decorativo. No entanto, o fruto foi propagado em outras regiões, como Estados Unidos, Austrália, Equador e Brasil (GUEVARA et al., 2019; MEINHART et al., 2019; MUTHU et al., 2016). Carambolas são ovais a elipsoides com 5 arestas longitudinais, possuem casca translúcida, fina e cerosa, e polpa de consistência rígida (ABDULLAH; FATHINUL-SYAHIR; MOHD-AZEMI, 2005; HAZARIKA et al., 2020). É rica em minerais e vitaminas, incluindo magnésio, potássio, antioxidantes e apresenta altas quantidades de fibra dietética insolúvel (CHAU; CHEN; LIN, 2004; MUTHU et al., 2016). A carambola tem sido considerada uma fruta medicinal e usada para tratar muitas doenças, incluindo vários distúrbios da pele, dores de cabeça, vômitos, tosse, diarreia, febre e diabetes (SAGHIR et al., 2013). O uso da fruta tem impulsionado as pesquisas, apresentando propriedades antioxidantes benéficas, anti-inflamatórias e antimicrobianas (CABRINI et al., 2011; MUTHU et al., 2016; SAGHIR et al., 2013). Todavia, por conter cristais de oxalato de cálcio que formam ácido oxálico presentes na polpa, este fruto tem sido direcionado a diversas pesquisas e é contraindicado a pacientes urêmicos (YANG et al., 2014).

Dietas ricas em frutas são amplamente vistas no mundo. Especialistas da Organização Mundial da Saúde (OMS) recomendam que a ingestão diária de frutas por pessoa seja de pelo menos 400 g, para garantir uma boa nutrição e boas condições de saúde. A presença dos frutos na alimentação tem resultado em um crescimento global totalizando volumes maiores que 870 milhões de toneladas de frutas colhidas ao ano (FAO, 2018). O mercado de frutas tropicais ou exóticas tem crescido ultimamente, tanto para importação quanto para exportação e a carambola tem se destacado como uma alternativa de cultivo (MUSTAFA et al., 2016). A Malásia tem se tornado um dos maiores produtores de carambola, com exportação para vários continentes, como a Europa, Oriente Médio e Leste Asiático (ZAINUDIN et al., 2014). No Brasil, a plantação é localizada principalmente na região Sudeste, em especial no Estado de São Paulo (MUSTAFA et al., 2016). Em 2014, estima-se cerca de 5.000.000 kg de carambolas frescas foram comercializadas na Companhia de Entrepósitos e Armazéns Gerais de São Paulo – CEAGESP, tendo ocorrido aumento nos últimos anos (MUSTAFA et al., 2016). A crescente procura pela carambola e suas propriedades, a adaptabilidade da planta às condições edafoclimáticas brasileiras, vida útil e sua reprodução, que permite serem colhidas muitas vezes ao ano, são parâmetros que favorecem o cultivo nacional (ARAÚJO; MINAMI, 2001; WANI et al., 2017).

### 3.2. PARÂMETROS DA QUALIDADE DA CARAMBOLA

Os aspectos de qualidade e segurança dos alimentos tornaram-se uma grande preocupação para os fornecedores e consumidores. A crescente necessidade de produção com alto rendimento tem gerado à indústria alimentícia, desafios referentes à gestão de qualidade e seu impacto na comercialização (CHANDRASEKARAN et al., 2019). A colheita dos frutos no estágio de maturação adequado atribui qualidade ao produto; frutas imaturas estão mais sujeitas a ficarem murchas e sofrerem danos mecânicos, o que as tornam de qualidade inferior às maduras. Frutas excessivamente maduras, por outro lado, tendem a se tornar macias, farinhentas, sem sabor e facilmente danificadas logo após a colheita (JIE; WEI, 2018; MADUWANTHI; MARAPANA, 2019). Atualmente, para evitar esses danos às carambolas, pesquisas sugerem que a sua colheita deve ser após o aparecimento da coloração amarela; após essa fase de “*colour-break*” o fruto atinge teor total de sólidos solúveis (TSS) desejáveis. Ademais, frutas colhidas muito cedo ou tarde demais são mais suscetíveis a deterioração microbiana, distúrbios fisiológicos e tem menor vida útil (CHEN et al., 2017; YAHAYA; OMAR, 2017).

#### 3.2.1. Maturação e parâmetros de classificação

Maturação é um termo amplamente definido como aquele estado em que o crescimento e o desenvolvimento natural são considerados completos, ou suficientes; processo pelo qual a fruta atinge seu sabor desejável, qualidade, cor e outras propriedades sensoriais (KYRIACOU; ROUPHAEL, 2018). Com o amadurecimento, a fruta se torna macia, produz compostos voláteis, diminui a acidez e aumenta o teor de sólido solúvel (TSS), que está relacionado a sua doçura (KYRIACOU; ROUPHAEL, 2018). As frutas podem ser não-climatéricas, quando não amadurecem após serem colhidas; ou podem ser climatéricas, quando continuam a amadurecer mesmo após serem colhidas. Para as não climatéricas, como as carambolas, a colheita deve ser feita quando o fruto já passou pelo processo de maturação, o que atribuirá qualidade e sabor ao fruto (CHEN et al., 2017; YAHAYA; OMAR, 2017).

Durante a fase de maturação ocorrem mudanças em parâmetros fundamentais para a qualidade, como sabor, cor e textura (PATHMANABAN; GNANAVEL; ANANDAN, 2019). As carambolas colhidas em estádios iniciais de maturação são destinadas à exportação devido ao maior tempo de prateleira, mas ainda não desenvolveram pleno sabor e têm uso limitado para guarnições, saladas e bebidas (FAMA, 2014). Por outro lado, em estádios avançados de maturação, a carambola pode ser comercializada para geleias, pudins, tortas e principalmente consumo direto (fruta fresca), embora sejam direcionadas apenas para entregas

em destinos próximos (FAMA, 2014; PAUZIAH et al., 2010).

A cor tem sido considerada o critério primário na escolha das carambolas; é o aspecto mais perceptível do amadurecimento e a característica usada, principalmente, na classificação de qualidade (LI; LECOURT; BISHOP, 2018; YAHAYA; OMAR, 2017). A cor influencia a percepção do paladar e pesquisas mostram sua correlação com o sabor, expresso em termos de doçura e acidez; essa transição de cor para as carambolas também explica o envelhecimento e a diminuição do frescor (ABDULLAH et al., 2006; LI; LECOURT; BISHOP, 2018). Para fins de exportação, a inspeção de qualidade da carambola tem sido embasada na Autoridade Federal de Marketing Agrícola (FAMA). Essa classificação é embasada na cor e sua distribuição no fruto, correlacionando o seu nível de maturação (YAHAYA; OMAR, 2017).

A transformação da cor, de verde da carambola para ligeiramente laranja como mostrado na Fig. 3.1, é atribuído a degradação da clorofila e o incremento de carotenoides (O'HARE, 1993). Essa mudança produzirá uma diferença perceptível na acidez e sólidos solúveis totais (YAHAYA; OMAR, 2017). Frutas imaturas possuem maior quantidade de ácido ascórbico e ácido oxálico, menor pH e menor doçura (YAHAYA; OMAR, 2017). Alguns pigmentos responsáveis pela mudança de cor, são dependentes do pH; com isso, gradualmente a fruta ficará mais doce e menos acida à medida que a cor muda de verde para amarelo e finalmente para laranja. No entanto, por se tratar de um fruto não-climatérico, uma vez colhido as concentrações de açúcar permanecerão constantes, mas a acidez pode diminuir durante o armazenamento e gerar alteração na cor do fruto (LI; LECOURT; BISHOP, 2018; YAHAYA; OMAR, 2017).



**Figura 3.1.** Carambolas em diferentes estádios de maturação.

Assim, têm-se classificado as frutas através de inspeções visuais subjetivas, conduzidas por avaliadores treinados, que avaliam a aparência externa, embasada em parâmetros físicos, como a cor. Entretanto, parâmetros como, maturação, sólidos solúveis e acidez, são fundamentais, mas que exigem análises laboriosas e destrutivas (KYRIACOU; ROUPHAEL, 2018; YAHAYA; OMAR, 2017). Ademais, mesmo havendo correlação da cor e os parâmetros físico-químicos referentes ao sabor na literatura, a cor das cascas não prevê o dia apropriado da colheita, a qualidade do fruto ou os níveis de sólidos solúveis totais e acidez titulável (AMIR et al., 2019; DAFNY-YALIN et al., 2010; KYRIACOU; ROUPHAEL, 2018). Além da alteração da cor, a natureza não uniforme do amadurecimento das carambolas também tem causado irregularidades e inconsistências na classificação (ABDULLAH et al., 2006). Com isso, a crescente demanda pelo controle de qualidade dos alimentos tem impulsionado as indústrias alimentícias a buscarem técnicas analíticas não destrutivas, inovadoras e com fácil manuseio, possibilitando a quantificação de vários compostos de forma rápida e precisa (JIMÉNEZ-CARVELO et al., 2019).

Experimentos recentes têm sido reportados utilizando análises não destrutivas em carambola. Dentre elas, correlação do suco de carambola com seu pH usando imagem digital (MUHAMMAD ZAHIR; YAHAYA; OMAR, 2021), predição de parâmetros de qualidade usando Vis/NIR (YAHAYA; OMAR, 2017), predição de pH e firmeza usando sensor de fibra óptica (OMAR; MATJAFRI, 2013) e predição da acidez usando espectrômetro portátil (OMAR; ATAN; MATJAFRI, 2012). Contudo, ainda não foi abordada a classificação dos estádios de maturação, ou a predição dos parâmetros de qualidade da carambola usando a espectroscopia no infravermelho próximo (NIR) com equipamento portátil.

### **3.3. TÉCNICAS ANALÍTICAS NÃO DESTRUTIVAS**

#### **3.3.1. Espectroscopia no infravermelho próximo (NIR).**

Entre os métodos existentes, a espectroscopia vibracional no intervalo de infravermelho próximo (NIR) tem sido relevante, por ser um método de análise rápido com grande aplicação em vários alimentos (BASANTIA; NOLLET; KAMRUZZAMAN, 2018). A espectroscopia NIR é baseada nas propriedades vibracionais das ligações químicas das moléculas orgânicas e em suas interações com a radiação no espectro do infravermelho (BASANTIA; NOLLET; KAMRUZZAMAN, 2018). A radiação NIR, refletida ou transmitida da luz incidente no material, é registrada pelo espectrômetro (detector) e transformada em absorção. A forma de espectro é caracterizada por tons e faixas combinadas de vibrações fundamentais que ocorrem na região do NIR de 780 a 2500 nm

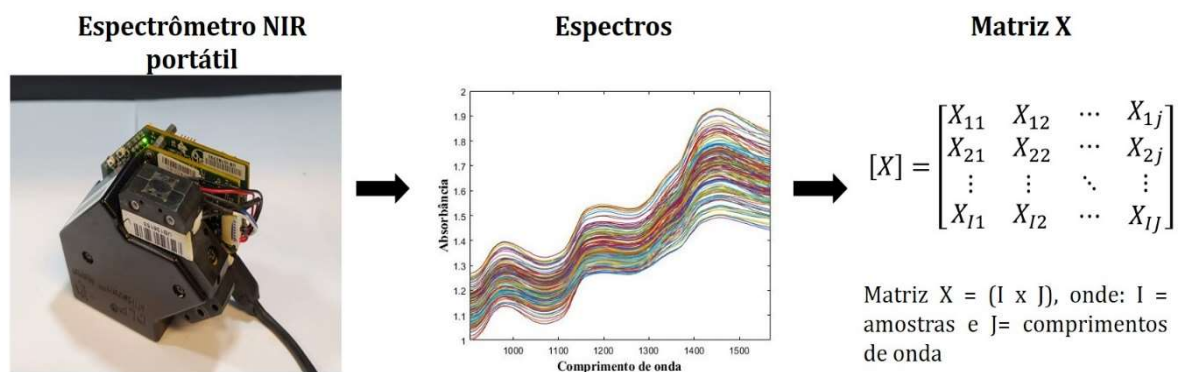
(CHANDRASEKARAN et al., 2019). É uma técnica com alto potencial para medição não destrutiva, aplicável em diferentes produtos agrícolas, que não envolve o uso de produtos químicos, demanda pouca ou nenhuma preparação de amostra e pode ser utilizado em ferramentas de monitoramento on-line, o que reduz o custo operacional (BARBIN et al., 2018; JAKUBÍKOVÁ et al., 2016; NICOLAÏ et al., 2007).

As indústrias de processamento pós-colheita requerem técnicas analíticas rápidas, não destrutivas e sem produtos químicos para implementação de medições on-line em linha. As aplicações do NIR são destinadas a pontos de entrega, produção, uso ou venda. Segundo Magwaza (2012), o NIR em relação a outras técnicas não destrutivas é mais avançado em termos de instrumentação, acessórios, disponibilidade de pacotes de software quimiométrico adequados. Além disso, o progresso tecnológico e a redução no tamanho dos componentes, tem permitido o desenvolvimento de novos espectrômetros, com design ergonômico, podendo ser transportado a diferentes locais de trabalho (CRUZ-TIRADO; MEDEIROS; BARBIN, 2021; PASQUINI, 2018).

O espectrômetro portátil na região do infravermelho próximo tem sido uma técnica analítica compacta, sensível e não destrutiva; que tem garantido confiabilidade e velocidade das análises (CORREIA et al., 2018; FRIEDRICH et al., 2014). O NIR portátil, como outros modelos de espectrômetros, tem a capacidade de classificar e quantificar vários compostos químicos simultaneamente. Suas vantagens incluem baixo custo, pouca ou nenhuma preparação amostral e principalmente portabilidade (CORREIA et al., 2018; FRIEDRICH et al., 2014). Os alimentos possuem intensidades e regiões de absorbância que correspondem a um espectro característico como uma “impressão digital” (WORKMAN; SHENK, 2015). Esse espectro pode ser utilizado para quantificação de compostos físico-químicos ou classificação embasados em fatores como condições da cultivo, origem, variedades e processo (COZZOLINO, 2016). Com isso, o NIR portátil tem sido utilizado em pesquisas de diversas frutas, direcionadas a previsão de parâmetros de qualidade de uva e pêssego (DONIS-GONZÁLEZ et al., 2020), análise de qualidade de tomates frescos no campo (BORBA et al., 2021), predição instantânea e simultânea de antocianinas e açúcar em framboesas frescas inteiras (GALES et al., 2021), a qualidade interna de laranjas, limões e tangerinas (SANTOS et al., 2021), predição de umidade e teor de sólidos solúveis em pera (MISHRA et al., 2021a), predição de sólidos solúveis, acidez titulável e firmeza de maçãs (PISSARD et al., 2021), e na avaliação do teor de sólidos solúveis em bananas (SRIPAURYA et al., 2021).

Os dados espectrais do NIR adquiridos dos produtos agrícolas podem ser

analisados por abordagens quantitativas e qualitativas da maturação; podem também, ser usados em análise de vários constituintes, simultaneamente, desses parâmetros de qualidade (LOHUMI et al., 2015). Essa quantificação, pela espectroscopia NIR, é baseada em distintos picos e vales amplos e altamente sobrepostos, o que dificulta a análise espectral e a extração de informações químicas e físicas do espectro (CHANDRASEKARAN et al., 2019). Deste modo, para extrair as informações, esses dados devem ser convertidos em uma matriz de dados que carece de uma combinação de ferramentas de análise multivariada, para extrair informações significativas (BIANCOLILLO et al., 2020), como demonstrado na Fig. 3.2.



**Figura 3.2.** Organização da matriz X a partir da aquisição espectral NIR.

### 3.4. QUIMIOMETRIA

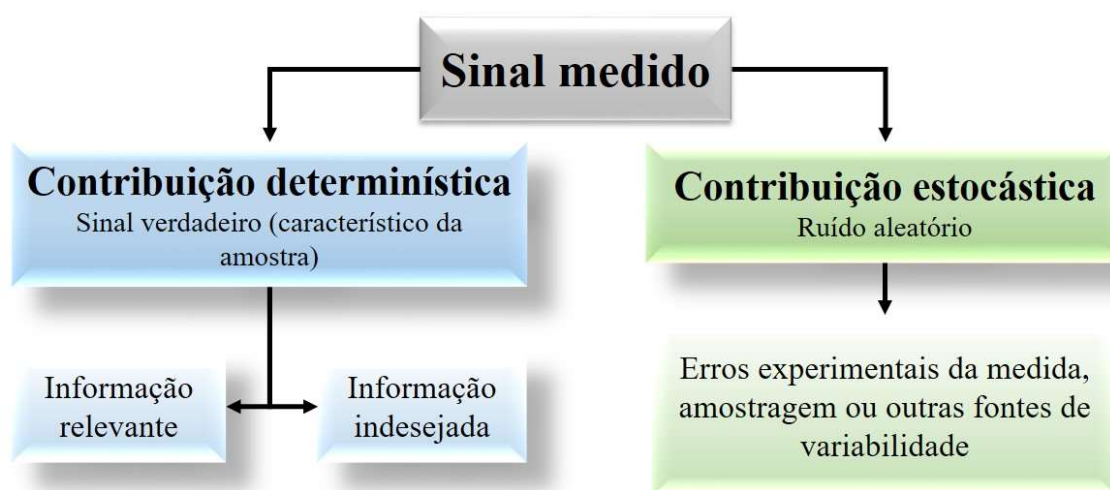
A origem da quimiometria se deu a partir da necessidade de adquirir informações químicas relacionadas à grande quantidade de dados obtidos por equipamentos sofisticados. Estes instrumentos não geram informações diretas com significado químico; para a interpretação e obtenção destes resultados é indispensável a utilização de ferramentas matemáticas e estatísticas que possam auxiliar na extração de informações relevantes (FORINA; CASALE; OLIVERI, 2009).

Os métodos quimiométricos estão divididos em dois grupos: os métodos “supervisionados” e os “não supervisionados” (FERREIRA, 2015). Os métodos supervisionados possuem amostras com valores de referências pré-estabelecidos, e durante a análise de dados e construção dos modelos essas informações são utilizadas. A técnica mais popular é o dos Mínimos Quadrados Parciais (PLS), que permite prever uma propriedade da amostra a partir das variáveis independentes originais (FERREIRA, 2015). Os métodos não supervisionados não solicitam essa informação pré-estabelecida, e por isso são ditos “métodos de análise exploratória de dados”. A Análise de Componentes Principais (PCA) é um método

não supervisionado, capaz de reduzir a dimensão de um conjunto de dados sem afetar a relação existente entre as amostras (FERREIRA, 2015).

### 3.4.1. Pré-tratamento

Os espectros apresentam duas contribuições; o sinal verdadeiro que representa a informação sistemática, correspondente a contribuição determinística do sinal e a outra é a contribuição estocástica oriunda de variações indesejadas, denominada de ruído (FERREIRA, 2015). Além do ruído aleatório, o sinal pode conter também fontes de variação de baixa frequência, que também são indesejáveis e não possuem relações com a composição química da amostra (FERREIRA, 2015). Essa variação de baixa frequência pode ser oriunda da contribuição determinística como mostrado no esboço da Fig. 3.3. Esses interferentes podem dificultar a interpretação e a criação de modelos robustos e eficientes. Para reduzir essas variações indesejadas e obter melhores resultados costuma-se aplicar algumas técnicas, conhecida por “pré-tratamentos” (ROGER et al., 2020).



**Figura 3.3.** Contribuição existente no sinal medido.

Os pré-tratamentos podem ser aplicados às amostras, denominados de transformação, ou às variáveis, chamados de pré-processamentos. Um exemplo de pré-processamento, é a centralização na média; que são direcionados às variáveis que não possuem o mesmo peso, como os dados de espectroscopia (FERREIRA, 2015). O processo se baseia no deslocamento do centro de dados para a origem do sistema de coordenadas. O

valor médio de cada variável é calculado, inicialmente, e em seguida essa média é subtraída de cada uma das variáveis do conjunto de dados (HIBBERT, 2016). A escolha adequada do pré-processamento, e assim a redução das variações indesejáveis, é essencial para o sucesso de qualquer análise multivariada de dados (ROGER et al., 2020).

Dados ruidosos, são perturbações indesejadas no espectro que podem ser causadas pela sensibilidade dos detectores, de instrumentos ou por fonte de excitação (TRYGG; GABRIELSSON; LUNDSTEDT, 2009). Para esses dados, o objetivo do pré-tratamento é reduzir a informação indesejada, baseado no aumento da razão sinal-ruído (SNR), gerando a separação entre o sinal real e o ruído (THEKKUDAN; RUTAN, 2009). Os métodos mais comuns, como a suavização, envolvem um segmento de pontos do espectro ("janela"), com tamanho definido, e filtros lineares que reduzem todos os pontos dessa janela a um ponto central. Então essa janela é deslocada e o processo é repetido, percorrendo integralmente o espectro (SAVITZKY, A.; GOLAY, 1964). A suavização reduz ruídos, mas também remove algumas informações que podem ser importantes, e, quanto maior a janela, maior será esse efeito. As principais técnicas de alisamento são: método da média, Savitzky-Golay e a suavização com filtro de Fourier (FERREIRA, 2015).

As linhas de base são efeitos estruturados de fundo ("contínuo" e de baixa frequência), e as razões, para a ocorrência nos sinais do NIR, é a perda da luz e sua interação com as partículas da amostra, levando à dispersão (MISHRA et al., 2020). Problemas instrumentais ou amostrais podem deslocar o espectro como um todo e assim deslocar a linha de base (no eixo y). Esse efeito aditivo pode gerar valores positivos ou negativos comparados ao zero de absorbância e pode ser corrigido aplicando a primeira derivada do espectro (ROGER et al., 2020). Para espectros que possuem inclinação que aumentam enquanto o número de onda decresce, a correção é feita com a segunda derivada. Savitzky-Golay é a técnica mais difundida dessas derivadas (FERREIRA, 2015; ROGER et al., 2020). Entretanto, essa técnica deve ser aplicada cuidadosamente, pois pode gerar decréscimo progressivo da razão sinal /ruído, ou seja, introduzir ruídos aos espectros (FERREIRA, 2015).

Além dos efeitos aditivos, vistos no deslocamento de base, há os efeitos multiplicativos que podem ser causados por fenômenos físicos, como dispersão de luz, no caso de espectroscopia óptica (BIANCOLILLO et al., 2020). Esses efeitos são causados pelo desvio forçado dos fótons de uma trajetória reta por não uniformidades presentes nas amostras (MISHRA et al., 2020). Os efeitos de espalhamento aditivo e multiplicativos podem ser oriundos de mudança no caminho ótico, sensibilidade do detector, variações na pressão e temperatura, e falta de homogeneidade no tamanho e forma das partículas (FERREIRA,

2015). A correção de dispersão pode ser por técnicas como *Standard Normal Variate* (SNV) (BARNES; DHANOA; LISTER, 1989) e *Multiplicative Scatter Correction* (MSC) (GELADI; MACDOUGALL; MARTENS, 1985). Ambas atuam de maneira semelhante: em cada uma das linhas da matriz de dados originais é feito um autoescalamamento, isso resulta em espectros com deslocamento de linha de base corrigidos. O processo remove os efeitos multiplicativos e mantém a forma original dos espectros, diferente da aplicação de derivadas como pré-tratamento (FERREIRA, 2015).

A aplicação de pré-tratamento causa modificações nos dados brutos e pode melhorar desempenho preditivo ou discriminatório. Entretanto, o conhecimento sobre as amostras, características do espectro e os processos químicos e físicos envolvidos na análise são fundamentais para a escolha adequada dessa aplicação (KEILANY, 1978). O excesso ou a aplicação incorreta do pré-tratamento pode causar perda de informações relevantes (PASQUINI, 2018). Estudos sobre o uso e combinações de pré-tratamentos têm sido realizados em diversos alimentos, em especial nas frutas (MISHRA et al., 2021a). Assim, as extensas pesquisas quimiométricas nas últimas décadas, e o estudo sobre a necessidade do uso de determinados pré-tratamentos para modelos preditivos em frutas, tem gerado avanço em técnicas recentemente disponíveis (MISHRA et al., 2020), como o *Variable sorting for normalization* (VSN) (RABATEL et al., 2020).

Após etapa de processamento, são conduzidos os métodos de regressão para quantificar propriedades analisadas, e os métodos de classificação para qualifica-las.

### 3.4.2. Análise exploratória

A análise exploratória é uma técnica não supervisionada, para reconhecimento de padrões, que permite a visualização de grupos ou tendências (Tibola et al., 2018). Essas semelhanças são embasadas nas amostras, variáveis e a relação das mesmas, representadas graficamente nesta análise. Além disso, com a análise exploratória é possível identificar a presença de amostras anômalas (*outliers*) (ABDI; WILLIAMS, 2010). Essas amostras influenciam fortemente o resultado das análises quimiométricas, podendo ser reavaliadas e retiradas do conjunto de dados, caso necessário (BRO; SMILDE, 2014).

Entre os métodos existentes, a Análise de Componentes Principais (PCA) do inglês *principal-component analysis* é o mais utilizado e tem como objetivo a redução da dimensionalidade da matriz (ABDI; WILLIAMS, 2010). O princípio é projetar uma tabela de dados que possuem semelhanças em um espaço com menor dimensionalidade, proporcionando o melhor ajuste da distribuição de dados. Esse processo visa a redução no

número de variáveis, maior interpretação dos dados e visualização de grupos ou classes (BRO; SMILDE, 2014). Desta forma, os dados podem ser direcionados para estudos de correlações entre os sinais obtidos por espectroscopias de infravermelho e análises de referência em uma dimensão reduzida e objetiva (BIANCOLILLO et al., 2020).

Inicialmente, a análise decompõe a matriz  $X$  (amostras x variáveis) em uma matriz de escores  $T$  (amostras x componentes principais) e uma matriz ortogonal de pesos  $L$  (variáveis x componentes principais), conforme a equação 1. Os scores são conjuntos de informações do espectro de cada amostra e os pesos são multiplicadores de vetores que atribuem pesos às variáveis (FERREIRA, 2015).

$$X = TL \quad (\text{Eq. 1})$$

O cálculo do PCA é realizado a partir do número de componentes principais escolhidos da matriz a ser usado (FERREIRA, 2015). As variáveis originais são combinadas em componentes principais (PC's) de forma linear. Elas condensam alto percentual informativo em poucas variáveis ou comprimentos de onda, o que reduz a dimensões dos dados sem perder informações químicas relevantes (KEMSLEY; DEFERNEZ; MARINI, 2019). A primeira componente (PC1) possui maior percentual de informação estatística, a segunda componente (PC2) gera informação não contidas na primeira componente principal e assim sucessivamente (FERREIRA, 2015). Assim,  $T_A$  representa o conjunto da nova matriz de escores, feito com as PC's selecionadas e  $L_A^T$  representa a nova matriz de pesos. Além dessas matrizes, é gerado um grupo que reúne as variâncias não informadas na escolha das PC's; esses dados são denominados de matriz de resíduos e aqui serão representados por  $\hat{E}$  (SAEYS et al., 2019). Assim, a matriz original também pode ser descrita conforme a equação 2.

$$X = T_A + L_A^T + \hat{E} \quad (\text{Eq.2})$$

Para espectroscopia, em geral, grande parte da variabilidade dos dados é projetada no espaço abrangido pelos dois primeiros componentes principais. Eles respondem por cerca de 90% da variância original, que corresponde a uma boa aproximação da matriz experimental (BIANCOLILLO et al., 2020).

### 3.4.3. Regressão

O uso de técnicas não destrutivas, em parte, é estimulado pela necessidade de quantificar uma propriedade de interesse. Isso é obtido com análises de dados multivariados através da construção de modelos matemáticos (regressão linear) que utilizam medidas experimentais (FERREIRA, 2015). Um dos métodos usualmente utilizados para a regressão é o método dos mínimos quadrados parciais do inglês, *partial least squares*, PLS. Trata-se de um método com ampla aplicabilidade em análises multivariadas, por apresentar bons resultados, ainda que o espectro contenha ruídos e colinearidade (MEHMOOD et al., 2012). O método busca mostrar uma relação entre a matriz  $X$  (amostras x variáveis) e vetor  $y$  (amostras x 1). Em espectroscopia, a matriz  $X$  são ( amostras x comprimentos de onda), e o vetor  $y$  são valores de referência, como os físico-químicos (BOSQUE-SENDRA et al., 2012). A regressão linear é calculada pelas Equações 3 e 4.

$$X = T_A L_A + E \quad (\text{Eq. 3})$$

$$Y = T_A q + e \quad (\text{Eq. 4})$$

Onde “ $T_A$ ” é a matriz de escores (scores), “ $L_A$ ” é a matriz de pesos (Loadings), “ $A$ ” representa quantas variáveis latentes (LVs) foram usadas para a construção do modelo, “ $E$ ” é a matriz de resíduo de  $X$ . Quanto a “ $y$ ” é o vetor com os parâmetros a serem analisados; “ $q$ ” é o vetor dos coeficientes de regressão; “ $e$ ” é vetor de resíduos de  $y$  (FERREIRA, 2015). As variáveis latentes (LV) também são um fator determinante à análise; as LVs escolhidas pelo analista irão descrever as variações importantes nos dados espectrais (ALAMPRESE et al., 2016).

#### 3.4.4. Discriminação

Quando a propriedade de interesse é categórica (discreta) busca-se a construção de modelos de classificação. Nesses estudos, cada amostra pertence a um conjunto de medidas experimentais e são classificadas conforme as propriedades de referência, sendo categorizadas como bom/ruim, falso/verdadeiro, ativo/não ativo etc. (FERREIRA, 2015). Abordagens discriminantes tem sido as soluções mais clássicas para lidar com problemas relacionados à classificação. Com esse conjunto de técnicas torna-se possível encontrar semelhanças e destacar diferenças entre amostras pertencentes a classes distintas (BEVILACQUA et al., 2013).

A Análise Discriminante pelo Método de Mínimos Quadrados Parciais (PLS-DA)

é um método supervisionado de classificação de padrões, embasada na regressão por mínimos quadrados parciais (PLSR). Esse método, encontra variáveis e direções no espaço multivariado contribuintes para a discriminação das classes obtidas no conjunto de calibração (BARKER; RAYENS, 2003). O método dos mínimos quadrados parciais busca relacionar os dados espectrais (matriz X) com propriedade (s) a serem analisadas (matriz Y). No conjunto de validação externa, é construído uma matriz de dados (amostras a serem classificadas x classes existentes). Inicialmente, é analisado a primeira coluna (primeira classe) e é atribuído o número (1) às amostras que pertencem à classe que está sendo analisada, e atribuído (0) a todas as outras amostras não pertencentes à classe (FERREIRA, 2015). Sucessivamente, cada uma das classes é analisada e os valores são atribuídos como esboçado na Fig. 3.4.

|          |   | Classes |   |     |   |
|----------|---|---------|---|-----|---|
|          |   | A       | B | ... | N |
| Amostras | 1 | 1       | 0 | ... | 0 |
|          | 2 |         |   |     |   |
|          | ⋮ |         |   |     |   |
|          | n |         |   |     |   |

**Figura 3.4.** Matriz da variáveis dependentes para N classes e n amostras.

Para a construção do modelo utilizando o método PLS, os valores da variável dependente não são números inteiros como visto anteriormente; são números reais estimados que se aproximam de 1 e 0, e com base neles, as classes são atribuídas (FERREIRA, 2015). Na PLS-DA, diferente do PCA, as (LVs) não são ortogonais e um fator importante é que os erros das matrizes X e Y, são levados em consideração e são igualmente distribuídos (BONA; MARÇO; VALDERRAMA, 2018; MARQUETTI et al., 2016).

### 3.4.5. Avaliação da performance dos modelos

#### 3.4.5.1. Modelos de regressão

A performance do modelo está relacionada com a capacidade preditiva, ou seja, menores erros geram modelos com bons desempenhos. O erro de validação cruzada do inglês *root mean square error for cross validation* (RMSEC), e o erro de predição, do inglês *root mean square error for prediction* (RMSEP) são valores utilizados na avaliação de performance conforme as equações 5 e 6 (BONA; MARÇO; VALDERRAMA, 2018;

FERREIRA, 2015).

$$RMSEP = \sqrt{\frac{\sum_{p=1}^P (y_p - \hat{y}_p)^2}{P}} \quad (\text{Eq. 5})$$

Onde, P é a quantidade de amostras usadas na previsão e  $\hat{y}_p$  e  $y_p$  são os valores previstos e medidos, respectivamente.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^I (y_i - \hat{y}_i)^2}{I}} \quad (\text{Eq. 6})$$

Onde, I é a quantidade de amostras usadas na validação interna e  $\hat{y}_i$  e  $y_i$  são os valores previstos e medidos, respectivamente.

O parâmetro de correlação também auxilia a performance, apresentando o quão ajustado o conjunto está. Coeficientes de correlação próximos a 1 representam que valores medidos e previstos são próximos. Outras medidas utilizadas para avaliação de desempenho em frutas é a taxa do desvio padrão, do inglês *ratio of the standard deviation* RPD (eq. 7) (SAEYS; MOUAZEN; RAMON, 2005) e a razão da faixa de erro, do inglês *Range Error Ratio* (RER) (eq.8) (RAMBO; AMORIM; FERREIRA, 2013).

$$RPD = \frac{SD}{RMSEP \text{ ou } RMSECV} \quad (\text{Eq. 7})$$

$$RER = \frac{Range}{RMSEP \text{ ou } RMSECV} \quad (\text{Eq. 8})$$

Onde: SD é o desvio padrão das amostras e range é a faixa de variação na concentração dos dados.

#### 3.4.5.2. Modelos de classificação

Nos modelos de classificação, além do RMSEC, RMSEP e a precisão definida na equação 11, temos outras figuras de méritos definidas a seguir. A sensibilidade representa a quantidade de amostra classificada corretamente, e é calculada pela equação 10 (BARKER; RAYENS, 2003). Quanto a seletividade, representa o quanto a amostra prevista não é classificada à classe em que de fato a amostra não é pertencente, e é calculada pela equação 11 (BONA; MARÇO; VALDERRAMA, 2018).

$$Precisão(\%) = \frac{(TP+TN)}{TOTAL} \times 100 \quad (\text{Eq. 9})$$

$$Sensibilidade = \frac{TP}{(TP+FN)} \quad (\text{Eq. 10})$$

$$Seletividade = \frac{TN}{(FP+TN)} \quad (\text{Eq. 11})$$

Onde, TP é a taxa de verdadeiro positivo, do inglês *true positive*; FP é a taxa de falso positivo, do inglês *false positive*; TN é a taxa de verdadeiro negativo, do inglês *true negative* e FN é a taxa de falso negativo, do inglês *false negative*.

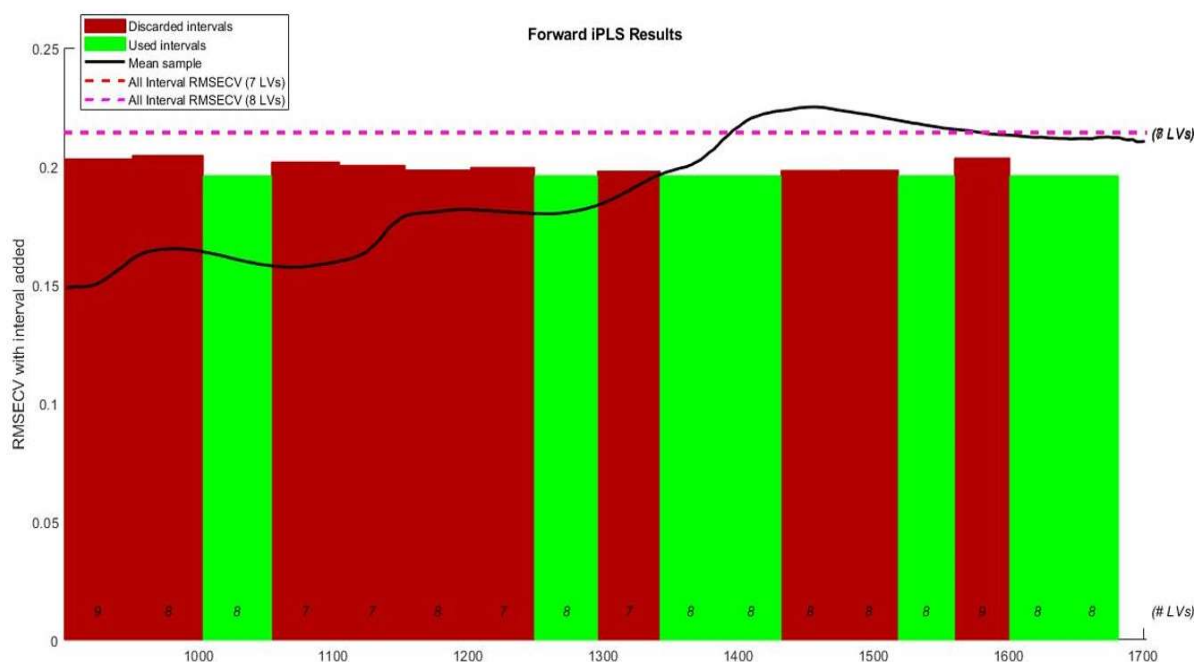
#### 3.4.6. Seleção de variáveis

Espectroscopia de infravermelho próximo, como tanto outras técnicas não destrutivas, em alguns casos, possuem informações irrelevantes ou ruídos que dificultam a construção e validação do modelo (TEÓFILO; MARTINS; FERREIRA, 2009). Para eliminar variáveis com informações irrelevantes, e assim melhorar a performance do modelo de previsão utiliza-se as técnicas de seleção de variáveis. Com a seleção é gerado um subconjunto com menos variáveis, mais correlacionadas com os parâmetros analisados e que resulta em modelos mais simples e com menos erros sistemáticos (PASQUINI, 2018). Assim, para a análise multivariada, a etapa de seleção de variáveis tornou-se relevante, em virtude dos progressos e aumento dos desempenhos apresentados (XIAOBO et al., 2010).

As variáveis selecionadas direcionam o analista à determinadas informações associadas a ligações químicas e regiões importantes aos parâmetros analisados (LAVINE, 2003). Essa ferramenta, além de dar suporte para a reafirmação da composição da amostra, é útil para a criação de equipamentos portáteis compactos que trabalham com um intervalo reduzido de comprimentos de onda (LAVINE, 2003; TEÓFILO; MARTINS; FERREIRA, 2009). O conhecimento dessas regiões de maior interesse, reduz custos computacionais e permite o direcionamento desses equipamentos às possíveis aplicações no campo e uso industrial (TEÓFILO; MARTINS; FERREIRA, 2009; YANG et al., 2008).

Há muitas técnicas de seleção de variáveis, e cada uma possui fundamentos e parâmetros de seleção diferentes (PASQUINI, 2018). Método dos mínimos quadrados por intervalo, iPLS (do inglês, *interval partial least square*), é uma das técnicas mais simples e difundida de seleção de variáveis. Ela fraciona o espectro em subintervalos menores equidistantes e depois constrói modelos PLS para cada divisão utilizando a mesma quantidade

de variáveis latentes como exemplificado na Fig. 3.5. Inicialmente, calcula-se o PLS do intervalo escolhido, de forma aleatória, e o compara com o modelo completo (todas as variáveis originais presentes) através do menor erro médio quadrático da validação cruzada (RMSECV) (MEHMOOD et al., 2012; NØRGAARD et al., 2000). Se o novo RMSECV for inferior ao modelo completo, o conjunto inicial é substituído, e subsequentemente, outro intervalo é selecionado.



**Figura 3.5.** Espectro de amostras de carambola com seleção de variáveis pelo método iPLS.

Algoritmo genético, GA (do inglês *genetic algorithm*) é outro método de seleção de variáveis bastante utilizado. O algoritmo é um método robusto, inspirado na teoria da evolução biológica e na seleção natural, com alta capacidade de alcançar variáveis que resultem em modelos mais preditivos (MEHMOOD et al., 2012). Durante a execução, são distribuídos valores binários, 0 e 1 às variáveis (LIU et al., 2005). O valor 0, exclui o dado espectral da seleção, por outro lado a atribuição de valor 1 inclui o dado na seleção (LIU et al., 2005). Os melhores desempenhos também são embasados no menor RMSECV.

Os algoritmos de seleção de variáveis têm sido direcionados a trabalhos com o uso de espectroscopia no infravermelho próximo aplicada em diversas frutas. Assis et al. (2018) aplicaram iPLS, GA-PLS em Dovyalis, afim de melhorar o desempenho de modelo preditivo para o teor de sólidos solúveis, pH e acidez usando espectroscopia de infravermelho próximo. A aplicação deles do iPLS não melhorou o desempenho do modelo, porém o GA-PLS teve melhor desempenho para TTA (RMSEP de 4,8,  $R^2_p$  de 0,40 e RPD de 1,67) e para pH

(RMSEP de 0,03,  $R_p^2$  de 0,90 e RPD de 6,67). Liu et al. (2010) com iPLS selecionou variáveis para previsão de sólidos solúveis e acidez titulável em mexerica, usando dispositivo de carga acoplada a espectroscopia de infravermelho próximo (CCD-NIRS). Entretanto, o algoritmo iPLS também não implementou melhoria nos modelos de TSS e TTA. Por outro lado, Xu et al. (2012) utilizaram iPLS e GA-PLS em predição de teor de açúcar em peras com NIR portátil apresentando melhoria das performances tanto de iPLS com ( $R_p^2 = 0.840$  e RMSEP 0.536) quanto para GA-iPLS com ( $R_p^2 = 0.878$  e RMSEP = 0.457). Inácio et al. (2013) também utilizaram iPLS e GA-PLS em determinação de antocianinas em açaí e a seleção implementou melhoria com iPLS (RPD=1.61) e GA-PLS (RPD=1.74). Por fim, Lu et al (2015) também utilizaram o GA-PLS em predição de firmeza e teor de sólidos solúveis em melão usando NIR, implementos estes que proporcionaram um modelo mais robusto com RMSECV de 0.99 e RPD próximo a 3.

A seguir são apresentados outros trabalhos que utilizaram espectroscopia NIR em frutas. A tabela 3.1 demonstra a relevância crescente do desenvolvimento das análises quimiométricas e dos espectrômetros, nos últimos 30 anos. Os resultados demonstraram a potencialidade da utilização de quimiometria acoplada a espectroscopia no infravermelho próximo para várias finalidades; o que tem tornado as análises de controle de qualidade, caracterização e discriminação mais rápidas, simples e de baixo custo.

**Tabela 3.1.** Trabalhos com o uso do NIR em diversas frutas.

| Fruta       | Parâmetro | Equipamento   | Quimiometria | Descrição                                                                                                                                                                                                                                                                                                                                 | Referência                        |
|-------------|-----------|---------------|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|
| Pêssego     | TSS       | 680 - 1235 nm | Regressão    | Utilizaram o modo de aquisição de interação (refletância) para quantificar o açúcar em pêssegos intactos. Apresentaram desempenho de ( $R^2 = 0.07$ ) e ( $SEP = 0.5$ ).                                                                                                                                                                  | (KAWANO; WATANABE; IWAMOTO, 1992) |
| Maçã        | TSS, pH   | 380 -1650 nm  | PCA e PLSR   | Usaram espectroscopia NIR visível (modo de refletância) para prever o teor de açúcar em mais de uma variedade de maçã. Também houve a aplicação de pré tratamentos como 1° e 2° derivada de Savitzky-Golay, e MSC. Apresentaram melhor desempenho para pH ( $SEP = 0.068$ ) e ( $R_p = 0,93$ ) e TSS ( $SEP = 0,61$ ) e ( $R_p = 0,82$ ). | (LAMMERT YN et al., 1998)         |
| Cereja doce | TSS       | 800 - 1700 nm | PLSR         | PLSR foi aplicado para prever teor de açúcar em cerejas usando espectroscopia NIR. Testaram seleção de bandas, mas sem haver melhoria no desempenho. Apresentaram desempenho de ( $R_p = 0.89$ ) e ( $SEP = 0,655$ ).                                                                                                                     | (LU, 2001)                        |
| Ameixa      | TSS       | 700 - 1100 nm | PLSR         | Usaram espectrômetro NIR de baixo custo (modo absorbância) para prever o teor de sólidos solúveis em ameixas inteiras. Usaram o método de validação cruzada e apresentaram desempenho de ( $R^2 = 0.96$ ) e ( $SEP = 1,02\%$ ).                                                                                                           | (SLAUGHTER; THOMPSON; TAN, 2003)  |

|          |          |                            |            |                                                                                                                                                                                                                                                                                   |                                                  |
|----------|----------|----------------------------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|
| Manga    | TSS      | 700 - 1100 nm              | PLSR       | Analisaram a qualidade e os estádios de maturação e previram o teor de sólidos solúveis. Apresentaram desempenho de (SEP = 0.55).                                                                                                                                                 | (SARANWO NG;<br>SORNSRIVI CHAI;<br>KAWANO, 2004) |
| Pera     | TSS      | 350-1800 nm                | PLSR       | Testaram as relações entre as medições do infravermelho próximo (Vis-NIR) por refletância difusa e os índices de qualidade interna da pera. Apresentaram desempenho ( $R_p = 0,912$ ) e (SEP = 0,662).                                                                            | (LIU; CHEN; OUYANG, 2008)                        |
| Laranjas | TTA, TSS | 350-2500 nm<br>Portátil    | PLSR       | Dois espectrômetros portáteis comerciais foram comparados para previsões de sólidos solúveis, acidez titulável e índice de maturidade, em polpa de laranja. Em modo de absorbância. Apresentaram melhores desempenho em para TSS (RPD = 1.03), TTA (RPD = 1.07) e pH (RPD = 0.8). | (CAYUELA; WEILAND, 2010)                         |
| Abacate  | Umidade  | 1000 - 1800 nm<br>Portátil | PLSR       | Usaram um espectrômetro NIR portátil para medir o teor de umidade do abacate (modo refletância). Apresentaram desempenho de ( $R = 0,96$ ) e (SEP = 2,78).                                                                                                                        | (BLAKEY; VAN ROOYEN, 2011)                       |
| Manga    | TSS pH   | 1200–2200 nm<br>Portátil   | PCA e PLSR | Uma fórmula foi desenvolvida para prever o índice de maturidade de cultivares de manga utilizando espectroscopia NIR. Em modo de refletância com desempenho para os modelos de pH ( $R_p = 0.703$ ) e                                                                             | (JHA et al., 2012)                               |

|          |                     |                                     |            |                                                                                                                                                                                                                                                                                              |                          |
|----------|---------------------|-------------------------------------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
|          |                     |                                     |            | TSS ( $R_p = 0.762$ ).                                                                                                                                                                                                                                                                       |                          |
| Maça     | Vitamina C e açúcar | 400–2500 nm                         | PLSR       | Avaliaram 150 genótipos de maçã através da técnica de espectroscopia VIS/NIR com modo de aquisição de refletância. Apresentaram melhor desempenho de ( $SEP = 4,9$ mg/100 g) e ( $R^2 = 0,8$ ).                                                                                              | (PISSARD et al., 2013)   |
| Melancia | TSS, umidade        | 500 –1010 nm<br>Portátil            | PLSR e PCA | Testaram o espectrômetro VIS-NIR portátil para avaliar a qualidade da melancia de forma não destrutiva em diferentes estádios de maturação. Predisseram valores de umidade e TSS com desempenho de ( $RMSEP = 0,717$ e $RPD = 1,83$ ) e ( $RMSEP = 0,276$ e $RPD = 2,79$ ), respectivamente. | (QI et al., 2014)        |
| Maracujá | Vitamina C          | 703 -1090 nm<br>Portátil            | PLSR       | Utilizaram espectroscopia VIS/NIR (modo de transmissão e interação) na região 703-1090 nm, para a quantificação do ácido ascórbico no maracujá. Apresentaram melhor desempenho de ( $R_p = 0.859$ e $SEP = 0.24$ ).                                                                          | (MANIWAR A et al., 2014) |
| Tomate   | TSS                 | 950 - 1800 nm<br>Portátil           | PLSR       | Realizaram a aquisição espectral no espectrômetro portátil NIR (pistola) (modo absorbância), durante o processo de maturação do início até a colheita dos tomates para prever o teor de sólidos solúveis. Apresentaram desempenho de ( $RPD = 2.2$ ) de validação.                           | (CAMPS et al., 2016)     |
| Maça     | brix                | 10,000 - 4000 $cm^{-1}$<br>Portátil | PLSR       | Estudaram a influência da variação da posição de medição do espectro em espectrômetro no infravermelho próximo (NIR) para quantificar teor de sólidos solúveis (SSC) em maçã. Os espectros foram coletados em torno das posições do caule e cálice para cada maçã e tiveram                  | (FAN et al., 2016)       |

desempenho de  $R_p = 0.977$  e  $RMSEP = 0.386$ .

|                                 |              |                            |                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                              |
|---------------------------------|--------------|----------------------------|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|
| Pera                            | TSS          | 600–1050 nm                | PLSR             | Um testador portátil, integrando espectroscopia no infravermelho próximo (NIR), foi desenvolvido para determinar a maturidade de peras asiáticas. Apresentaram desempenho para os modelos de TSS de ( $R_p = 0.96$ , $SEP = 0.29$ ).                                                                                                                                                                                                                     | (CHOI et al., 2017)          |
| Bananito                        | Maturação    | 360 –740<br>Portátil       | PCA e PLS-<br>DA | Analísaram Bananito ( <i>Musa acuminata</i> , AA) em três estádios de maturação (estádios 2, 4 e 6). A análise discriminante de mínimos quadrados parciais (PLS-DA), foram aplicados para a classificação do estádio de maturidade com precisão de 93.3%.                                                                                                                                                                                                | (PU et al., 2019)            |
| Sapota                          | Maturação    | 350 –2500 nm<br>Portátil   | PCA e<br>PLS-DA  | Analísaram a discriminação entre sapotas com maturação natura e forçada. Utilizaram espectrômetro NIR portátil na em modo absorbância e tiveram precisão de 100%.                                                                                                                                                                                                                                                                                        | (MAHANTI; CHAKRABORTY, 2020) |
| Laranja, limao, lima, tangerina | AA, TSS, TTA | 900 – 1,700 nm<br>Portátil | PLSR             | Utilizaram um sensor de espectroscopia infravermelho próximo portátil (NIRS) em combinação com a quimiometria para desenvolver modelos robustos e precisos para estudar a maturação de várias frutas cítricas (laranjas, limões, clementinas, tangerinas, limas Tahiti) e suas respectivas vitaminas C. A medições foram feitas em modo absorbância e o desempenho para AA ( $R^2_{CV} = 0.87$ ), TSS ( $R^2_{CV} = 0.97$ ) e TTA ( $R^2_{CV} = 0.92$ ). | (SANTOS et al., 2021)        |

---

|        |               |                                      |      |                                                                                                                                                                                                                                                                                                                     |                        |
|--------|---------------|--------------------------------------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| Pera   | Umidade e TSS | 400 -1000 e 900 –1700 nm<br>Portátil | PLSR | Foram utilizadas estratégias sequenciais de fusão de dados para fundir os dados de dois espectrômetros portáteis, melhorando a previsão de umidade e teor de sólidos solúveis em peras. As medições foram feitas em modo absorbância e o desempenho para TSS foi de (RMSEP = 0.47) e umidade foi de (RMSEP = 0.69). | (MISHRA et al., 2021a) |
| Tomate | TTA e TSS     | 1295 –2611 nm<br>Portátil            | PLSR | Um espectrômetro de infravermelho próximo (NIR) portátil com tamanho reduzido (equipado com um sistema micro-eletromecânico MEMs) foi usado para desenvolver modelos de predição para avaliar parâmetros de qualidade do tomate. O desempenho para TSS foi de ( $R_p$ = 0.88) e para TTA foi de ( $R_p$ = 0.94).    | (BORBA et al., 2021)   |

---

Pesquisas sugerem que alguns exemplos, como esses aqui demonstrados, motivam o uso de seleção de variáveis para melhorar desempenho de estimativa/previsão. A seleção de variáveis pode melhorar o desempenho do modelo, mas ao mesmo tempo pode eliminar alguma redundância útil do mesmo (MEHMOOD et al., 2012).

**CAPÍTULO 4 - ONLINE MEASUREMENT OF CARAMBOLA (*AVERRHOA CARAMBOLA* L.) PHYSICOCHEMICAL PROPERTIES AND ESTIMATION OF MATURITY STAGES USING A PORTABLE NIR SPECTROMETER**

---

**Online measurement of carambola (*Averrhoa carambola* L.) physicochemical properties and estimation of maturity stages using a portable NIR spectrometer**

Ingrid A. de Moraes<sup>a</sup>; J.P. Cruz-Tirado<sup>a</sup>; Douglas Barbin<sup>a,\*</sup>

<sup>a</sup> Department of Food Engineering, School of Food Engineering, University of Campinas, Campinas, SP, Brazil

\* Corresponding author: [dfbarbin@unicamp.br](mailto:dfbarbin@unicamp.br)

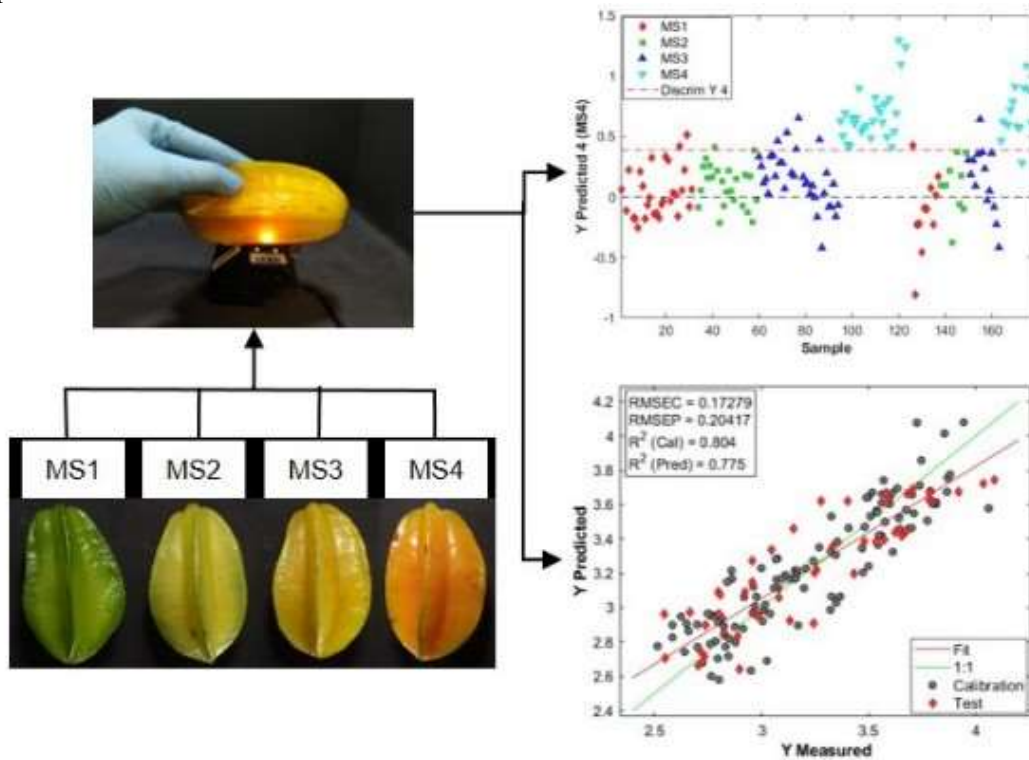
**Abstract**

Carambola is a tropical fruit with rising value in developed countries due to its nutritional value and exotic aspect. It is important to assess carambola quality in different maturity stages, to estimate a “fair price” and to assign fruit for specific applications and markets. This work reported the use of portable NIR spectrometer in the range of 900 to 1700 nm as non-destructive, chemical-free technique for determination of carambola physicochemical properties, according to maturity stage. Colour, total soluble solids, ascorbic acid, moisture, pH and titratable acidity analysis were performed for 177 fruit from two clones and four maturity stages (MS1, MS2, MS3 and MS4). PLS-DA and PLSR models were built to classify carambola according to maturity stage and to estimate its physicochemical properties, respectively. Genetic algorithm (GA) and interval partial least square (iPLS) were tested for improving model performance. The PLS-DA model based on important variables selected by iPLS achieved the best performance with 81.3 % accuracy to classify carambola according to maturity stage. Variable selection (iPLS and GA-PLS) allowed an improvement in the performance of the PLSR models, with pH and moisture content achieving ( $R^2_P$  of 0.78 and 0.74), (RMSEP of 0.2 and 0.87), (RPD of 2.01 and 2.23) and (RER of 8.02 and 10.38), respectively, which is acceptable for screening. Portable NIR spectrometer, which can be considered low-cost when compared to other non-destructive techniques, in tandem with

chemometrics can be a promising tool to assess the composition and to classify carambola according to maturity stage.

**Keywords:** machine learning; start fruit; ripeness; spectroscopy

**Graphical abstract**



## Highlights

- Portable NIR spectrometer was used for *on-line* prediction of maturity stage.
- Performance of NIR spectrometer to predict physico-chemical properties was tested.
- PLSR model for moisture content prediction showed a RER > 10 and RPD > 2.
- PLS-DA achieved 81.3 % for correct classification of carambola according maturity.
- GA and iPLS improved regression models performance with reduction of RMSEP.

---

List of abbreviations and symbols

*GA* – Genetic algorithm

*iPLS* – Interval partial least squares

*LV* – latent variables

*MC* – mean centering

*PCA* – principal component analysis

*PLS-R* – Partial least squares regression

*PLS-DA* – Partial least square discriminant analysis

*RMSEC* – Root mean square error of calibration

*RMSECV* – Root mean square error of cross validation

*RMSEP* – Root mean square error of prediction

*RER* – ratio error

*RPD* – ratio of standard deviation

$R^2_C$  – coefficient of determination of calibration

$R^2_{CV}$  – coefficient of determination of cross validation

$R^2_P$  – coefficient of determination of prediction

*S-G* – Savitzky–Golay

*SNV* – standard normal variate

*VSN* – variable sorting for normalization

#### 4.1. INTRODUCTION

Carambola (*Averrhoa carambola* L.), also known as starfruit, is a rich source of minerals such as magnesium and potassium, and natural antioxidants such as vitamin C and carotenoids (Jia et al., 2017; Muthu et al., 2016). Carambola is mainly sold as fresh fruit, but it has been consumed in different ways and directed by regional medicine to various treatments (Dasgupta et al., 2013). For this reason, it is important to assess carambola quality and physicochemical properties in different maturity stages, in order to estimate a “fair price” and to derivate fruit for specific applications (Yahaya and Omar, 2017).

Harvesting the fruit at the appropriate maturity stage is essential for the best quality and desired flavour (Jie and Wei, 2018; Maduwanthi and Marapana, 2019). During the ripening phase, the fruit becomes soft, produces volatile compounds, decreases acidity and increases total soluble solid (SSC) (Yahaya and Omar, 2017). Fruit harvested in early stages contain less sugar and are prone to wilting and suffering mechanical damage. On the other hand, overripe fruit produce compounds with undesirable flavours, tend to become soft, floury, and easily damaged soon after harvest (Jie and Wei, 2018; Maduwanthi and Marapana, 2019). Carambolas harvested in early maturity stages are destined to exportation due to their longer shelf life, but have not yet developed full flavour and have limited use for garnishes, salads and drinks (FAMA, 2014). On the other hand, in advanced maturity stages, carambola can be marketed for jellies, puddings, pies and especially fresh, although they are directed only to deliveries to nearby destinations (FAMA, 2014; Pauziah et al., 2010). Therefore, it is important to know physicochemical properties of carambola at maturity stages.

Currently, carambola quality inspection is based on colour distribution on the fruit, size and shape (FAMA, 2014; Stan, 1981). Fruit classification is carried out by visual inspection of trained evaluators, generating a great variability in the results, since this task is subjective and does not allow adequate standardization (Yahaya and Omar, 2017). On the other hand, even with a correlation between colour and physicochemical parameters related to flavor in the literature, skin colour does not predict the appropriate day of harvest, fruit quality or sweetness and acidity levels (Muhammad Zahir et al., 2021). Therefore, the quick and reliable prediction of physico-chemical properties such as total soluble solids, titratable acidity, pH and vitamin C can improve correct classification of carambola (Amir et al., 2019; Kyriacou and Roupheal, 2018).

Traditional analytical techniques used to assess physico-chemical properties related to the quality of the fruit are destructive, slow and involve the use of chemicals (Alander et al., 2013). Fruit industries require fast, non-destructive and chemical free

analytical techniques for implementing *on-* and *in-line* measurements. Near Infrared Spectroscopy (NIRS) requires a minimum of sample preparation, is chemical free and allows quick characterization and quantification of several attributes in agricultural products, gaining interest in recent years (Correia et al., 2018; Friedrich et al., 2014). In addition, according to Magwaza (2012), NIRS is more advanced compared to other non-destructive techniques in terms of instrumentation, accessories, availability of suitable chemometric software packages and even applications in the food sector. Recently, the reduction in components size, such as leds and chips, allowed to develop handheld/portable NIR spectrometers (Pasquini, 2018). Portable NIR spectrometers have an ergonomic design, are transportable for different environmental work and, when well calibrated, can be an alternative for benchtop devices (Cruz-Tirado et al., 2021). Portable NIR spectrometers have demonstrated good performance to estimate the quality of fresh tomatoes in the field (Borba et al., 2021), instantaneous and simultaneous prediction for anthocyanins and sugar in whole fresh raspberries (Gales et al., 2021), the internal quality of oranges, lemons and tangerines (Santos et al., 2021), prediction of moisture and soluble solids content in pear (Mishra et al., 2021), prediction of soluble solids, titratable acidity and firmness of apples (Pissard et al., 2021), and the evaluation of the content of soluble solids in bananas (Sripaurya et al., 2021). However, to the best of our knowledge, there is no research available in the literature on the evaluation of the quality of carambola using portable NIR spectrometer.

Therefore, the objective of this work is to develop a methodology based on a portable NIR spectrometer for the online determination of physicochemical properties and the maturity stage of carambola.

## **4.2. MATERIAL AND METHODS**

### **4.2.1. Carambola samples**

Several batches of fruits were obtained from local commerce (CEASA supply station, Campinas-SP) in different dates, totalling 177 samples from B17 and B10 varieties. These are the most popular clones that are commercially grown: ‘B17’ are sweeter, less acidic and has better eating quality as compared to ‘B10’, but it has shorter shelf-life (Abd Rahman and Ahmad Hafiz, 2013). First, fruits were disinfected in Chlorine-based solution ( $0.2 \text{ g L}^{-1}$ ) and stored at  $23 \text{ }^{\circ}\text{C}$  until the time of analysis to standardize temperature. After colour analysis, samples were classified in 7 levels, according to the Federal Agricultural Marketing Authority (FAMA, 2014). However, since the colour change between the stages was very subtle, a new classification standard with 4 stages maturity (MS) was proposed. The

fruits were classified based on the hue angle ( $h^\circ$ ) into MS1 (maturity stage 1, green hue  $> 100^\circ$ ), MS2 (maturity stage 2, green/yellow,  $92^\circ < \text{hue} \leq 100^\circ$ ), MS3 (maturity stage 3, yellow,  $83^\circ < \text{hue} \leq 92^\circ$ ) and MS4 (maturity stage 4, yellow/orange hue, ( $\text{hue} \leq 83^\circ$ ))(Fig. 4.1).

#### 4.2.2. Physical and physico-chemical analyses

The physical and physico-chemical parameters were measured after the acquisition of the spectra. Colour parameters were measured using a colorimeter (CM-2600D, illuminant D65 Konica Minolta Sensing Inc., Osaka, Japan), calibrated with a white ceramic standard, and the result was expressed in terms of lightness ( $L^*$ ), green-red ( $a^*$ ) and blue-yellow ( $b^*$ ) of the CIELAB system (CIE, 1976; McLAREN, 2008). In addition, the amount of Chroma ( $C^*_{ab}$ ) (Eq. 1) and hue ( $h_{ab}$ ) (Eq. 2) were calculated to compare the maturity stages (Theanjumol et al., 2019).

$$C^*_{ab} = \sqrt[3]{a^{*2} + b^{*2}} \quad (\text{Eq. 1})$$

$$h_{ab} = \tan^{-1}\left(\frac{b}{a}\right) \quad (\text{Eq. 2})$$

Moisture content, total soluble solids content (TSS), ascorbic acid content (AA) - vitamin C, pH and titratable acidity (TTA) were analysed. The moisture content of the samples were obtained by the gravimetric method after drying 4 g of sample in a vacuum oven at  $60^\circ\text{C}$  until constant weight (AOAC, 2006). The content of total soluble solids was determined using a manual refractometer model (KASVI, K52-032), on a scale from 0 to 32 %, and calibrated with distilled water AOAC 932.12. The pH was determined using a pH meter (model MB-10; Marte, São Paulo, Brazil), titratable acidity and vitamin C were measured from the juice of the fruit (Nielsen, 2017).

#### 4.2.3. Acquisition of NIR spectra

Acquisition of NIR spectra in the 902 - 1700 nm range, with 4 nm intervals was performed on the whole fruit in absorbance mode, directly on the peel, using a portable spectrometer (DLPR NIRscan™ Nano, Texas Instruments, USA), with a 10 W halogen lamp, and NIRscan™ Nano software. For each sample, three measurements (and 10 replicates for the same measurement) were recorded, at different positions along the fruit, in order to obtain a better representation of the samples.

##### 4.2.3.1. Spectra pre-processing

The original data set was randomly divided into two subsets; 70 % of the data (123) were used for calibration and the remaining 30 % (54) for prediction (external test set). The spectra were mean centered and then pre-processing methods were tested: smoothing Savitzky-Golay (S-G) utilizing a window size of 7 to 13 by performing a local polynomial regression (order 0) to remove noise. First and second derivatives were used to correct the baseline, the derivatives were calculated using the Savitzky-Golay algorithm, with a window size of 7 to 15 data points. Multiplicative scatter correction (MSC) and the standard normal variate (SNV) were applied to correct the effects of light scattering (Barnes et al., 1989; Martens and Geladi, 1983) . The variable sorting for normalization (VSN) was also tested. This novel algorithm, when introduced in SNV pre-processing, significantly improves signal shape and model interpretation. The method assumes that not all bands are equally altered by the unwanted effects and, consequently, assigns a weight to each variable in the interval [0.1] corresponding to its probability of being affected only by the dispersion. This estimates the extent to which a wavelength is affected by size effects (additive and multiplicative displacements) rather than shape effects (resources attributable to chemically relevant contributions) (Rabatel et al., 2020).

#### 4.2.3.2. Exploratory analysis

As an exploratory technique, principal component analysis (PCA) was performed identify whether the spectra was affected by carambola maturity stages. PCA models were developed with the full wavelength range, using the singular value decomposition algorithm (SVD) (95 % confidence level). For the construction of the models, outlier's (anomalous spectra) were detected and eliminated using the amount of  $Q_{\text{residual}}$  and Hotelling  $T^2$ .

#### 4.2.3.3. Classification of maturity stage

Partial least squares discriminant analysis (PLS-DA) was applied to classify carambola according to its maturity stage considering four classes (MS1, MS2, MS3 and MS4). The number of latent variables (LV) for the PLS-DA models was selected using the lowest root mean squared error for cross-validation (RMSECV), which was performed using *ten subsets randomly constituted* (Barbin et al., 2013). This procedure was similar for partial least squares regression (PLSR). The performance of the classification models was expressed by Sensitivity (Eq. 3), Specificity (Eq. 4), Error rate (Eq. 5) and Accuracy (Eq. 6) in the calibration, cross-validation and external validation data sets.

$$\text{Sensitivity ( \% )} = \frac{TP}{(TP+TN)} \times 100 \quad (\text{Eq. 3})$$

$$\text{Specificity ( \% )} = \frac{TN}{(TN+FP)} \times 100 \quad (\text{Eq. 4})$$

$$\text{Error rate ( \% )} = \frac{(FP+FN)}{TOTAL} \times 100 \quad (\text{Eq. 5})$$

$$\text{Accuracy ( \% )} = \frac{(TP+TN)}{TOTAL} \times 100 \quad (\text{Eq. 6})$$

Where, TP = true positive; FN = false negative; TN = true negative; and FP = false positive.

#### 4.2.3.4. Prediction of physicochemical properties

Partial Least-squares Regression (PLSR) was applied using NIR spectra as predictors for each characteristic analysed. The spectral region from 902 – 913 nm and 1649 – 1698 nm was removed to avoid noise. The performance of the regression models was assessed by coefficient of determination ( $R^2$ ), number of latent variables (LV) and root mean squared error (RMSE) for calibration (RMSEC), cross-validation (RMSECV) and prediction (RMSEP) (Skibsted et al., 2004), residual predictive deviation (RPD) and the range error ratio (RER) (Barbin et al., 2015). RPD is useful to assess overall performance of prediction models, where  $RPD < 1.5$  indicates unusable model,  $1.5 < RPD < 2.0$  states that the model can distinguish between high and low values,  $2.0 < RPD < 2.5$  suggests a model with approximate quantitative predictions possible,  $2.5 < RPD < 3.0$  indicates a good prediction model and,  $RPD > 3$ , indicates excellent predictive capacity of the model (Saeys et al., 2005). According to the American Cereal Chemicals Association (AACC), the PLSR models performance also can be accessed by RER values:  $RER \geq 4$  is qualified for screening calibration, when the  $RER \geq 10$  the model is acceptable for quality control, and if the  $RER \geq 15$  the model is very good for quantification (Rambo et al., 2013).

#### 4.2.3.5. Variable Selection

##### Iterative partial least square (iPLS)

The iPLS variable selection method is an extension developed for PLS, where a partial square linear regression is performed in each equidistant interval along the entire

spectrum. The results are presented in the form of a graph, variance versus the number of latent variables, in order to facilitate comparison across the spectral range. In the application of the iPLS algorithm, the spectra were divided into equidistant intervals by automatic configuration with tested sub-windows from 1 to 60 (Xiaobo et al., 2010).

#### Genetic algorithm

The genetic algorithm (GA) is a random search and global optimization algorithm and is very effective to select the characteristic wavelength. The RMSECV was used as a model fitness scale, and window were tested with sub-windows from 1 to 50, double crossover, maximum number of generations of 100, maximum number of latent variables of 20, and one iteration were used to obtain the wavelengths associated with best correlation between spectra, and reference values (Rady and Guyer, 2015).

#### 4.2.4. Statistical analysis

The physico-chemical results for each class were compared using the analysis of variance (ANOVA). Subsequently, the Tukey multiple comparison test was performed ( $p < 0.05$ ), in order to analyse the difference between the sample sets of each maturity stage. Multivariate statistical analysis was performed using Matlab R2019a (Mathworks, Natick, USA) using PLSToolbox from Eigenvector Research, Inc. (Manson, WA, USA).

### 4.3. RESULTS AND DISCUSSION

#### 4.3.1. Physicochemical properties

After classification of maturity stages, it was observed samples with a similar value for hue angle, but with discrepant values of soluble solids content, pH and TTA. This makes the spectroscopy technique relevant, as besides the classification of fruits based on their hue, the technique also allows for the prediction of physicochemical parameters. Thus, it is possible to identify ripe fruits with low sweetness and high acidity. During ripening, carambola becomes soft, produces volatile compounds, increases its sugar content and decreases the water content and acidity (Yahaya and Omar, 2017). The results showed, as expected, relevant changes between the extreme maturity stages in all physico-chemical properties (Tukey's test,  $p < 0.05$ ), highlighting the values of pH, TSS (Table 4.1) and colour (Fig. 4.1).

**Table 4.1.** Statistical results of the physicochemical properties of the carambola samples in four different maturity stages.

| Class/<br>Analyse                       | MS1                |                   | MS2               |                     | MS3               |                    | MS4           |                    |
|-----------------------------------------|--------------------|-------------------|-------------------|---------------------|-------------------|--------------------|---------------|--------------------|
|                                         | Range              | Mean $\pm$ SD     | Range             | Mean $\pm$ SD       | Range             | Mean $\pm$ SD      | Range         | Mean $\pm$ SD      |
| AA ( $10^{-2}$ g Gallic acid $L^{-1}$ ) | 1.94 - 41.31       | 17.28D $\pm$ 9.71 | 2.80 - 53.85      | 21.63BC $\pm$ 12.56 | 5.32 - 63.85      | 28.03B $\pm$ 15.48 | 10.38 - 75.77 | 37.38A $\pm$ 14.00 |
| pH                                      | 2.51 - 3.08        | 2.80D $\pm$ 0.14  | 2.72 - 3.56       | 3.09C $\pm$ 0.27    | 2.85 - 4.15       | 3.40B $\pm$ 0.27   | 3.49 - 4.13   | 3.76A $\pm$ 0.16   |
| TTA (% oxalic acid)                     | 0.36 - 0.73        | 0.56C $\pm$ 0.10  | 0.26 - 0.65       | 0.40BC $\pm$ 0.08   | 0.22 - 0.58       | 0.37B $\pm$ 0.07   | 0.22 - 0.57   | 0.34A $\pm$ 0.08   |
| TSS (%)                                 | 6.60 - 10.00       | 8.15D $\pm$ 0.84  | 7.59 - 14.60      | 9.51C $\pm$ 1.69    | 7.99 - 13.60      | 10.87B $\pm$ 1.57  | 8.60 - 15.89  | 11.65A $\pm$ 1.32  |
| Moisture (%)                            | 86.95 - 92.37      | 89.44A $\pm$ 1.23 | 83.89 - 91.90     | 88.81AB $\pm$ 2.27  | 83.57 - 91.99     | 88.17B $\pm$ 2.20  | 85.19 - 92.58 | 88.36B $\pm$ 1.32  |
| L*                                      | 34.59 - 48.00      | 42.22A $\pm$ 3.66 | 33.94 - 58.38     | 43.60A $\pm$ 4.92   | 35.01 - 49.94     | 42.84A $\pm$ 4.14  | 29.31 - 40.60 | 36.90B $\pm$ 2.27  |
| a*                                      | (-11.27) - (-4.92) | -8.19D $\pm$ 1.57 | (-6.84) - (-1.15) | -3.31C $\pm$ 1.41   | (-1.12) - (-4.52) | 1.16B $\pm$ 1.62   | 4.29 - 12.18  | 8.00A $\pm$ 1.99   |
| b*                                      | 25.88 - 40.83      | 35.02B $\pm$ 3.66 | 24.64 - 42.67     | 32.79C $\pm$ 4.12   | 29.99 - 42.39     | 36.30AB $\pm$ 2.50 | 32.59 - 43.49 | 37.69A $\pm$ 2.21  |

\* The data correspond to the mean  $\pm$  SE of two repetitions. Different letters for the same parameter analysed

indicate significant differences, between the maturity stages, by ANOVA ( $P < 0.05$ ). (TSS) content of total soluble solids, (AA) content of ascorbic acid (vitamin C), (TTA) titratable acidity.

| Class                                                                               | MS1                                                                                 |                    | MS2                                                                                 |                   | MS3                                                                                  |                    | MS4                                                                                   |                   |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|--------------------|-------------------------------------------------------------------------------------|-------------------|--------------------------------------------------------------------------------------|--------------------|---------------------------------------------------------------------------------------|-------------------|
|  |  |                    |  |                   |  |                    |  |                   |
| Analyse                                                                             | Range                                                                               | Mean $\pm$ SD      | Range                                                                               | Mean $\pm$ SD     | Range                                                                                | Mean $\pm$ SD      | Range                                                                                 | Mean $\pm$ SD     |
| $h_{ab}$                                                                            | 100.14 - 107.12                                                                     | 103.10A $\pm$ 1.69 | 92.02 - 99.74                                                                       | 95.84B $\pm$ 2.49 | 83.03 - 91.87                                                                        | 88.24C $\pm$ 78.60 | 71.55 - 82.95                                                                         | 78.06D $\pm$ 2.87 |
| $C^*_{ab}$                                                                          | 26.34 - 42.33                                                                       | 35.98B $\pm$ 3.84  | 24.88 - 42.89                                                                       | 32.99C $\pm$ 4.10 | 29.99 - 42.12                                                                        | 36.36B $\pm$ 2.52  | 33.30 - 44.12                                                                         | 38.58A $\pm$ 2.28 |

**Figure 4.1.** Maturity stages for carambola according to hue ( $h_{ab}$ ) value

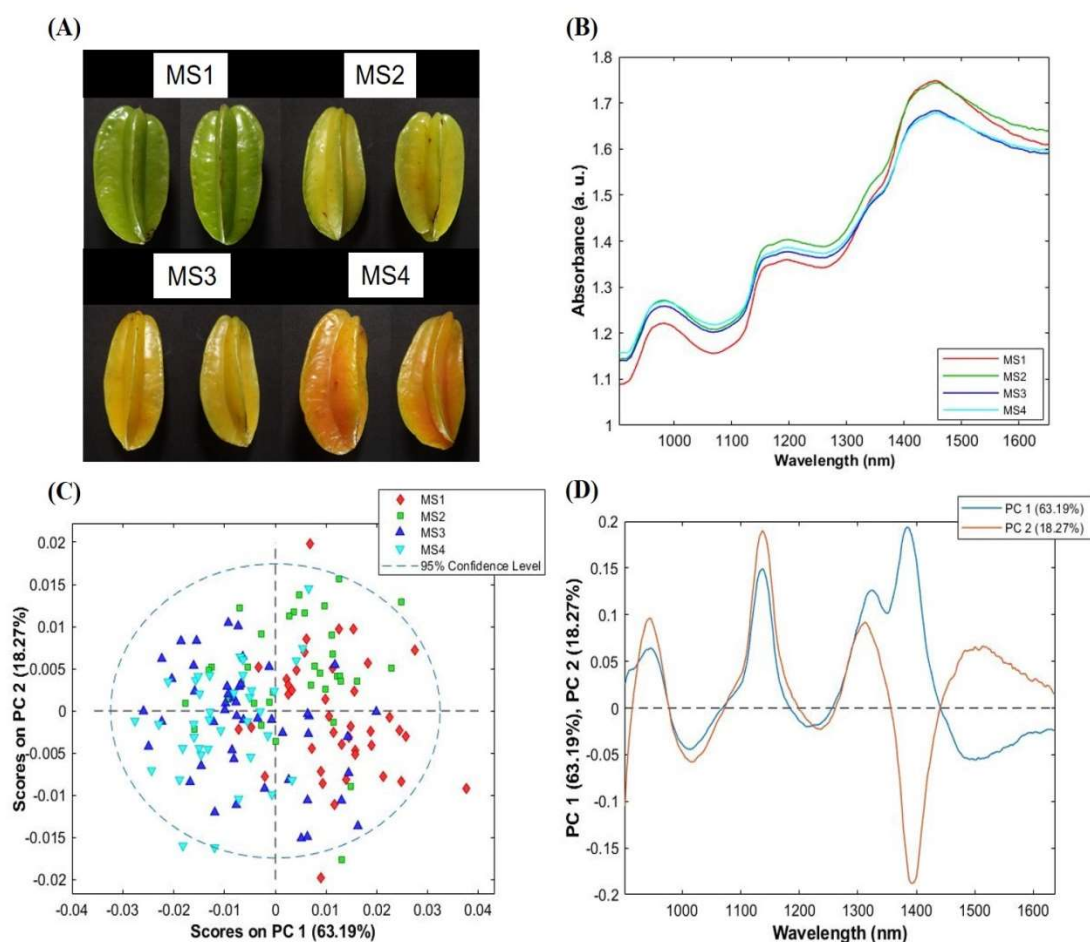
The colour change from green (with the reduction of  $a^*$  values) to yellow (increase in  $b^*$  values) of the carambola (Fig. 4.2A) during ripening is due to the degradation of chlorophyll and the increase in carotenoids (O'Hare, 1993). In addition, the last stage (MS4) showed lower values of  $L^*$  than the other stages ( $p < 0.05$ , Tukey's test), thus revealing less luminosity in the surface. Changes in parameters  $a^*$  and  $b^*$  also influenced saturation ( $C^*_{ab}$ ) and tone ( $h_{ab}$ ). Hue greater than  $90^\circ$  represent greener fruit, while values close to and less than  $90^\circ$  are yellow (Boonyaritthongchai and Kanlayanarat, 2012). Besides, some authors have reported that this development of the colour of the fruit may be correlated with the increase in the level of sugars such as glucose, fructose and sucrose (Rahman, 2007; Soumya and Nair, 2014).

The levels of TSS (6.60 to 15.89 %) are directly related to different stages of ripeness (Table 4.1), similar to results reported by Warren and Sargent (2011), and being lower for MS1 and higher for MS4. It was observed that TSS levels varied in different regions of the fruit, as factors such as maturation, growing conditions and fruit size can also interfere in TSS content (Magwaza et al., 2012). In contrast, moisture content had less variation and was similar to those reported in the literature (Yeasmin et al., 2021). Carambola on MS1 and MS2 stages had a lower pH ( $p < 0.05$ , Tukey's test) and higher TTA compared to MS3 and MS4 fruit. The pH values and TTA were similar to previous pH (3.31 to 3.51) and TTA (0.58 to 0.83 %) values reported (Ding et al., 2007; Huang et al., 2018). The main organic acids in carambola are the oxalic acid followed by malic acid, which act as intermediaries in the metabolic processes and are directly involved in growth, maturation and senescence of fruit (Soumya and Nair, 2014). Organic acids determine the characteristic flavor of the fruit and play an important role in fruit quality and consumer acceptance (Soumya and Nair, 2014). On the other hand, AA content showed an increase with the ripening of carambola from 17.28 g Gallic acid  $L^{-1}$  of juice in MS1 stage to 37.38 g Gallic acid  $L^{-1}$  juice in MS4. These values were similar to the values (21 to 40.2  $10^{-3}$  g Gallic  $g^{-1}$  sample) reported by Ali and Yunus Jaafar (1992). High values of AA (or vitamin C) are important for health due to its ability to eliminate free radicals in biological systems, preventing diseases (Fageeh et al., 2021).

#### 4.3.2. Spectra profile and exploratory analysis

Figure 4.2C shows the spectra profile and PCA analysis for carambola in different

maturity stages (MS1-4). Fruit contain about 80–90 % water and the remaining fraction is composed of carbohydrates, organic acids, vitamins and minerals (Magwaza et al., 2012). The NIR spectra of samples, with high percentage of water such as carambola (Figure 4.2A), have a combination of overtones and fundamental vibrations, mainly attributed to the organic bonds (O - H, C - H and N - H), that make absorption bands relatively broad and complex (Magwaza et al., 2012). The spectral regions of 970 nm and 1450 nm associated to the second and first harmonic of O-H, respectively, showed differences in absorption for the classes (Nicolai et al., 2007), which are related to moisture content. The bands near 1445 nm and 1000 nm are related to the organic acids present in the carambola, such as oxalic and ascorbic. These regions correspond to the first and second overtone regions of O – H bonds, respectively and are difficult to see because they overlap with the bands assigned to water. The difference in the absorbance value of these regions may be associated with a drop in the levels of organic acids and an increase in the pH seen during the maturation process (Weyer and Lo, 2006). The absorption bands associated to second overtones (O–H) in 920 nm, third overtones (C-H) in 910 nm and the second connection (C - H) in 1190 nm are associated to sugars such as glucose and fructose in carambola (Golic et al., 2003; Osborne et al., 1993). These regions showed differences in absorption for the stages and may be related to the increase in soluble solids contents from the stages MS1 to MS4. These bands are very close to the regions where the water has strong absorption, which makes it difficult to visualize and generates overlapping (Delwiche et al., 2008).



**Figure 4.2.** A) Carambola in different maturity stages (MS1, MS2, MS3 and MS4), B) Mean spectra of carambola in different maturity stages, C) Scores plot based on 1<sup>st</sup> S-G derivative + smoothing S-G spectra (902–1637 nm), and D) Loadings plot of PCA analysis.

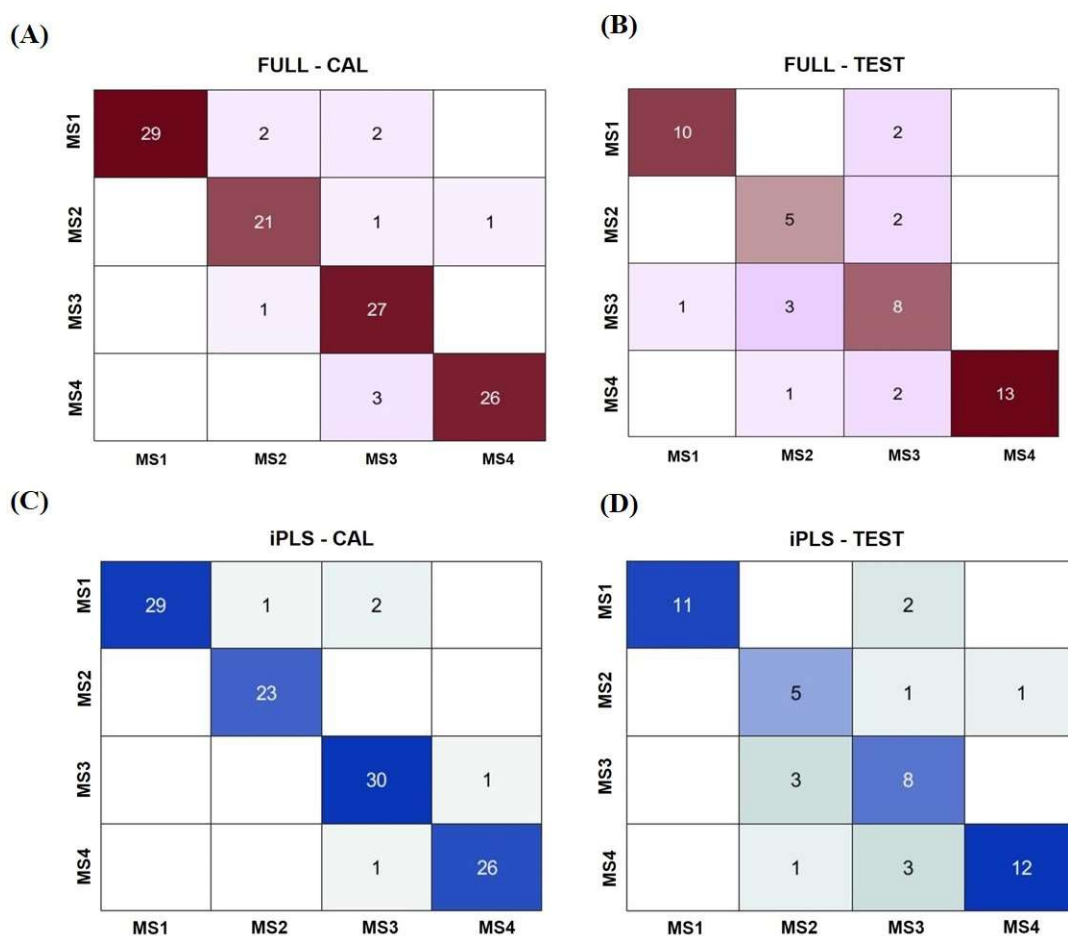
The principal component analysis (PCA) was performed using spectral region from 902 to 1637 nm. The spectral region from 1637 – 1698 nm was removed because it presented some noise. The raw spectral data was mean centered (MC) and then pre-processed using smoothing S-G + 1<sup>st</sup> S-G derivative. Figure 4.2C shows the score plot, where the first two principal components explained 81.46 % of variability for carambola in different maturity stages. The scores for the carambola moved from positive PC1 scores to negative PC1 scores, which is from MS1 stage (unripe) to MS4 stage (ripe). MS1 fruit are clearly localized on positive region of PC1 scores, while MS4 are located on negative region of PC1 scores. Both MS2 and MS3 are overlapped and they are in the center of score plot, indicating the transition phase from MS1 stage to MS4 stage. The loadings plot in Fig. 4.2D showed that the

peaks associated with water (970 and 1450 nm), organic acids (1000 and 1445 nm) and sugars (910, 920 and 1190 nm) are important for class discrimination.

#### 4.3.3. Discrimination model

The confusion matrix of PLS-DA models for carambola classification according to maturity stages is shown in Fig. 4.3. Numbers of correctly classified maturity stages are shown diagonally, while numbers outside the diagonal denote erroneous classifications. Smoothing S-G with 11-point (order 0; window 11) + 1<sup>st</sup> S-G derivative (order 2; window 7) showed the best performance. The best PLS-DA model reached accuracy of 97.22 % for samples from MS1 (green), 75.15 % for samples from MS2 (green/yellow), 74.03 % for samples with from (MS3 (yellow, and 90.27 % for samples from MS4 (yellow/orange) (Table 4.3, in supplementary material). The accuracy obtained in prediction set were similar to accuracy for calibration set, indicating that models are free from over and under fitting (Fig.4.3). The loadings of the PLS-DA model (Fig. 4.6, in supplementary material) showed that regions related to the contents of water, acids and sugars are important for classification of carambola. PLS-DA model was less sensitive and specific for MS2 and MS3 samples than MS1 and MS4 samples in both calibration and prediction set (Table 4.3, in supplementary material). In addition, these intermediate stages had more samples incorrectly classified, which may be related to the smooth and gradual change of some components during the maturation process, such moisture, TSS, TTA and AA (Table 4.1)(Yahaya and Omar, 2017). Based on the results, it is possible to assume that the discriminant models obtained are robust to discriminate between MS1 and MS4 classes, MS1 and MS2, and MS3 and MS4, but they have lower classification performance for MS2 and MS3 classes. Costa et al. (2019) classified grapes in three stages of ripeness and also obtained a lower performance of distinction in the stage of intermediate ripening, which was due to the smooth change of colour of grape during its ripening process.

Variable selection was performed using both genetic algorithm (GA) and interval partial least square (iPLS). The GA did not improve classification model. In contrast, iPLS algorithm selected the spectral regions from 905 to 1640 nm (eliminating 21 variables), which showed better performance than full PLS-DA model, with an increasing in accuracy of 11.94 % for MS1 and 1.84 % in MS3 in the test set (Fig. 4.3). In addition, for the calibration set, the MS2 and MS3 increased accuracy of 4.79 % and 7.57 %, respectively, but accuracy in test set was not improved.



**Figure 4.3.** Confusion matrix of PLS-DA models for calibration (CAL) set and validation (TEST) set using full spectra (A and B) and reduced spectra by iPLS (C and D).

#### 4.3.4. Prediction of physico-chemical properties by PLSR

PLSR regression was used to model the relationships between the NIR spectra and the physico-chemical properties of carambola (TSS, TTA, moisture, pH and AA) (Fig. 4.4). The spectral region from 902 – 913 nm and 1649 – 1698 nm was removed to avoid noise. In addition, the iPLS and GA-PLS variable selection methods were applied and the window number was tested until the procedure with the best performance was found in each model. In addition, Fig. 4.4 shows the best PLSR model for each physico-chemical measured parameter.

**Table 4.2.** Statistics for the calibration and prediction sets for physical and physico-chemical analysis in carambola using portable NIR and PLSR, iPLS and GA-PLS.

| Physico-chemical parameter | Pre-processing                                                                       |     | Model | Nubmer of variables | LV | R <sup>2</sup> C | RMSEC | R <sup>2</sup> CV | RMSECV | R <sup>2</sup> P | RMSEP | RPD  | RER  |
|----------------------------|--------------------------------------------------------------------------------------|-----|-------|---------------------|----|------------------|-------|-------------------|--------|------------------|-------|------|------|
| AA                         | Smoothing (order 0; window 15) + 1 <sup>st</sup> S-G derivative (order 2; window 13) | S-G | Full  | 208                 | 9  | 0.67             | 7.72  | 0.52              | 9.43   | 0.53             | 8.47  | 1.85 | 8.72 |
|                            |                                                                                      |     | iPLS  | 175                 | 9  | 0.52             | 7.54  | 0.52              | 9.42   | 0.54             | 8.42  | 1.86 | 8.77 |
|                            |                                                                                      |     | GA    | 145                 | 12 | 0.69             | 7.44  | 0.50              | 9.68   | 0.54             | 8.40  | 1.87 | 8.79 |
| pH                         | Smoothing (order 0; window 11) + 1 <sup>st</sup> S-G derivative (order 2; window 11) | S-G | none  | 208                 | 9  | 0.82             | 0.17  | 0.71              | 0.21   | 0.74             | 0.21  | 1.91 | 7.63 |
|                            |                                                                                      |     | iPLS  | 81                  | 9  | 0.80             | 0.17  | 0.73              | 0.20   | 0.78             | 0.20  | 2.01 | 8.02 |
|                            |                                                                                      |     | GA    | 173                 | 11 | 0.81             | 0.17  | 0.66              | 0.23   | 0.76             | 0.21  | 1.95 | 7.76 |
| TTA                        | Smoothing (order 0; window 9) S-G                                                    | S-G | Full  | 222                 | 9  | 0.69             | 0.06  | 0.56              | 0.08   | 0.66             | 0.07  | 1.79 | 7.55 |
|                            |                                                                                      |     | iPLS  | 74                  | 9  | 0.70             | 0.06  | 0.57              | 0.08   | 0.68             | 0.07  | 1.81 | 7.66 |
|                            |                                                                                      |     | GA    | 120                 | 9  | 0.70             | 0.06  | 0.54              | 0.08   | 0.69             | 0.06  | 1.84 | 7.80 |

|          |                                                                                                   |      |     |   |      |      |      |      |      |      |      |       |
|----------|---------------------------------------------------------------------------------------------------|------|-----|---|------|------|------|------|------|------|------|-------|
| TTS      | Smoothing S-G<br>(order 0; window 13)<br>+ 1 <sup>st</sup> S-G derivative<br>(order 2; window 13) | none | 208 | 7 | 0.69 | 1.00 | 0.61 | 1.12 | 0.64 | 1.12 | 1.75 | 8.33  |
|          |                                                                                                   | iPLS | 104 | 7 | 0.70 | 0.99 | 0.63 | 1.10 | 0.68 | 1.05 | 1.85 | 8.81  |
|          |                                                                                                   | GA   | 18  | 7 | 0.71 | 0.97 | 0.66 | 1.05 | 0.66 | 1.08 | 1.80 | 8.61  |
| Moisture | Smoothing S-G<br>(order 0; window 13)<br>+ 1 <sup>st</sup> S-G derivative<br>(order 2; window 9)  | none | 208 | 9 | 0.78 | 0.85 | 0.68 | 1.04 | 0.72 | 0.90 | 2.14 | 9.98  |
|          |                                                                                                   | iPLS | 56  | 9 | 0.75 | 0.90 | 0.68 | 1.02 | 0.74 | 0.89 | 2.18 | 10.16 |
|          |                                                                                                   | GA   | 133 | 9 | 0.74 | 0.91 | 0.63 | 1.11 | 0.74 | 0.87 | 2.23 | 10.38 |

The  $R^2$  (Cal and Test set) for pH and moisture prediction models were superior to TTA, TSS and AA parameter, as show in Table 4.1. Previous works reported similar  $R^2$  values of 0.656 for TTA and 0.808 for pH in carambola using Vis/NIR spectroscopy (Yahaya and Omar, 2017), and  $R^2 = 0.58$  for TTA in strawberry (Amodio et al., 2017). This could be related to the short range of values for some physico-chemical properties such as TTA, and low concentrations, such as i.e ascorbic acid.

On the other hand, in general, the variable selection by iPLS (TSS and pH) or GA ( % moisture, TTA and AA) improved PLSR model (Table 4.2). Magwaza (2012) have argued that models involving a few carefully selected wavelengths were effective for prediction of chemical composition in citrus fruit. The iPLS algorithm seeks a particularly informative spectral range in relation to the analysed parameter to generate simple models and increase its predictive capacity (Nørgaard et al., 2000). The comparison between the selected range and the full spectrum models help to interpret the relationship between the model and the sample compositions. For its part, GA algorithms also have a great ability to improve results, as it is a random search algorithm and to guarantee the model's reliability, the calculation procedure is repeated many times, until the best variables are found. This decreases the values RMSEP and RMSECV without reducing prediction capacity (Whitley, 1994).

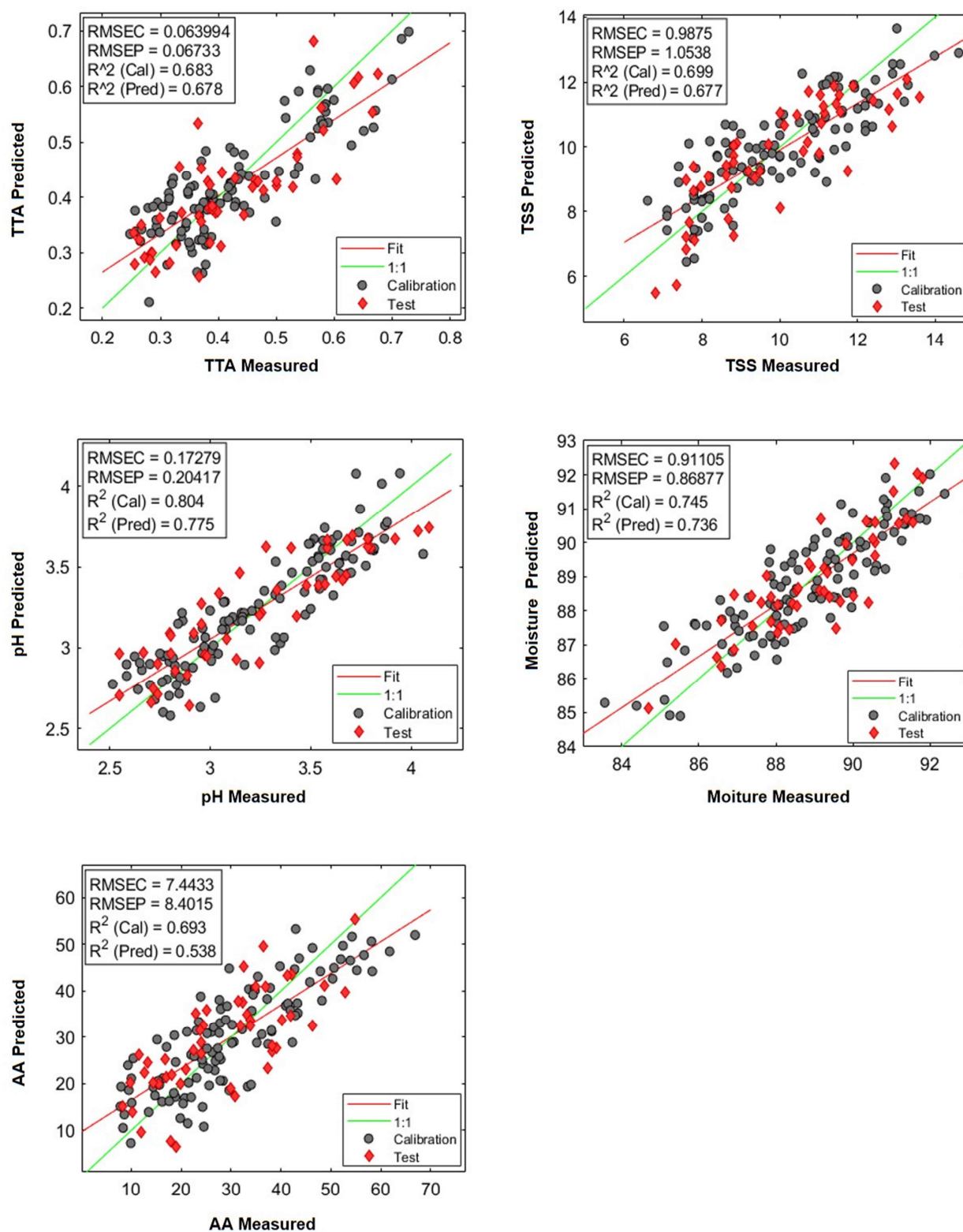
The best PLS model, after the variables selection, for TSS (104 selected variables), for TTA (120 selected variables) and for AA (145 selected variables) showed  $1.5 < \text{RPD} < 2$  values, which it suggests that the model has the ability to distinguish between high and low values (Saeys et al., 2005), and  $4 < \text{RER} < 10$ , which it represents a model for screening calibration (Rambo et al., 2013). According to Nicolăi et al. (2007) the prediction of TTA using NIRS is more difficult, since organic acids tend to be found in relatively lower concentrations, near the limit of detection of the method. However , RPD values in this work for TTA prediction were superior than reported for strawberries (RPD = 1.33) (Amodio et al., 2017), and grape (RPD = 1.10) (Fernández-Novales et al., 2009). On the other hand, Yahaya e Omar (2017) reported similar values  $R^2_P = 0.656$  for prediction of TSS in carambola with Vis/NIR spectroscopy. In addition, it was reported better prediction models for AA in apple (RPD = 2) (Pissard et al., 2013) TTA in purple passion fruit (RPD = 1.8) (Maniwaru et al., 2019) and mango (RMSEC = 0.73 and RMSECV = 0.72) (Jha et al., 2012)

Prediction models with selected variables for pH (81 selected variables) and moisture (133 selected variables) showed  $\text{RPD} > 2$ , which suggests a model with approximate quantitative predictions possible (Saeys et al., 2005). The pH model reached RER of 8.02,

showing that the model can be used for screening. Models for moisture prediction presented  $RER = 10.38$ , suggesting that the model is acceptable for quality control (Rambo et al., 2013). According to Walsh, Mcglone and Han (2020) the water absorption bands dominate the spectrum of fruit, which probably influence the results for moisture prediction. Similar results were reported for pH prediction Yahaya and Omar (2017) with  $RMSEP = 0.188$  for carambola using VIS / NIR spectroscopy.

Carambola contains a high content of water and other chemicals, such as vitamins and soluble solids, and is composed of a thin, nutrient-rich shell (Muthu et al., 2016). The thickness of the shell varies according to the maturity stage, and chemical components may have influenced the prediction models since spectra were acquired directly on the shell. Also, some chemical and physical parameters can be affected by some external factors, such as temperature and relative humidity at the time of harvest, which also affects the accuracy and robustness of prediction (Kusumiyati et al., 2021; Walsh et al., 2020). Previous works reported that the growth conditions, the size of the fruit cells and the tissue structure, in some cases translucent as a star fruit (Fig. 4.7 supplementary material), could interfere with the performance of the model in NIR. This is because they are less predictable interferences related to the transmission of light at the time of acquisition of the spectra (Xie et al., 2021). On the other hand, the use of more than one variety, as used here, can assign robustness to the prediction model in NIR (Walsh et al., 2020; Xie et al., 2021).

The need for alternative methods, which do not generate large amounts of chemical residues and / or costly analyses, makes the NIR method an adequate alternative. The determination of five properties through predictive models capable of distinguishing between high and low values, in this study, would economize long periods for traditional analyses. Research on some fruit, such as mango, has generated great advances; Australian industry has adopted the use of portable NIRS to evaluate fruit in the field, assisting in the decision to harvest and classify after harvest (Walsh et al., 2020). In this line, research on the prediction of physico-chemical properties and classification according to maturity stage of carambola using portable NIR and chemometrics should be addressed and standardized to help fruit producers.



**Figure 4.4.** PLSR models with best performance for total titratable acidity (TTA), total of soluble solids (TSS), pH, moisture and ascorbic acid (AA).

#### 4.4. CONCLUSION

In order to supply high-quality carambolas to customers, it is essential to know the maturity stage and physico-chemical properties of carambola. During ripening, the content of total soluble solids, ascorbic acid and pH increased and the content of titratable acidity and moisture content decreased. The PLS-DA method showed that the fruit spectra contains information that allows correct classification of carambola according to their maturity stage, with an overall accuracy of 94.4 % in the calibration set and 81.3 % in the test set. For physico-chemical properties, the variable selection by iPLS and GA-PLS showed improvements in the performance of the prediction models. Portable NIR based-models allowed the prediction of the content of TSS, AA, TAA, pH and moisture content, which can be used online analysis in industry for carambola quality control.

### Acknowledgment

This study was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code [001](#); and São Paulo Research Foundation (FAPESP) (project number 2015/24351-2). J. P. Cruz-Tirado acknowledges scholarship funding from FAPESP, grant n 2020/09198-1. Ingrid Alves de Moraes acknowledges scholarship funding from FAPESP, grant n° 2019/12625-1.

### 4.5. REFERENCES

ABD RAHMAN, M.; AHMAD HAFIZ, B. GENETIC IMPROVEMENT OF FRUIT QUALITY TRAITS IN STARFRUIT (AVERRHOA CARAMBOLA) HYBRIDS. **Acta Horticulturae**, v. 1012, n. 1012, p. 259–264, nov. 2013.

ALANDER, J. T. et al. A Review of Optical Nondestructive Visual and Near-Infrared Methods for Food Quality and Safety. **International Journal of Spectroscopy**, v. 2013, p. 1–36, 24 mar. 2013.

ALI, S. H.; JAAFAR, M. Y. Effect of harvest maturity on physical and chemical characteristics of carambola ( *Averrhoa carambola* L.). **New Zealand Journal of Crop and Horticultural Science**, v. 20, n. 2, p. 133–136, abr. 1992.

AMIR, R. et al. The biodiversity of different traits of pomegranate fruit peels from a broad collection of diverse cultivars. **Scientia Horticulturae**, v. 246, p. 842–848, 27 fev. 2019.

AMODIO, M. L. et al. Potential of NIR spectroscopy for predicting internal

quality and discriminating among strawberry fruits from different production systems. **Postharvest Biology and Technology**, v. 125, p. 112–121, mar. 2017.

BARBIN, D. F. et al. Non-destructive determination of chemical composition in intact and minced pork using near-infrared hyperspectral imaging. **Food Chemistry**, v. 138, n. 2–3, p. 1162–1171, jun. 2013.

BARBIN, D. F. et al. Prediction of chicken quality attributes by near infrared spectroscopy. **Food Chemistry**, v. 168, p. 554–560, 2015.

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. **Applied Spectroscopy**, v. 43, n. 5, p. 772–777, 31 jul. 1989.

BOONYARITTHONGCHAI, P.; KANLAYANARAT, S. EFFECT OF MODIFIED ATMOSPHERE STORAGE ON THE POSTHARVEST QUALITY OF CARAMBOLA FRUITS. **Acta Horticulturae**, v. 934, n. 934, p. 221–226, jun. 2012.

BORBA, K. R. et al. Portable near Infrared Spectroscopy as a Tool for Fresh Tomato Quality Control Analysis in the Field. **Applied Sciences**, v. 11, n. 7, p. 3209, 2 abr. 2021.

CIE, (COMMISSION INTERNATIONALE DE L'ÉCLAIRAGE ). **Colorimetry**. Vienna, Switzerland: 1976Disponível em: <<https://www.scirp.org/reference/referencespapers.aspx?referenceid=2673731>>. Acesso em: 24 maio. 2021

CORREIA, R. M. et al. Portable near infrared spectroscopy applied to quality control of Brazilian coffee. **Talanta**, v. 176, p. 59–68, 1 jan. 2018.

CRUZ-TIRADO, J. P.; LUCIMAR DA SILVA MEDEIROS, M.; BARBIN, D. F. On-line monitoring of egg freshness using a portable NIR spectrometer in tandem with machine learning. **Journal of Food Engineering**, v. 306, p. 110643, 1 out. 2021.

DASGUPTA, P.; CHAKRABORTY, P.; BALA, N. N. Averrhoa Carambola: An Updated Review. **International Journal of Pharma Research & Review**, v. 2, n. 7, p. 54, 2013.

DELWICHE, S. R.; MEKWATANAKARN, W.; WANG, C. Y. Soluble Solids and Simple Sugars Measurement in Intact Mango Using Near Infrared Spectroscopy. **HortTechnology**, v. 18, n. 3, p. 410–416, jan. 2008.

DING, P.; AHMAD, S. H.; GHAZALI, H. M. Changes in selected quality characteristics of minimally processed carambola (*Averrhoa carambola* L.) when treated with ascorbic acid. **Journal of the Science of Food and Agriculture**, v. 87, n. 4, p. 702–709, mar.

2007.

DOS SANTOS COSTA, D. et al. Development of predictive models for quality and maturation stage attributes of wine grapes using vis-nir reflectance spectroscopy. **Postharvest Biology and Technology**, v. 150, n. December 2018, p. 166–178, abr. 2019.

FAGEEH, H. N. et al. Efficacy of vitamin C supplementation as an adjunct in the non-surgical management of periodontitis: a systematic review. **Systematic Reviews**, v. 10, n. 1, p. 5, 4 dez. 2021.

FAMA. **Official Portal FEDERAL AGRICULTURAL MARKETING AUTHORITY Ministry of Agriculture and Food Industry**. Disponível em: <<http://www.fama.gov.my/documents/20143/0/Belimbing1.pdf/a9fc1eb2-cbcf-a870-df14-08789f7275ff>>. Acesso em: 9 abr. 2021.

FERNÁNDEZ-NOVALES, J. et al. Assessment of quality parameters in grapes during ripening using a miniature fiber-optic near-infrared spectrometer. **International Journal of Food Sciences and Nutrition**, v. 60, n. sup7, p. 265–277, 21 jan. 2009.

FRIEDRICH, D. M. et al. **Miniature near-infrared spectrometer for point-of-use chemical analysis**. (Y. G. Soskind, C. Olson, Eds.) Photonic Instrumentation Engineering. **Anais...SPIE**, 8 mar. 2014 Disponível em: <<http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2040669>>. Acesso em: 14 nov. 2020

GALES, O. et al. Application of near <scp>infra-red</scp> spectroscopy as an instantaneous and simultaneous prediction tool for anthocyanins and sugar in whole fresh raspberry. **Journal of the Science of Food and Agriculture**, v. 101, n. 6, p. 2449–2454, 17 abr. 2021.

GOLIC, M.; WALSH, K.; LAWSON, P. Short-Wavelength Near-Infrared Spectra of Sucrose, Glucose, and Fructose with Respect to Sugar Concentration and Temperature. **Applied Spectroscopy**, v. 57, n. 2, p. 139–145, fev. 2003.

HUANG, H.-W.; CHEN, B.-Y.; WANG, C.-Y. Comparison of high pressure and high temperature short time processing on quality of carambola juice during cold storage. **Journal of Food Science and Technology**, v. 55, n. 5, p. 1716–1725, 5 maio 2018.

JHA, S. N. et al. Non-destructive prediction of sweetness of intact mango using near infrared spectroscopy. **Scientia Horticulturae**, v. 138, p. 171–175, 2012.

JIA, X. et al. Non-flavonoid phenolics from *Averrhoa carambola* fresh fruit. **Journal of Functional Foods**, v. 32, p. 419–425, 1 maio 2017.

JIE, D.; WEI, X. Review on the recent progress of non-destructive detection

technology for internal quality of watermelon. **Computers and Electronics in Agriculture**, v. 151, p. 156–164, 1 ago. 2018.

KUSUMIYATI; ARIP MUNAWAR, A.; SUHANDY, D. Fast, simultaneous and contactless assessment of intact mango fruit by means of near infrared spectroscopy. **AIMS Agriculture and Food**, v. 6, n. 1, p. 172–184, 2021.

KYRIACOU, M. C.; ROUPHAEL, Y. Towards a new definition of quality for fresh fruits and vegetables. **Scientia Horticulturae**, v. 234, p. 463–469, abr. 2018.

MADUWANTHI, S. D. T.; MARAPANA, R. A. U. J. Induced Ripening Agents and Their Effect on Fruit Quality of Banana. **International Journal of Food Science**, v. 2019, p. 1–8, 2 maio 2019.

MAGWAZA, L. S. et al. NIR Spectroscopy Applications for Internal and External Quality Analysis of Citrus Fruit — A Review. p. 425–444, 2012.

MANIWARA, P. et al. Evaluation of NIRS as non-destructive test to evaluate quality traits of purple passion fruit. **Scientia Horticulturae**, v. 257, n. July, p. 108712, nov. 2019.

MARTENS, J. H. S.; GELADI, P. Multivariate Linearity Transformation for Near-Infrared Reflectance Spectrometry. **Proc. Nordic Symp. on Applied Statistics**, p. 205–234, 1983.

MCLAREN, K. XIII-The Development of the CIE 1976 ( $L^*$   $a^*$   $b^*$ ) Uniform Colour Space and Colour-difference Formula. **Journal of the Society of Dyers and Colourists**, v. 92, n. 9, p. 338–341, 22 out. 2008.

MISHRA, P. et al. Sequential fusion of information from two portable spectrometers for improved prediction of moisture and soluble solids content in pear fruit. **Talanta**, v. 223, p. 121733, 1 fev. 2021.

MUHAMMAD ZAHIR, S. A. D. D.; YAHAYA, O. K. M. M.; OMAR, A. F. Correlating the natural color of tropical fruit juice with its  $\text{pH}$ . **Color Research & Application**, v. 46, n. 2, p. 467–476, 23 abr. 2021.

MUTHU, N. et al. Nutritional, Medicinal and Toxicological Attributes of Star-Fruits (*Averrhoa carambola* L.): A Review. **Bioinformation**, v. 12, n. 12, p. 420–424, 4 dez. 2016.

NICOLAÏ, B. M. et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. **Postharvest Biology and Technology**, v. 46, n. 2, p. 99–118, nov. 2007.

NIELSEN, S. S. **Food Analysis Laboratory Manual**. Cham: Springer

International Publishing, 2017.

NØRGAARD, L. et al. Interval Partial Least-Squares Regression ( i PLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. **Applied Spectroscopy**, v. 54, n. 3, p. 413–419, 31 mar. 2000.

O'HARE, T. J. Postharvest physiology and storage of carambola (starfruit): a review. **Postharvest Biology and Technology**, v. 2, n. 4, p. 257–267, 1 abr. 1993.

OSBORNE, B. G.; FEARN, T.; HINDLE, P. H. **Practical NIR spectroscopy with applications in food and beverage analysis**. 2<sup>a</sup> ed. Harlow: Longman Scientific and Technical, 1993.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, p. 8–36, 5 out. 2018.

PAUZIAH, M. et al. QUALITY OF STARFRUIT HARVESTED AT ADVANCED MATURITY STAGE. **Acta Horticulturae**, v. 880, n. 880, p. 231–235, nov. 2010.

PISSARD, A. et al. Non-destructive measurement of vitamin C, total polyphenol and sugar content in apples using near-infrared spectroscopy. **Journal of the Science of Food and Agriculture**, v. 93, n. 2, p. 238–244, 30 jan. 2013.

PISSARD, A. et al. Evaluation of a handheld ultra-compact NIR spectrometer for rapid and non-destructive determination of apple fruit quality. **Postharvest Biology and Technology**, v. 172, p. 111375, 1 fev. 2021.

RABATEL, G. et al. VSN: Variable sorting for normalization. **Journal of Chemometrics**, v. 34, n. 2, p. 1–16, fev. 2020.

RADY, A. M.; GUYER, D. E. Evaluation of sugar content in potatoes using NIR reflectance and wavelength selection techniques. **Postharvest Biology and Technology**, v. 103, p. 17–26, 1 maio 2015.

RAHMAN, M. S. (ED.). **Handbook of Food Preservation**. [s.l.] CRC Press, 2007.

RAMBO, M. K. D.; AMORIM, E. P.; FERREIRA, M. M. C. Potential of visible-near infrared spectroscopy combined with chemometrics for analysis of some constituents of coffee and banana residues. **Analytica Chimica Acta**, v. 775, p. 41–49, maio 2013.

SAEYS, W.; MOUAZEN, A. M.; RAMON, H. Potential for Onsite and Online Analysis of Pig Manure using Visible and Near Infrared Reflectance Spectroscopy. **Biosystems Engineering**, v. 91, n. 4, p. 393–402, ago. 2005.

SANTOS, C. S. P. et al. Non-Destructive Measurement of the Internal Quality of

Citrus Fruits Using a Portable NIR Device. **Journal of AOAC INTERNATIONAL**, v. 104, n. 1, p. 61–67, 5 mar. 2021.

SKIBSTED, E. T. S. et al. New Indicator for Optimal Preprocessing and Wavelength Selection of Near-Infrared Spectra. **Applied Spectroscopy**, v. 58, n. 3, p. 264–271, 31 mar. 2004.

SOUMYA, S. L.; NAIR, B. R. Changes in the biochemical profile of fruits of two species of *Averrhoa* during development. **International Journal of Pharmacy and Pharmaceutical Sciences**, v. 6, n. 4, p. 572–577, 2014.

SRIPAURYA, T. et al. Gros Michel banana soluble solids content evaluation and maturity classification using a developed portable 6 channel NIR device measurement. **Measurement**, v. 173, p. 108615, 1 mar. 2021.

STAN, C. CODEX STAN 82 Page 1 of 4. p. 1–4, 1981.

THEANJUMPOL, P. et al. Non-destructive identification and estimation of granulation in ‘Sai Num Pung’ tangerine fruit using near infrared spectroscopy and chemometrics. **Postharvest Biology and Technology**, v. 153, n. December 2018, p. 13–20, jul. 2019.

WALSH, K. B.; MCGLONE, V. A.; HAN, D. H. The uses of near infra-red spectroscopy in postharvest decision support: A review. **Postharvest Biology and Technology**, v. 163, n. August 2019, p. 111139, maio 2020.

WARREN, O.; SARGENT, S. A. Carambola ( *Averrhoa carambola* L.). In: **Postharvest Biology and Technology of Tropical and Subtropical Fruits**. [s.l.] Elsevier, 2011. p. 397-414e.

WEYER, L. G.; LO, S.-C. Spectra- Structure Correlations in the Near-Infrared. In: GRIFFITHS, P. R. (Ed.). **Handbook of Vibrational Spectroscopy**. Chichester, UK: John Wiley & Sons, Ltd, 2006. v. 32p. 140–141.

WHITLEY, D. A genetic algorithm tutorial. **Statistics and Computing**, v. 4, n. 2, jun. 1994.

XIAOBO, Z. et al. Variables selection methods in near-infrared spectroscopy. **Analytica Chimica Acta**, v. 667, n. 1–2, p. 14–32, 2010.

XIE, D.; LIU, D.; GUO, W. Relationship of the optical properties with soluble solids content and moisture content of strawberry during ripening. **Postharvest Biology and Technology**, v. 179, p. 111569, 1 set. 2021.

YAHAYA, O. K. M.; OMAR, A. F. **SPECTROSCOPY OF TROPICAL FRUITS Sala Mango and B10 Carambola**. [s.l.: s.n.].

YEASMIN, F. et al. The optimization of the drying process and vitamin C retention of carambola: An impact of storage and temperature. **Journal of Food Processing and Preservation**, v. 45, n. 1, p. 1–14, 6 jan. 2021.

### SUPPLEMENTARY MATERIAL

**Table 4.3** shows the results for sensitivity, specificity, classification error and accuracy for PLS-DA models based on full or reduced spectra pre-processed by Smoothing and 1<sup>st</sup> S-G derivative (individual or in combination). The variable selection, performed by iPLS and GA, was performed on the best full spectra model. iPLS showed best performance than GA, and the merit figures are show in Table 4.3.

**Table 4.3.** PLS-DA models based on spectra data to discriminate four maturation level (MS1, MS2, MS3 and MS4).

| Stages                 |     | Pre-processing                                                                                |      |      |                                                                                               |      |      |                                       |      |      |             |      |      |
|------------------------|-----|-----------------------------------------------------------------------------------------------|------|------|-----------------------------------------------------------------------------------------------|------|------|---------------------------------------|------|------|-------------|------|------|
|                        |     | smoothing S-G (order 2;<br>window 11) + 1 <sup>st</sup> S-G derivative<br>(order 2; window 7) |      |      | smoothing S-G (order 2;<br>window 11) + 1 <sup>st</sup> S-G derivative<br>(order 2; window 7) |      |      | smoothing S-G (order 2;<br>window 13) |      |      | Mean center |      |      |
| Model                  |     | iPLS                                                                                          |      |      | Full                                                                                          |      |      | Full                                  |      |      | Full        |      |      |
| Spectral<br>range (nm) |     | 905-1640                                                                                      |      |      | 905-1652                                                                                      |      |      | 913-1655                              |      |      | 913-1668    |      |      |
| LV                     |     | 14                                                                                            |      |      | 14                                                                                            |      |      | 11                                    |      |      | 7           |      |      |
| Set                    |     | Cal                                                                                           | CV   | Pred | Cal                                                                                           | CV   | Pred | Cal                                   | CV   | Pred | Cal         | CV   | Pred |
| Sensitivity            | MS1 | <b>1.00</b>                                                                                   | 0.79 | 1.00 | <b>1.00</b>                                                                                   | 0.79 | 1.00 | <b>0.90</b>                           | 0.86 | 0.73 | <b>0.93</b> | 0.83 | 0.75 |
|                        | MS2 | <b>0.96</b>                                                                                   | 0.63 | 0.56 | <b>0.92</b>                                                                                   | 0.79 | 0.44 | <b>0.81</b>                           | 0.71 | 0.90 | <b>0.76</b> | 0.64 | 0.60 |
|                        | MS3 | <b>0.85</b>                                                                                   | 0.58 | 0.57 | <b>0.85</b>                                                                                   | 0.58 | 0.57 | <b>0.76</b>                           | 0.64 | 0.71 | <b>0.63</b> | 0.47 | 0.83 |
|                        | MS4 | <b>0.89</b>                                                                                   | 0.78 | 0.92 | <b>1.00</b>                                                                                   | 0.85 | 0.92 | <b>0.96</b>                           | 0.93 | 1.00 | <b>0.93</b> | 0.89 | 0.77 |
| Specificity            | MS1 | <b>0.94</b>                                                                                   | 0.85 | 0.78 | <b>0.93</b>                                                                                   | 0.85 | 0.75 | <b>0.93</b>                           | 0.87 | 0.92 | <b>0.92</b> | 0.92 | 0.91 |
|                        | MS2 | <b>0.92</b>                                                                                   | 0.82 | 0.95 | <b>0.92</b>                                                                                   | 0.82 | 0.95 | <b>0.87</b>                           | 0.82 | 0.90 | <b>0.84</b> | 0.80 | 0.84 |
|                        | MS3 | <b>0.90</b>                                                                                   | 0.63 | 0.79 | <b>0.83</b>                                                                                   | 0.71 | 0.70 | <b>0.73</b>                           | 0.63 | 0.77 | <b>0.76</b> | 0.68 | 0.74 |
|                        | MS4 | <b>0.92</b>                                                                                   | 0.80 | 0.85 | <b>0.90</b>                                                                                   | 0.81 | 0.94 | <b>0.90</b>                           | 0.84 | 0.86 | <b>0.86</b> | 0.81 | 0.91 |
| Class. Error           | MS1 | <b>0.03</b>                                                                                   | 0.18 | 0.11 | <b>0.04</b>                                                                                   | 0.18 | 0.13 | <b>0.09</b>                           | 0.14 | 0.18 | <b>0.07</b> | 0.12 | 0.17 |
|                        | MS2 | <b>0.06</b>                                                                                   | 0.28 | 0.25 | <b>0.08</b>                                                                                   | 0.19 | 0.30 | <b>0.16</b>                           | 0.23 | 0.10 | <b>0.20</b> | 0.28 | 0.28 |

|                         |            |              |       |       |              |       |       |              |       |       |              |       |       |
|-------------------------|------------|--------------|-------|-------|--------------|-------|-------|--------------|-------|-------|--------------|-------|-------|
|                         | <b>MS3</b> | <b>0.13</b>  | 0.40  | 0.32  | <b>0.16</b>  | 0.36  | 0.37  | <b>0.26</b>  | 0.37  | 0.26  | <b>0.31</b>  | 0.43  | 0.21  |
|                         | <b>MS4</b> | <b>0.10</b>  | 0.21  | 0.11  | <b>0.05</b>  | 0.17  | 0.07  | <b>0.07</b>  | 0.11  | 0.07  | <b>0.10</b>  | 0.15  | 0.16  |
| <b>Accuracy<br/>(%)</b> | <b>MS1</b> | <b>98.21</b> | 80.25 | 97.22 | <b>98.24</b> | 80.91 | 86.86 | <b>90.05</b> | 84.17 | 88.21 | <b>89.80</b> | 85.88 | 80.48 |
|                         | <b>MS2</b> | <b>97.92</b> | 67.30 | 75.15 | <b>93.44</b> | 75.26 | 75.21 | <b>84.76</b> | 72.46 | 83.68 | <b>78.67</b> | 69.33 | 74.59 |
|                         | <b>MS3</b> | <b>94.83</b> | 67.90 | 74.03 | <b>88.16</b> | 71.44 | 72.69 | <b>84.67</b> | 65.04 | 82.77 | <b>74.27</b> | 61.52 | 73.33 |
|                         | <b>MS4</b> | <b>97.57</b> | 79.95 | 90.27 | <b>97.70</b> | 80.14 | 90.55 | <b>94.62</b> | 87.46 | 89.01 | <b>87.68</b> | 77.09 | 85.52 |

**Table 4.4** shows the performance of PLSR models built for predict physico-chemical properties of carambola fruit: % moisture content, pH, total titratable acid (TTA), total soluble solids (TSS) and ascorbic acid (AA). PLSR models were built using different pre-processing techniques, including Variable Sorting Normalization (VSN) and Standard Variate Normalization (SNV) to correct light scattering effects, S-G smoothing to correct noisy, and 1<sup>st</sup> and 2<sup>nd</sup> S-G derivative to correct baseline. The result showed models pre-processed with VSN was better performance than SNV models in all proprieties analyzed. However, models pre-processed with S-G smoothing and 1<sup>st</sup> S-G derivative showed the best performance in all cases.

**Table 4.4.** PLSR models based on spectra data to the parameters analysed in carambola fruit.

| Properties | Pre-processing                                                                              | Model | Number |           | LV   | R <sup>2</sup> <sub>C</sub> | RMSE | R <sup>2</sup> <sub>CV</sub> | RMSE | R <sup>2</sup> <sub>P</sub> | RMSE | RPD  | RER |
|------------|---------------------------------------------------------------------------------------------|-------|--------|-----------|------|-----------------------------|------|------------------------------|------|-----------------------------|------|------|-----|
|            |                                                                                             |       | of     | variables |      |                             |      |                              |      |                             |      |      |     |
| AA         | Smoothing S-G (order 2; window 15)<br>+ 1 <sup>st</sup> S-G derivative (order 2; window 13) | Full  | 208    | 9         | 0.67 | 7.72                        | 0.52 | 9.43                         | 0.53 | 8.47                        | 1.85 | 8.72 |     |
|            | Smoothing S-G (order 2; window 15)<br>+ 1 <sup>st</sup> S-G derivative (order 2; window 13) | iPLS  | 175    | 9         | 0.52 | 7.54                        | 0.52 | 9.42                         | 0.54 | 8.42                        | 1.86 | 8.77 |     |
|            | Smoothing S-G (order 2; window 15)<br>+ 1 <sup>st</sup> S-G derivative (order 2; window 13) | GA    | 145    | 12        | 0.69 | 7.44                        | 0.50 | 9.68                         | 0.54 | 8.40                        | 1.87 | 8.79 |     |
|            | Smoothing S-G (order 2; window 15)<br>+ SNV                                                 | Full  | 220    | 10        | 0.67 | 7.40                        | 0.44 | 9.88                         | 0.44 | 9.40                        | 1.67 | 7.85 |     |
|            | Smoothing S-G (order 2; window 15)<br>+ VSN                                                 | Full  | 211    | 9         | 0.66 | 7.50                        | 0.47 | 9.61                         | 0.46 | 9.00                        | 1.74 | 8.20 |     |
|            | Smoothing S-G (order 2; window 11)<br>+ 1 <sup>st</sup> S-G derivative (order 2; window 11) | Full  | 208    | 9         | 0.82 | 0.17                        | 0.71 | 0.21                         | 0.74 | 0.21                        | 1.91 | 7.63 |     |
|            | Smoothing S-G (order 2; window 11)                                                          | iPLS  | 81     | 9         | 0.80 | 0.17                        | 0.73 | 0.20                         | 0.78 | 0.20                        | 2.01 | 8.02 |     |

|     |                                                                                          |      |     |    |      |       |      |      |      |      |      |      |
|-----|------------------------------------------------------------------------------------------|------|-----|----|------|-------|------|------|------|------|------|------|
| pH  | + 1 <sup>st</sup> S-G derivative (order 2; window 11)                                    |      |     |    |      |       |      |      |      |      |      |      |
|     | Smoothing S-G (order 2; window 11) + 1 <sup>st</sup> S-G derivative (order 2; window 11) | GA   | 173 | 11 | 0.81 | 0.17  | 0.66 | 0.23 | 0.76 | 0.21 | 1.95 | 7.76 |
|     | Smoothing S-G (order 2; window 13) + SNV                                                 | Full | 209 | 10 | 0.78 | 0.18  | 0.64 | 0.23 | 0.71 | 0.24 | 1.75 | 6.98 |
|     | Smoothing S-G (order 2; window 13) + VSN                                                 | Full | 214 | 9  | 0.80 | 0.18  | 0.67 | 0.23 | 0.74 | 0.23 | 1.82 | 7.26 |
| TTA | Smoothing S-G (order 2; window 9)                                                        | Full | 222 | 9  | 0.69 | 0.06  | 0.56 | 0.08 | 0.66 | 0.07 | 1.79 | 7.55 |
|     | Smoothing S-G (order 2; window 9)                                                        | iPLS | 74  | 9  | 0.70 | 0.06  | 0.57 | 0.08 | 0.68 | 0.07 | 1.81 | 7.66 |
|     | Smoothing S-G (order 2; window 9)                                                        | GA   | 120 | 9  | 0.70 | 0.06  | 0.54 | 0.08 | 0.69 | 0.06 | 1.84 | 7.80 |
|     | Smoothing S-G (order 2; window 11) + SNV                                                 | Full | 223 | 6  | 0.53 | 0.069 | 0.43 | 0.08 | 0.51 | 0.08 | 1.41 | 5.97 |
|     | Smoothing S-G (order 2; window 11) + VSN                                                 | Full | 219 | 6  | 0.57 | 0.07  | 0.47 | 0.08 | 0.48 | 0.08 | 1.61 | 6.81 |
|     | Smoothing S-G (order 2; window 13) + 1 <sup>st</sup> S-G derivative (order 2; window 13) | Full | 208 | 7  | 0.69 | 1.00  | 0.61 | 1.12 | 0.64 | 1.12 | 1.75 | 8.33 |

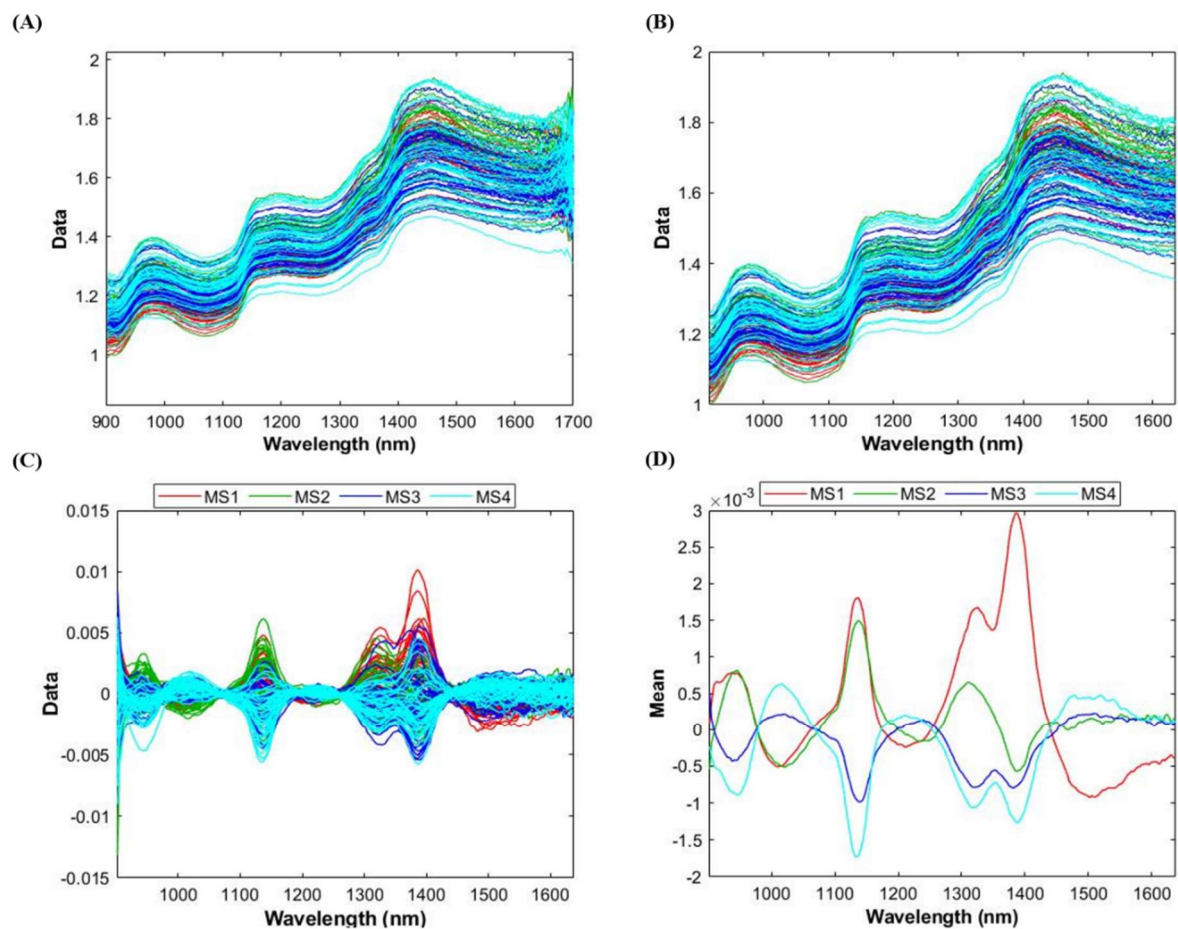
|          |                                                                                                   |      |     |   |      |      |      |      |      |      |      |       |
|----------|---------------------------------------------------------------------------------------------------|------|-----|---|------|------|------|------|------|------|------|-------|
| TSS      | Smoothing S-G (order 2; window 13)<br>+ 1 <sup>st</sup> S-G derivative (order 2; window 13)       | iPLS | 104 | 7 | 0.70 | 0.99 | 0.63 | 1.10 | 0.68 | 1.05 | 1.85 | 8.81  |
|          | Smoothing S-G (order 2; window 13)<br>+ 1 <sup>st</sup> S-G derivative (order 2; window 13)       | GA   | 18  | 7 | 0.71 | 0.97 | 0.66 | 1.05 | 0.66 | 1.08 | 1.80 | 8.61  |
|          | Smoothing S-G (order 2; window 13)<br>+ 1 <sup>st</sup> S-G derivative (order 2; window 11) + SNV | Full | 208 | 7 | 0.68 | 0.97 | 0.55 | 1.15 | 0.58 | 1.15 | 1.69 | 8.07  |
|          | Smoothing S-G (order 2; window 13)<br>+ 1 <sup>st</sup> S-G derivative (order 2; window 11) + VSN | Full | 205 | 8 | 0.68 | 1.00 | 0.56 | 1.16 | 0.57 | 1.15 | 1.69 | 8.08  |
| Moisture | Smoothing S-G (order 2; window 13)<br>+ 1 <sup>st</sup> S-G derivative (order 2; window 9)        | Full | 208 | 9 | 0.78 | 0.85 | 0.68 | 1.04 | 0.72 | 0.90 | 2.14 | 9.98  |
|          | Smoothing S-G (order 2; window 13)<br>+ 1 <sup>st</sup> S-G derivative (order 2; window 9)        | iPLS | 56  | 9 | 0.75 | 0.90 | 0.68 | 1.02 | 0.74 | 0.89 | 2.18 | 10.16 |
|          | Smoothing S-G (order 2; window 13)<br>+ 1st S-G derivative (order 2; window 9)                    | GA   | 133 | 9 | 0.74 | 0.91 | 0.63 | 1.11 | 0.74 | 0.87 | 2.23 | 10.38 |

---

|                                             |      |     |   |      |      |      |      |      |      |      |      |
|---------------------------------------------|------|-----|---|------|------|------|------|------|------|------|------|
| Smoothing S-G (order 2; window 13)<br>+ SNV | Full | 219 | 8 | 0.74 | 0.90 | 0.64 | 1.09 | 0.66 | 0.93 | 2.09 | 9.74 |
| Smoothing S-G (order 2; window 13)<br>+ VSN | Full | 214 | 8 | 0.76 | 0.89 | 0.68 | 1.02 | 0.71 | 0.90 | 2.14 | 9.98 |

---

Fig. 4.5A shows the raw spectra of carambola fruits. The spectral region between 1637 – 1700 nm was removed because showed highly noisy (Fig. 4.5B). Fig. 4.5C and Fig. 4.5D shows preprocessed spectra with smoothing and 1<sup>st</sup> derivative for full data and mean spectra, respectively.



**Figure 4.5.** A) Raw spectra, B) Spectra with region 913–1637 nm, C) full spectral data and D) mean spectra pre-processing with 1<sup>st</sup> S-G derivative + smoothing S-G spectra.

Fig. 4.6 shows loadings plots for PLS-DA model pre-processed with smoothing (order 2; window 13) + 1<sup>st</sup> S-G derivative (order 2; window 9) to classify carambola according to maturity stage: MS1, MS2, MS3 and MS4.

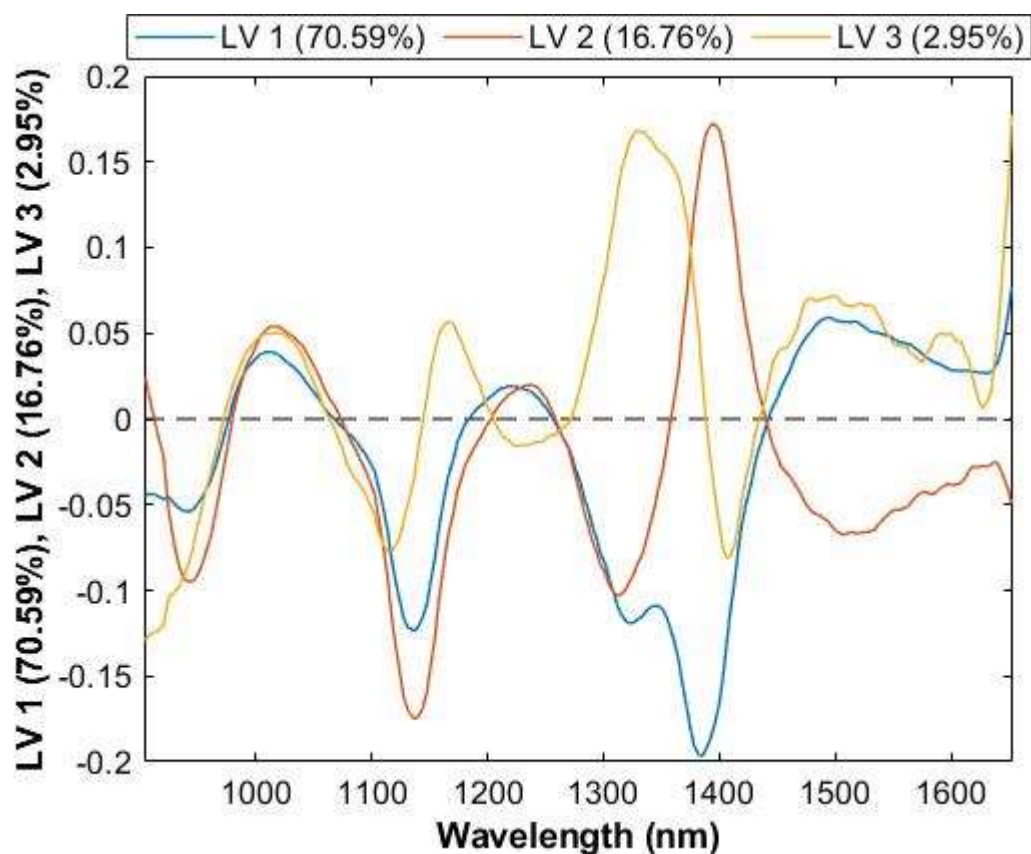
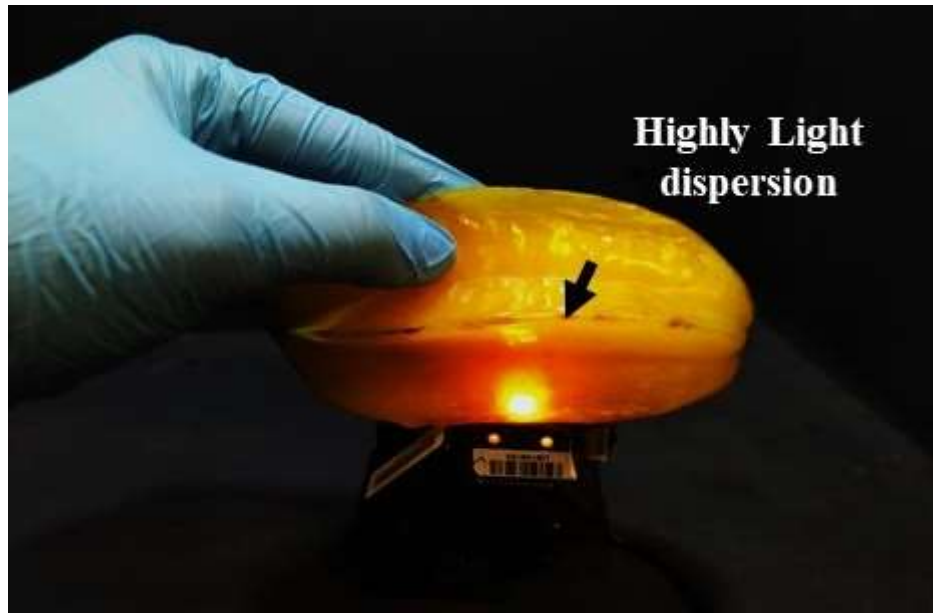


Figure 4.6. Loadings of PLS-DA model in spectral range of 905–1640 nm.

**Fig. 4.7** shows the spectral acquisition of carambola fruit. The figure shows a highly light dispersion inside carambola fruit, which had a great influence in model performance.



**Figure 4.7.** High light dispersion.

## **CAPÍTULO 5 – DISCUSSÃO GERAL**

---

### 5.1. DISCUSSÃO GERAL

O equipamento portátil de espectroscopia NIR foi utilizado para quantificar parâmetros físico-químicos e classificar a carambola quanto ao estágio de maturação. Inicialmente, como esperado, foi observado que as frutas apresentaram diferenças entre os estádios de maturação, em relação aos parâmetros analisados. Com o amadurecimento, as frutas aumentaram a concentração de sólidos solúveis, diminuíram o pH, aumentaram teor de ácido ascórbico e diminuíram o percentual de umidade. Os espectros dos quatro estádios apresentaram picos em regiões semelhantes, que podem ser atribuídas a água, ácidos orgânicos e açúcares. Entretanto, quando comparadas, as classes apresentaram picos com diferentes intensidades, o que pode estar ligado a mudança nos parâmetros analisados durante o amadurecimento. A análise de componentes principais, selecionando dois PCs (PC1 e PC2), explicou 81,46% da variabilidade das carambolas e também foram observados que a mudança das pontuações de PC1 positiva para PC1 negativas estava associada a transição dos estádios de maturação de MS1 (não maduras) para MS4(maduras). O gráfico de pesos mostrou que os picos associados a água, ácidos orgânicos e açúcares foram importantes para a discriminação das classes.

Com base nos modelos de PLS-DA, foi possível supor que os modelos discriminantes obtidos são robustos para discriminar entre todas as classes, porém com desempenho inferior para as classes MS2 e MS3. Além disso, o algoritmo iPLS selecionou as regiões espectrais de 905 a 1640 nm (eliminando 21 variáveis), que apresentaram melhor desempenho do que o modelo PLS-DA completo. A seleção aumentou a precisão (de 86,9 % para 97,2%) em MS1 e (80,1 % para 90,5 %) em MS3, resultando em um aumento de 3,5 % (de 81,3 % para 84,2 %) da precisão do conjunto de teste comparado ao modelo completo.

A regressão PLSR foi usada para prever as propriedades físico-químicas da carambola (TSS, TTA, umidade, pH e AA), utilizando os espectros NIR como preditores. A seleção das variáveis por iPLS (TSS e pH) ou GA (% umidade, TTA e AA) melhorou o modelo de PLSR. Os valores de  $R^2_P$ , RMSEP, RPD e RER revelaram que os modelos de pH e umidade foram superiores aos parâmetros TTA, TSS e AA. Os resultados revelaram modelos com alta precisão, sendo indicados para controle de qualidade. Assim, a determinação das propriedades por meio de modelos preditivos capazes de distinguir entre valores altos e baixos, deste estudo, economizaria longos períodos destinados às análises tradicionais.

## **CAPÍTULO 6 – CONCLUSÃO GERAL**

---

### 6.1. CONCLUSÃO GERAL

De maneira geral, o equipamento portátil de espectroscopia NIR em combinação com as ferramentas quimiométricas, permitiu classificar as carambolas de acordo com o grau de maturação. O PCA permitiu a visualização de agrupamentos das classes MS1 e MS4. A seleção de variáveis (GA-PLS e iPLS) aprimorou os desempenhos tanto para os modelos de PLS-DA quanto os de PLSR, o que viabiliza a aplicação de espectrômetros NIR com baixos custos em aplicações online industriais. Os modelos PLS-DA apresentaram bom desempenho para distinção entre os estádios de maturação. Os modelos PLSR, podem ser utilizados para auxiliar a classificação da qualidade de forma mais rápida e precisa, ao predizer os teores de pH, umidade, TTA, TSS e AA, independente do estágio de maturação da fruta. Portanto, essa técnica pode ser promissora na detecção e classificação de parâmetros físico-químicos e estádios de maturação de carambola.

## CAPÍTULO 7 – REFERÊNCIAS

---

### 7.1. REFERÊNCIAS

ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 2, n. 4, p. 433–459, 2010.

ABDULLAH, M. Z. et al. Discrimination and classification of fresh-cut starfruits (*Averrhoa carambola* L.) using automated machine vision system. **Journal of Food Engineering**, v. 76, n. 4, p. 506–523, out. 2006.

ABDULLAH, M. Z.; FATHINUL-SYAHIR, A. S.; MOHD-AZEMI, B. M. N. Automated inspection system for colour and shape grading of starfruit (*Averrhoa carambola* L.) using machine vision sensor. **Transactions of the Institute of Measurement and Control**, v. 27, n. 2, p. 65–87, 2 jun. 2005.

ALAMPRESE, C. et al. Identification and quantification of turkey meat adulteration in fresh, frozen-thawed and cooked minced beef by FT-NIR spectroscopy and chemometrics. **Meat Science**, v. 121, p. 175–181, 1 nov. 2016.

ALANDER, J. T. et al. A Review of Optical Nondestructive Visual and Near-Infrared Methods for Food Quality and Safety. **International Journal of Spectroscopy**, v. 2013, p. 1–36, 24 mar. 2013.

AMIR, R. et al. The biodiversity of different traits of pomegranate fruit peels from a broad collection of diverse cultivars. **Scientia Horticulturae**, v. 246, p. 842–848, 27 fev. 2019.

ARAÚJO, P. S. R. DE; MINAMI, K. Seleção de caramboleiras pelas características biométricas e físico-químicas dos frutos. **Scientia Agricola**, v. 58, n. 1, p. 91–99, mar. 2001.

BARBIN, D. F. et al. Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses. **Journal of Food Science and Technology**, v. 55, n. 7, p. 2457–2466, 16 jul. 2018.

BARKER, M.; RAYENS, W. Partial least squares for discrimination. **Journal of Chemometrics**, v. 17, n. 3, p. 166–173, 1 mar. 2003.

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. **Applied Spectroscopy**, v. 43, n. 5, p. 772–777, 31 jul. 1989.

BASANTIA, N. C.; NOLLET, L. M. L.; KAMRUZZAMAN, M. **Hyperspectral Imaging Analysis and Applications for Food Quality**. [s.l.: s.n.].

BEVILACQUA, M. et al. Classification and Class-Modelling. In: **Data Handling in Science and Technology**. [s.l.] Elsevier Ltd, 2013. v. 28p. 171–233.

BIANCOLILLO, A. et al. Chemometric Strategies for Spectroscopy-Based Food Authentication. **Applied Sciences**, v. 10, n. 18, p. 6544, 18 set. 2020.

BLAKEY, R. J.; VAN ROOYEN, Z. Non-destructive measurement of moisture content in avocado's using handheld near-infrared spectroscopy. **Proceedings of the Seventh World Avocado Congress**, n. May, p. 9–11, 2011.

BONA, E.; MARÇO, P. H.; VALDERRAMA, P. Chemometrics Applied to Food Control. In: **Food Control and Biosecurity**. [s.l.] Elsevier Inc., 2018. v. 16p. 105–133.

BORBA, K. R. et al. Portable near Infrared Spectroscopy as a Tool for Fresh Tomato Quality Control Analysis in the Field. **Applied Sciences**, v. 11, n. 7, p. 3209, 2 abr. 2021.

BOSQUE-SENDRA, J. M. et al. **Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triacylglycerol compositional data-A review***Analytica Chimica Acta*Elsevier, , 27 abr. 2012.

BRO, R.; SMILDE, A. K. Principal component analysis. **Analytical Methods**, v. 6, n. 9, p. 2812–2831, 2014.

CABRINI, D. A. et al. Analysis of the Potential Topical Anti-Inflammatory Activity of *Averrhoa carambola* L. in Mice. **Evidence-Based Complementary and Alternative Medicine**, v. 2011, p. 1–7, 2011.

CAMPS, C. et al. Using the soluble solids accumulation in tomatoes from fruit setting until harvest for the construction of a predictive model by hand-held NIR spectroscopy. **Acta Horticulturae**, v. 1119, n. June, p. 321–328, 2016.

CAYUELA, J. A.; WEILAND, C. Intact orange quality prediction with two portable NIR spectrometers. **Postharvest Biology and Technology**, v. 58, n. 2, p. 113–120, 1 nov. 2010.

CHANDRASEKARAN, I. et al. Potential of Near-Infrared (NIR) Spectroscopy and Hyperspectral Imaging for Quality and Safety Assessment of Fruits: an Overview. **Food Analytical Methods**, v. 12, n. 11, p. 2438–2458, 8 nov. 2019.

CHAU, C.-F.; CHEN, C.-H.; LIN, C.-Y. Insoluble fiber-rich fractions derived from *Averrhoa carambola*: hypoglycemic effects determined by in vitro methods. **LWT - Food Science and Technology**, v. 37, n. 3, p. 331–335, maio 2004.

CHEN, S.-W. et al. Effect of harvest season, maturity and storage temperature on storability of carambola 'Honglong' fruit. **Scientia Horticulturae**, v. 220, p. 42–51, 16 jun. 2017.

CHOI, J. H. et al. Portable, non-destructive tester integrating VIS/NIR reflectance spectroscopy for the detection of sugar content in Asian pears. **Scientia Horticulturae**, v. 220, p. 147–153, 16 jun. 2017.

CORREIA, R. M. et al. Portable near infrared spectroscopy applied to quality control of Brazilian coffee. **Talanta**, v. 176, p. 59–68, 1 jan. 2018.

COZZOLINO, D. Authentication of cereals and cereal products. In: **Advances in Food Authenticity Testing**. [s.l.] Elsevier Inc., 2016. p. 441–457.

CRUZ-TIRADO, J. P.; LUCIMAR DA SILVA MEDEIROS, M.; BARBIN, D. F. On-line monitoring of egg freshness using a portable NIR spectrometer in tandem with machine learning. **Journal of Food Engineering**, v. 306, p. 110643, 1 out. 2021.

DAFNY-YALIN, M. et al. Color, Sugars and Organic Acids Composition in Aril Juices and Peel Homogenates Prepared from Different Pomegranate Accessions. **Journal of Agricultural and Food Chemistry**, v. 58, n. 7, p. 4342–4352, 14 abr. 2010.

DE ASSIS, M. W. D. et al. PLS, iPLS, GA-PLS models for soluble solids content, pH and acidity determination in intact dovyalis fruit using near-infrared spectroscopy. **Journal of the Science of Food and Agriculture**, v. 98, n. 15, p. 5750–5755, 2018.

DONIS-GONZÁLEZ, I. R. et al. Performance Evaluation of Two Commercially Available Portable Spectrometers to Non-Invasively Determine Table Grape and Peach Quality Attributes. **Agronomy**, v. 10, n. 1, p. 148, 19 jan. 2020.

FAMA. **Official Portal FEDERAL AGRICULTURAL MARKETING AUTHORITY Ministry of Agriculture and Food Industry**. Disponível em: <<http://www.fama.gov.my/documents/20143/0/Belimbing1.pdf/a9fc1eb2-cbcf-a870-df14-08789f7275ff>>. Acesso em: 9 abr. 2021.

FAN, S. et al. Effect of spectrum measurement position variation on the robustness of NIR spectroscopy models for soluble solids content of apple. **Biosystems Engineering**, v. 143, p. 9–19, 2016.

FAO. **FAO**. Disponível em: <[http://www.fao.org/faostat/en/#rankings/countries\\_by\\_commodity](http://www.fao.org/faostat/en/#rankings/countries_by_commodity)>. Acesso em: 25 nov. 2020.

FAO. Major Tropical Fruits. p. 24, 2019.

FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. [s.l.] Editora da Unicamp, 2015.

FORINA, M.; CASALE, M.; OLIVERI, P. Application of Chemometrics to Food Chemistry. In: **Comprehensive Chemometrics**. [s.l.] Elsevier, 2009. v. 4p. 75–128.

FRIEDRICH, D. M. et al. **Miniature near-infrared spectrometer for point-of-use chemical analysis**. (Y. G. Soskind, C. Olson, Eds.) Photonic Instrumentation Engineering. **Anais...SPIE**, 8 mar. 2014 Disponível em: <<http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2040669>>. Acesso em: 14 nov. 2020

GALES, O. et al. Application of near <scp>infra-red</scp> spectroscopy as an instantaneous and simultaneous prediction tool for anthocyanins and sugar in whole fresh raspberry. **Journal of the Science of Food and Agriculture**, v. 101, n. 6, p. 2449–2454, 17 abr. 2021.

GELADI, P.; MACDOUGALL, D.; MARTENS, H. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. **Applied Spectroscopy**, v. 39, n. 3, p. 491–500, 1985.

GUEVARA, M. et al. Chemical Composition and Antioxidant Activity of the Main Fruits Consumed in the Western Coastal Region of Ecuador as a Source of Health-Promoting Compounds. **Antioxidants**, v. 8, n. 9, p. 387, 10 set. 2019.

HAZARIKA, B. et al. An overview of the unexplored underutilized fruit crops of Assam, India. **Journal of Applied and Natural Science**, v. 12, n. 3, p. 442–453, 12 set. 2020.

HIBBERT, D. B. Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations 2016). **Pure and Applied Chemistry**, v. 88, n. 4, p. 407–443, 2016.

INÁCIO, M. R. C. et al. Total anthocyanin content determination in intact açaí (*Euterpe oleracea* Mart.) and palmitero-juçara (*Euterpe edulis* Mart.) fruit using near infrared spectroscopy (NIR) and multivariate calibration. **Food Chemistry**, v. 136, n. 3–4, p. 1160–1164, 15 fev. 2013.

JAKUBÍKOVÁ, M. et al. Near-infrared spectroscopy for rapid classification of fruit spirits. **Journal of Food Science and Technology**, v. 53, n. 6, p. 2797–2803, 28 jun. 2016.

JHA, S. N. et al. Non-destructive prediction of sweetness of intact mango using near infrared spectroscopy. **Scientia Horticulturae**, v. 138, p. 171–175, 2012.

JIA, X. et al. Non-flavonoid phenolics from *Averrhoa carambola* fresh fruit. **Journal of Functional Foods**, v. 32, p. 419–425, 1 maio 2017.

JIE, D.; WEI, X. Review on the recent progress of non-destructive detection technology for internal quality of watermelon. **Computers and Electronics in Agriculture**, v. 151, p. 156–164, 1 ago. 2018.

JIMÉNEZ-CARVELO, A. M. et al. Alternative data mining/machine learning

methods for the analytical evaluation of food quality and authenticity – A review. **Food Research International**, v. 122, p. 25–39, ago. 2019.

KAWANO, S.; WATANABE, H.; IWAMOTO, M. Determination of Sugar Content in Intact Peaches by Near Infrared Spectroscopy with Fiber Optics in Interactance Mode. **Engei Gakkai zasshi**, v. 61, n. 2, p. 445–451, 1992.

KEILANY, Z. Book Reviews: Book Reviews. **Review of Social Economy**, v. 36, n. 2, p. 228–229, 1978.

KEMSLEY, E. K.; DEFERNEZ, M.; MARINI, F. **Multivariate statistics: Considerations and confidences in food authenticity problems** Food Control Elsevier Ltd, , 1 nov. 2019.

KYRIACOU, M. C.; ROUPHAEL, Y. Towards a new definition of quality for fresh fruits and vegetables. **Scientia Horticulturae**, v. 234, p. 463–469, abr. 2018.

LAMMERTYN, J. et al. NON-DESTRUCTIVE MEASUREMENT OF ACIDITY, SOLUBLE SOLIDS, AND FIRMNESS OF JONAGOLD APPLES USING NIR-SPECTROSCOPY. **Transactions of the ASAE**, v. 41, n. 4, p. 1089–1094, 1998.

LAVINE, B. A user-friendly guide to multivariate calibration and classification , Tomas Naes, Tomas Isakson, Tom Fearn and Tony Davies, NIR Publications, Chichester, 2002, ISBN 0-9528666-2-5, £45.00. **Journal of Chemometrics**, v. 17, n. 10, p. 571–572, out. 2003.

LI, B.; LECOURT, J.; BISHOP, G. Advances in Non-Destructive Early Assessment of Fruit Ripeness towards Defining Optimal Time of Harvest and Yield Prediction—A Review. **Plants**, v. 7, n. 1, p. 3, 10 jan. 2018.

LIM, T. K. **Edible Medicinal and Non-Medicinal Plants**. Dordrecht: Springer Netherlands, 2012. v. 1

LIU, Y. et al. **Application of genetic algorithms in fundamental study of nondestructive measurement of internal quality with FT-NIR spectroscopy**. Optical Sensors and Sensing Systems for Natural Resources and Food Safety and Quality. **Anais...SPIE**, 9 nov. 2005Disponível em: <<https://www.spiedigitallibrary.org/conference-proceedings-of-spie/5996/599618/Application-of-genetic-algorithms-in-fundamental-study-of-nondestructive-measurement/10.1117/12.630417.full>>. Acesso em: 2 jun. 2021

LIU, Y. et al. Nondestructive measurement of internal quality of Nanfeng mandarin fruit by charge coupled device near infrared spectroscopy. **Computers and Electronics in Agriculture**, v. 71, n. SUPPL. 1, p. S10–S14, 1 abr. 2010.

LIU, Y.; CHEN, X.; OUYANG, A. Nondestructive determination of pear internal

quality indices by visible and near-infrared spectrometry. **LWT - Food Science and Technology**, v. 41, n. 9, p. 1720–1725, 1 nov. 2008.

LOHUMI, S. et al. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. **Trends in Food Science & Technology**, v. 46, n. 1, p. 85–98, 1 nov. 2015.

LU et al. Measurement of optical properties of fruits and vegetables: A review. **Postharvest Biology and Technology**, v. 159, p. 111003, 1 jan. 2020a.

LU, J. et al. Nondestructive determination of soluble solids and firmness in mix-cultivar melon using near-infrared CCD spectroscopy. **Journal of Innovative Optical Health Sciences**, v. 8, n. 6, 1 nov. 2015.

LU, R. PREDICTING FIRMNESS AND SUGAR CONTENT OF SWEET CHERRIES USING NEAR-INFRARED DIFFUSE REFLECTANCE SPECTROSCOPY. **Transactions of the ASAE**, v. 44, n. 5, 2001.

LU, Y. et al. Biologia e Tecnologia pós-harvest Tecnologia de imagem hiperespectral para avaliação de qualidade e segurança de produtos hortícolas: Uma revisão e celebração do progresso dos últimos 20 anos Destaques Abstrata. 2020b.

MADUWANTHI, S. D. T.; MARAPANA, R. A. U. J. Induced Ripening Agents and Their Effect on Fruit Quality of Banana. **International Journal of Food Science**, v. 2019, p. 1–8, 2 maio 2019.

MAGWAZA, L. S. et al. NIR Spectroscopy Applications for Internal and External Quality Analysis of Citrus Fruit — A Review. p. 425–444, 2012.

MAHANTI, N. K.; CHAKRABORTY, S. K. Application of chemometrics to identify artificial ripening in sapota (Manilkara Zapota) using visible near infrared absorbance spectra. **Computers and Electronics in Agriculture**, v. 175, n. December 2019, p. 105539, 2020.

MANIWARA, P. et al. The use of visible and near infrared spectroscopy for evaluating passion fruit postharvest quality. **Journal of Food Engineering**, v. 143, p. 33–43, 1 dez. 2014.

MAPA, M. DA A. P. E A. **Fruticultura brasileira ganha mais um aliado no combate à mosca das frutas — Português (Brasil)**. Disponível em: <<https://www.gov.br/agricultura/pt-br/assuntos/noticias/fruticultura-brasileira-ganha-mais-um-aliado-no-combate-a-mosca-das-frutas>>. Acesso em: 9 jun. 2021.

MARQUETTI, I. et al. Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee.

**Computers and Electronics in Agriculture**, v. 121, p. 313–319, 1 fev. 2016.

MEHMOOD, T. et al. **A review of variable selection methods in Partial Least Squares Regression Chemometrics and Intelligent Laboratory Systems** Elsevier, , 15 ago. 2012.

MEINHART, A. D. et al. Chlorogenic and caffeic acids in 64 fruits consumed in Brazil. **Food Chemistry**, v. 286, p. 51–63, 15 jul. 2019.

MISHRA, P. et al. New data preprocessing trends based on ensemble of multiple preprocessing techniques. **TrAC - Trends in Analytical Chemistry**, v. 132, p. 116045, 2020.

MISHRA, P. et al. Sequential fusion of information from two portable spectrometers for improved prediction of moisture and soluble solids content in pear fruit. **Talanta**, v. 223, p. 121733, 1 fev. 2021a.

MISHRA, P. et al. Chemometric pre-processing can negatively affect the performance of near-infrared spectroscopy models for fruit quality prediction. **Talanta**, v. 229, p. 122303, 1 jul. 2021b.

MUHAMMAD ZAHIR, S. A. D. D.; YAHAYA, O. K. M. M.; OMAR, A. F. Correlating the natural color of tropical fruit juice with its <sc>pH</sc>. **Color Research & Application**, v. 46, n. 2, p. 467–476, 23 abr. 2021.

MUSTAFA, M. A. et al. Enhancing the antioxidant content of carambola (*Averrhoa carambola*) during cold storage and methyl jasmonate treatments. **Postharvest Biology and Technology**, v. 118, p. 79–86, 2016.

MUTHU, N. et al. Nutritional, Medicinal and Toxicological Attributes of Star-Fruits (*Averrhoa carambola* L.): A Review. **Bioinformation**, v. 12, n. 12, p. 420–424, 4 dez. 2016.

NICOLAÏ, B. M. et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. **Postharvest Biology and Technology**, v. 46, n. 2, p. 99–118, nov. 2007.

NØRGAARD, L. et al. Interval Partial Least-Squares Regression ( i PLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. **Applied Spectroscopy**, v. 54, n. 3, p. 413–419, 31 mar. 2000.

O'HARE, T. J. Postharvest physiology and storage of carambola (starfruit): a review. **Postharvest Biology and Technology**, v. 2, n. 4, p. 257–267, 1 abr. 1993.

OMAR, A. F.; ATAN, H.; MATJAFRI, M. Z. Visible Spectral Linearisation, Gradient Shift and Normalisation in Quantifying Carambola Acidity. **Food Biophysics**, v. 7, n. 4, p. 289–295, 31 dez. 2012.

OMAR, A. F.; MATJAFRI, M. Z. Specialized optical fiber sensor for nondestructive intrinsic quality measurement of Averrhoa Carambola. **Photonic Sensors**, v. 3, n. 3, p. 272–282, 14 set. 2013.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, p. 8–36, 5 out. 2018.

PATHMANABAN, P.; GNANAVEL, B. K.; ANANDAN, S. S. Recent application of imaging techniques for fruit quality assessment. **Trends in Food Science & Technology**, v. 94, p. 32–42, 1 dez. 2019.

PAUZIAH, M. et al. QUALITY OF STARFRUIT HARVESTED AT ADVANCED MATURITY STAGE. **Acta Horticulturae**, v. 880, n. 880, p. 231–235, nov. 2010.

PISSARD, A. et al. Non-destructive measurement of vitamin C, total polyphenol and sugar content in apples using near-infrared spectroscopy. **Journal of the Science of Food and Agriculture**, v. 93, n. 2, p. 238–244, 30 jan. 2013.

PISSARD, A. et al. Evaluation of a handheld ultra-compact NIR spectrometer for rapid and non-destructive determination of apple fruit quality. **Postharvest Biology and Technology**, v. 172, p. 111375, 1 fev. 2021.

PU, Y. Y. et al. Ripeness Classification of Bananito Fruit (*Musa acuminata*, AA): a Comparison Study of Visible Spectroscopy and Hyperspectral Imaging. **Food Analytical Methods**, v. 12, n. 8, p. 1693–1704, 2019.

QI, S. et al. Establishment of a comprehensive indicator to nondestructively analyze watermelon quality at different ripening stages. **Journal of Innovative Optical Health Sciences**, v. 7, n. 4, 2014.

RABATEL, G. et al. VSN: Variable sorting for normalization. **Journal of Chemometrics**, v. 34, n. 2, p. 1–16, fev. 2020.

RAMBO, M. K. D.; AMORIM, E. P.; FERREIRA, M. M. C. Potential of visible-near infrared spectroscopy combined with chemometrics for analysis of some constituents of coffee and banana residues. **Analytica Chimica Acta**, v. 775, p. 41–49, maio 2013.

ROGER, J.-M. et al. **Pre-processing Methods**. Second Edi ed. [s.l.] Elsevier, 2020. v. 3

SAEYS, W. et al. **Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review** *Postharvest Biology and Technology* Elsevier B.V., 1 dez. 2019.

SAEYS, W.; MOUAZEN, A. M.; RAMON, H. Potential for Onsite and Online

Analysis of Pig Manure using Visible and Near Infrared Reflectance Spectroscopy. **Biosystems Engineering**, v. 91, n. 4, p. 393–402, ago. 2005.

SAGHIR, A. M. et al. Star fruit (*Averrhoa carambola* L.): From traditional uses to pharmacological activities [Fruta de la estrella (*Averrhoa carambola* L.): Desde los usos tradicionales a las actividades farmacológicas]. 2013.

SANTOS, C. S. P. et al. Non-Destructive Measurement of the Internal Quality of Citrus Fruits Using a Portable NIR Device. **Journal of AOAC INTERNATIONAL**, v. 104, n. 1, p. 61–67, 5 mar. 2021.

SARANWONG, S.; SORNSRIVICHAI, J.; KAWANO, S. Prediction of ripe-stage eating quality of mango fruit from its harvest quality measured nondestructively by near infrared spectroscopy. **Postharvest Biology and Technology**, v. 31, n. 2, p. 137–145, 2004.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation. **Anal. Chem**, v. 36, n. 8, p. 1627–1639, 1964.

SLAUGHTER, D. C.; THOMPSON, J. F.; TAN, E. S. Nondestructive determination of total and soluble solids in fresh prune using near infrared spectroscopy. **Postharvest Biology and Technology**, v. 28, n. 3, p. 437–444, 2003.

SRIPAURYA, T. et al. Gros Michel banana soluble solids content evaluation and maturity classification using a developed portable 6 channel NIR device measurement. **Measurement**, v. 173, p. 108615, 1 mar. 2021.

TEÓFILO, R. F.; MARTINS, J. P. A.; FERREIRA, M. M. C. Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. **Journal of Chemometrics**, v. 23, n. 1, p. 32–48, jan. 2009.

THEKKUDAN, D. F.; RUTAN, S. C. **Denoising and Signal-to-Noise Ratio Enhancement: Classical Filtering**. Second Edi ed. [s.l.] Elsevier, 2009. v. 3

TRYGG, J.; GABRIELSSON, J.; LUNDSTEDT, T. Background Estimation, Denoising, and Preprocessing. **Comprehensive Chemometrics**, v. 2, p. 1–8, 2009.

WANI, S. M. et al. Minimal Processing of Tropical and Subtropical Fruits, Vegetables, Nuts, and Seeds. In: **Food Engineering Series**. [s.l.] Springer, 2017. p. 469–512.

WORKMAN, J.; SHENK, J. Understanding and Using the Near-Infrared Spectrum as an Analytical Method. In: [s.l.: s.n.]. p. 1–10.

XIAOBO, Z. et al. Variables selection methods in near-infrared spectroscopy. **Analytica Chimica Acta**, v. 667, n. 1–2, p. 14–32, 2010.

XU, H. et al. Variable selection in visible and near-infrared spectra: Application to on-line determination of sugar content in pears. **Journal of Food Engineering**, v. 109, n. 1, p.

142–147, 1 mar. 2012.

YAHAYA, O. K. M.; OMAR, A. F. **SPECTROSCOPY OF TROPICAL FRUITS Sala Mango and B10 Carambola**. [s.l: s.n.].

YANG, D. et al. Two tetrahydroisoquinoline alkaloids from the fruit of *Averrhoa carambola*. **Phytochemistry Letters**, v. 7, n. 1, p. 217–220, 1 fev. 2014.

YANG, J.-B. et al. **Feature selection via sensitivity analysis of MLP probabilistic outputs**. 2008 IEEE International Conference on Systems, Man and Cybernetics. **Anais...IEEE**, out. 2008Disponível em: <<http://ieeexplore.ieee.org/document/4811372/>>. Acesso em: 2 jun. 2021

ZAINUDIN, M. A. M. et al. Variation of bioactive compounds and antioxidant activity of carambola (*Averrhoa carambola* L.) fruit at different ripening stages. **Scientia Horticulturae**, v. 172, p. 325–331, 9 jun. 2014.