



**UNIVERSIDADE FEDERAL DE PERNAMBUCO**

**CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA – CCEN**

**DEPARTAMENTO DE QUÍMICA FUNDAMENTAL – DQF**

**EMANUEL JOSÉ NASCIMENTO MARQUES**

**DESENVOLVIMENTO DE MÉTODOS ANALÍTICOS NÃO  
DESTRUTIVOS PARA AVALIAÇÃO DA QUALIDADE DE  
FRUTAS UTILIZANDO ESPECTROSCOPIA E IMAGEM  
HIPERESPECTRAL NA REGIÃO DO INFRAVERMELHO  
PRÓXIMO**

**Recife**

**2017**

EMANUEL JOSÉ NASCIMENTO MARQUES

DESENVOLVIMENTO DE MÉTODOS ANALÍTICOS NÃO DESTRUTIVOS PARA  
AVALIAÇÃO DA QUALIDADE DE FRUTAS UTILIZANDO ESPECTROSCOPIA  
E IMAGEM HIPERESPECTRAL NA REGIÃO DO INFRAVERMELHO PRÓXIMO

Tese de doutorado apresentada ao Programa de Pós-graduação em Química da Universidade Federal de Pernambuco, como parte dos requisitos para a obtenção do título de Doutor em Química.

**Orientadora:** Profa. Dra. Maria Fernanda Pimentel Avelar

**Coorientador:** Dr. Sérgio Tonetto de Freitas

Recife

2017

Catálogo na fonte  
Bibliotecário Jefferson Luiz Alves Nazareno CRB4-1758

M357d Marques, Emanuel José Nascimento.  
Desenvolvimento de métodos analíticos não destrutivos para avaliação da qualidade de frutas utilizando espectroscopia e imagem hiperespectral na região do infravermelho próximo / Emanuel José Nascimento Marques . – 2017.  
184 f., fig.: tab.

Orientadora: Maria Fernanda Pimentel Avelar.  
Tese (doutorado) – Universidade Federal de Pernambuco. CCEN. Química. Recife, 2017.  
Inclui referências e apêndice.

1. Química analítica. 2. Espectroscopia. 3. Frutas I. Avelar, Maria Fernanda Pimentel. (Orientadora) II. Título.

543                    CDD (22. ed.)                    UFPE-FQ 2018-35

**EMANUEL JOSÉ NASCIMENTO MARQUES**

***DESENVOLVIMENTO DE MÉTODOS ANALÍTICOS NÃO  
DESTRUTIVOS PARA AVALIAÇÃO DA QUALIDADE DE FRUTAS  
UTILIZANDO ESPECTROSCOPIA E IMAGEM HIPERESPECTRAL  
NA REGIÃO DO INFRAVERMELHO PRÓXIMO***

Tese de doutorado apresentada ao Programa de Pós-graduação em Química da Universidade Federal de Pernambuco, como parte dos requisitos para a obtenção do título de Doutor em Química.

**Aprovado em: 30/08/2017.**

**BANCA EXAMINADORA**

---

**Profa. Maria Fernanda Pimentel Avelar (Orientadora)**

Departamento de Engenharia Química  
Universidade Federal de Pernambuco

---

**Profa. Fernanda Araújo Honorato**

Departamento de Engenharia Química  
Universidade Federal de Pernambuco

---

**Profa. Claudete Fernandes Pereira**

Departamento de Química Fundamental  
Universidade Federal de Pernambuco

---

**Prof. Everaldo Paulo de Medeiros**

Empresa Brasileira de Pesquisa Agropecuária

---

**Prof. Ricardo Oliveira da Silva**

Departamento de Química Fundamental  
Universidade Federal de Pernambuco

Dedico este trabalho à minha esposa Aline, por sempre estar ao meu lado, pelo seu incentivo, confiança e apoio incondicionais.

## AGRADECIMENTOS

Primeiramente, agradeço à minha orientadora, Profa. Dra. Maria Fernanda Pimentel, que me concedeu a oportunidade de desenvolver o trabalho em seu grupo de pesquisa, pela excelente orientação, pela compreensão e paciência ao longo do período de desenvolvimento do doutorado e, principalmente, pelo seu exemplo de profissionalismo e ética. Além, é claro, de seus valiosos conselhos.

Ao meu coorientador, Dr. Sérgio Tonetto de Freitas, pela orientação, entusiasmo e discussões, que sempre foram muito construtivas para o desenvolvimento do trabalho.

À minha esposa Aline, que sempre me apoiou e contribuiu com discussões e sugestões durante o desenvolvimento deste trabalho.

Ao CNPq pela bolsa de doutorado concedida.

À CAPES pela bolsa concedida na modalidade “Doutorado Sanduíche”.

Ao Prof. Dr. Celio Pasquini, pelas contribuições significativas e pela oportunidade e apoio oferecidos para a realização do “Doutorado Sanduíche”.

Ao Prof. Dr. Ivo Milton Raimundo Jr., pelo apoio oferecido durante o período em que realizei o “Doutorado Sanduíche”.

Ao meu supervisor no exterior, Dr. Vincent Baeten, pela oportunidade, orientação e todo o suporte oferecido durante o período em que estive na Bélgica e desenvolvi o trabalho no *Centre Walloon de Recherches Agronomiques (CRA-W)*.

Ao pessoal do Laboratório de Combustíveis (LAC), que contribuíram direta ou indiretamente para o desenvolvimento do trabalho. Em especial: Carol, Neirivaldo, Leandro, Francielson, Alianda, Eduardo, Vítor, Rafaela, Marcela, Jéssica, Marcondes, Carol Maia e Fernanda V. Cruz, pela amizade, ajuda e valiosa troca de conhecimentos.

Aos pesquisadores e funcionários do CRA-W, que contribuíram direta ou indiretamente para a realização deste trabalho. Em especial: Dr. Pierre Dardenne, Dr. Juan Antonio Fernandez Pierna, Olivier Minet, Bernard Lecler, Damien Vincke, Philippe Vermeulen, Audrey Pissard e Dra. Ouissan Abbas.

À Dra. Cristina Malegori, pela excelente parceria nos estudos relacionados à espectroscopia NIR, imagens hiperespectrais e qualidade de frutas.

Ao Prof. Dr. José Manuel Amigo Rubio, pelo excelente suporte e ajuda em assuntos relacionados à Quimiometria, saquê e cachaças mineiras.

Aos membros da banca examinadora, Profa. Dra. Fernanda Araújo Honorato, Profa. Dra. Claudete Fernandes Pereira, Prof. Dr. Ricardo Oliveira da Silva e Dr. Everaldo

Paulo de Medeiros, que contribuíram muito para a versão final desta tese, com suas correções e sugestões.

Aos funcionários da Secretaria de Pós-Graduação, Patrícia e Maurilio, que me ajudaram muito durante todo o período de desenvolvimento do doutorado.

Aos meus familiares, pelo incentivo, confiança e apoio.

Ao Filipe, pelas aulas de inglês emergenciais antes da viagem à Bélgica.

Às babás Laile e Daniela, que cuidaram do trio de bebês sapecas (Matheusinho, Julinha e Gui-Gui) enquanto eu escrevia esta tese.

A todos aqueles que participaram de alguma forma para minha formação acadêmica e pessoal, agradeço.

## RESUMO

As frutas e seus derivados representam uma importante classe de produtos agroindustriais. Atualmente, o Brasil é o terceiro maior produtor de frutas, sendo responsável por cerca de 6% da produção mundial. Neste cenário, a mangicultura tem se destacado e nos últimos anos a região Nordeste tem contribuído de forma significativa para a produção de frutas destinadas à exportação e à comercialização no mercado interno. Entretanto, o amadurecimento não uniforme de frutas de um mesmo lote é um dos principais problemas existentes na cadeia produtiva de manga e tem impacto direto no consumo e na comercialização do produto. De maneira similar ao que ocorre no Brasil, a produção de frutas também representa uma atividade agrícola importante para alguns países europeus, como é o caso da produção de maçãs na Bélgica, onde o principal desafio para os produtores é fornecer frutas de alta qualidade para um mercado consumidor altamente exigente. Neste trabalho, avaliou-se o uso da espectroscopia na região do infravermelho próximo (NIR, do inglês *Near Infrared*), por meio de um espectrômetro portátil, para a determinação de parâmetros de qualidade em mangas e maçãs, de forma rápida e não destrutiva. Foram construídos modelos de calibração empregando-se diferentes técnicas de calibração multivariada, pré-processamento espectral e seleção de variáveis. De modo geral, verificou-se que o espectrômetro NIR portátil pode ser utilizado para a determinação da maioria dos parâmetros avaliados neste estudo em mangas (sólidos solúveis, matéria seca, acidez titulável e firmeza de polpa) e maçãs (sólidos solúveis, acidez titulável, firmeza de polpa, índice de amido e índice Streif), com exatidão aceitável para atestar a qualidade dos frutos. Em uma segunda etapa, realizou-se a transferência de calibração entre espectrômetros NIR de bancada e portátil, visando à determinação de parâmetros de qualidade de maçãs. Foi utilizada a técnica de Padronização Direta, a qual foi empregada com êxito para realizar o procedimento de transferência de calibração. Adicionalmente, foi realizado um estudo preliminar para avaliar a possibilidade de detecção precoce de danos mecânicos em mangas utilizando imagens hiperespectrais na região do NIR, obtidas por meio de uma câmera hiperespectral.

**Palavras-chave:** Frutas. Espectroscopia NIR. Imagem Hiperespectral. Calibração Multivariada. Transferência de Calibração. PLS. SPA-MLR. LS-SVM.

## ABSTRACT

Fruits and their derivatives represent an important class of agro-industrial products. Currently, Brazil is the third largest fruit producer, accounting for about 6% of the world production. In this scenario, the mango production has stood out and in recent years, the Northeast region has contributed significantly to the production of fruits for overseas and domestic market. However, non-uniform maturity of fruits from the same batch is one of the main problems in the mango production chain, which has a direct impact on the consumption and commercialization of the product. Similar to what happens in Brazil, fruit production is also an important agricultural activity in some European countries, as the case of apple production in Belgium, where the main challenge for the growers is to provide high quality fruit to a consumer market that demands high quality products. In this study, the use of Near Infrared (NIR) Spectroscopy was evaluated by means of a portable spectrometer for rapid and non-destructive determination of quality parameters in mangoes and apples. Calibration models were constructed using different multivariate calibration techniques, spectral preprocessing and variable selection. According to the results, the portable NIR spectrometer can be used to determine the majority of quality parameters evaluated in mangoes (solids, soluble solids, dry matter, titratable acidity and pulp firmness) and apples (soluble solids, titratable acidity, pulp firmness, starch index and Streif index) with acceptable accuracy to attest the quality of the fruit. In a second step, the calibration transfer was performed between the benchtop and portable NIR spectrometers, aiming to determine quality parameters in apples. The Direct Standardization technique was successfully used to carry out the calibration transfer procedure. In addition, a preliminary study was carried out to evaluate the possibility of early detection of mechanical damage in mangoes using hyperspectral images in the NIR region, obtained with a hyperspectral camera.

**Keywords:** Fruit. NIR. Spectroscopy. Hyperspectral Image. Multivariate Calibration. Calibration Transfer. PLS. SPA-MLR. LS-SVM.

## LISTA DE FIGURAS

- Figura 1 – Princípio de funcionamento de um filtro linear variável (LVF) associado a um arranjo de fotodiodos. Neste sistema, a radiação incidente é dispersa em suas frequências individuais em função da posição ao longo do comprimento do filtro. Em seguida, a radiação proveniente do filtro é direcionada para o conjunto de sensores (adaptado de PEDERSON et. al., 2014)..... 30
- Figura 2 – Espectrômetro NIR portátil modelo MicroNIR 1700 (VIAVI, CA, EUA): exemplo de equipamento portátil dispersivo, que emprega um filtro linear variável (LVF) associado a um arranjo de fotodiodos para efetuar a decomposição do feixe luminoso e viabilizar a seleção dos comprimentos de onda específicos. .... 30
- Figura 3 – Principais modos de medida utilizados em espectroscopia NIR: a) transmitância; b) transflectância; c) reflectância difusa; d) interactância; e e) transmitância através de meio dispersivo (onde, “F” representa a fonte de radiação; e “D” representa o detector) (adaptado de PASQUINI, 2003). .... 32
- Figura 4 – (A) Cubo hiperespectral, onde os eixos  $x$  e  $y$  representam a informação espacial, enquanto o eixo  $\lambda$  representa a informação espectral; (B) espectros de pixels específicos selecionados no plano da imagem ( $xy$ ); e (C) imagem da amostra em um comprimento de onda específico (adaptado de AMIGO et al., 2013)..... 35
- Figura 5 – Desdobramento do cubo hiperespectral ao longo de cada comprimento de onda. Por meio deste procedimento, cada pixel é acomodado em uma linha da nova matriz, que terá dimensões iguais a  $x.y$  pixels (número de linhas) e  $\lambda$  variáveis (número de colunas) (adaptado de AMIGO et al., 2013)..... 35
- Figura 6 – Exemplo da complexidade de um espectro NIR: o espectro da amostra registrado pelo equipamento (linha azul de maior espessura) é resultado da sobreposição de inúmeros picos individuais correspondentes aos componentes da amostra. Neste caso, o material analisado é composto majoritariamente por água, açúcares, celulose e proteína (adaptado de <https://www.wur.nl/en/show/NIR-spectroscopy-1.htm>). .... 38
- Figura 7 – Organização dos dados em calibração multivariada: a matriz  $X$  contém os dados espectrais registrados, de forma que suas linhas ( $m$ ) representam as amostras e as colunas ( $n$ ) as variáveis; enquanto o vetor  $y$  (com  $m$  linhas) contém os valores do parâmetro de interesse..... 44
- Figura 8 – Superfície de resposta utilizada para a otimização simultânea dos parâmetros  $\gamma$  e  $\sigma^2$  em função dos valores de RMSECV durante a construção do modelo de calibração por LS-SVM (CODIGILL & DARDENNE, 2004)..... 57
- Figura 9 – Procedimentos realizados na primeira iteração do algoritmo, considerando um exemplo didático contendo três variáveis: (A) escolha do primeiro vetor ( $x_1$ ); (B) determinação do subespaço ortogonal ao vetor escolhido; (C) projeção dos vetores remanescentes ( $x_2$  e  $x_3$ ) neste subespaço; e (D) seleção do vetor (que representa uma variável) associado à maior projeção (adaptado de GALVÃO & ARAÚJO, 2015)..... 60
- Figura 10 – Escala visual utilizada para classificação dos estádios de maturação de mangas da variedade “Tommy Atkins” (SANTOS et al., 2008)..... 72

Figura 11 – (A) Mangas da variedade “Tommy Atkins” sendo avaliadas com o espectrômetro MicroNIR 1700 (VIAVI, Santa Rosa, CA, EUA); e (B) detalhe mostrando as dimensões reduzidas deste equipamento. ....	73
Figura 12 – Aquisição de espectros de uma fatia inteira de manga (~ 31 mm). As medidas foram realizadas posicionando-se o espectrômetro diretamente sobre a casca de uma da fatia e um suporte de polietileno do lado oposto .....	74
Figura 13 – Pontos (P1 a P6) distribuídos ao longo da região axial de faces opostas do fruto, utilizados para aquisição dos espectros na região do NIR.....	75
Figura 14 – Etapas de preparo das amostras para as análises de referência: (A) fruto inteiro; (B) fruto com parte da epiderme removida, utilizado para análise de FP; (C) corte e extração de uma das faces do fruto; (D) separação de amostra de uma das faces do fruto, com dimensões aproximadas de 5 x 2 x 1 cm (comprimento x largura x profundidade); (E) amostras extraídas das duas faces opostas do fruto; (F) amostra representativa do fruto, composta por pequenos pedaços provenientes das duas faces; (G) fração da amostra utilizada para determinação de MS; (H) fração da amostra utilizada para prensagem; e (I) fração líquida resultante da prensagem, utilizada para determinação de SS e AT. ....	76
Figura 15 – Determinação de SS empregando-se um refratômetro digital (Atago, Tóquio, Japão). ....	77
Figura 16 – Determinação de AT empregando-se uma bureta digital (modelo Digitrate Pro, Jencons, Franklin, KY, EUA).....	78
Figura 17 – Determinação de FP empregando-se um texturômetro digital (modelo TA.XT Plus, Extralab, Jarinu, SP, Brasil).....	79
Figura 18 – Sistema de aquisição de imagens SisuCHEMA SWIR (Specim Spectral Imaging Ltd., Oulu, Finlândia).....	81
Figura 19 – (A) Amostra de manga com dimensões aproximadas de 50 x 30 x 20 mm (comprimento x largura x profundidade), posicionada sobre um suporte de poliestireno (PS); (B) determinação da espessura da amostra por meio de um paquímetro digital (King-Tools, China) antes do registro da imagem; (C) amostra de manga com espessura de aproximadamente 21,00 mm; e (D) amostra de manga com espessura de aproximadamente 0,60 mm.....	82
Figura 20 – Frutos em diferentes estádios de maturação utilizados no experimento para detecção precoce de danos mecânicos. ....	83
Figura 21 – (A) Peça cilíndrica com dimensões de 9,5 x 3,0 cm (comprimento x diâmetro) e massa igual a 130 g utilizada para induzir os danos mecânicos nos frutos; e (B) cano de PVC com dimensões de 60 x 4 cm (comprimento x diâmetro) utilizado para direcionar a peça cilíndrica durante a queda para o local determinado para o impacto (centro da amostra). ....	84
Figura 22 – Espectros pré-processados por SNV, de fatias de manga com diferentes espessuras. As regiões espectrais próximas a 1070 e 1205 nm foram destacadas nos quadros A e B para melhorar a visualização dos espectros. ....	86
Figura 23 – (A) Gráfico de escores em PC1 e PC2 mostrando a separação de grupos de fatias de manga em função de suas espessuras, correspondendo aos números 1, 2, 3, 4 e 5; e (B) gráfico de <i>loadings</i> para PC1 e PC2 mostrando as principais regiões espectrais relacionadas à distribuição dos escores nestas componentes. ....	88

- Figura 24 – Espectros de absorvância (sem pré-processamento) provenientes de 250 mangas (“Tommy Atkins”) intactas, na faixa entre 950 e 1650 nm. .... 89
- Figura 25 – Valores obtidos pelo método de referência *versus* os valores previstos pelo modelo de calibração nas etapas de calibração e validação externa para os parâmetros: (A) sólidos solúveis (SS), (B) matéria seca (MS), (C) acidez titulável (AT) e (D) firmeza de polpa (FP). A linha sólida representa a bissetriz. Os resultados das etapas de calibração e validação externa são representados por círculos preenchidos e círculos vazios, respectivamente. ... 93
- Figura 26 – Valores previstos pelos modelos de calibração para: (A) sólidos solúveis (SS), (B) matéria seca (MS), (C) acidez titulável (AT) e (D) firmeza de polpa (FP). Estes parâmetros foram determinados em mangas “Tommy Atkins” ao longo do período de armazenamento, que correspondeu a 14 dias (360 h) em temperatura de 24°C ( $\pm 1^\circ\text{C}$ ). Apenas três mangas representativas do conjunto são mostradas na figura, a fim de permitir melhor visualização dos resultados. .... 97
- Figura 27 – (A) Espectros pré-processados, utilizando primeira derivada com filtro de Savitzky-Golay (polinômio de 2ª ordem e janela 15 pontos), de 20 pixels aleatórios provenientes das imagens de 5 amostras de manga de diferentes espessuras; e (B) imagens destas 5 amostras (B-1 correspondendo à amostra menos espessa e B-5 correspondendo à amostra mais espessa) em pseudocores no comprimento de onda de 1143 nm. As cores azul e vermelha correspondem aos valores mínimos e máximos de absorvância em cada pixel no comprimento de onda de 1143 nm, respectivamente. .... 99
- Figura 28 – Gráfico de escores (PC1 *versus* PC2) resultante da PCA realizada com pixels provenientes das imagens das amostras de manga de diferentes espessuras mantidas sobre o suporte de PS. Os símbolos em vermelho referem-se à amostra menos espessa, enquanto os outros símbolos referem-se às demais amostras. .... 100
- Figura 29 – Gráfico exibindo os *loadings* em PC1 (em azul), o espectro médio do suporte de PS (em vermelho) e os espectros médios das amostras de manga de diferentes espessuras (amostra de menor espessura em verde e demais amostras em preto)..... 101
- Figura 30 – (A) Espectros pré-processados, utilizando primeira derivada com filtro de Savitzky-Golay (polinômio de 2ª ordem e janela 15 pontos), de 20 pixels aleatórios provenientes das imagens de 5 amostras de manga de diferentes espessuras; e (B) imagens destas 5 amostras (B-1 correspondendo à amostra menos espessa e B-5 correspondendo à amostra mais espessa) em pseudocores no comprimento de onda de 1231 nm. As cores azul e vermelha correspondem aos valores mínimos e máximos de absorvância em cada pixel no comprimento de onda de 1231 nm, respectivamente. .... 102
- Figura 31 – Gráfico de escores (PC1 *versus* PC2) resultante da PCA realizada com pixels provenientes das imagens das amostras de manga de diferentes espessuras mantidas sobre o suporte de PEAD. Os símbolos em vermelho referem-se à amostra menos espessa, enquanto os outros símbolos referem-se às demais amostras. .... 103
- Figura 32 – Gráfico exibindo os *loadings* em PC1 (em azul), o espectro médio do suporte de PEAD (em vermelho) e os espectros médios das amostras de manga de diferentes espessuras (amostra de menor espessura em verde e demais amostras em preto)..... 105
- Figura 33 – Resultados obtidos após a aplicação da técnica de agrupamento *K-means* (utilizando 6 clusters) nas imagens provenientes de um fruto em estágio de maturação pouco avançado (FV), registradas em 8 momentos distintos: antes do impacto (1), imediatamente após o

- impacto (2) e depois de 1 h (3), 2 h (4), 4 h (5), 24 h (6), 48 h (7) e 72 h (8) após o impacto. .... 107
- Figura 34 – Resultados obtidos após a aplicação da técnica de agrupamento *K-means* (utilizando 6 clusters) nas imagens provenientes de um fruto em estágio de maturação intermediário (FI), registradas em 8 momentos distintos: antes do impacto (1), imediatamente após o impacto (2) e depois de 1 h (3), 2 h (4), 4 h (5), 24 h (6), 48 h (7) e 72 h (8) após o impacto. .... 108
- Figura 35 – Resultados obtidos após a aplicação da técnica de agrupamento *K-means* (utilizando 6 clusters) nas imagens provenientes de um fruto em estágio de maturação avançado (FM), registradas em 8 momentos distintos: antes do impacto (1), imediatamente após o impacto (2) e depois de 1 h (3), 2 h (4), 4 h (5), 24 h (6), 48 h (7) e 72 h (8) após o impacto. .... 109
- Figura 36 – (A) espectrômetro NIR de bancada modelo XDS Rapid Content Analyzer (FOSS NIRSystems, MD, EUA); e (B) espectrômetro NIR portátil modelo MicroNIR 1700 (VIAVI, CA, EUA). .... 117
- Figura 37 – (A) Análise de maçãs utilizando-se o espectrômetro XDS; (B) detalhe do posicionamento da amostra diretamente sobre a janela de leitura deste equipamento, utilizando-se um adaptador “tipo íris”; (C) análise de maçãs utilizando-se o espectrômetro MicroNIR; e (D) detalhe do posicionamento da amostra sobre a janela de leitura da mesa desenvolvida especialmente para acomodar este equipamento. Em todos os casos, os espectros foram obtidos a partir de leituras realizadas em 4 pontos equidistantes distribuídos ao longo da região equatorial de cada fruto. .... 118
- Figura 38 – Espectros de absorvância (sem pré-processamento) de 1861 maçãs intactas (provenientes dos 5 projetos), registrados por meio do espectrômetro XDS na faixa entre 400 e 2500 nm. A região destacada na figura corresponde à faixa espectral registrada pelo espectrômetro MicroNIR, entre 1100 e 1600 nm. .... 125
- Figura 39 – Espectros registrados com o espectrômetro XDS na faixa espectral de 400 a 2500 nm e submetidos a diferentes técnicas de pré-processamento: (A) sem pré-processamento; (B) SNV; (C) 1DSG\_P2\_J15; e (D) 2DSG\_P2\_J15. .... 129
- Figura 40 – Espectros registrados com o espectrômetro XDS na faixa espectral de 1100 a 1600 nm e submetidos a diferentes técnicas de pré-processamento: (A) sem pré-processamento; (B) SNV; (C) 1DSG\_P2\_J15; e (D) 2DSG\_P2\_J15. .... 130
- Figura 41 – Espectros registrados com o espectrômetro MicroNIR na faixa espectral de 1100 a 1600 nm e submetidos a diferentes técnicas de pré-processamento: (A) sem pré-processamento; (B) SNV; (C) 1DSG\_P2\_J15; e (D) 2DSG\_P2\_J15. .... 131
- Figura 42 – Gráficos de escores em PC1 e PC2 relativos à PCA realizada para seleção e exclusão de amostras que apresentaram menor similaridade espectral em relação ao conjunto de amostras de transferência. Esta PCA foi realizada com as amostras provenientes do projeto 1 (símbolos na cor azul) e do conjunto de transferência (símbolos na cor bege). .... 135
- Figura 43 – Gráficos de escores em PC1 e PC2 relativos à PCA realizada para seleção e exclusão de amostras que apresentaram menor similaridade espectral em relação ao conjunto de amostras de transferência. Esta PCA foi realizada com as amostras provenientes do projeto 2 (símbolos na cor vermelha) e do conjunto de transferência (símbolos na cor bege). .... 136
- Figura 44 – Gráficos de escores em PC1 e PC2 relativos à PCA realizada para seleção e exclusão de amostras que apresentaram menor similaridade espectral em relação ao conjunto de

- amostras de transferência. Esta PCA foi realizada com as amostras provenientes do projeto 3 (símbolos na cor verde) e do conjunto de transferência (símbolos na cor bege). ..... 136
- Figura 45 – Gráficos de escores em PC1 e PC2 relativos à PCA realizada para seleção e exclusão de amostras que apresentaram menor similaridade espectral em relação ao conjunto de amostras de transferência. Esta PCA foi realizada com as amostras provenientes do projeto 4 (símbolos na cor azul) e do conjunto de transferência (símbolos na cor bege). ..... 137
- Figura 46 – Gráficos de escores em PC1 e PC2 relativos à PCA realizada para seleção e exclusão de amostras que apresentaram menor similaridade espectral em relação ao conjunto de amostras de transferência. Esta PCA foi realizada com as amostras provenientes do projeto 5 (símbolos na cor lilás) e do conjunto de transferência (símbolos na cor bege). ..... 137
- Figura 47 – Gráficos de escores em PC1 e PC2 (A) e PC1 e PC3 (B) obtidos incluindo-se todas as amostras remanescentes (símbolos de diferentes cores) e as amostras de transferência (símbolos na cor preta), após a exclusão das amostras que apresentaram menor similaridade espectral em relação ao conjunto de amostras de transferência..... 139
- Figura 48 – Valores obtidos pelo método de referência *versus* valores previstos pelos modelos de calibração, nas etapas de calibração (símbolos em preto) e validação externa (símbolos em vermelho) para o parâmetro sólidos solúveis (SS), empregando-se SNV como pré-processamento espectral e diferentes técnicas de calibração multivariada: (A) PLS, (B) SPA-MLR, e (C) LS-SVM. A linha preta representa a bissetriz. .... 141
- Figura 49 – Valores obtidos pelo método de referência *versus* valores previstos pelos modelos de calibração, nas etapas de calibração (símbolos em preto) e validação externa (símbolos em vermelho) para o parâmetro acidez titulável (AT), empregando-se SNV como pré-processamento espectral e diferentes técnicas de calibração multivariada: (A) PLS, (B) SPA-MLR, e (C) LS-SVM. A linha preta representa a bissetriz. .... 144
- Figura 50 – Valores obtidos pelo método de referência *versus* valores previstos pelos modelos de calibração, nas etapas de calibração (símbolos em preto) e validação externa (símbolos em vermelho) para o parâmetro firmeza de polpa (FP), empregando-se SNV como pré-processamento espectral e diferentes técnicas de calibração multivariada: (A) PLS, (B) SPA-MLR, e (C) LS-SVM. A linha preta representa a bissetriz. .... 147
- Figura 51 – Valores obtidos pelo método de referência *versus* valores previstos pelos modelos de calibração, nas etapas de calibração (símbolos em preto) e validação externa (símbolos em vermelho) para o parâmetro índice de amido (IA), empregando-se SNV como pré-processamento espectral e diferentes técnicas de calibração multivariada: (A) PLS, (B) SPA-MLR, e (C) LS-SVM. A linha preta representa a bissetriz. .... 150
- Figura 52 – Valores obtidos pelo método de referência *versus* valores previstos pelos modelos de calibração, nas etapas de calibração (símbolos em preto) e validação externa (símbolos em vermelho) para o parâmetro índice Streif (IS), empregando-se SNV como pré-processamento espectral e diferentes técnicas de calibração multivariada: (A) PLS, (B) SPA-MLR, e (C) LS-SVM. A linha preta representa a bissetriz. .... 153
- Figura 53 – Conjunto de espectros originais (sem pré-processamento), provenientes de 93 maçãs registrados nos equipamentos primário (A) e secundário (B). ..... 155
- Figura 54 – Conjunto de espectros, pré-processados por SNV, provenientes de 93 maçãs registrados nos equipamentos primário (A) e secundário (B), e diferenças espectrais existentes entre os dois espectrômetros (C). ..... 156

Figura 55 – Valores de  $RMSEP^{S-DS}_p$  (linha preta) obtidos em função do número de amostras de transferência ( $n = 2$  até  $n = 30$ ). Os gráficos A, B e C referem-se aos modelos de calibração para determinação de sólidos solúveis, construídos por PLS, SPA-MLR e LS-SVM, respectivamente; enquanto os gráficos D, E e F referem-se aos modelos de calibração para determinação de acidez titulável, construídos por PLS, SPA-MLR e LS-SVM, respectivamente. As linhas verde e vermelha tracejadas representam os valores de  $RMSEP^P_p$  e  $RMSEP^S_p$ , respectivamente..... 160

Figura 56 – Valores de  $RMSEP^{S-DS}_p$  (linha preta) obtidos em função do número de amostras de transferência ( $n = 2$  até  $n = 30$ ). Os gráficos A, B e C referem-se aos modelos de calibração para determinação de firmeza de polpa, construídos por PLS, SPA-MLR e LS-SVM, respectivamente; enquanto os gráficos D, E e F referem-se aos modelos de calibração para determinação de índice de amido, construídos por PLS, SPA-MLR e LS-SVM, respectivamente. As linhas verde e vermelha tracejadas representam os valores de  $RMSEP^P_p$  e  $RMSEP^S_p$ , respectivamente..... 161

Figura 57 – Valores de  $RMSEP^{S-DS}_p$  (linha preta) obtidos em função do número de amostras de transferência ( $n = 2$  até  $n = 30$ ). Os gráficos A, B e C referem-se aos modelos de calibração para determinação de índice Streif, construídos por PLS, SPA-MLR e LS-SVM, respectivamente. As linhas verde e vermelha tracejadas representam os valores de  $RMSEP^P_p$  e  $RMSEP^S_p$ , respectivamente..... 162

## LISTA DE TABELAS

Tabela 1 – Funções núcleo normalmente empregadas em SVM e LS-SVM, suas respectivas equações matemáticas e os parâmetros que devem ser otimizados pelo usuário destas técnicas.....	52
Tabela 2 – Parâmetros estatísticos relativos aos dados de referência. ....	90
Tabela 3 – Parâmetros estatísticos relativos às etapas de calibração e validação externa para determinação de sólidos solúveis (SS), matéria seca (MS), acidez titulável (AT) e firmeza de polpa (FP) em manga “Tommy Atkins”, empregando-se SNV como pré-processamento espectral e duas diferentes faixas espectrais para construção dos modelos de calibração: (1) 950-1650 nm (PLS/Full); e (2) somente os comprimentos de onda selecionados pelo algoritmo <i>Jack-Knife</i> (PLS/JK). ....	92
Tabela 4 – Informações referentes às bases de dados utilizadas neste trabalho, descrevendo o período de realização dos projetos, o número de variedades de maçãs, o número de amostras analisadas e os parâmetros de qualidade determinados por meio de análises de referência.....	116
Tabela 5 – Parâmetros estatísticos relativos aos parâmetros de qualidade determinados pelos métodos de referência.....	126
Tabela 6 – Matriz de correlação entre os parâmetros de qualidade determinados pelos métodos de referência.....	128
Tabela 7 – Resultados obtidos na etapa de validação interna para os parâmetros de qualidade de maçãs avaliados utilizando o espectrômetro XDS (em duas faixas espectrais distintas) e o espectrômetro MicroNIR. Foram utilizadas as seguintes técnicas de pré-processamento espectral: SNV; primeira derivada com filtro de Savitzky-Golay utilizando polinômio de 2ª ordem e janela de 15 pontos (1DSG_P2_J15); e segunda derivada com filtro de Savitzky-Golay utilizando polinômio de 2ª ordem e janela de 15 pontos (2DSG_P2_J15). ....	132
Tabela 8 – Parâmetros estatísticos relativos às etapas de calibração e validação externa para determinação de sólidos solúveis (SS), empregando-se SNV como pré-processamento espectral e os métodos de regressão PLS, SPA-MLR e LS-SVM para construção dos modelos de calibração multivariada.....	140
Tabela 9 – Parâmetros estatísticos relativos às etapas de calibração e validação externa para determinação de acidez titulável (AT), empregando-se SNV como pré-processamento espectral e os métodos de regressão PLS, SPA-MLR e LS-SVM para construção dos modelos de calibração multivariada.....	143
Tabela 10 – Parâmetros estatísticos relativos às etapas de calibração e validação externa para determinação de firmeza de polpa (FP), empregando-se SNV como pré-processamento espectral e os métodos de regressão PLS, SPA-MLR e LS-SVM para construção dos modelos de calibração multivariada.....	146
Tabela 11 – Parâmetros estatísticos relativos às etapas de calibração e validação externa para determinação de índice de amido (IA), empregando-se SNV como pré-processamento espectral e os métodos de regressão PLS, SPA-MLR e LS-SVM para construção dos modelos de calibração multivariada.....	149

Tabela 12 – Parâmetros estatísticos relativos às etapas de calibração e validação externa para determinação de índice Streif (IS), empregando-se SNV como pré-processamento espectral e os métodos de regressão PLS, SPA-MLR e LS-SVM para construção dos modelos de calibração multivariada..... 152

Tabela 13 – Valores de RMSEP obtidos pelos métodos PLS, SPA-MLR e LS-SVM para os diferentes parâmetros de qualidade avaliados..... 158

## LISTA DE ABREVIATURAS

ANN	Redes Neurais Artificiais ( <i>Artificial Neural Networks</i> )
AOTF	Filtros Óptico-Acústicos Sintonizáveis ( <i>Acousto-Optical Tunable Filters</i> )
AT	Acidez Titulável
DS	Padronização Direta ( <i>Direct Standardization</i> )
FC	Função Custo
FI	Frutos Intermediários (em estágio de maturação intermediário)
FM	Frutos Maduros (em estágio de maturação avançado)
FP	Firmeza de Polpa
FV	Frutos Verdes (em estágio de maturação pouco avançado)
GA	Algoritmo Genético ( <i>Genetic Algorithm</i> )
IA	Índice de Amido
IS	Índice Streif
JK	Algoritmo Jack Knife ( <i>Jack Knife Algorithm</i> )
KKT	Condições de Karush-Kuhn-Tucker ( <i>Karush-Kuhn-Tucker conditions</i> )
KS	Algoritmo de Kennard-Stone ( <i>Kennard-Stone Algorithm</i> )
LED	Diodos Emissores de Luz ( <i>Light Emitting Diodes</i> )
LS-SVM	Máquinas de Vetores de Suporte por Mínimos Quadrados ( <i>Least Squares Support Vector Machines</i> )
LVF	Filtro Linear Variável ( <i>Linear Variable Filter</i> )
MIR	Infravermelho Médio ( <i>Mid Infrared</i> )
MLR	Regressão Linear Múltipla ( <i>Multiple Linear Regression</i> )
MS	Matéria Seca
MSC	Correção Multiplicativa de Sinal ( <i>Multiplicative Signal Correction</i> )
NIR	Infravermelho Próximo ( <i>Near Infrared</i> )
OSC	Correção de Sinal Ortogonal ( <i>Orthogonal Signal Correction</i> )
PC	Componentes Principais ( <i>Principal Components</i> )
PCA	Análise de Componentes Principais ( <i>Principal Component Analysis</i> )

PCR	Regressão em Componentes Principais ( <i>Principal Component Regression</i> )
PEAD	Polietileno de Alta Densidade
PLS	Mínimos Quadrados Parciais ( <i>Partial Least Squares</i> )
PS	Poliestireno
RBF	Função Gaussiana de Base Radial ( <i>Radial Basis Function</i> )
RMSECV	Raiz Quadrada do Erro Médio Quadrático de Validação Cruzada ( <i>Root Mean Square Error of Cross Validation</i> )
RMSEP	Raiz Quadrada do Erro Médio Quadrático de Predição ( <i>Root Mean Square Error of Prediction</i> )
RMSEP <sub>p</sub>	RMSEP obtido pelo conjunto de predição do equipamento primário utilizando o modelo calculado neste equipamento
RMSEP <sup>S-DS</sup> <sub>p</sub>	RMSEP obtido pelo conjunto de predição do equipamento secundário padronizado por DS, aplicado ao modelo calculado no primário
RMSEP <sub>p</sub> <sup>S</sup>	RMSEP obtido pelo conjunto de predição do equipamento secundário utilizando o modelo calculado no primário
RMSEV	Raiz Quadrada do Erro Médio Quadrático de Validação ( <i>Root Mean Square Error of Validation</i> )
SNV	Varição Normal Padrão ( <i>Standard Normal Variate</i> )
SPA	Algoritmo das Projeções Sucessivas ( <i>Successive Projections Algorithm</i> )
SPXY	Algoritmo SPXY ( <i>Sample set Partitioning based on joint X-y distances algorithm</i> )
SS	Sólidos Solúveis
SVM	Máquinas de Vetores de Suporte ( <i>Support Vector Machines</i> )
WT	Transformada Wavelet ( <i>Wavelet Transformation</i> )

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO GERAL .....</b>	<b>23</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>26</b>
2.1	ESPECTROSCOPIA NA REGIÃO DO INFRAVERMELHO PRÓXIMO (NIR)..	26
<b>2.1.1</b>	<b>Aspectos gerais.....</b>	<b>26</b>
<b>2.1.2</b>	<b>Instrumentação.....</b>	<b>27</b>
2.1.2.1	Espectrômetros baseados em filtros ópticos .....	27
2.1.2.2	Espectrômetros baseados em Diodos Emissores de Luz (LED).....	28
2.1.2.3	Espectrômetros baseados em redes de difração .....	28
2.1.2.3.1	<i>Espectrômetros baseados em rede de difração móvel .....</i>	<i>28</i>
2.1.2.3.2	<i>Espectrômetros baseados em rede de difração fixa.....</i>	<i>29</i>
2.1.2.3.3	<i>Espectrômetros baseados em filtro linear variável.....</i>	<i>29</i>
2.1.2.4	Espectrômetros baseados em Filtros Óptico-Acústicos Sintonizáveis (AOTF) .....	31
2.1.2.5	Espectrômetros baseados em Transformada de Fourier (interferométricos) .....	31
<b>2.1.3</b>	<b>Modos de medida.....</b>	<b>32</b>
2.2	IMAGENS HIPERESPECTRAIS NA REGIÃO DO INFRAVERMELHO PRÓXIMO .....	33
<b>2.2.1</b>	<b>Instrumentação para a obtenção de imagens hiperespectrais.....</b>	<b>36</b>
2.3	MÉTODOS QUIMIOMÉTRICOS .....	37
<b>2.3.1</b>	<b>Pré-processamento de dados espectrais.....</b>	<b>38</b>
2.3.1.1	Suavização.....	39
2.3.1.2	Correção de linha de base.....	40
2.3.1.3	Derivada.....	40
2.3.1.4	Correção de sinal multiplicativa (MSC) .....	41
2.3.1.5	Varição normal padrão (SNV) .....	41
<b>2.3.2</b>	<b>Análise e calibração multivariada .....</b>	<b>41</b>
2.3.2.1	Análise Exploratória.....	41
2.3.2.2	Calibração Multivariada .....	43
2.3.2.2.1	<i>Regressão linear múltipla (MLR) .....</i>	<i>47</i>
2.3.2.2.2	<i>Regressão por mínimos quadrados parciais (PLS).....</i>	<i>48</i>
2.3.2.2.3	<i>Máquinas de vetores de suporte por mínimos quadrados (LS-SVM) .....</i>	<i>50</i>
<b>2.3.3</b>	<b>Seleção de variáveis .....</b>	<b>58</b>
2.3.3.1	Algoritmo <i>Jack Knife</i> (JK) .....	58
2.3.3.2	Algoritmo das projeções sucessivas (SPA).....	59

<b>2.3.4</b>	<b>Seleção de amostras</b> .....	62
2.3.4.1	Algoritmo Kennard-Stone (KS).....	62
2.3.4.2	Algoritmo SPXY .....	63
<b>2.3.5</b>	<b>Transferência de calibração</b> .....	63
2.3.5.1	Padronização das respostas espectrais por Padronização Direta (DS) .....	65
<b>3</b>	<b>AVALIAÇÃO DA QUALIDADE DE MANGAS “TOMMY ATKINS” UTILIZANDO UM ESCPECTRÔMETRO NIR PORTÁTIL E IMAGENS HIPERESPECTRAIS NA REGIÃO DO NIR</b> .....	<b>67</b>
3.1	INTRODUÇÃO.....	67
3.2	OBJETIVOS .....	71
<b>3.2.1</b>	<b>Objetivo Geral</b> .....	71
<b>3.2.2</b>	<b>Objetivos específicos</b> .....	71
3.3	MATERIAL E MÉTODOS.....	72
<b>3.3.1</b>	<b>Parte 1: avaliação da qualidade de mangas “Tommy Atkins” utilizando um espectrômetro NIR portátil</b> .....	72
3.3.1.1	Amostras .....	72
3.3.1.2	Instrumentação e aquisição espectral.....	73
3.3.1.3	Penetração da radiação na região do NIR através dos tecidos epidérmicos e subepidérmicos de mangas “Tommy Atkins” utilizando o espectrômetro NIR portátil MicroNIR 1700 .....	74
3.3.1.4	Desenvolvimento de modelos de calibração multivariada para determinação de parâmetros de qualidade de manga .....	75
3.3.1.4.1	<i>Análises de referência</i> .....	76
3.3.1.5	Monitoramento não destrutivo da qualidade dos frutos durante o amadurecimento.	79
3.3.1.6	Análise de dados e software .....	79
<b>3.3.2</b>	<b>Parte 2: identificação de danos mecânicos em mangas “Tommy Atkins” utilizando imagens hiperespectrais na região do NIR</b> .....	80
3.3.2.1	Amostras .....	80
3.3.2.2	Instrumentação e aquisição das imagens .....	81
3.3.2.3	Penetração da radiação na região do NIR através dos tecidos epidérmicos e subepidérmicos de mangas “Tommy Atkins” utilizando a câmera hiperespectral SisuCHEMA SWIR.....	81
3.3.2.4	Deteção precoce de danos mecânicos induzidos artificialmente em mangas “Tommy Atkins” por meio de imagens hiperespectrais na região do NIR.....	83
3.3.2.5	Análise de dados e software .....	84
3.4	RESULTADOS E DISCUSSÃO.....	85

<b>3.4.1</b>	<b>Parte 1: avaliação da qualidade de mangas “Tommy Atkins” utilizando um espectrômetro NIR portátil.....</b>	<b>85</b>
3.4.1.1	Penetração da radiação na região do NIR através dos tecidos epidérmicos e subepidérmicos de mangas “Tommy Atkins” utilizando o espectrômetro NIR portátil MicroNIR 1700.....	85
3.4.1.2	Desenvolvimento de modelos de calibração multivariada para determinação de parâmetros de qualidade de manga.....	88
3.4.1.2.1	<i>Atribuição de bandas espectrais.....</i>	89
3.4.1.2.2	<i>Desenvolvimento dos modelos de calibração.....</i>	90
3.4.1.3	Monitoramento não destrutivo da qualidade dos frutos durante o amadurecimento.....	95
<b>3.4.2</b>	<b>Parte 2: Identificação de danos mecânicos em mangas “Tommy Atkins” utilizando imagens hiperespectrais na região do NIR.....</b>	<b>98</b>
3.4.2.1	Penetração da radiação na região do NIR através dos tecidos epidérmicos e subepidérmicos de mangas “Tommy Atkins” utilizando a câmera hiperespectral SisuCHEMA SWIR.....	98
3.4.2.2	Detecção precoce de danos mecânicos induzidos artificialmente em mangas “Tommy Atkins” por meio de imagens hiperespectrais na região do NIR.....	106
3.5	CONCLUSÕES.....	110
3.6	PERSPECTIVAS FUTURAS.....	111
<b>4</b>	<b>TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE ESPECTRÔMETROS NIR DE BANCADA E PORTÁTIL PARA DETERMINAÇÃO DA QUALIDADE DE MAÇÃS.....</b>	<b>113</b>
4.1	INTRODUÇÃO.....	113
4.2	OBJETIVOS.....	115
<b>4.2.1</b>	<b>Objetivo geral.....</b>	<b>115</b>
<b>4.2.2</b>	<b>Objetivos específicos.....</b>	<b>115</b>
4.3	MATERIAL E MÉTODOS.....	115
<b>4.3.1</b>	<b>Amostras.....</b>	<b>115</b>
<b>4.3.2</b>	<b>Instrumentação e aquisição espectral.....</b>	<b>116</b>
<b>4.3.3</b>	<b>Análises de referência.....</b>	<b>118</b>
4.3.3.1	Determinação de sólidos solúveis (SS).....	119
4.3.3.2	Determinação de acidez titulável (AT).....	119
4.3.3.3	Determinação de firmeza de polpa (FP).....	119
4.3.3.4	Determinação de índice de amido (IA).....	120
4.3.3.5	Determinação de índice Streif (IS).....	120

<b>4.3.4</b>	<b>Estudo 1: Comparação entre o desempenho dos espectrômetros XDS e MicroNIR para a determinação de parâmetros de qualidade de maçãs .....</b>	<b>120</b>
4.3.4.1	Análise de dados e software .....	121
<b>4.3.5</b>	<b>Estudo 2: Transferência de calibração entre os espectrômetros NIR de bancada (XDS) e portátil (MicroNIR).....</b>	<b>121</b>
4.3.5.1	Análise de dados e software .....	122
4.4	RESULTADOS E DISCUSSÃO .....	124
<b>4.4.1</b>	<b>Atribuição de bandas espectrais .....</b>	<b>124</b>
<b>4.4.2</b>	<b>Análises de referência.....</b>	<b>125</b>
<b>4.4.3</b>	<b>Estudo 1: Comparação entre o desempenho dos espectrômetros XDS e MicroNIR para a determinação de parâmetros de qualidade de maçãs .....</b>	<b>128</b>
<b>4.4.4</b>	<b>Estudo 2: Transferência de calibração entre os espectrômetros NIR de bancada (XDS) e portátil (MicroNIR).....</b>	<b>134</b>
4.4.4.1	Modelos de calibração multivariada .....	134
4.4.4.2	Transferência de calibração .....	155
4.5	CONCLUSÕES .....	163
4.6	PERSPECTIVAS FUTURAS .....	164
	<b>REFERÊNCIAS .....</b>	<b>166</b>
	<b>APÊNDICE A – PRODUÇÃO CIENTÍFICA .....</b>	<b>182</b>

## 1 INTRODUÇÃO GERAL

As frutas e seus derivados representam uma importante classe de produtos agroindustriais para o Brasil. Atualmente o país é o terceiro maior produtor, sendo responsável por cerca de 6% da produção mundial de frutas (IBRAF, 2014). Neste cenário, a produção de mangas (mangicultura) ocupa uma posição de destaque, já que o país se encontra entre os sete maiores produtores e ocupa a terceira posição no ranking mundial de exportadores da fruta (ANUÁRIO BRASILEIRO DA FRUTICULTURA, 2017).

No cenário atual, embora a mangicultura seja uma atividade importante para o desenvolvimento nacional, alguns problemas relacionados à cadeia produtiva de manga ainda devem ser solucionados. O amadurecimento não uniforme dos frutos de um mesmo lote é um dos principais problemas enfrentados pela agroindústria de manga, resultando na comercialização de frutos com diferentes estádios de maturação e qualidade (SIGRIST, 2004). Esse fato promove o aumento nos índices de perdas e reduz o consumo e comercialização de manga, tanto no mercado interno como externo. A oferta de frutos heterogêneos no mercado é resultado da inadequação dos métodos utilizados para a determinação do estágio de maturação e qualidade de consumo dos frutos, normalmente baseados em avaliações visuais (THANGARAJ & IRULAPPAN, 1989).

De maneira similar ao que ocorre no Brasil, a produção de frutas também representa uma atividade agrícola importante para alguns países europeus, como é o caso da produção de maçãs na Bélgica, que atingiu 250.000 toneladas em 2016 (STATISTA, 2017). O mercado consumidor de maçãs é altamente exigente quanto à qualidade, o que demanda um sistema de beneficiamento que permita realizar uma seleção criteriosa das frutas com potencial de mercado *in natura*. Assim, antes de serem comercializadas, as frutas devem atender a requisitos mínimos de qualidade, que incluem atributos relacionados com características externas, tais como: tamanho, formato, cor, entre outros; bem como internas, tais como: teor de açúcares, acidez, firmeza, entre outros (PENG & LU, 2008). A avaliação desses atributos é importante, pois afetam diretamente o sabor e a textura das frutas e, conseqüentemente, a aceitação dos consumidores. Neste contexto, as técnicas analíticas não destrutivas são especialmente adequadas para serem utilizadas para esta finalidade.

A espectroscopia na região do infravermelho próximo (NIR, do inglês *Near Infrared Spectroscopy*) é uma das técnicas analíticas não destrutivas que tem sido propostas para a avaliação da qualidade de diversas frutas, tais como mangas e maçãs. Essa técnica analítica permite a quantificação de diversos parâmetros de qualidade, tais como: sólidos solúveis, matéria seca, acidez titulável, firmeza de polpa, amido, cor de polpa, polifenóis, índices de maturação e vitamina C (SARANWONG et al., 2004; SUBEDI et al., 2007; PISSARD et al., 2012; GIOVANELLI et al., 2014; FAN et al., 2016; MARQUES et al., 2016). Adicionalmente, esta técnica também tem sido sugerida para prever a qualidade de consumo e vida pós-colheita de frutas a partir de análises realizadas no momento da colheita (SARANWONG et al., 2004; SUBEDI et al., 2007; PISSARD et al., 2012; NASSUR, 2013).

A espectroscopia NIR apresenta a vantagem de ser rápida, não destrutiva, não necessitar de preparo de amostras ou uso de reagentes químicos, estando, portanto, alinhada aos conceitos da Química Verde. Entretanto, seu uso em condições de campo foi limitado durante muitos anos, em função de restrições impostas pelo tamanho e pouca robustez dos equipamentos disponíveis no mercado. Apenas recentemente, o desenvolvimento de novas tecnologias utilizadas na construção dos espectrômetros NIR têm permitido a redução significativa do tamanho e do custo desses equipamentos. Assim, os espectrômetros NIR portáteis representam atualmente uma alternativa atrativa para as agroindústrias e produtores realizarem o monitoramento da qualidade de seus produtos. Alguns estudos prévios demonstraram a capacidade desses equipamentos para esse tipo de aplicação (SARANWONG et al., 2003; BETEMPS et al., 2011; JHA et al., 2012; JHA et al., 2014; MARQUES et al., 2016).

Além da espectroscopia NIR, outras técnicas analíticas não destrutivas estão disponíveis para a avaliação da qualidade de frutas e de outros produtos agrícolas, dentre as quais vêm se destacando aquelas baseadas em imagens hiperespectrais (LU, 2008; QIN & LU, 2008; LEIVA-VALENZUELA et al., 2013; RIVERA et al., 2014; WANG et al., 2015). Estas imagens são obtidas por meio de câmeras hiperespectrais e exibem características típicas encontradas nas imagens digitais e em espectros provenientes de espectrômetros convencionas, fornecendo simultaneamente informações espectrais e espaciais da amostra sob análise (GOWEN et al., 2007; LU, 2008). Na área agrícola, as imagens hiperespectrais têm sido utilizadas principalmente em sensoriamento remoto de campos de produção e, mais recentemente, em aplicações relativas à pós-colheita de frutas. Nesta área podem ser encontrados na literatura estudos relacionados à detecção de

danos mecânicos (RIVERA et al., 2014), determinação de parâmetros de qualidade (TALLADA et al., 2006; ELMASRY et al., 2007; MENDOZA et al., 2011; WANG et al., 2015), detecção de danos superficiais (BARANOWSKI et al., 2012) e detecção de infestação de insetos (WANG et al., 2011).

Neste contexto, o presente trabalho aborda inicialmente (capítulo 3) um problema relacionado ao amadurecimento não uniforme de frutos de um mesmo lote, que impacta a cadeia produtiva de manga no Brasil. Assim, propõem-se o uso da espectroscopia NIR, por meio de um espectrômetro portátil, como técnica analítica rápida e não destrutiva para a determinação de parâmetros de qualidade de mangas e determinação precisa do estágio de maturação dos frutos no momento da colheita. Neste estudo, avaliou-se também a possibilidade de utilização de imagens hiperespectrais na região do NIR para realizar a detecção precoce de danos mecânicos presentes nos frutos.

Em um segundo momento (capítulo 4) são apresentados estudos sobre a transferência de calibração entre espectrômetros NIR de bancada e portátil, visando à determinação de parâmetros de qualidade de maçãs cultivadas na Bélgica; além da comparação prévia entre o desempenho desses espectrômetros quando utilizados na avaliação da qualidade destas frutas.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 ESPECTROSCOPIA NA REGIÃO DO INFRAVERMELHO PRÓXIMO (NIR)

#### 2.1.1 Aspectos gerais

A espectroscopia NIR é um tipo de espectroscopia vibracional caracterizada pelo uso da radiação eletromagnética compreendida entre 780 e 2500 nm. A energia dos fótons emitidos nesta faixa espectral não é suficiente para promover transições eletrônicas nas moléculas, mas é maior do que o necessário para promover as moléculas para os seus estados vibracionais excitados mais baixos (via transição vibracional fundamental), como ocorre na região do infravermelho médio (MIR, do inglês, *mid infrared*) (PASQUINI, 2003).

Na espectroscopia NIR, a amostra é irradiada utilizando-se uma fonte de radiação apropriada, que em geral é uma lâmpada de filamento de tungstênio simples ou halógena, e a radiação refletida ou transmitida é medida por meio de um detector. A intensidade de absorção da radiação eletromagnética em cada comprimento de onda dentro da faixa do NIR é registrada pelo espectrômetro, resultando no espectro da amostra analisada. O espectro é dependente de propriedades físicas e da composição química da amostra, bem como de suas características em relação ao espalhamento da radiação, que depende diretamente de sua microestrutura.

A maioria das bandas de absorção observadas no espectro NIR são resultantes de sobretons e combinações provenientes de transições vibracionais fundamentais que ocorrem na região do MIR, principalmente em ligações N-H, C-H, O-H e S-H (SKOOG et al, 2006; NICOLAI et al., 2007). As bandas de absorção na região do NIR são extremamente sobrepostas e cerca de 10 a 1000 vezes menos intensas que as bandas na região do MIR. O uso de fontes de radiação intensas e detectores de alta eficiência para a faixa espectral do NIR permite que a relação sinal ruído não seja comprometida (HONORATO, 2006).

Uma das vantagens da espectroscopia NIR em relação à espectroscopia MIR ou outras técnicas analíticas não destrutivas, é que a instrumentação utilizada apresenta custo relativamente baixo, pois os instrumentos operados nessa região podem ser construídos

utilizando sistemas ópticos semelhantes aos que operam na região do UV-Vis (SKOOG et al, 2006). Além disso, os espectrômetros NIR são mais robustos, já que possuem sistema óptico menos sensível à umidade do ar.

Apesar da espectroscopia NIR atualmente ter ampla aceitação pela comunidade científica e indústrias de diferentes setores, seu uso foi negligenciado durante um longo período de tempo. Enquanto a espectroscopia MIR ganhou ampla aceitação pelos espectroscopistas, a região do NIR foi pouco utilizada desde sua descoberta, em 1800 pelo astrônomo inglês William Herschel, até poucas décadas atrás (PASQUINI, 2003). A principal razão para esse fato é atribuída à complexidade dos espectros NIR, que apresentam bandas largas resultantes da sobreposição de picos individuais, dificultando sua interpretação. Assim, as primeiras publicações envolvendo a espectroscopia NIR associada a métodos multivariados só surgiram em 1965 com o trabalho pioneiro de Karl Norris, sendo que o amplo uso da técnica ocorreu somente a partir da década de 1980 (BLANCO et al, 2002; PASQUINI, 2003). O desenvolvimento da microeletrônica, dos computadores e da quimiometria foi essencial para o estabelecimento e difusão da tecnologia NIR.

### **2.1.2 Instrumentação**

Em geral, os espectrômetros NIR são constituídos pelos seguintes componentes: fonte de radiação (usualmente uma lâmpada halógena), compartimento de amostra, monocromador, detector e componentes ópticos diversos, tais como lentes, colimadores, divisores de feixe, esferas de integração e fibras ópticas. Normalmente, esses instrumentos são classificados de acordo com o tipo de monocromador utilizado, que é o componente responsável pela decomposição do feixe luminoso, permitindo a seleção de comprimentos de onda específicos (PASQUINI, 2003; SKOOG et al., 2006; NICOLAI et al., 2007). As principais classes de espectrômetros NIR, classificados em função do tipo de monocromador, são descritos brevemente a seguir (item 2.1.2.1 até 2.1.2.5)

#### **2.1.2.1 Espectrômetros baseados em filtros ópticos**

Nesses espectrômetros, filtros ópticos são usados para realizar a seleção de comprimentos de onda. Esses instrumentos são normalmente desenvolvidos para aplicações específicas (instrumentos dedicados). Em geral, esses equipamentos

apresentam baixo custo, mas possuem resolução espectral limitada (PASQUINI, 2003; SKOOG et al., 2006).

#### 2.1.2.2 Espectrômetros baseados em Diodos Emissores de Luz (LED)

Os espectrômetros baseados em Diodos Emissores de Luz (LED, do inglês *Light Emitting Diodes*) seguem a tendência atual de redução de custo e tamanho, visando principalmente aplicações em campo. Esses instrumentos podem utilizar um conjunto de LEDs para fornecer radiação na região do NIR sob a forma de bandas estreitas em comprimentos de onda específicos ou utilizá-los para fornecer radiação policromática, que pode ser dispersa por meio de filtros ópticos ou grades de difração. Assim como ocorre com os instrumentos baseados em filtros ópticos, os espectrômetros baseados em LED são normalmente equipamentos dedicados, utilizados em aplicações específicas (PASQUINI, 2003; SKOOG et al., 2006; FONSECA & RAIMUNDO Jr., 2007).

#### 2.1.2.3 Espectrômetros baseados em redes de difração

Nesse tipo de espectrômetro, uma rede de difração ou um prisma é utilizado como monocromador para separar a radiação na região do NIR em suas frequências individuais. Esses instrumentos dispersivos podem ser baseados em redes de difração móveis ou fixas.

##### 2.1.2.3.1 Espectrômetros baseados em rede de difração móvel

Nesse tipo de instrumento, uma rede de difração côncava ou plana se movimenta constantemente, permitindo que a radiação incidente seja dispersada em suas frequências individuais e direcionadas para o detector de forma sequencial. Assim, esses equipamentos são também conhecidos como espectrômetros de varredura. Os espectrômetros baseados nesse tipo de tecnologia foram os primeiros instrumentos fabricados para operar na região do NIR e continuam sendo utilizados atualmente. Esses espectrômetros tiveram papel fundamental no desenvolvimento e consolidação da espectroscopia NIR como ferramenta analítica.

As principais vantagens oferecidas pelos espectrômetros baseados em redes de difração móveis estão relacionadas ao custo desses equipamentos quando comparados a outros espectrômetros baseados em tecnologias mais modernas, além de possibilidade de

obtenção de espectros com alta resolução. Entretanto, apresentam como desvantagem a baixa velocidade de varredura, que resulta em maior tempo de análise; e o fato de possuírem partes móveis, que limita o uso desses instrumentos para aplicações em campo ou em ambientes considerados agressivos, tais como locais sujeitos a vibrações (PASQUINI, 2003; SKOOG et al., 2006; NICOLAI et al., 2007).

#### *2.1.2.3.2 Espectrômetros baseados em rede de difração fixa*

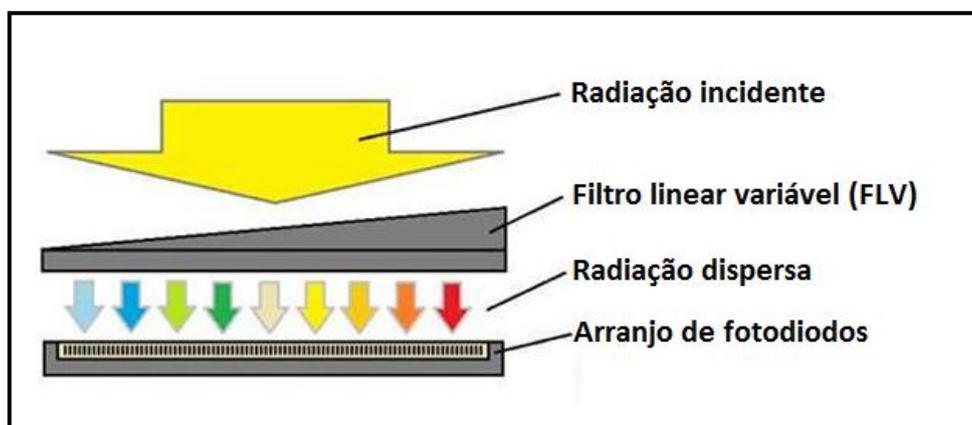
Os espectrômetros baseados nesse tipo de tecnologia possuem uma rede de difração fixa, côncava ou plana, na qual a radiação incidente é dispersada em suas frequências individuais e direcionadas ao mesmo tempo para o detector, que neste caso é um arranjo de fotodiodos. Esse tipo de detector é constituído por um arranjo linear de sensores de silício, ativo na faixa de 350 a 1100 nm, ou de arseneto de gálio e índio (InGaAs), ativo na faixa de 1100 a 2500 nm, que podem conter até 256 elementos independentes.

Os instrumentos baseados em redes de difração fixa permitem superar algumas limitações características dos instrumentos baseados em redes de difração móveis, já que possibilitam a obtenção de espectros em poucos milissegundos e não possuem partes móveis. Assim, esses equipamentos são adequados para uso em diversas aplicações em campo e, adicionalmente, permitem o monitoramento de amostras em tempo real. Esta tecnologia é empregada frequentemente em espectrômetros portáteis, pois permite miniaturização e maior robustez pela ausência de partes móveis. Aplicações típicas para esse tipo de espectrômetro incluem linhas de classificação de frutas, bem como equipamentos montados diretamente sobre máquinas agrícolas, viabilizando o monitoramento de parâmetros de qualidade de diferentes culturas diretamente no campo (PASQUINI, 2003; SKOOG et al., 2006; NICOLAI et al., 2007).

#### *2.1.2.3.3 Espectrômetros baseados em filtro linear variável*

Um avanço recente relacionado à tecnologia de rede de difração fixa consiste no uso de arranjos de fotodiodos como detectores associados a um tipo especial de filtro óptico, chamado de filtro linear variável (LVF, do inglês *Linear Variable Filter*). Um LVF é um tipo de filtro óptico que apresenta fisicamente o formato de cunha. Esse dispositivo é um elemento óptico dispersivo dependente da posição espacial da radiação

incidente, permitindo a seleção de comprimentos de onda específicos ao longo de seu comprimento. Nos instrumentos que usam esta tecnologia, o LVF é posicionado diretamente sobre o arranjo de fotodiodos, permitindo que a radiação dispersa proveniente do filtro seja direcionada diretamente para o conjunto de sensores (PEDERSON et. al., 2014). O princípio de funcionamento desse sistema é mostrado na Figura 1. Um exemplo típico de instrumento baseado nesta tecnologia é o espectrômetro NIR portátil modelo MicroNIR 1700 (VIAVI, CA, EUA), mostrado na Figura 2 a título de ilustração.



**Figura 1** – Princípio de funcionamento de um filtro linear variável (LVF) associado a um arranjo de fotodiodos. Nesse sistema, a radiação incidente é dispersa em suas frequências individuais em função da posição ao longo do comprimento do filtro. Em seguida, a radiação proveniente do filtro é direcionada para o conjunto de sensores (adaptado de PEDERSON et. al., 2014).



**Figura 2** – Espectrômetro NIR portátil modelo MicroNIR 1700 (VIAVI, CA, EUA): exemplo de equipamento portátil dispersivo, que emprega um filtro linear variável (LVF) associado a um arranjo de fotodiodos para efetuar a decomposição do feixe luminoso e viabilizar a seleção dos comprimentos de onda específicos.

#### 2.1.2.4 Espectrômetros baseados em Filtros Óptico-Acústicos Sintonizáveis (AOTF)

Os instrumentos baseados em Filtros Óptico-Acústicos Sintonizáveis (AOTF, do inglês, *Acousto-Optical Tunable Filters*) são espectrômetros de varredura, que utilizam uma tecnologia diferente daquela empregada nos instrumentos baseados em redes de difração para separar a radiação policromática em suas frequências individuais. O monocromador desses equipamentos é um dispositivo constituído por um cristal birrefringente ao qual é acoplado um transdutor feito de material piezoelétrico, que recebe um sinal de rádio frequência e o converte em onda acústica. Esta onda acústica se propaga através do cristal e produz uma variação periódica em seu índice de refração, que é função da faixa de rádio frequência utilizada. Assim, a radiação emitida por uma fonte apropriada incide sobre o cristal e interage com as ondas acústicas, levando o cristal a refratar seletivamente bandas estreitas de comprimento de onda específicos.

Esta tecnologia permite a construção de espectrômetros sem partes móveis e capazes de realizar varreduras extremamente rápidas utilizando a região espectral do NIR, sendo possível selecionar até 2000 comprimentos de onda distintos por segundo. Em geral, a velocidade de varredura é limitada pelo tempo de resposta dos detectores utilizados (PASQUINI, 2003).

#### 2.1.2.5 Espectrômetros baseados em Transformada de Fourier (interferométricos)

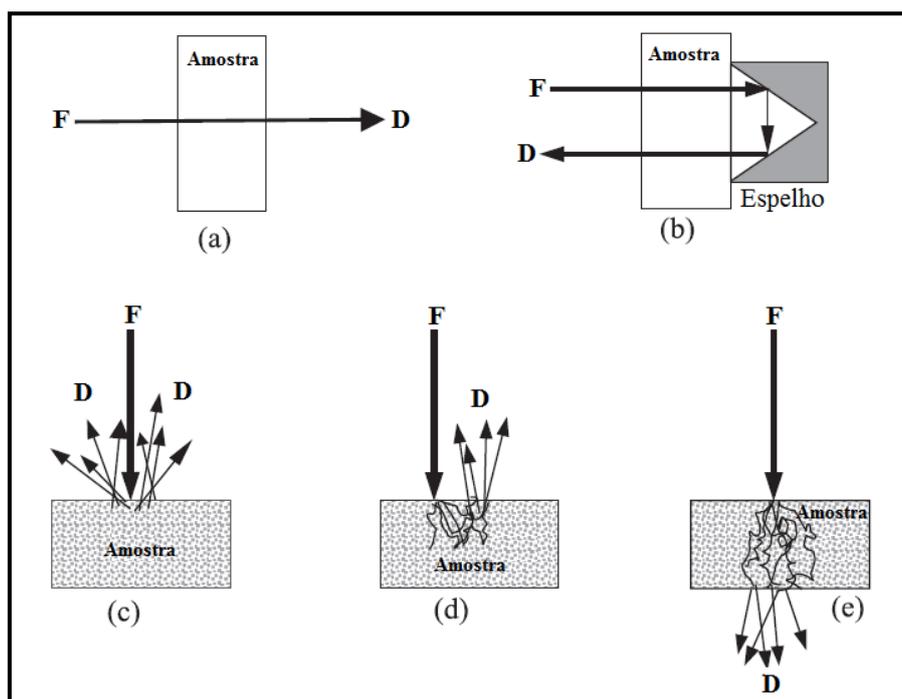
Os espectrômetros baseados nesta tecnologia não possuem qualquer componente óptico que separe a radiação policromática proveniente da fonte em suas frequências individuais, como ocorre nos instrumentos dispersivos. O funcionamento desse tipo de espectrômetro é baseado na geração de um interferograma por meio de um interferômetro do tipo “Michelson” ou de configuração similar. Os espectros são obtidos pelo cálculo da transformada de Fourier deste interferograma, que é o processo matemático pelo qual o interferograma, inicialmente no domínio do tempo, é transformado em seus componentes de frequência com suas amplitudes correspondentes, originando o espectro NIR característico.

Dentre os tipos de espectrômetros disponíveis atualmente, os instrumentos interferométricos apresentam características superiores em termos de precisão e exatidão de comprimento de onda, permitindo obter valores menores que 0,05 nm, bem como alcançar resolução menor que 1 nm. Além disso, apresentam alta relação sinal/ruído e

permitem realizar varreduras rápidas, já que, nesse tipo de instrumento, toda a radiação proveniente da fonte chega ao detector ao mesmo tempo (PASQUINI, 2003; SKOOG et al., 2006; NICOLAI et al., 2007).

### 2.1.3 Modos de medida

Nas técnicas espectroscópicas, o resultado de uma análise é decorrente da interação entre radiação e matéria. Assim, durante a realização de uma medida, a radiação que incide sobre a amostra pode ser refletida, absorvida ou transmitida. A contribuição relativa de cada fenômeno para a medida final depende da composição química e de características físicas da amostra. Em espectroscopia NIR a aquisição dos espectros pode ser realizada utilizando-se diferentes modos de medida (Figura 3), dentre os quais destacam-se os modos de transmitância e reflectância difusa, que são utilizados com maior frequência (PASQUINI, 2003; SKOOG et al., 2006; NICOLAI et al., 2007).



**Figura 3** – Principais modos de medida utilizados em espectroscopia NIR: a) transmitância; b) transflectância; c) reflectância difusa; d) interactância; e) transmitância através de meio dispersivo (onde, “F” representa a fonte de radiação; e “D” representa o detector) (adaptado de PASQUINI, 2003).

Em medidas realizadas por transmitância a fonte de luz é posicionada em frente ao detector, de modo que a radiação passe através da amostra. É o modo de medida frequentemente utilizado em espectroscopia UV/Vis, onde amostras líquidas transparentes são analisadas em cubetas de vidro ou quartzo, normalmente com caminhos ópticos de 1 a 50 mm.

Em medidas realizadas por reflectância difusa a fonte de luz e o detector são posicionados em ângulo específico, normalmente 45°, para evitar reflexão especular. Nesse modo de medida, normalmente utilizado para análise de materiais sólidos, o espalhamento e a absorção da radiação pelas partículas são fenômenos que contribuem de forma significativa para a variação da intensidade do sinal registrado pelo detector. Assim, é recomendado que sejam utilizadas técnicas de pré-processamento de dados espectrais, visando o tratamento adequado dos dados obtidos (PASQUINI, 2003; SKOOG et al., 2006; NICOLAI et al., 2007).

Kubelka e Munk (1931) desenvolveram um método adequado para realizar o tratamento do sinal obtido por reflectância difusa, originando uma equação para substituir a Lei de Lambert-Beer, que só é válida para materiais transparentes e homogêneos. Entretanto, esta equação é pouco utilizada na prática, pois a medida deve ser realizada utilizando-se uma esfera de integração. Assim, utiliza-se frequentemente a equação apresentada abaixo (Equação 2.1), que, embora simplificada, pode ser utilizada na grande maioria dos casos envolvendo o desenvolvimento de métodos analíticos baseados em medidas de reflectância difusa (PASQUINI, 2003; NICOLAI et al., 2007).

$$f(C) = \text{Log} \frac{1}{R} \quad (\text{Equação 2.1})$$

Onde,

C = Concentração;

R = Sinal obtido por reflectância difusa.

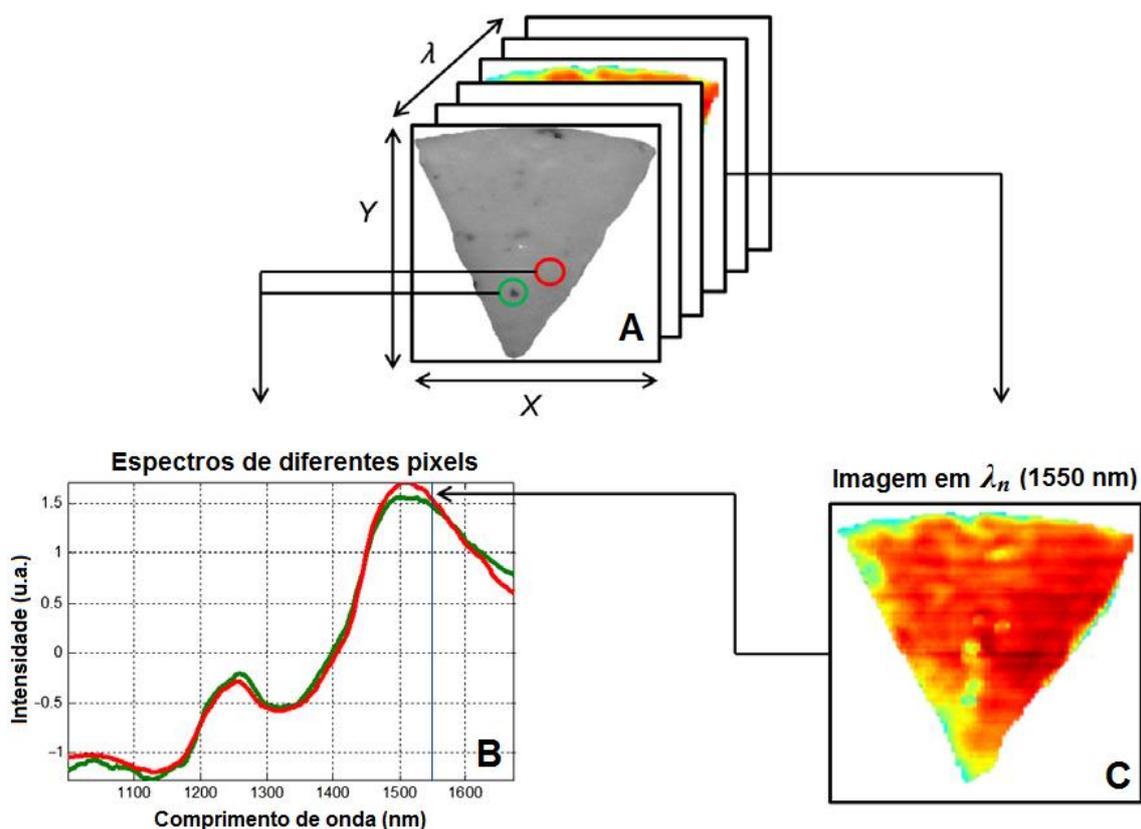
## 2.2 IMAGENS HIPERESPECTRAIS NA REGIÃO DO INFRAVERMELHO PRÓXIMO

As imagens hiperespectrais na região do NIR podem ser consideradas uma combinação entre as imagens digitais e a espectroscopia NIR e, desta forma, permitem a

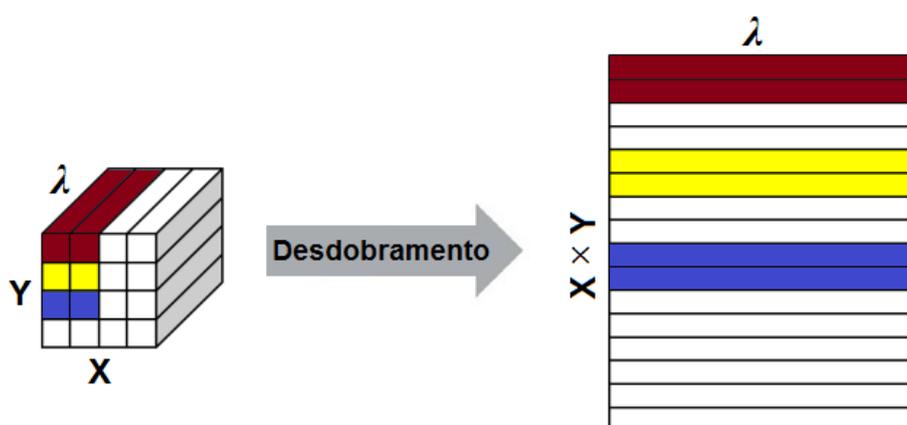
obtenção simultânea de informações espectrais e espaciais da amostra sob análise (GOWEN et al., 2007; LU, 2008; CHENG & SUN, 2014). Diferentemente do que ocorre com as imagens digitais, baseadas no sistema RGB (onde as imagens são formadas por três canais distintos: vermelho, verde e azul), as imagens hiperespectrais são formadas por um número muito maior de canais, que correspondem aos comprimentos de onda distintos registrados por uma câmera hiperespectral. Em geral, define-se que as imagens hiperespectrais devem apresentar mais do que 100 canais (GELADI et al., 2007).

As imagens hiperespectrais apresentam uma estrutura tridimensional, que é normalmente denominada de cubo hiperespectral, onde dois eixos ( $x$  e  $y$ ) representam a informação espacial e o terceiro eixo ( $\lambda$ ) representa a informação espectral (LU, 2008; AMIGO et al., 2013; WU & SUN, 2013). Desta forma, cada imagem hiperespectral é formada por um conjunto de pontos, chamados de pixels, distribuídos no plano formado pelos eixos  $x$  e  $y$ , sendo que cada pixel possui um espectro completo (ao longo do eixo  $\lambda$ ) (Figura 4-A). Logo, a seleção de um único pixel permite a visualização do espectro registrado em uma localização espacial específica, fornecendo a assinatura espectral dos componentes químicos presentes nessa parte da imagem (Figura 4-B). Caso seja selecionada uma área específica no plano ( $xy$ ) ou a imagem inteira e um valor específico de comprimento de onda ( $\lambda$ ), serão obtidos os valores de intensidade para todos os pixels, resultando em uma imagem monocromática (em um único comprimento de onda) (Figura 4-C).

A análise das imagens hiperespectrais é realizada após a transformação do cubo de dados em uma matriz que possa ser manipulada matematicamente. Para tal, é realizado o desdobramento do cubo hiperespectral ao longo de cada comprimento de onda, conforme mostrado na Figura 5. Por meio deste procedimento, cada pixel é acomodado em uma linha da nova matriz, que terá dimensões iguais a  $x.y$  pixels (número de linhas) e  $\lambda$  variáveis (número de colunas). O desdobramento do cubo de dados em uma matriz é uma estratégia simples, que possibilita a aplicação subsequente de diversos métodos quimiométricos para a extração de informações das imagens hiperespectrais (AMIGO et al., 2013).



**Figura 4** – (A) Cubo hiperespectral, onde os eixos  $x$  e  $y$  representam a informação espacial, enquanto o eixo  $\lambda$  representa a informação espectral; (B) espectros de pixels específicos selecionados no plano da imagem ( $xy$ ); e (C) imagem da amostra em um comprimento de onda específico (adaptado de AMIGO et al., 2013).



**Figura 5** – Desdobramento do cubo hiperespectral ao longo de cada comprimento de onda. Por meio deste procedimento, cada pixel é acomodado em uma linha da nova matriz, que terá dimensões iguais a  $x.y$  pixels (número de linhas) e  $\lambda$  variáveis (número de colunas) (adaptado de AMIGO et al., 2013).

A visualização das imagens hiperespectrais pode ser realizada empregando-se uma escala de cinza ou uma escala de cores para representar a intensidade dos pixels, resultando assim nas chamadas imagens em pseudocores (GELADI et al., 2007). É importante ressaltar que as pseudocores não estão relacionadas diretamente à composição química da amostra, mas correspondem à intensidade de absorção da radiação em um comprimento de onda específico, como já mencionado. Em geral, cores como azul e vermelho são utilizadas para representar os valores mínimos e máximos de absorbância em cada pixel no comprimento de onda selecionado.

### 2.2.1 Instrumentação para a obtenção de imagens hiperespectrais

Existem três configurações básicas de sistemas de aquisição de imagens hiperespectrais, cujas denominações estão relacionadas com a forma de aquisição das imagens, que pode ser realizada por ponto, linha ou plano. O resultado final obtido com qualquer uma destas configurações será sempre um cubo hiperespectral de dados (AMIGO et al., 2013). Tais configurações são descritas brevemente a seguir (GELADI et al., 2007; AMIGO et al., 2013):

(1) Aquisição de imagens por ponto (*Point Scan*): nesse tipo de configuração apenas o espectro completo de um ponto da amostra (representado por um pixel) é registrado por vez. Em seguida, a amostra é reposicionada e um novo espectro é obtido. Utilizando esta configuração é possível obter imagens com alta resolução espacial, entretanto, o tempo para aquisição das imagens é longo, sendo necessárias várias horas ou até mesmo dias para a obtenção de uma imagem completa.

(2) Aquisição de imagens por linha (*Line Scan*): nesse tipo de configuração são registrados os espectros de uma linha contendo vários pontos de uma única vez. Em seguida, a amostra é movimentada pelo sistema (normalmente por meio de uma plataforma) e um novo conjunto de espectros é registrado. Utilizando esta configuração a varredura de uma amostra pode ser realizada em poucos segundos ou minutos.

(3) Aquisição de imagens por plano (*Plane Scan*): nesse tipo de configuração não há movimentação da amostra ou do detector, e as imagens da amostra inteira são registradas em cada comprimento de onda por vez. Isto é realizado por meio do registro sequencial de imagens utilizando diferentes comprimentos de onda, selecionados por meio de um filtro ajustável adequado para esta finalidade (ex.: Filtros Óptico-Acústicos

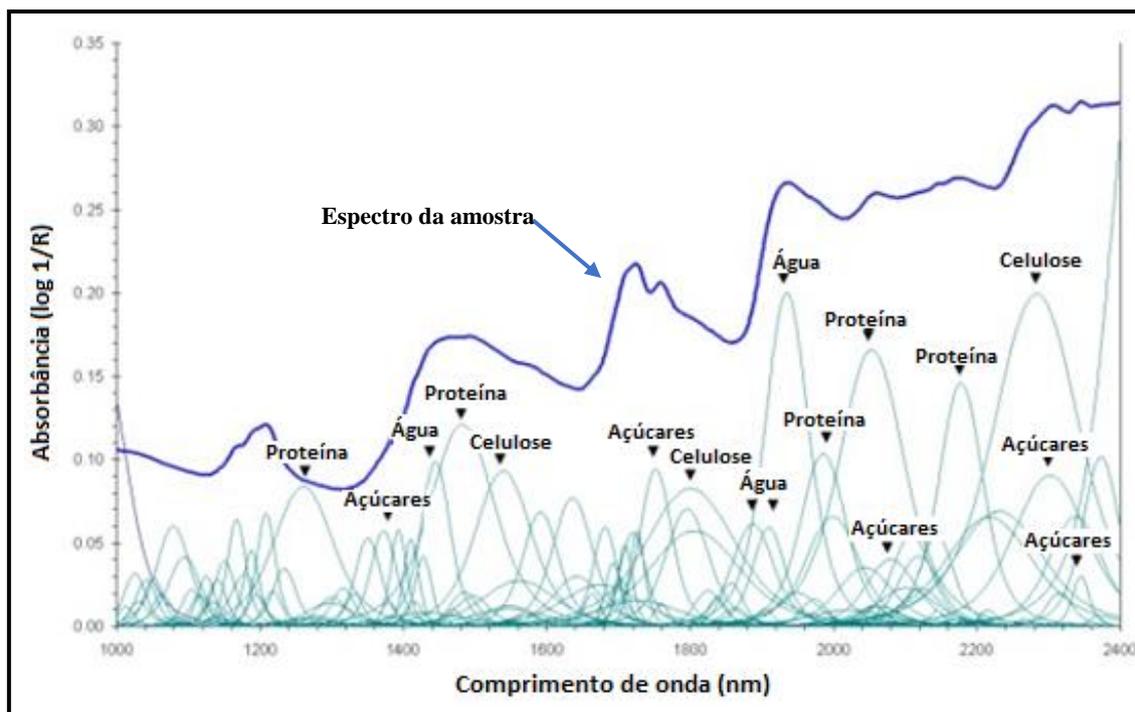
Sintonizáveis). Após o registro das imagens em cada comprimento de onda, obtém-se a imagem hiperespectral da amostra.

### 2.3 MÉTODOS QUIMIOMÉTRICOS

A quimiometria é uma área da química baseada na aplicação de métodos estatísticos e matemáticos para extração de informações relevantes de dados analíticos multivariados, tais como os dados provenientes de espectros na região do NIR (BARROS NETO et al., 2006). Os métodos e algoritmos pertencentes à quimiometria foram essenciais para o estabelecimento e difusão da espectroscopia NIR como uma poderosa ferramenta de análise em diversas aplicações qualitativas e quantitativas na química analítica (COSTA FILHO et al, 2002).

É importante destacar que a diferença entre dados univariados e multivariados é que nos primeiros, a análise dos dados é realizada utilizando-se uma única variável de cada vez, que pode ser, por exemplo, a concentração de uma espécie de interesse ou uma propriedade físico-química, tal como densidade, viscosidade, ponto de ebulição, entre outras. Nos dados multivariados, são analisadas mais de uma variável simultaneamente, sendo possível identificar a correlação entre elas. Assim, é possível obter um entendimento mais completo e sistematizado dos resultados analíticos

Os métodos quimiométricos são amplamente utilizados no tratamento de dados em diversas situações, tais como: planejamento e otimização de experimentos, pré-processamento de dados espectrais, reconhecimento de padrões e classificação de amostras, seleção de amostras e variáveis espectrais, e calibração multivariada (BEEBE et al, 1998). Estes métodos são indispensáveis para a interpretação adequada dos espectros NIR, pois estes apresentam bandas largas resultantes da sobreposição de picos individuais, que os tornam extremamente complexos. A Figura 6 ilustra a complexidade de um espectro NIR por meio de um espectro simulado de um material composto majoritariamente por água, açúcares, celulose e proteína. Pode-se verificar claramente que o espectro da amostra registrado pelo equipamento (linha azul de maior espessura) é resultado da sobreposição de inúmeros picos individuais correspondentes aos componentes da amostra.



**Figura 6** – Exemplo da complexidade de um espectro NIR: o espectro da amostra registrado pelo equipamento (linha azul de maior espessura) é resultado da sobreposição de inúmeras bandas individuais correspondentes aos componentes da amostra. Neste caso, o material analisado é composto majoritariamente por água, açúcares, celulose e proteína (adaptado de <https://www.wur.nl/en/show/NIR-spectroscopy-1.htm>).

### 2.3.1 Pré-processamento de dados espectrais

O pré-processamento de dados espectrais pode ser definido como qualquer manipulação matemática realizada antes da construção dos modelos preditivos (BEEBE et al, 1998).

Em geral, os dados provenientes de espectros NIR são submetidos a diferentes técnicas de pré-processamento antes de serem utilizados no desenvolvimento de modelos qualitativos ou quantitativos, visando remoção ou redução de fontes de variação sistemáticas ou aleatórias que não estejam relacionadas ao problema de interesse. As variações ocorrem normalmente devido à instabilidade do instrumento, ao espalhamento da radiação durante a realização das medidas, ou ainda devido à variabilidade das propriedades físicas das amostras (SIMÕES, 2008). Os problemas mais comuns estão associados ao espalhamento da radiação por amostras sólidas em medidas realizadas no

modo de reflectância difusa ou outros fenômenos relacionados com desvios na linha de base do espectro (PASQUINI, 2003).

É importante ressaltar que a escolha da técnica de pré-processamento é uma etapa crítica da modelagem, pois a mesma pode influenciar tanto positiva quanto negativamente no desempenho do modelo. Assim, é necessário que haja uma interação entre a técnica de pré-processamento utilizada e o reexame dos dados originais, além do conhecimento da natureza dos dados, bem como do processo químico ou físico do sistema em estudo (BEEBE et al, 1998).

As técnicas de pré-processamento de dados podem ser aplicadas tanto às amostras quanto às variáveis. As técnicas mais comuns aplicadas às variáveis são: centralização dos dados na média, que corresponde a subtrair o valor de cada elemento do vetor coluna (que corresponde a cada variável) pelo valor médio dos elementos da coluna; e escalonamento, onde os valores originais em cada coluna são subtraídos das respectivas médias e divididos pelo desvio padrão (BEEBE et al, 1998). Centralizar os dados na média equivale geometricamente, a efetuar uma translação do sistema de eixos ao longo do vetor das médias, para o centro do conjunto de dados; e escalonar os dados tem como objetivo colocar as variáveis em uma mesma escala quando tais variáveis apresentam unidades diferentes entre si ou quando a faixa de variação dos dados é grande. Quando esses dois pré-processamentos são realizados simultaneamente, é dito que as colunas da matriz estão autoescalonadas pela variância, ou somente, que a matriz está autoescalonada (TEÓFILO, 2013).

Dentre as diversas técnicas de pré-processamento aplicadas às amostras destacam-se: suavização, derivada, correção multiplicativa de sinal (MSC, do inglês *Multiplicative Signal Correction*), variação normal padrão (SNV, do inglês *Standard Normal Variate*), correção de sinal ortogonal (OSC, do inglês *Orthogonal Signal Correction*) e transformada wavelet (WT, do inglês *Wavelet Transformation*). Algumas destas técnicas são descritas brevemente a seguir (item 2.2.1.1 até 2.2.1.5).

#### 2.3.1.1 Suavização

A suavização é utilizada para remover matematicamente ruídos aleatórios de alta frequência, aumentando a relação sinal-ruído. Para esse procedimento existem diferentes filtros digitais, tais como: média móvel, Savitzky-Golay (BEEBE et al, 1998),

transformada de Fourier (CERQUEIRA et al, 2000) e transformada Wavelet (GALVÃO et al, 2001).

A suavização com filtro de Savitzky-Golay é uma das técnicas mais utilizadas. Esse filtro baseia-se na divisão do sinal analítico em uma série de intervalos de tamanho  $n$  (janela), de modo sequencial e ponto a ponto. Posteriormente, um polinômio de baixa ordem é ajustado aos pontos de uma janela (intervalo) pelo método dos mínimos quadrados (CERQUEIRA et al, 2000). Deve-se atentar para que o número de pontos da janela não seja elevado, caso contrário algumas bandas podem ser removidas ou distorcidas, resultando em perdas de informações presentes nos espectros (BEEBE et al, 1998).

#### 2.3.1.2 Correção de linha de base

A correção de linha de base é aplicada para minimizar variações sistemáticas de linha de base. Nesta técnica o espectro pré-processado é obtido subtraindo-se os dados originais a partir de uma função que represente a linha de base, que pode ser uma função linear (BEEBE et al, 1998). Pode também ser feita simplesmente subtraindo-se o sinal analítico em uma região espectral onde não há bandas ou picos. Esta técnica também pode ser associada a técnicas de suavização para minimização de ruído, tais como filtro de média móvel ou filtro de Savitzky-Golay.

#### 2.3.1.3 Derivada

A derivada é uma das técnicas de pré-processamento mais utilizadas em dados provenientes de espectroscopia NIR. A primeira derivada é utilizada para corrigir variações de linha de base dos espectros (DE LIRA, 2010). Esta técnica remove deslocamentos constantes de linha de base, além de possibilitar uma melhor visualização de picos existentes nos sinais originais (PONTES, 2009). A segunda derivada é utilizada para remover variações lineares de linha de base, normalmente devido a efeitos de espalhamento. É importante ressaltar, que o uso de derivadas resulta em uma diminuição na relação sinal-ruído dos espectros, tornando-se necessário utilizar filtros de suavização, tais como Savitzky-Golay (NAES et al., 2002; DE LIRA, 2010).

#### 2.3.1.4 Correção de sinal multiplicativa (MSC)

A MSC baseia-se na regressão de cada espectro contra um espectro de referência, que normalmente é o espectro médio do conjunto de calibração. Os espectros são corrigidos subtraindo-se dos valores originais o coeficiente linear das respectivas retas de regressão e dividindo os valores resultantes pela inclinação destas mesmas retas (HONORATO et al., 2007). Em geral a MSC simplifica o modelo de calibração, reduzindo o número de componentes, e melhora sua linearidade (NAES et al., 2002). Esta técnica é frequentemente utilizada para reduzir efeitos de espalhamento.

#### 2.3.1.5 Variação normal padrão (SNV)

O principal objetivo da SNV, de maneira similar a MSC, é eliminar ou reduzir efeitos de espalhamento, que ocorrem frequentemente em análise de amostras sólidas. Nesta técnica, calcula-se a média e o desvio padrão utilizando todos os pontos do espectro. Em seguida, o valor de cada ponto do espectro original é subtraído da média e dividido pelo desvio padrão. A SNV é aplicada para cada espectro individualmente, diferentemente do que ocorre com a MSC, onde a correção é baseada no espectro médio do conjunto de calibração. Em geral, a SNV apresenta maior capacidade de correção de efeitos de espalhamento (WANG et al., 2015).

### **2.3.2 Análise e calibração multivariada**

#### 2.3.2.1 Análise Exploratória

É importante a realização de uma avaliação prévia do conjunto de dados antes do início dos procedimentos de construção dos modelos de calibração. Esta avaliação prévia é normalmente conhecida como análise exploratória de dados e, em geral, tem como objetivo a visualização e/ou avaliação da existência de agrupamentos ou algum tipo de tendência natural na distribuição dos dados sob análise. Um dos métodos mais frequentemente aplicados para esta finalidade é a Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*), que consiste em uma técnica de

ortogonalização baseada em mudança de base vetorial (BEEBE et al., 1998; NAES et al., 2002).

A PCA é uma técnica matemática que emprega uma transformação ortogonal linear para converter o conjunto das variáveis originais, que são correlacionadas, em um conjunto menor de novas variáveis linearmente não correlacionadas, preservando o máximo de informação. Esse processo é realizado por meio da combinação linear das variáveis originais, que constituem um novo sistema de eixos ortogonais entre si (não correlacionados), denominados componentes principais (PC, do inglês *Principal Components*), fatores, ou variáveis latentes (BEEBE et al, 1998; FERREIRA et al., 1999). Esta transformação é definida de forma que a primeira PC apresente a maior variância possível, sendo, portanto, responsável pela máxima variabilidade dos dados. Sequencialmente, cada PC seguinte representa a máxima variabilidade dos dados remanescentes, respeitando a restrição de serem ortogonais às PC anteriores.

As coordenadas das amostras originais no novo sistema de eixos das PC são chamadas de escores. A contribuição que cada variável original exerce sobre uma determinada PC é denominada peso ou *loading*, que corresponde ao cosseno do ângulo entre o eixo da variável original e o eixo da PC. Assim, cada amostra passa a ser representada por um ponto no espaço multidimensional, do qual é possível extrair informações sobre a mesma e/ou sobre agrupamentos ou classes que apresentem características semelhantes (BEEBE et al, 1998; FERREIRA et al., 1999; BRERETON, 2000).

Em análise exploratória de dados é também frequente o uso de técnicas de agrupamento ou *clustering*, que compreendem um conjunto de métodos computacionais cujo propósito consiste em separar objetos em grupos, baseando-se nas características que esses objetos apresentam. A ideia básica consiste em colocar em um mesmo grupo objetos que sejam similares de acordo com algum critério pré-determinado. Em geral, o critério adotado é baseado em uma função de dissimilaridade, que descreve a distância entre dois objetos. Idealmente, os grupos devem apresentar alta homogeneidade interna e alta separação (heterogeneidade externa). Logo, os elementos de um determinado conjunto devem ser mutuamente similares e, preferencialmente, muito diferentes dos elementos pertencentes aos outros conjuntos (LINDEN, 2009; DUBEY et al., 2013).

Dentre diversos métodos de agrupamento ou de classificação não-supervisionados existentes, pode-se destacar o algoritmo *K-means*. Esse algoritmo é frequentemente utilizado para solucionar problemas de classificação e/ou agrupamento de dados, devido

a sua simplicidade, velocidade e flexibilidade (LINDEN, 2009; SUNU, 2014). No contexto da análise de imagens, o algoritmo *K-means* permite realizar a segmentação da imagem por meio do agrupamento dos pixels em classes distintas em função da similaridade apresentada entre seus espectros (SELIM et al., 1984; DUBEY et al., 2013).

No algoritmo *K-means* busca-se minimizar a distância dos elementos a um conjunto de  $k$  centros dado por  $\mu = \{x_1, x_2, \dots, x_k\}$  de forma iterativa. A distância entre um ponto  $p_i$  e um conjunto de clusters, dada por  $d(p_i, \mu)$ , é definida como sendo a distância do ponto ao centro mais próximo dele. Assim, a função a ser minimizada é dada por (LINDEN, 2009; DUBEY et al., 2013):

$$d(P, \mu) = \frac{1}{n} \sum_{i=1}^n d(p_i, \mu)^2 \quad (\text{Equação 2.2})$$

O algoritmo depende do número de *clusters*, que deve ser definido inicialmente pelo usuário. Em alguns casos esse fato pode ser um problema, já que normalmente não se sabe quantos clusters existem *a priori* (LINDEN, 2009).

### 2.3.2.2 Calibração Multivariada

O termo calibração, do ponto de vista da quimiometria, refere-se à construção de um modelo matemático visando estabelecer uma relação quantitativa entre os dados de saída de um instrumento (medidas instrumentais multivariadas, tais como espectros NIR) e os parâmetros físicos ou químicos de interesse previamente determinados por uma técnica analítica de referência (FERREIRA et al., 1999).

Em calibração multivariada utiliza-se frequentemente o conceito de modelo inverso. Neste caso, as respostas instrumentais são definidas como variáveis independentes (matriz  $\mathbf{X}$ ) e as concentrações ou propriedades de interesse são definidas como variáveis dependentes (vetor  $\mathbf{y}$ ) (TEÓFILO, 2013).

A organização dos dados é realizada da seguinte maneira: os dados espectrais são representados sob a forma de uma matriz  $\mathbf{X}$ , de dimensão  $m \times n$ , onde as linhas representam as amostras, cujos espectros foram registrados pelo instrumento, e as colunas representam as variáveis espectrais (comprimentos de onda); enquanto os parâmetros de interesse, determinados pelo método analítico de referência, são representados por um vetor  $\mathbf{y}$  (coluna) de dimensão  $m$  (Figura 7).

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mn} \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

Variáveis independentes
Variável dependente

**Figura 7** – Organização dos dados em matrizes e vetores utilizados em calibração multivariada: a matriz  $\mathbf{X}$  contém os dados espectrais registrados, de forma que suas linhas ( $m$ ) representam as amostras e as colunas ( $n$ ) as variáveis; enquanto o vetor  $\mathbf{y}$  (com  $m$  linhas) contém os valores do parâmetro de interesse.

O processo de desenvolvimento de modelos de calibração multivariados pode ser dividido em duas fases: calibração e validação externa (também chamada de predição). Adicionalmente, a fase de calibração é ainda dividida em duas etapas: a calibração, propriamente dita, que tem como objetivo o desenvolvimento de um modelo matemático, que é obtido relacionando-se a matriz de dados das variáveis instrumentais medidas ( $\mathbf{X}$ ) com o vetor contendo os valores da propriedade de interesse ( $\mathbf{y}$ ) determinados pelo método de referência; e a validação interna, também conhecida como validação cruzada, que tem como objetivo principal determinar o número ideal de variáveis que devem ser incluídas no modelo, bem como estimar o seu erro médio. Pode-se considerar que esta etapa permite a otimização do modelo construído.

Os procedimentos utilizados na validação cruzada são baseados no conjunto de amostras já utilizado na etapa de calibração. Neste caso, uma ou mais amostras são removidas do conjunto de calibração e um modelo é construído com as amostras restantes. A amostra ou o subconjunto de amostras removido têm os valores de suas propriedades de interesse previstos empregando-se um número crescente de variáveis. Esse processo é repetido para cada amostra ou subconjunto de amostras e um erro médio é calculado para cada número de variáveis utilizado, empregando-se, para tal, a raiz quadrada do erro médio quadrático de validação cruzada (RMSECV, do inglês *Root Mean Square Error of Cross Validation*). O número de variáveis a ser utilizado no modelo é determinado no momento em que não há mais variação significativa no valor de RMSECV a medida em que são introduzidas novas variáveis (BEEBE et al., 1998; PASQUINI, 2003; SOUZA et

al., 2013). O valor de RMSECV é calculado por meio da Equação 2.3 (BEEBE et al., 1998):

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n_{cal}} (y_{i,cal} - \hat{y}_{i,cal})^2}{n_{cal}}} \quad (\text{Equação 2.3})$$

Onde,

$n_{cal}$  = número de amostras do conjunto de calibração;

$y_{i,cal}$  = valor do parâmetro de interesse da i-ésima amostra do conjunto de calibração;

$\hat{y}_{i,cal}$  = valor do parâmetro estimado pelo modelo.

Após a determinação do número de variáveis que devem ser incluídas, o modelo de calibração pode ser construído, devendo este ser validado por meio de um conjunto de amostras independentes (amostras que não participaram da fase de calibração) e que tenham os valores da propriedade de interesse previamente conhecidos. Esta fase é conhecida como validação externa e tem como principal objetivo determinar a real capacidade de predição do modelo. Assim, espera-se que os valores previstos para as amostras do conjunto de validação externa sejam concordantes com os valores reais, determinados pelo método de referência (SOUZA et al., 2013). Em geral, a capacidade preditiva do modelo de calibração é avaliada nesta fase utilizando-se o valor da raiz quadrada do erro médio quadrático de previsão (RMSEP, do inglês *Root Mean Square Error of Prediction*), que deve ser o menor possível. O valor de RMSEP é calculado por meio da Equação 2.4 (BEEBE et al, 1998).

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_{prev}} (y_{i,prev} - \hat{y}_{i,prev})^2}{n_{prev}}} \quad (\text{Equação 2.4})$$

Onde,

$n_{prev}$  = número de amostras do conjunto de previsão (conjunto de validação externa);

$y_{i,prev}$  = valor do parâmetro de interesse da  $i$ -ésima amostra do conjunto de previsão (conjunto de validação externa);

$\hat{y}_{i,prev}$  = valor do parâmetro estimado pelo modelo.

Adicionalmente, os valores dos parâmetros estatísticos  $R^2$  (coeficiente de determinação) e *bias* são também utilizados com frequência na avaliação do desempenho preditivo dos modelos de calibração. O valor de  $R^2$  permite a avaliação do ajuste linear entre os valores determinados pelo método de referência e os valores previstos pelo modelo, além de representar a fração da variação em torno da média que é explicada pela regressão. O valor de *bias* permite verificar a existência de erros sistemáticos, que afetam os valores previstos pelo modelo sempre em um mesmo sentido, resultando em estimativas acima ou abaixo do esperado.

Por fim, depois da realização das fases de calibração e validação externa, o modelo pode ser utilizado para prever propriedades de amostras desconhecidas, que possuam a mesma natureza das amostras do conjunto de calibração, a partir do seu sinal espectral (PERALTA-ZAMORA et al., 1997; FERREIRA et al., 1999). É importante ressaltar, que o desempenho preditivo do modelo desenvolvido deve ser avaliado periodicamente quando este for utilizado em análises de rotina, a fim de verificar a necessidade de atualização do mesmo. Em geral, a atualização é realizada por meio da inclusão de novas amostras de referência que representem adequadamente as variações ocorridas na matriz das amostras e/ou que permitam a compensação de flutuações instrumentais (PASQUINI, 2003).

A construção de modelos de calibração multivariados a partir de dados espectrais pode ser realizada por meio de diversos métodos matemáticos, dentre os quais pode-se destacar: regressão linear múltipla (MLR, do inglês *Multiple Linear Regression*), regressão em componentes principais (PCR, do inglês *Principal Component Regression*), regressão por mínimos quadrados parciais (PLS, do inglês *Partial Least Squares*) (BEEBE et al., 1998; NAES et al., 2002), redes neurais artificiais (ANN, do inglês *Artificial Neural Networks*) (CERQUEIRA et al., 2001) e máquinas de vetores de suporte por mínimos quadrados (LS-SVM, do inglês *Least Squares Support Vector Machines*) (CODGILL & DARDENNE, 2004; THISSEN et al., 2004). Em geral, os métodos MLR, PCR e PLS são utilizados para a construção de modelos de calibração em casos onde há uma relação linear entre as variáveis independentes ( $\mathbf{X}$ ) e as variáveis dependentes ( $\mathbf{Y}$ ); enquanto os métodos ANN e LS-SVM podem ser utilizados para a modelagem de

sistemas não lineares. Os métodos MLR, PLS e SVM são descritos em mais detalhes a seguir (item 2.2.2.1 até 2.2.2.3).

É importante ressaltar, que métodos matemáticos de seleção de variáveis e amostras devem ser utilizados durante o desenvolvimento dos modelos de calibração, a fim de melhorar o desempenho preditivo destes. Os métodos de seleção de variáveis permitem encontrar variáveis que contenham informações relevantes para o problema de interesse. Enquanto, os métodos de seleção de amostras permitem separar adequadamente as amostras entre os conjuntos de calibração e validação externa, de modo que estas sejam representativas do conjunto original e que os valores de referência se encontrem dentro de uma faixa adequada para calibração e validação.

#### 2.3.2.2.1 Regressão linear múltipla (MLR)

A MLR é considerada um dos métodos mais simples e de fácil interpretação em calibração multivariada. Nesse método, os dados originais (informação contida na matriz  $\mathbf{X}$ ) são utilizados diretamente na modelagem do parâmetro de interesse. Assim, cada variável dependente (vetor  $\mathbf{y}$ ) é expressa como uma combinação linear das variáveis independentes contidas na matriz  $\mathbf{X}$  e um vetor  $\mathbf{b}$ , que contém os coeficientes de regressão, conforme mostrado na relação abaixo (Equação 2.5) (BEEBE et al, 1998):

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad (\text{Equação 2.5})$$

Onde,

$\mathbf{X}$  = matriz de dados das amostras de calibração;

$\mathbf{b}$  = vetor dos coeficientes de regressão;

$\mathbf{e}$  = vetor dos resíduos.

O vetor  $\mathbf{b}$  dos coeficientes lineares é estimado na etapa de calibração utilizando-se o critério dos mínimos quadrados, que minimiza o vetor dos resíduos ( $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$ ), sendo calculado pela relação a seguir (Equação 2.6) (BEEBE et al., 1998):

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (\text{Equação 2.6})$$

As estimativas para previsão dos parâmetros de interesse de amostras desconhecidas são obtidas de acordo com a relação abaixo (Equação 2.7) (BEEBE et al., 1998):

$$\hat{y} = \mathbf{X}'\mathbf{b} + \mathbf{e} \quad (\text{Equação 2.7})$$

Onde,

$\mathbf{X}'$  = matriz de dados das novas amostras (desconhecidas);

$\mathbf{b}$  = vetor dos coeficientes de regressão;

$\mathbf{e}$  = vetor dos resíduos.

A MLR é um método de regressão simples e de grande utilidade em aplicações relacionadas com calibração multivariada. Entretanto, dois problemas podem limitar sua aplicação prática. O primeiro corresponde a exigência de que o número de amostras seja maior que o número de variáveis, pois como o modelo é baseado na resolução de um sistema de equações lineares, esta condição é necessária para que o sistema tenha solução matemática. Caso contrário, a matriz  $\mathbf{X}^T\mathbf{X}$  não poderá ser invertida e o sistema se tornará indeterminado. O segundo diz respeito à colinearidade entre as variáveis espectrais, pois o cálculo da inversa da matriz  $\mathbf{X}^T\mathbf{X}$  pode ser realizado somente se o determinante da inversa for diferente de zero. No entanto, a existência de colunas na matriz  $\mathbf{X}$  altamente correlacionadas faz com que o valor do determinante da matriz  $(\mathbf{X}^T\mathbf{X})^{-1}$  seja igual ou próximo de zero, causando problemas de instabilidade numérica e propagação de erros na determinação de  $\mathbf{y}$  (VANDEGINSTE et al., 1998; BRERETON, 2000; SENA et al., 2000). Na prática, em função destas limitações, o uso da MLR é normalmente associado a técnicas de seleção de variáveis.

#### 2.3.2.2.2 Regressão por mínimos quadrados parciais (PLS)

A regressão PLS é um dos métodos mais utilizados em calibração multivariada e representa uma alternativa para contornar os problemas apresentados pela MLR. Este método baseia-se em um procedimento matemático similar ao utilizado na PCA, entretanto, tanto as informações contidas na matriz de dados espectrais  $\mathbf{X}$ , como também aquelas contidas no vetor  $\mathbf{y}$ , que representa o parâmetro de interesse, são utilizadas nos cálculos. A regressão PLS resolve os dois principais problemas da MLR, uma vez que

promove a redução do número original de variáveis sem ocasionar perda significativa da informação. Assim, são solucionados os problemas relativos a existência de alta colinearidade entre as colunas da matriz  $\mathbf{X}$  e a necessidade de se utilizar um número elevado de amostras para a construção do modelo de calibração (BEEBE et al, 1998; BRERETON, 2000).

Os cálculos se iniciam por meio da decomposição simultânea da matriz  $\mathbf{X}$  e do vetor  $\mathbf{y}$  em uma soma de  $h$  variáveis latentes, conforme mostrado nas relações a seguir (Equações 2.8 e 2.9) (BEEBE et al., 1998).

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum \mathbf{t}_h \mathbf{p}_h^T + \mathbf{E} \quad (\text{Equação 2.8})$$

$$\mathbf{y} = \mathbf{UQ}^T + \mathbf{F} = \sum \mathbf{u}_h \mathbf{q}_h^T + \mathbf{E} \quad (\text{Equação 2.9})$$

Onde,

$\mathbf{T}$  e  $\mathbf{U}$  = matrizes dos escores para a matriz  $\mathbf{X}$  e para o vetor  $\mathbf{y}$ , respectivamente;

$\mathbf{P}$  e  $\mathbf{Q}$  = matrizes dos pesos para a matriz  $\mathbf{X}$  e para o vetor  $\mathbf{y}$ , respectivamente;

$\mathbf{E}$  = matriz dos resíduos espectrais;

$\mathbf{F}$  = matriz dos resíduos dos parâmetros de interesse.

Esta decomposição em escores e pesos pode ser realizada por meio de diferentes algoritmos, os quais irão produzir resultados similares ao final do processo. Um exemplo de algoritmo empregado para esta finalidade é o NIPALS (do inglês *Non-linear Iterative Partial Least Squares*) (GELADI & KOWALSKI, 1986). Em seguida, os escores  $\mathbf{t}_h$  da matriz  $\mathbf{X}$  e os escores  $\mathbf{u}_h$  do vetor  $\mathbf{y}$  são relacionados por meio de uma simples correlação linear, conforme mostrado na Equação 2.10:

$$\mathbf{u}_h = b_h \times \mathbf{t}_h \quad (\text{Equação 2.10})$$

Os valores de  $b_h$  são agrupados em uma matriz diagonal  $\mathbf{B}$  para “ $h$ ” variáveis latentes. Esta matriz  $\mathbf{B}$  contém os coeficientes de regressão entre a matriz de escores  $\mathbf{U}$  de  $\mathbf{y}$  e a matriz de escores  $\mathbf{T}$  de  $\mathbf{X}$ . A melhor relação linear possível entre os escores de  $\mathbf{X}$  e  $\mathbf{y}$  é obtida por meio de pequenas rotações nas variáveis latentes destes dois blocos. Em

seguida, o vetor  $\mathbf{y}$  pode ser calculado conforme mostrado na relação abaixo (Equação 2.11):

$$\mathbf{y} = \mathbf{TBQ}^T + \mathbf{F} \quad (\text{Equação 2.11})$$

A última etapa consiste em obter as estimativas das propriedades de interesse de novas amostras (desconhecidas). Este procedimento é realizado a partir dos novos escores da matriz  $\mathbf{T}'$  (obtidos a partir dos espectros das novas amostras em análise), substituídos na a equação mostrada a seguir (Equação 2.12).

$$\mathbf{y} = \mathbf{T}'\mathbf{BQ}^T \quad (\text{Equação 2.12})$$

Em modelos de calibração desenvolvidos utilizando-se regressão PLS, o número ótimo de variáveis latentes é normalmente definido por meio de validação cruzada, empregando-se as amostras do conjunto de calibração. Pode-se considerar que este procedimento permite a otimização do modelo construído, conforme discutido anteriormente neste capítulo (item 2.2.2).

#### 2.3.2.2.3 Máquinas de vetores de suporte por mínimos quadrados (LS-SVM)

A relação entre as variáveis independentes ( $\mathbf{X}$ ) e as variáveis dependentes ( $\mathbf{Y}$ ) pode, em alguns casos, seguir um comportamento não-linear. Em situações que apresentem suaves não-linearidades, ainda é possível desenvolver modelos de calibração utilizando PLS. Entretanto, é comum que seja necessário o uso de um número excessivo de variáveis latentes, além de serem obtidos frequentemente modelos sobre-ajustados (WENTZELL & MONTOTO, 2003). Assim, torna-se necessário o uso de métodos que permitam a modelagem adequada de sistemas não-lineares, como é o caso das máquinas de vetores de suporte (SVM, do inglês *Support Vector Machines*).

As SVM são técnicas de aprendizagem estatística que foram desenvolvidas inicialmente para serem utilizadas em problemas de classificação binária, mas seus princípios foram estendidos para aplicações em calibração multivariada (THISSEN et al., 2004). As SVM em sua presente forma foram desenvolvidas na década de 1990 e, portanto, são consideradas técnicas quimiométricas relativamente novas (VAPNIK, 1998; FERRÃO et al., 2007).

Estas técnicas apresentam algumas características que as tornam ferramentas especialmente úteis em calibração multivariada, dentre as quais pode-se destacar (FERRÃO et al., 2007): (1) elevada capacidade de generalização, que corresponde à medida da eficiência na previsão de parâmetros de interesse de amostras não pertencentes aos conjuntos de calibração e/ou validação, evitando-se o sobre-ajuste (*overfitting*) do modelo; (2) robustez em grandes dimensões, permitindo sua aplicação em dados de alta dimensionalidade, tais como em análise de imagens; (3) convexidade da função objetivo, pois o uso das SVM implica na otimização de uma função quadrática (que possui apenas um mínimo), conferindo uma vantagem em relação às redes neurais, nas quais a função objetivo apresenta vários mínimos locais; e (4) teoria bem definida, pois as SVM são baseadas na Teoria de Aprendizado Estatístico, que já está bem estabelecida dentro da matemática e da estatística.

Uma característica inconveniente das SVM é o fato de serem baseadas em programação quadrática, que apresenta grande complexidade matemática e despende muito tempo computacional. Assim, para superar algumas limitações práticas associadas às SVM, SUYKENS et al. (2002) desenvolveram as LS-SVM, introduzindo algumas modificações nos algoritmos originais. As LS-SVM utilizam um sistema de equações lineares, a função custo por mínimos quadrados e restrições de igualdade, que resultam em menor custo computacional em relação às SVM, sem perda na qualidade das soluções, já que ambas as técnicas são baseadas nos mesmos princípios fundamentais (BORIN, 2007; FERRÃO et al., 2007). As principais etapas e cálculos realizados durante a aplicação das LS-SVM são apresentados de forma simplificada a seguir.

#### *Funções Núcleo e regressão pelo método dos mínimos quadrados*

Inicialmente, os dados espectrais contidos na matriz  $\mathbf{X}$  são mapeados para um espaço de características de alta dimensão. Este processo pode ser realizado por meio de uma função linear ou não-linear, que são conhecidas como funções núcleo (*kernel*) ou funções de linearização, representadas neste capítulo pelo termo  $\varphi$ . O método dos mínimos quadrados é aplicado neste espaço de características (CODIGILL & DARDENNE, 2004; BORIN, 2007).

As SVM apresentam diferentes tipos de funções núcleo, as quais caracterizam seu modo de reconhecimento de padrões. A escolha do tipo de função núcleo, bem como o ajuste de seus respectivos parâmetros, são fatores que apresentam grande influência sobre

os resultados obtidos pela técnica. Dentre as funções mais utilizadas, pode-se destacar as seguintes: linear, utilizada para dados lineares; polinomial, cujo grau pode ser definido pelo usuário; gaussiana de base radial (RBF, do inglês *radial basis function*), que corresponde a um espaço de características de dimensão infinita; e sigmoïdal, que confere à SVM um comportamento similar ao de uma rede neural (BORIN, 2007; FERRÃO et al., 2007). Na Tabela 1 são mostradas as principais funções núcleo, suas equações matemáticas correspondentes e os parâmetros que devem ser otimizados pelo usuário da técnica (SUYKENS et al., 2002).

**Tabela 1** – Funções núcleo normalmente empregadas em SVM e LS-SVM, suas respectivas equações matemáticas e os parâmetros que devem ser otimizados pelo usuário destas técnicas.

Funções núcleo ( $\varphi$ )	Equação	Parâmetros a serem otimizados
Linear	$x_i^T x_j$	-
Polinomial	$(x_i^T x_j + a)^b$	a, b
Gaussiana RBF	$e^{-\ x_i - x_j\ ^2 / 2\sigma^2}$	$\sigma^2$
Sigmoïdal	$\tanh(\beta_0 x_i^T x_j + \beta_1)$	$\beta_0, \beta_1$

Assim, considerando um conjunto de calibração genérico pertencente ao espaço euclidiano n-dimensional ( $R^n$ ) (BORIN, 2007):

$$[(x_1, \dots, x_n); x_1, \dots, x_n \in R^n] \quad (\text{Equação 2.13})$$

Uma função núcleo é utilizada para mapear o espaço da entrada original  $R^n$  para um espaço de características de alta dimensão:

$$\varphi(x) = [\varphi(x_1), \varphi(x_2), \dots, \varphi(x_n)] \quad (\text{Equação 2.14})$$

Em seguida, aplica-se a regressão linear no hiperespaço de características, por meio do método dos mínimos quadrados. Nesta etapa, é obtida a relação linear entre as

variáveis espectrais ( $\mathbf{x}$ ) e o parâmetro de interesse ( $y$ ) (THISSEN et al., 2004; BORIN, 2007; FERRÃO et al., 2007):

$$y(x_i) = \mathbf{w}^T \varphi(x_i) + b \quad (\text{Equação 2.15})$$

Onde:

$\mathbf{w}$  = vetor de pesos;

$b$  = coeficiente linear da reta de regressão.

A princípio, pode-se pensar que a solução para a Equação 2.15 poderia ser obtida por meio de um procedimento similar ao aplicado para a MLR (Equação 2.6), já discutido anteriormente. Entretanto, no caso da MLR trabalha-se com os dados originais provenientes da matriz de dados espectrais ( $\mathbf{X}$ ), enquanto que no caso da LS-SVM, a solução para a Equação 2.15 deve ser obtida por meio de dados provenientes de uma nova matriz, que foi obtida após a aplicação da função núcleo às variáveis espectrais originais (CODIGILL & DARDENNE, 2004).

#### *Função Custo*

É importante recordar que no método de regressão por mínimos quadrados busca-se obter uma função linear que minimize o a soma do quadrado dos erros de predição, ou seja, busca-se satisfazer a seguinte condição (CODIGILL & DARDENNE, 2004):

$$\min(e^2) = \min[\sum(y_i - \hat{y}_i)^2] \quad (\text{Equação 2.16})$$

Onde:

$e^2$  = soma do quadrado dos erros de predição;

$y_i$  = valor do parâmetro de interesse da  $i$ -ésima amostra do conjunto de previsão;

$\hat{y}_i$  = valor do parâmetro estimado pelo modelo.

No caso das SVM, é introduzida uma modificação no processo de otimização do ajuste da função linear. Além, da minimização dos erros de predição ( $e^2$ ), busca-se também minimizar a magnitude dos coeficientes de regressão ( $b$ ). Isto é realizado por meio da introdução de um novo termo na relação mostrada pela Equação 2.16, originando

a chamada “Função Custo” (FC) (Equação 2.17) (CODIGILL & DARDENNE, 2004; THISSEN et al., 2004; FERRÃO et al., 2007).

$$FC = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^n e_i^2 \quad (\text{Equação 2.17})$$

Onde,

$\mathbf{w}$  = vetor dos pesos relacionados à reta de regressão;

$\gamma$  = parâmetro de regularização “gama”, que penaliza erros altos (deve ser otimizado pelo usuário);

$\mathbf{e}$  = vetor de mínimos quadrados.

A primeira parte da FC é responsável por regularizar e penalizar os pesos relacionados à reta de regressão com valores elevados. Em função desta regularização, os pesos convergem para valores menores. Este procedimento é necessário, pois pesos elevados resultam em excessiva variância do modelo, prejudicando a capacidade de generalização das LS-SVM. Além disso, podem contribuir para a obtenção de modelos sobre-ajustados. A segunda parte da FC representa os erros de regressão considerando todo o conjunto de calibração. O parâmetro  $\gamma$  (gama), que deve ser otimizado pelo usuário, fornece o peso relativo da segunda parte da FC quando comparada com a primeira parte da mesma (CODIGILL & DARDENNE, 2004; THISSEN et al., 2004; FERRÃO et al., 2007).

A FC deve ser minimizada, pois este procedimento evita o sobre-ajuste do modelo, permitindo que este apresente boa capacidade de generalização. Os cálculos são realizados aplicando-se uma restrição de igualdade à FC, conforme mostrado na Equação 2.18 (BORIN, 2007; FERRÃO et al., 2007):

$$y_i = \mathbf{w}^T \varphi(x_i) + b + e_i \quad (\text{Equação 2.18})$$

Onde,

$\mathbf{w}$  = vetor de pesos;

$b$  = coeficiente linear da regressão.

$e_i$  = erros mínimos em relação à reta de regressão.

A Equação 2.17 juntamente com sua restrição dada pela Equação 2.18 correspondem a um problema de otimização convexa, que pode ser resolvido por meio do método dos multiplicadores de Lagrange, conforme mostrado pela Equação 2.19 (THISSEN et al., 2004; FERRÃO et al., 2007). Este método permite encontrar máximos e mínimos de uma função de uma ou mais variáveis suscetíveis a uma ou mais restrições. A otimização convexa da FC apresenta como vantagem a obtenção de uma solução única, pois permite a determinação de apenas um mínimo global. Diferentemente do que ocorre com outros métodos, como é o caso das redes neurais, que possuem diferentes mínimos locais (FERRÃO et al., 2007).

$$L = \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^N e_i^2 - \sum_{i=1}^n \alpha_i [w^T \varphi(x_i) + b + e_i - y_i] \quad (\text{Equação 2.19})$$

Neste caso, temos:

$$y_i = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, e_i = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}, \alpha_i = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

Onde,

$y_i$  = vetor dos parâmetros de interesse;

$e_i$  = vetor de erros;

$\alpha_i$  = vetor de multiplicadores de Lagrange (conhecidos como vetores de suporte).

O ponto ótimo da otimização convexa é obtido calculando-se as derivadas parciais  $\frac{\partial L(w,b,e_i,\alpha_i)}{\partial w}$ ,  $\frac{\partial L(w,b,e_i,\alpha_i)}{\partial b}$ ,  $\frac{\partial L(w,b,e_i,\alpha_i)}{\partial e_i}$ ,  $\frac{\partial L(w,b,e_i,\alpha_i)}{\partial \alpha_i}$  da função dada pela Equação 2.19 e depois igualando-as à zero, conforme mostrado nas Equações 2.20 e 2.21 abaixo (ARFKEN & WEBER, 1995; FERRÃO et al., 2007):

$$\frac{\partial L(w,b,e_i,\alpha_i)}{\partial w} = w - \sum_{i=1}^n \alpha_i \varphi(x_i) = 0 \quad \therefore w = \sum_{i=1}^n \alpha_i \varphi(x_i) \quad (\text{Equação 2.20})$$

$$\frac{\partial L(w,b,e_i,\alpha_i)}{\partial e_i} = w - \sum_{i=1}^n \gamma e - \alpha_i = 0 \quad \therefore \alpha_i = \gamma e \quad (\text{Equação 2.21})$$

A partir da Equação 2.20 pode-se verificar que os pesos ( $w$ ) podem ser escritos como uma combinação linear dos multiplicadores de Lagrange ( $\alpha$ ), correspondentes às amostras de calibração. Em seguida, substitui-se o resultado da Equação 2.20 na equação original de regressão linear (Equação 2.15), obtendo-se o seguinte resultado (Equação 2.22) (THISSEN et al., 2004; FERRÃO et al., 2007):

$$y(x) = \sum_{i=1}^n a_i \varphi(x_i)^T \varphi(x_i) + b \quad (\text{Equação 2.22})$$

Os vetores  $\alpha$  são obtidos a partir da resolução de um sistema linear de equações na forma matricial (Equação 2.23), seguindo as condições de Karush-Kuhn-Tucker (KKT), que permitem a obtenção de uma resposta única para o sistema em questão (CRISTIANINI & SHAWE-TAYLOR, 2000; SUYKENS et al., 2002; CODIGILL & DARDENNE, 2004).

$$\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & K(x_1, x_1) + 1/\gamma & \dots & K(x_1, x_n) + 1/\gamma \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_n, x_1) + 1/\gamma & \dots & K(x_n, x_n) + 1/\gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \quad (\text{Equação 2.23})$$

A resolução da Equação 2.23 resulta no modelo LS-SVM, conforme mostrado pela Equação 2.24. Os termos  $\alpha_i$  e  $b$  são soluções do sistema linear nas condições de KKT, enquanto o termo  $\mathbf{K}$  representa a matriz núcleo de transformação, cujos elementos foram obtidos após a aplicação de uma função núcleo apropriada aos dados espectrais originais (pertencentes à matriz  $\mathbf{X}$ ) (BORIN, 2007; THISSEN et al., 2004; FERRÃO et al., 2007):

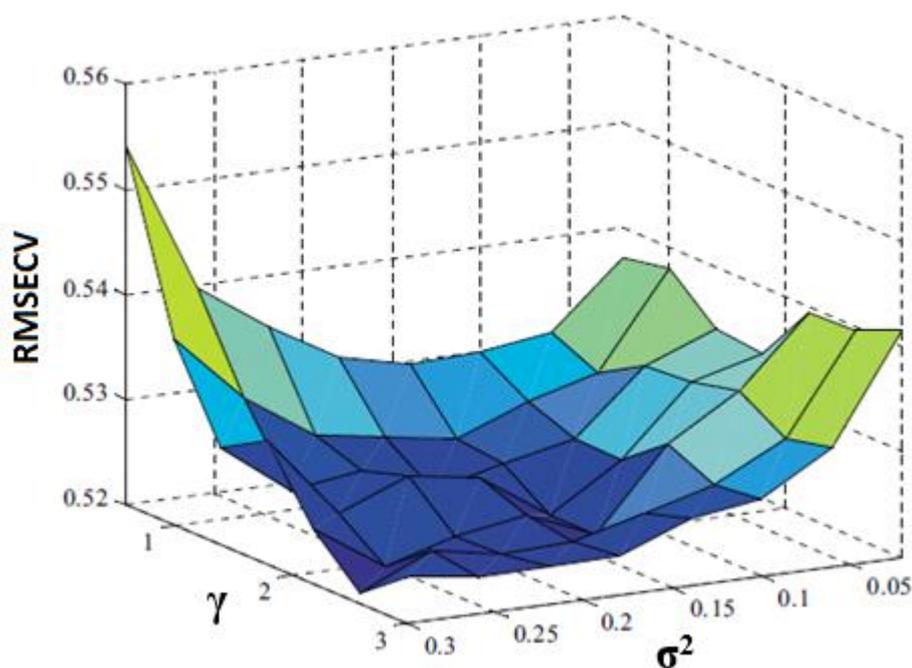
$$y(x) = \sum_{i=1}^n a_i K(x, x_i) + b \quad (\text{Equação 2.24})$$

#### *Otimização da função núcleo e da função custo durante a construção de modelos de calibração por LS-SVM*

A construção de modelos de calibração por meio das LS-SVM deve passar obrigatoriamente por uma etapa de otimização, que pode ser considerada correspondente à determinação do número ótimo de variáveis latentes em modelos PLS. Os parâmetros que devem ser otimizados dependem do tipo de função núcleo utilizada, conforme

mostrado anteriormente na Tabela 2.1. Adicionalmente, sempre será necessário otimizar o parâmetro de regularização  $\gamma$  (gama) da FC, independentemente do tipo de função núcleo empregada (CODIGILL & DARDENNE, 2004; 2004; FERRÃO et al., 2007).

Neste contexto, para uma LS-SVM com função núcleo RBF, são otimizados 2 parâmetros:  $\gamma$  e  $\sigma^2$  (sigma quadrático, que corresponde à largura da função gaussiana). Neste caso, os parâmetros são otimizados simultaneamente por meio de uma superfície de resposta, que permite a obtenção de um valor mínimo para o erro do modelo (normalmente considera-se o valor de RMSECV) em função dos valores de  $\gamma$  e  $\sigma^2$ . Na Figura 8 é mostrado um exemplo de superfície de resposta utilizada para a otimização destes parâmetros. Adicionalmente, devem ser observadas outras características associadas à LS-SVM utilizada, tais como: tipo de SVM (Epsilon-SVR ou Nu-SVR) e o uso de técnicas de compressão de dados (normalmente PCA ou PLS) (CODIGILL & DARDENNE, 2004; BORIN, 2007; FERRÃO et al., 2007).



**Figura 8** – Superfície de resposta utilizada para a otimização simultânea dos parâmetros  $\gamma$  e  $\sigma^2$  em função dos valores de RMSECV durante a construção do modelo de calibração por LS-SVM (CODIGILL & DARDENNE, 2004).

### 2.3.3 Seleção de variáveis

Existem diversas técnicas de seleção de variáveis que podem ser utilizadas durante a construção de modelos de calibração visando melhorar o desempenho preditivo destes. Em geral, busca-se encontrar variáveis que contenham informações relevantes para o problema de interesse. Adicionalmente, as técnicas de seleção de variáveis permitem solucionar problemas relacionados à multicolinearidade (variáveis altamente correlacionadas), que podem limitar a aplicação de algumas técnicas de calibração multivariada. Dentre os diversos algoritmos disponíveis para realização da seleção de variáveis pode-se destacar: algoritmo genético (GA, do inglês *Genetic Algorithm*) (LUCASIU et al., 1994; JOUAN-RIMBAUD et al., 1995), *stepwise* (FORINA et al, 2007), transformada wavelet (JOUAN-RIMBAUD et al, 1997), regressão por mínimos quadrados parciais por intervalo (iPLS) (NØRGAARD et al, 2000), algoritmo *Jack Knife* (JK) (NAES et al, 2002) e algoritmo das projeções sucessivas (SPA, do inglês *Successive Projections Algorithm*) (ARAÚJO et al, 2001). Os algoritmos JK e SPA são descritos em mais detalhes a seguir (itens 2.2.3.1 e 2.2.3.2).

#### 2.3.3.1 Algoritmo *Jack Knife* (JK)

O algoritmo JK é utilizado para selecionar as regiões do espectro ou as variáveis mais informativas para a construção do modelo de calibração, baseando-se no uso dos coeficientes de regressão. Neste caso, os erros padrão das estimativas dos coeficientes de regressão obtidos pelo modelo são avaliados por meio de uma série de subconjuntos de dados. Em seguida, são calculados os intervalos de confiança para os coeficientes de regressão e avalia-se a significância de cada variável utilizada no modelo (NAES et al., 2002; HONORATO et al, 2007).

Este algoritmo é normalmente utilizado em modelos de calibração baseados em regressão PLS. O algoritmo JK é iniciado pela construção de um modelo PLS utilizando-se todas as amostras disponíveis, obtendo-se, assim, os coeficientes de regressão ( $\mathbf{b}$ ). Em seguida, as amostras são divididas, de forma aleatória ou por meio de algum método sistemático apropriado, em  $M$  subconjuntos que serão utilizados na etapa de validação cruzada durante o desenvolvimento do modelo de calibração. Nesta etapa, são obtidos os coeficientes de regressão para cada modelo proveniente de cada subconjunto ( $\mathbf{b}_m$ ) (WESTAD & MARTENS, 2010; KARAMAN et al., 2013). A variância de cada

coeficiente de regressão ( $s^2(b)$ ) é calculada de acordo com a seguinte relação (Equação 2.25) (KARAMAN et al., 2013):

$$s^2(b) = [\sum_{m=1}^M (b - b_{-m})^2] \left( \frac{M-1}{M} \right) \quad (\text{Equação 2.25})$$

Onde,

$\mathbf{b}$  = coeficientes de regressão provenientes do modelo de calibração contendo todas as amostras;

$\mathbf{b}_m$  = coeficientes de regressão provenientes dos modelos de calibração obtidos de cada subconjunto;

$M$  = número de subconjuntos de amostras utilizados na validação cruzada.

Utilizando os valores destas variâncias aplica-se o teste t de Student por meio da relação mostrada na Equação 2.26. O teste é realizado considerando-se  $M$  graus de liberdade (número de subconjuntos), a fim de avaliar se o valor do coeficiente de regressão é significativamente diferente de zero. Neste caso, são obtidos p-valores associados a cada coeficiente, que são considerados significativos se apresentarem valores menores que um *threshold* pré-definido (KARAMAN et al., 2013):

$$t_b = \frac{b}{s(b)} \quad (\text{Equação 2.26})$$

Onde,

$\mathbf{b}$  = coeficientes de regressão provenientes do modelo de calibração contendo todas as amostras;

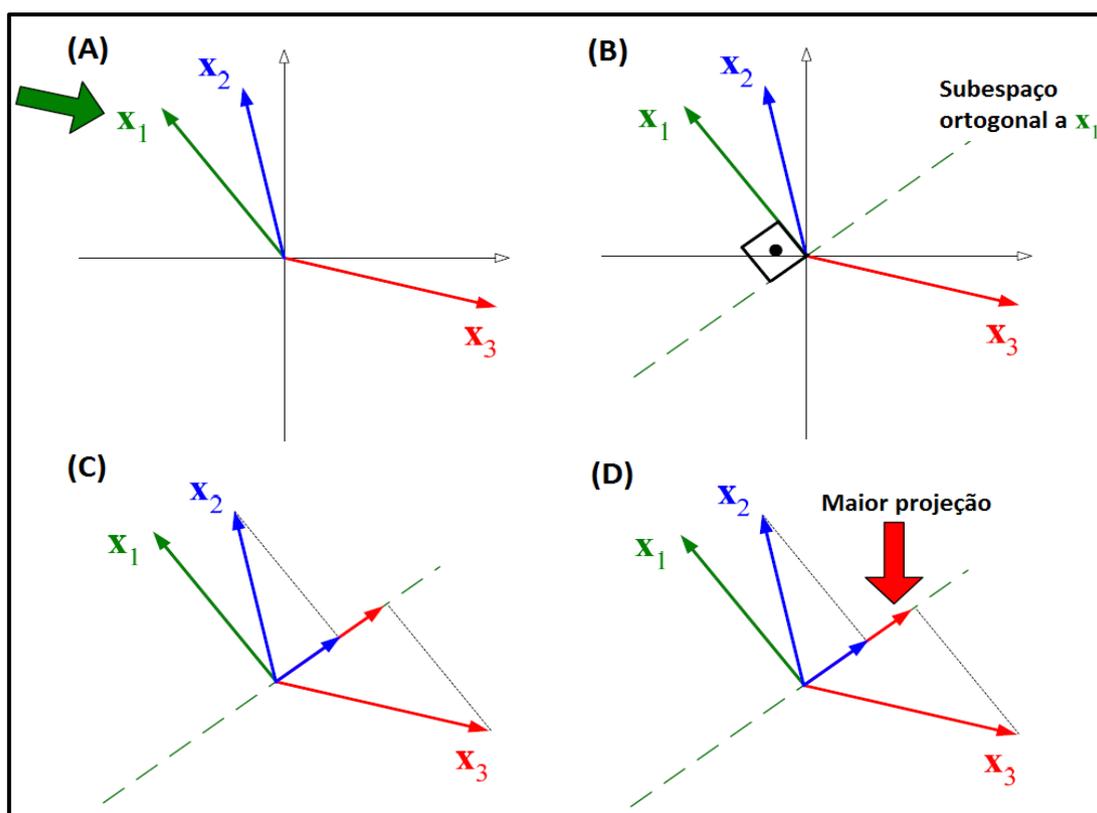
$s(\mathbf{b})$  = raiz quadrada da variância dos coeficientes de regressão.

### 2.3.3.2 Algoritmo das projeções sucessivas (SPA)

O algoritmo SPA, proposto por Araújo et al. (2001), é um método iterativo de seleção de variáveis desenvolvido para minimizar efeitos de multicolinearidade, especificamente quando se utiliza MLR para a construção de modelos de calibração. Seu objetivo é selecionar um subconjunto representativo de variáveis espectrais cujo conteúdo de informação seja minimamente redundante, o que é realizado por meio de três etapas

distintas: 1) construção de subconjuntos de variáveis espectrais; 2) avaliação dos subconjuntos visando selecionar a combinação ótima de variáveis (por meio de validação cruzada ou de um conjunto independente de validação); e 3) eliminação de variáveis não-informativas (ARAÚJO et al., 2001; GALVÃO et al., 2008).

Inicialmente, a construção dos subconjuntos é realizada por meio de uma sequência de operações de projeção envolvendo as colunas da matriz de respostas instrumentais, que resulta na obtenção de vários subconjuntos de variáveis espectrais (que são representadas por vetores). O princípio básico utilizado pelo algoritmo consiste em determinar o subespaço ortogonal a um dos vetores, e neste, projetar os outros vetores remanescentes. O vetor associado à maior projeção é então selecionado e o procedimento repetido com todos os vetores restantes. A Figura 9 ilustra os procedimentos realizados na primeira iteração do algoritmo, considerando um exemplo didático contendo três variáveis.



**Figura 9** – Procedimentos realizados na primeira iteração do algoritmo, considerando um exemplo didático contendo três variáveis: (A) escolha do primeiro vetor ( $x_1$ ); (B) determinação do subespaço ortogonal ao vetor escolhido; (C) projeção dos vetores remanescentes ( $x_2$  e  $x_3$ ) neste subespaço; e (D) seleção do vetor (que representa uma variável) associado à maior projeção (adaptado de GALVÃO & ARAÚJO, 2015).

Em seguida, os subconjuntos são avaliados de acordo com o desempenho preditivo dos modelos MLR resultantes. Para tal, utiliza-se o valor da raiz quadrada do erro médio quadrático de validação (RMSEV, do inglês *Root Mean Square Error of Validation*) (Equação 2.27), que deve ser o menor possível (GALVÃO et al., 2007).

$$RMSEV = \sqrt{\frac{\sum_{i=1}^{n_{val}} (y_{i,val} - \hat{y}_{i,val})^2}{n_{val}}} \quad (\text{Equação 2.27})$$

Onde,

$n_{val}$  = número de amostras do conjunto de validação;

$y_{i,val}$  = valor do parâmetro de interesse da  $i$ -ésima amostra do conjunto de validação;

$\hat{y}_{i,val}$  = valor do parâmetro estimado pelo modelo.

Em geral, esta avaliação é realizada utilizando-se um conjunto de amostras de validação independente. No entanto, Galvão et al. (2007) realizaram uma implementação no algoritmo SPA permitindo sua utilização na etapa de validação cruzada, dispensando o uso de amostras de validação independentes. Neste caso, a escolha das variáveis é baseada no menor valor do RMSECV. Assim, o melhor subconjunto de variáveis é aquele que, após a construção do modelo MLR, apresentar o menor valor do RMSECV.

A última etapa do algoritmo consiste em eliminar as variáveis consideradas não informativas, já que estas não contribuem significativamente para melhorar a capacidade preditiva do modelo de calibração. Este procedimento é também baseado no uso dos valores de RMSEV ou RMSECV. Neste caso, as variáveis selecionadas na etapa 2 são dispostas de acordo com um índice de relevância e o valor de RMSEV ou RMSECV é avaliado em função do número de variáveis mantidas nos modelos MLR (GALVÃO et al., 2008).

Em geral, os modelos SPA-MLR são mais simples que os modelos PLS, pois são utilizadas poucas variáveis espectrais na modelagem. Adicionalmente, diversos estudos têm demonstrado que a capacidade preditiva dos modelos SPA-MLR é comparável, ou até superior aos modelos PLS utilizando todas as variáveis (DI NEZIO et al, 2007; PISTONESI et al, 2010; PONTES et al, 2011).

### 2.3.4 Seleção de amostras

As amostras que serão utilizadas para a construção do modelo de calibração devem ser previamente separadas entre os conjuntos de calibração e validação externa. Esta separação deve ser realizada de modo que as amostras selecionadas sejam representativas do conjunto original e que os valores de referência se encontrem dentro de uma faixa adequada para calibração e validação. Dentre os métodos utilizados para esta finalidade pode-se destacar: métodos baseados em seleção aleatória (HAMZAÇEBI & KUTAY, 2006), método duplex (SNEE, 1977), algoritmo Kennard-Stone (KS) (KENNARD e STONE, 1969) e algoritmo SPXY (do inglês *Sample set Partitioning based on joint X-y distances*) (GALVÃO et al., 2005). Os algoritmos KS e SPXY são descritos em mais detalhes a seguir (itens 2.2.4.1 e 2.2.4.2). O algoritmo KS é frequentemente utilizado para seleção de amostras em problemas de classificação e transferência de calibração, enquanto o algoritmo SPXY é geralmente utilizado em problemas de calibração multivariada (GALVÃO et al., 2005).

#### 2.3.4.1 Algoritmo Kennard-Stone (KS)

O algoritmo KS tem como objetivo selecionar um subconjunto representativo em um conjunto de  $N$  amostras. Neste algoritmo, as distâncias euclidianas entre os vetores das respostas instrumentais ( $\mathbf{X}$ ) das amostras selecionadas são maximizadas (GALVÃO et al, 2005). A uniformidade de distribuição de cada subconjunto ao longo do espaço de dados  $\mathbf{X}$  é assegurada por meio do procedimento *stepwise*, no qual uma nova seleção é realizada excluindo-se as amostras já selecionadas. A distância Euclidiana  $d_{\mathbf{x}}(a,b)$  entre os vetores  $\mathbf{X}$  de cada par (a,b) de amostras é calculado conforme a relação mostrada na Equação 2.28:

$$d_{\mathbf{x}}(a, b) = \sqrt{\sum_{j=1}^J [x_a(j) - x_b(j)]^2}; \quad a, b \in [1, N] \quad (\text{Equação 2.28})$$

Onde,

$x_a(j)$  e  $x_b(j)$  = respostas instrumentais no  $j$ -ésimo comprimento de onda para as amostras “a” e “b”, respectivamente;

$j$  = número de comprimentos de onda do espectro.

A seleção começa tomando-se o par  $(a_1, b_1)$  de amostras nas quais a distância  $d_x(a_1, b_1)$  é maior. A cada iteração subsequente, o algoritmo seleciona a amostra que exibe a maior distância mínima em relação a todas as amostras já selecionadas. Estes procedimentos são repetidos até que o número de amostras especificado inicialmente seja alcançado (GALVÃO et al., 2005).

#### 2.3.4.2 Algoritmo SPXY

Este método é uma adaptação do algoritmo KS, que maximiza as distâncias euclidianas entre os vetores das respostas instrumentais ( $\mathbf{x}$ ). O algoritmo SPXY, diferentemente deste, considera tanto as diferenças em  $\mathbf{X}$  (matriz de dados espectrais) quanto em  $\mathbf{y}$  (vetor dos valores de referência) no cálculo das distâncias inter-amostras. Neste caso, cada distância  $d_y(a, b)$  é calculada de acordo com a relação mostrada a seguir (Equação 2.29) (GALVÃO et al., 2005):

$$d_y(a, b) = \sqrt{(y_a - y_b)^2} = |y_a - y_b|; \quad a, b \in [1, N] \quad (\text{Equação 2.29})$$

Para garantir a igualdade de importância para a distribuição das amostras no espaço de  $\mathbf{X}$  e  $\mathbf{y}$ , as distâncias  $d_x(a, b)$  e  $d_y(a, b)$  são divididas pelos seus valores máximos no conjuntos de dados, obtendo-se assim uma distância  $\mathbf{X}$ - $\mathbf{y}$  normalizada, conforme mostrado na Equação 2.30 (GALVÃO et al., 2005):

$$d_{xy}(a, b) = \frac{d_x(a, b)}{\max_{a, b \in [1, N]} d_x(a, b)} + \frac{d_y(a, b)}{\max_{a, b \in [1, N]} d_y(a, b)}; \quad a, b \in [1, N] \quad (\text{Equação 2.30})$$

Neste caso, a uniformidade de distribuição de cada subconjunto ao longo dos espaços de dados  $\mathbf{X}$  e  $\mathbf{y}$  é assegurada por meio de um procedimento de seleção *stepwise* utilizando-se  $d_{xy}(a, b)$  ao invés de  $d_x(a, b)$ , como ocorre no algoritmo KS.

#### 2.3.5 Transferência de calibração

A impossibilidade de se compartilhar modelos de calibração entre dois ou mais espectrômetros é um dos principais fatores que limitam o amplo uso da espectroscopia NIR

como ferramenta analítica (ALAMAR et al., 2007). O uso de um modelo desenvolvido em um equipamento (chamado de primário) com espectros registrados em um outro equipamento (chamado de secundário) ou sob novas condições ambientais e/ou instrumentais exige atenção. Este fato ocorre, pois, diferenças existentes entre as respostas instrumentais de diferentes espectrômetros e mudanças na resposta de um mesmo equipamento ao longo do tempo, podem levar o modelo original a fornecer previsões errôneas (HONORATO et al., 2007).

Um modelo torna-se inválido quando variações não incluídas na etapa de calibração são introduzidas nos novos espectros. Em geral, existem três situações que podem levar a esta condição (FEUDALE et al., 2002; HONORATO et al., 2007): (1) mudanças na constituição química e/ou física das amostras; (2) mudanças no ambiente do equipamento; e (3) mudanças na função de resposta do equipamento.

A aquisição de novos espectros utilizando o equipamento secundário, seguida da construção de novos modelos de calibração é uma forma de solucionar o problema apresentado anteriormente. Entretanto, a “recalibração completa” não é uma solução prática, sendo até mesmo inviável em algumas situações, pois requer tempo e investimentos significativos. Isto ocorre, pois, o emprego desta prática implica na necessidade de aquisição de novas amostras, realização de análises de referência, registro de espectros e tratamento matemático adequado dos dados.

Neste contexto, uma alternativa eficaz baseia-se na aplicação de técnicas quimiométricas, conhecidas como técnicas de transferência de calibração, para efetuar a correção das diferenças instrumentais e/ou ambientais, tornando os dados provenientes do equipamento secundário compatíveis com os modelos de calibração desenvolvidos com o equipamento primário. Assim, os modelos de calibração tornam-se transferíveis entre dois ou mais espectrômetros e evita-se a “recalibração completa” (FEUDALE et al., 2002; SALGUERO-CHAPARRO et al., 2013).

Em geral, as estratégias empregadas para realizar a transferência de calibração podem ser divididas em três grupos: (1) técnicas de padronização das respostas espectrais; (2) técnicas de padronização dos valores previstos; e (3) técnicas de aumento da robustez do modelo de calibração. No item 2.3.5.1 será detalhada a técnica de padronização das respostas espectrais empregando o método de Padronização Direta (DS, do inglês *Direct Standardization*), que foi a estratégia empregada nessa tese.

### 2.3.5.1 Padronização das respostas espectrais por Padronização Direta (DS)

As técnicas de transferência de calibração baseadas na padronização das respostas espectrais são aplicadas antes da construção dos modelos de calibração, e consistem em manipulações matemáticas dos dados espectrais provenientes dos equipamentos primário e secundário. Nestas técnicas, a função de resposta do equipamento secundário é modificada para se tornar similar à função de resposta do equipamento primário, no qual o modelo de calibração foi desenvolvido. Este procedimento é realizado utilizando-se espectros provenientes de um conjunto de amostras (conhecidas como amostras de transferência), que são registrados nos dois equipamentos, não havendo a necessidade de se utilizar os valores de referência das propriedades de interesse ( $y$ ) (HONORATO et al., 2007).

A técnica DS baseia-se na relação linear entre as matrizes espectrais provenientes dos equipamentos primário e secundário, que é representada pela Equação 2.31 (FEUDALE et al., 2002):

$$\mathbf{X}_P = \mathbf{X}_S \times \mathbf{F} \quad (\text{Equação 2.31})$$

Onde,

$\mathbf{X}_P$  = matriz espectral das amostras de transferência obtida com o equipamento primário;

$\mathbf{X}_S$  = matriz espectral das amostras de transferência obtida com o equipamento secundário;

$\mathbf{F}$  = matriz de transformação.

A matriz  $\mathbf{F}$  pode ser obtida diretamente se a matriz  $\mathbf{X}_S$  for inversível, conforme mostrado pela Equação 2.32. Entretanto, essa exigência dificilmente é satisfeita, já que normalmente existem mais variáveis que amostras, resultando em uma matriz  $\mathbf{X}_S$  retangular. Assim, a matriz  $\mathbf{F}$  é frequentemente calculada utilizando-se métodos multivariados, tais como PCR ou PLS. Neste caso, as matrizes  $\mathbf{X}_P$  e  $\mathbf{X}_S$  são projetadas sobre o espaço coluna reduzido de  $\mathbf{X}_P$ , e os escores obtidos nessas projeções são utilizados ao invés das matrizes originais (FEUDALE et al., 2002; HONORATO et al., 2007).

$$\mathbf{F} = (\mathbf{X}_S)^{-1} \times \mathbf{X}_P \quad (\text{Equação 2.32})$$

A padronização dos novos espectros registrados no equipamento secundário é realizada por meio da relação mostrada a seguir na Equação 2.33. Neste caso, o vetor contendo a resposta espectral de uma amostra desconhecida registrada no equipamento secundário ( $x_S$ ) é padronizado após ser multiplicado pela matriz de transformação ( $F$ ) (FEUDALE et al., 2002; HONORATO et al., 2007).

$$x^a = x_S \times F \quad (\text{Equação 2.33})$$

Onde,

$x^a$  = vetor resposta esperado no equipamento primário;

$x_S$  = vetor resposta de uma amostra desconhecida, medida no equipamento secundário;

$F$  = matriz de transformação.

Na transferência de calibração por DS, considera-se que as diferenças entre os espectros registrados nos equipamentos primário e secundário sejam provenientes apenas de variações instrumentais. Entretanto, variações na composição física e/ou química das amostras, bem como variações ambientais (temperatura, umidade, entre outras) são incluídas na modelagem e podem representar uma fonte de erros (IGNE, 2009). Assim, recomenda-se que as amostras de referência sejam analisadas nos dois equipamentos exatamente nas mesmas condições, garantindo que a matriz  $F$  represente apenas as variações instrumentais (HONORATO et al., 2007).

### **3 AVALIAÇÃO DA QUALIDADE DE MANGAS “TOMMY ATKINS” UTILIZANDO UM ESCPECTRÔMETRO NIR PORTÁTIL E IMAGENS HIPERESPECTRAIS NA REGIÃO DO NIR**

#### **3.1 INTRODUÇÃO**

A produção de mangas (mangicultura) representa uma importante atividade agrícola para o Brasil, já que o país se encontra entre os sete maiores produtores e ocupa a sexta posição no ranking mundial de exportadores da fruta (ANUÁRIO BRASILEIRO DA FRUTICULTURA, 2017). Em 2016, foram produzidas 976.815 toneladas de mangas, sendo 154.211 toneladas destinadas à exportação, resultando em receita de aproximadamente 180 milhões de dólares (ANUÁRIO BRASILEIRO DA FRUTICULTURA, 2017).

Nos últimos anos a região Nordeste tem contribuído de forma significativa para o desenvolvimento da fruticultura nacional. As condições climáticas favoráveis desta região, caracterizadas pela intensa insolação durante todo o ano, e a boa aptidão do solo em relação à irrigação são fatores que contribuem para promover a qualidade da produção irrigada de frutas (LACERDA et al., 2004; PINTO, 2008). Neste cenário, destaca-se a região do Vale do São Francisco, notadamente o polo frutícola de Petrolina-PE e Juazeiro-BA, o qual é atualmente responsável por mais de 90% das exportações de manga do país (ANUÁRIO BRASILEIRO DA FRUTICULTURA, 2013; ANUÁRIO BRASILEIRO DA FRUTICULTURA, 2017).

A manga é uma fruta climatérica e, portanto, durante a fase de maturação apresenta uma elevação na taxa respiratória, que resulta em aumento na velocidade de reações que levam ao seu amadurecimento e envelhecimento fisiológico. Assim, os frutos podem ser colhidos em um estágio de maturação pouco avançado, pois o processo de amadurecimento ocorre normalmente após a colheita (FACHINELLO & NACHTIGAL, 2009). Para garantir que as mangas cheguem aos consumidores e indústrias processadoras com qualidade adequada, é importante determinar o estágio de maturação ideal para a colheita, que dependerá do tempo necessário para que o fruto seja consumido ou industrializado. Para frutas transportadas e/ou armazenadas por longos períodos, é necessário realizar a colheita em estádios de maturação pouco avançados, mas que já tenham atingido sua maturidade fisiológica (SARANWONG et al., 2004; JHA et al.,

2007; SANTOS et al., 2008; JHA et al., 2014). Para frutas transportadas e/ou armazenadas por curtos períodos, essas podem ser colhidas em estádios de maturação mais avançados o que garantirá uma maior qualidade de consumo (JHA et al., 2007; SANTOS et al., 2008; JHA et al., 2014).

Os frutos colhidos em estádios de maturação pouco avançados amadurecem de forma irregular. Se colhidos muito cedo, podem não amadurecer e tornar-se impróprios para o consumo. Os frutos em estádios de maturação menos avançados apresentam maior sensibilidade a baixas temperaturas de armazenamento e maior tendência à perda de água por transpiração. Adicionalmente, apresentam alterações na cor, firmeza, teores de açúcares e acidez durante o período de amadurecimento. Em contrapartida, frutos colhidos em estádios de maturação mais avançados são mais sensíveis a danos mecânicos e ao ataque de microrganismos, e possuem menor tempo útil pós-colheita (SANTOS et al., 2008).

Embora a mangicultura seja uma atividade importante para o desenvolvimento nacional, alguns problemas relacionados à cadeia produtiva de manga ainda devem ser solucionados. O amadurecimento não uniforme dos frutos de um mesmo lote é um dos principais problemas encontrados pela agroindústria de manga, resultando na comercialização de frutos com diferentes estádios de maturação e qualidade (SIGRIST, 2004). Esse fato promove o aumento nos índices de perdas e reduz o consumo e comercialização de manga, tanto no mercado interno como externo. A oferta de frutos heterogêneos no mercado é resultado da inadequação dos métodos utilizados para determinação do estágio de maturação e qualidade de consumo dos frutos, normalmente baseados em avaliações visuais (THANGARAJ & IRULAPPAN, 1989).

Algumas características visuais consideradas são a tonalidade e a cerosidade da casca, assim como o formato do ápice e do “ombro” do fruto (GENÚ E PINTO, 2002; SIGRIST, 2004). Logo, o julgamento dessas características é dependente da experiência do produtor, contribuindo para o caráter subjetivo desse tipo de avaliação. Neste sentido, alguns estudos têm mostrado que o uso exclusivo das características visuais para a determinação do estágio de maturação, é ineficiente, resultando na colheita de frutos com diferentes estádios de maturação (BETEMPS et al., 2011).

A determinação do estágio de maturação no momento da colheita é, portanto, fundamental para que os produtores, principalmente os exportadores, possam planejar estratégias adequadas de armazenamento, transporte e comercialização de seus produtos. Além disso, as perdas pós-colheita podem ser reduzidas, já que a vida útil dos frutos é

dependente deste parâmetro (KIENZLE et al., 2012; JHA et al., 2014). Neste contexto, vários métodos analíticos não destrutivos têm sido propostos para a determinação de parâmetros de qualidade de mangas de forma que possam ser utilizados na avaliação do estágio de maturação, na determinação do momento ideal de colheita, ou ainda, para a classificação pós-colheita dos frutos.

Atualmente, diversas técnicas analíticas não destrutivas estão disponíveis para a avaliação da qualidade de frutas e de outros produtos agrícolas, tais como: ressonância magnética nuclear de baixo campo (relaxometria), tomografia utilizando raios-x, colorimetria, fluorescência, técnicas acústicas, narizes eletrônicos, espectroscopia NIR, entre outras (JHA & MATSUOKA, 2000; WANG et al., 2015). Dentre os métodos disponíveis, a espectroscopia NIR tem se destacado como técnica analítica não destrutiva para a avaliação da qualidade de mangas e determinação precisa do estágio de maturação, já que não sofre interferência da pigmentação da casca ou devido à presença do caroço dos frutos (SARANWONG et al., 2001; SUBEDI et al., 2007). A pigmentação da casca é considerada um parâmetro de pouca robustez para a determinação do estágio de maturação das mangas. Outra vantagem da espectroscopia NIR é a capacidade de efetuar medidas extremamente rápidas, sem uso de reagentes químicos ou qualquer tipo de preparo de amostra, necessitando apenas do desenvolvimento prévio de modelos de calibração para cada material genético e condições de cultivo (SLAUGHTER, 1995; SUBEDI et al., 2007).

Há diversos estudos publicados desde 1997, que demonstram o potencial da espectroscopia NIR (utilizando instrumentos de bancada) associada a modelos de calibração multivariados para a quantificação de diversos parâmetros físico-químicos em manga, tais como sólidos solúveis (SS), matéria seca (MS), acidez titulável (AT), firmeza de polpa (FP) e teor de amido (GUTHRIE & WALSH, 1997; SCHMILOVITCH et al., 2000; SARANWONG et al., 2001; SARANWONG et al., 2003; MAHAYOTHEE et al., 2004; SARANWONG et al., 2004; SUBEDI et al., 2007; DELWICHE et al., 2008; VALENTE et al., 2009; BETEMPS et al., 2011; SUBEDI & WALSH, 2011; JHA et al., 2012; JHA et al., 2014; MARQUES et al., 2016). Além do potencial para a determinação do estágio de maturação e qualidade dos frutos, a espectroscopia NIR também tem sido sugerida como uma técnica para prever a qualidade de consumo e tempo útil pós-colheita de mangas a partir de análises realizadas no momento da colheita (SARANWONG et al., 2004; SUBEDI et al., 2007; NASSUR, 2013).

Recentemente, alguns estudos têm demonstrado que espectrômetros NIR portáteis podem ser utilizados adequadamente para a determinação de parâmetros de qualidade de mangas, resultando em desempenho similar aos espectrômetros NIR de bancada para este tipo de aplicação (SARANWONG et al., 2003; BETEMPS et al., 2011; JHA et al., 2012; JHA et al., 2014; MARQUES et al., 2016). Nos últimos anos, o desenvolvimento de novas tecnologias utilizadas na construção dos espectrômetros NIR permitiu a redução significativa do tamanho e do custo destes equipamentos. Assim, os espectrômetros NIR portáteis representam uma alternativa atrativa e viável para as agroindústrias e pequenos produtores, já que possuem características de tamanho e desempenho compatíveis com suas necessidades, além de possuírem menores preços quando comparados aos equipamentos de bancada. Logo, esses equipamentos possuem um grande potencial para o monitoramento da qualidade de frutos na cadeia produtiva de manga.

Além dos espectrômetros NIR portáteis, as câmeras hiperespectrais (operando na região Vis/NIR ou NIR) são equipamentos cada vez mais utilizados para a avaliação da qualidade de produtos agrícolas (LU, 2008; QIN & LU, 2008; LEIVA-VALENZUELA et al., 2013; RIVERA et al., 2014; WANG et al., 2015). Esses instrumentos combinam características típicas das câmeras digitais e dos espectrômetros convencionais, permitindo a obtenção simultânea de informações espectrais e espaciais da amostra sob análise (GOWEN et al., 2007; LU, 2008).

Na área agrícola, uma possível aplicação para as imagens hiperespectrais é a detecção precoce de danos mecânicos em frutos. Os danos mecânicos, normalmente ocasionados por impacto, podem ocorrer no momento da colheita ou durante o período pós-colheita, sendo um problema frequente na cadeia produtiva de diferentes tipos de frutos (MARTÍNEZ-ROMERO et al., 2004; LI & THOMAS, 2014). A detecção e/ou quantificação precoce (pouco tempo após a ocorrência do impacto) destes danos representa um desafio para a maioria dos produtores, pois, em geral, esses tornam-se visíveis somente após determinado período de tempo, normalmente quando os frutos já se encontram na etapa de comercialização. Em geral, os danos mecânicos estão relacionados com a perda de firmeza dos frutos e representam um problema importante na cadeia produtiva de manga, principalmente para os exportadores, que necessitam atender a padrões de qualidade mais rigorosos (HAHN, 2004; RIVERA et al., 2014).

Na literatura, há poucos estudos abordando a detecção precoce de danos mecânicos em frutas utilizando imagens hiperespectrais. Rivera et al. (2014) utilizaram imagens hiperespectrais para inspeção pós-colheita de mangas da variedade “Manila”

visando a detecção precoce de danos mecânicos induzidos nos frutos. Foi utilizado um sistema de aquisição de imagens operando na faixa espectral entre 650 e 1100 nm. Neste estudo, os autores avaliaram o desempenho de diferentes métodos de agrupamento e classificação na tentativa de discriminar e classificar pixels pertencentes às áreas sadias e danificadas do fruto. O melhor resultado foi alcançado empregando o algoritmo *K-Nearest Neighbours*, que permitiu classificar corretamente 97,9% dos pixels. NAGATA et al. (2006) utilizaram um sistema de aquisição de imagens hiperespectrais operando na faixa espectral entre 650 e 1100 nm com o objetivo de detectar danos mecânicos na superfície de morangos. Os autores obtiveram os melhores resultados utilizando o método denominado “diferença normalizada” para realizar a discriminação dos pixels. BARANOWSKY et al. (2012) utilizaram um sistema equipado com diferentes sensores, capaz de registrar imagens na faixa espectral entre 400 e 5000 nm, visando a detecção de danos na superfície de maçãs. Neste estudo, os autores verificaram que o uso de uma faixa espectral abrangente melhorou de forma significativa a capacidade de detecção precoce dos danos mecânicos presentes nos frutos.

## **3.2 OBJETIVOS**

### **3.2.1 Objetivo Geral**

O objetivo geral deste estudo foi avaliar o potencial de um espectrômetro NIR portátil para a determinação rápida e não destrutiva de parâmetros de qualidade de mangas da variedade “Tommy Atkins”, e a possibilidade utilizar imagens hiperespectrais na região do NIR para a detecção precoce de danos mecânicos induzidos artificialmente nos frutos.

### **3.2.2 Objetivos específicos**

Os objetivos específicos foram os seguintes: (1) determinar a profundidade de penetração da radiação eletromagnética na região do NIR (entre 950 e 1650 nm) através dos tecidos epidérmicos e subepidérmicos de mangas “Tommy Atkins” utilizando o espectrômetro NIR portátil MicroNIR 1700 (VIAVI, CA, EUA); (2) desenvolver modelos de calibração multivariada para a determinação dos seguintes parâmetros de qualidade de mangas “Tommy Atkins”: sólidos solúveis (SS), matéria seca (MS), acidez titulável (AT)

e firmeza de polpa (FP); (3) utilizar os modelos de calibração multivariada para determinar de forma não destrutiva e individual a qualidade dos frutos durante o amadurecimento; (4) determinar a profundidade de penetração da radiação eletromagnética na região do NIR (entre 930 e 2500 nm) através dos tecidos epidérmicos e subepidérmicos de mangas “Tommy Atkins” utilizando a câmera hiperespectral SisuCHEMA SWIR (Specim Spectral Imaging Ltd, Oulu, Finlândia); e (5) avaliar a possibilidade de realizar a detecção precoce de danos mecânicos induzidos artificialmente em mangas “Tommy Atkins” por meio de imagens hiperespectrais na região do NIR.

### 3.3 MATERIAL E MÉTODOS

Este capítulo foi dividido em duas partes, que se referem diretamente aos objetivos gerais do trabalho, apresentados previamente no item “3.2.1 – Objetivos Gerais”.

#### 3.3.1 Parte 1: avaliação da qualidade de mangas “Tommy Atkins” utilizando um espectrômetro NIR portátil

##### 3.3.1.1 Amostras

Foram colhidas 400 mangas da variedade “Tommy Atkins”, nos estádios de maturação entre 1 e 2 (SANTOS et al., 2008) (Figura 10), em um pomar comercial localizado em Juazeiro-BA (latitude 9°24’S, longitude 40°21’W) em março de 2014. Em seguida, os frutos foram transportados para o Laboratório de Pós-colheita da Embrapa Semiárido, localizada em Petrolina-PE, onde foram pré-selecionados, lavados, secos e armazenados em temperatura de 24°C ( $\pm 1^\circ\text{C}$ ) durante 18 dias.



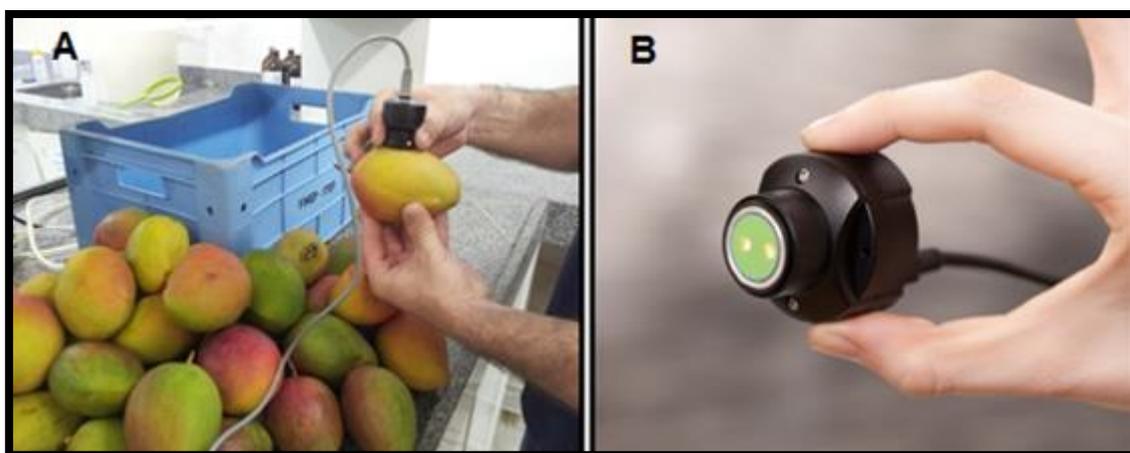
**Figura 10** – Escala visual utilizada para classificação dos estádios de maturação de mangas da variedade “Tommy Atkins” (SANTOS et al., 2008).

A seleção prévia dos frutos foi realizada visando a eliminação de frutos com características não adequadas para a realização dos experimentos. Foram eliminadas amostras que apresentaram formato irregular, tamanho muito abaixo ou acima da média, danos físicos aparentes e indícios de doenças. Todos os frutos utilizados nos experimentos descritos na “Parte 1” do presente trabalho (“AVALIAÇÃO DA QUALIDADE DE MANGAS “TOMMY ATKINS” UTILIZANDO UM ESPECTRÔMETRO NIR PORTÁTIL”) foram provenientes deste processo de seleção.

### 3.3.1.2 Instrumentação e aquisição espectral

Os espectros foram registrados com o espectrômetro portátil MicroNIR 1700 (VIAVI, Santa Rosa, CA, EUA) (Figura 11) em ambiente com temperatura controlada em 24°C ( $\pm 1^\circ\text{C}$ ). Esse espectrômetro possui um filtro linear variável como elemento dispersivo, o qual é posicionado diretamente sobre o detector, que é um arranjo linear de fotodiodos de arseneto de gálio e índio (InGaAs). O equipamento possui duas pequenas lâmpadas de tungstênio, interface USB para alimentação e transferência de dados, e dimensões bastante reduzidas: 45 mm de diâmetro, 42 mm de altura, pesando cerca de 60 g.

A aquisição dos espectros foi realizada no modo de refletância ( $\log 1/R$ ), na faixa espectral entre 950 e 1650 nm em intervalos de 6,25 nm. O tempo de integração foi de 10 ms, sendo que cada espectro registrado correspondeu a média de 50 varreduras, resultando em um tempo total de medida de 0,5 s.



**Figura 11** – (A) Mangas da variedade “Tommy Atkins” sendo avaliadas com o

espectrômetro MicroNIR 1700 (VIAVI, Santa Rosa, CA, EUA); e (B) detalhe mostrando as dimensões reduzidas deste equipamento.

### 3.3.1.3 Penetração da radiação na região do NIR através dos tecidos epidérmicos e subepidérmicos de mangas “Tommy Atkins” utilizando o espectrômetro NIR portátil MicroNIR 1700

Os frutos foram cortados ao meio e os caroços descartados, resultando em duas fatias espessas para cada fruto. Os espectros foram registrados posicionando-se o espectrômetro diretamente sobre a casca de uma das fatias e um suporte de polietileno do lado oposto. Inicialmente, foram registrados dois espectros da fatia inteira (Figura 12). Em seguida, utilizou-se uma faca para remover uma fatia de aproximadamente 1 mm do lado oposto ao da casca (mesocarpo do fruto) e novamente dois espectros foram registrados (sobre a casca). Esse procedimento foi repetido diversas vezes, resultando em fatias de manga com espessuras variando de 31,0 até 0,7 mm. A espessura de cada fatia foi determinada por meio de um paquímetro digital (King-Tools, China) antes de cada aquisição espectral. Os cálculos foram realizados utilizando-se os espectros médios obtidos para cada fatia.

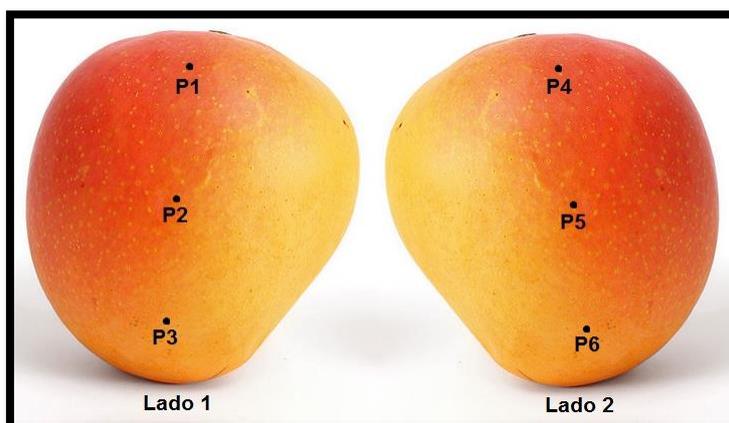


**Figura 12** – Aquisição de espectros de uma fatia inteira de manga (~ 31 mm). As medidas foram realizadas posicionando-se o espectrômetro diretamente sobre a casca de uma da fatia e um suporte de polietileno do lado oposto

### 3.3.1.4 Desenvolvimento de modelos de calibração multivariada para determinação de parâmetros de qualidade de manga

Um total de 250 mangas “Tommy Atkins” foram divididas aleatoriamente em 10 grupos de 25 frutos. A aquisição dos espectros e as análises de referência foram realizadas no dia da colheita e depois a cada dois dias, utilizando-se grupos de frutos distintos em cada dia de avaliação. Os frutos foram armazenados durante 18 dias sob temperatura de 24°C ( $\pm 1^\circ\text{C}$ ).

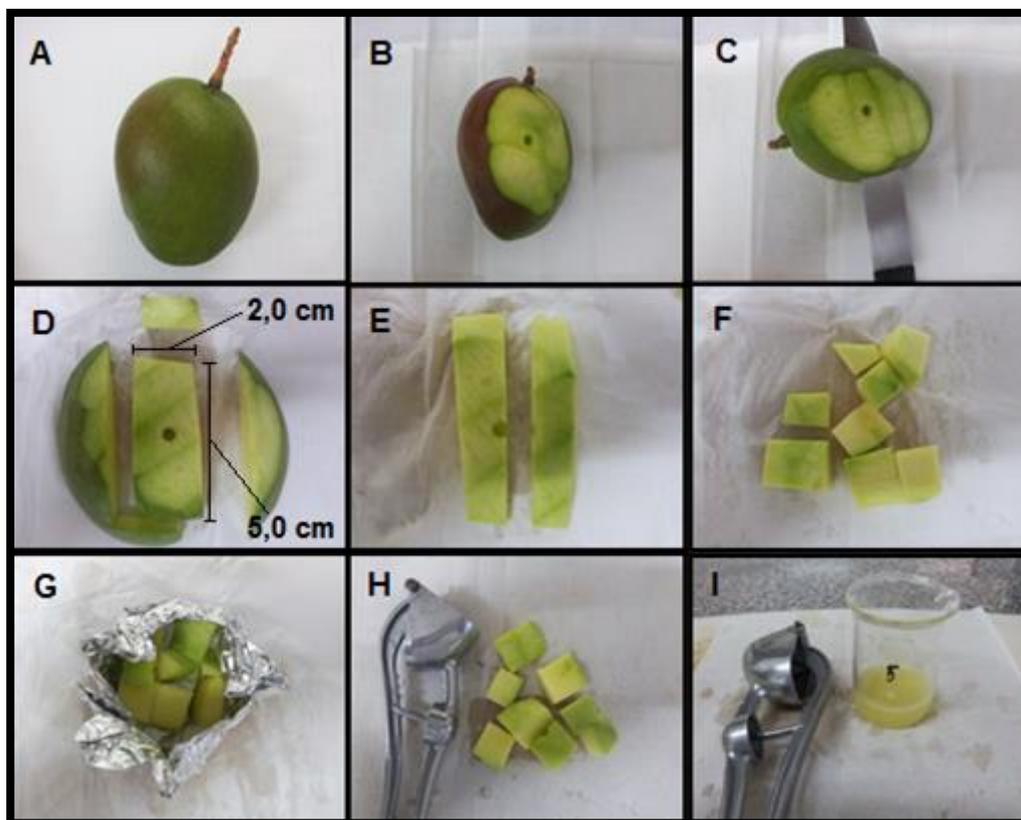
Os espectros foram registrados em seis pontos distribuídos ao longo da região axial, sendo três pontos localizados em cada face do fruto (Figura 13). As medidas foram realizadas posicionando-se o espectrômetro diretamente sobre a casca dos frutos. Os modelos de calibração foram desenvolvidos utilizando-se os espectros médios de cada fruto.



**Figura 13** – Pontos (P1 a P6) distribuídos ao longo da região axial de faces opostas do fruto, utilizados para aquisição dos espectros na região do NIR.

As análises de referência foram realizadas empregando-se as regiões dos frutos correspondentes às áreas previamente utilizadas para a aquisição dos espectros. Inicialmente, uma pequena área ( $\sim 1 \text{ cm}^2$ ) da epiderme das duas faces de cada fruto foi removida para determinação da firmeza de polpa (FP). Em seguida, uma amostra com as dimensões aproximadas de 5 x 2 x 1 cm (comprimento x largura x profundidade) foi retirada de cada uma das faces de cada fruto. Essas amostras foram cortadas em pequenos pedaços, misturadas e divididas em duas frações: (1) utilizada para determinação de

matéria seca; e (2) utilizada para determinação de sólidos solúveis (SS) e acidez titulável (AT) empregando os métodos de referência. A fração 2 foi prensada por meio de um extrator de sucos manual e a fração líquida resultante foi utilizada para as análises. Na Figura 14 são mostradas as etapas de preparo das amostras para as análises de referência.



**Figura 14** – Etapas de preparo das amostras para as análises de referência: (A) fruto inteiro; (B) fruto com parte da epiderme removida, utilizado para análise de FP; (C) corte e extração de uma das faces do fruto; (D) separação de amostra de uma das faces do fruto, com dimensões aproximadas de 5 x 2 x 1 cm (comprimento x largura x profundidade); (E) amostras extraídas das duas faces opostas do fruto; (F) amostra representativa do fruto, composta por pequenos pedaços provenientes das duas faces; (G) fração da amostra utilizada para determinação de MS; (H) fração da amostra utilizada para prensagem; e (I) fração líquida resultante da prensagem, utilizada para determinação de SS e AT.

#### 3.3.1.4.1 Análises de referência

##### *Determinação de sólidos solúveis (SS)*

A determinação de SS foi realizada por meio de um refratômetro digital portátil (Atago, Tóquio, Japão) com compensação automática de temperatura e faixa de operação entre 0 e 53 °Brix (Figura 15). As leituras foram realizadas utilizando-se aproximadamente 1 mL de suco. Os resultados obtidos foram expressos em °Brix.



**Figura 15** – Determinação de SS empregando-se um refratômetro digital (Atago, Tóquio, Japão).

#### *Determinação de Matéria Seca (MS)*

A MS foi determinada pela diferença entre a massa da amostra úmida e seca. As amostras foram secas em estufa, em temperatura de 65 °C, até que atingissem massa constante (AOAC, 1995). Os resultados obtidos foram expressos como porcentagem de MS.

#### *Determinação de acidez titulável (AT)*

A determinação da AT foi realizada por meio da titulação de 1 mL de suco com uma solução de NaOH 0,1 mol L<sup>-1</sup>, utilizando-se uma bureta digital (modelo Digitrate Pro, Jencons, Franklin, KY, EUA) (Figura 16). Uma solução de fenolftaleína foi utilizada como indicador do ponto de viragem (AOAC, 1995). As determinações foram realizadas em duas repetições e os valores médios foram utilizados para a construção dos modelos de calibração. Os resultados obtidos foram expressos como porcentagem (m/v) de ácido cítrico.



**Figura 16** – Determinação de AT empregando-se uma bureta digital (modelo Digitrator Pro, Jencons, Franklin, KY, EUA).

*Determinação de firmeza de polpa (FP)*

A determinação da FP foi realizada utilizando-se um texturômetro digital (modelo TA.XT Plus, Extralab, Jarinu, SP, Brasil) equipado com ponteira de 6 mm de diâmetro (Figura 17). Este parâmetro foi determinado nos frutos inteiros e os resultados representam a força necessária para efetuar 10 mm de penetração da ponteira através da polpa do fruto. Foram realizadas duas medidas por fruto, em lados opostos na região equatorial, após a remoção de uma pequena área da epiderme ( $\sim 1 \text{ cm}^2$ ), e os valores médios foram utilizados para a construção dos modelos de calibração. Os resultados obtidos foram expressos em Newtons (N).



**Figura 17** – Determinação de FP empregando-se um texturômetro digital (modelo TA.XT Plus, Extralab, Jarinu, SP, Brasil).

#### 3.3.1.5 Monitoramento não destrutivo da qualidade dos frutos durante o amadurecimento.

O amadurecimento dos frutos foi monitorado de forma não destrutiva por meio do espectrômetro MicroNIR 1700. Um total de 22 mangas da variedade “Tommy Atkins” foram armazenadas em temperatura de 24°C ( $\pm 1^\circ\text{C}$ ) durante 14 dias. Os espectros dos frutos foram registrados no momento da colheita e duas vezes ao dia durante o período de armazenamento, utilizando o mesmo procedimento descrito anteriormente (item 3.3.1.2). A cada avaliação, foram registrados os espectros de todos os frutos, permitindo o monitoramento de mudanças físico-químicas resultantes do processo de amadurecimento em cada fruto individualmente.

#### 3.3.1.6 Análise de dados e software

Inicialmente, diferentes técnicas de pré-processamento espectral foram avaliadas: SNV, primeira e segunda derivadas com filtro de Savitzky-Golay utilizando polinômio de 2ª ordem e janela de diferentes tamanhos, MSC e a combinação de SNV e derivadas utilizando filtro de Savitzky-Golay. As amostras foram divididas nos conjuntos de

calibração e validação externa utilizando-se o algoritmo SPXY (GALVÃO et al., 2005) para cada parâmetro de qualidade avaliado.

Os modelos de calibração foram construídos utilizando-se regressão PLS e duas estratégias diferentes: (1) utilizando todas as variáveis contidas na faixa espectral de 950 a 1650 nm (PLS/Full); e (2) utilizando somente as variáveis selecionadas pelo algoritmo *Jack-Knife* (PLS/JK). O número de variáveis latentes mantidas nos modelos de calibração foi determinado de acordo com o valor de RMSECV, utilizando-se validação cruzada completa. O desempenho preditivo dos modelos foi avaliado na etapa de validação externa (predição), considerando-se os valores de RMSEP,  $R^2$  e bias.

Os cálculos relacionados às etapas de pré-processamento, seleção de amostras, seleção de variáveis e construção dos modelos de calibração multivariados foram realizados utilizando-se os softwares The Unscrambler versão 9.8 (CAMO, Oslo, Noruega) e MATLAB versão 2010a (Mathworks, Natick, MA, EUA).

### **3.3.2 Parte 2: identificação de danos mecânicos em mangas “Tommy Atkins” utilizando imagens hiperespectrais na região do NIR**

#### **3.3.2.1 Amostras**

Foram colhidas 40 mangas da variedade “Tommy Atkins”, nos estádios de maturação entre 1 e 4 (SANTOS et al., 2008) (Figura 3.1), em um pomar comercial localizado em Juazeiro-BA (latitude 9°24’S, longitude 40°21’W) em outubro de 2015. Os frutos foram transportados para o Laboratório de Pós-colheita da Embrapa Semiárido, localizado em Petrolina-PE, onde foram lavados, secos e acondicionados em caixas de papelão apropriadas para transporte. Em seguida, foram enviados por via aérea para o laboratório do Grupo de Instrumentação e Automação (GIA) da UNICAMP, localizado em Campinas-SP, onde permaneceram pelo período de 5 dias em temperatura de 22°C ( $\pm 1^\circ\text{C}$ ).

Antes de iniciar os experimentos, realizou-se uma seleção prévia dos frutos conforme descrito anteriormente no item “3.3.1.1 – Amostras”. Assim, todos os frutos utilizados nos experimentos descritos na “Parte 2” do presente trabalho (“IDENTIFICAÇÃO DE DANOS MECÂNICOS EM MANGAS “TOMMY ATKINS” UTILIZANDO IMAGENS HIPERESPECTRAIS NA REGIÃO DO NIR”) foram provenientes deste processo de seleção.

### 3.3.2.2 Instrumentação e aquisição das imagens

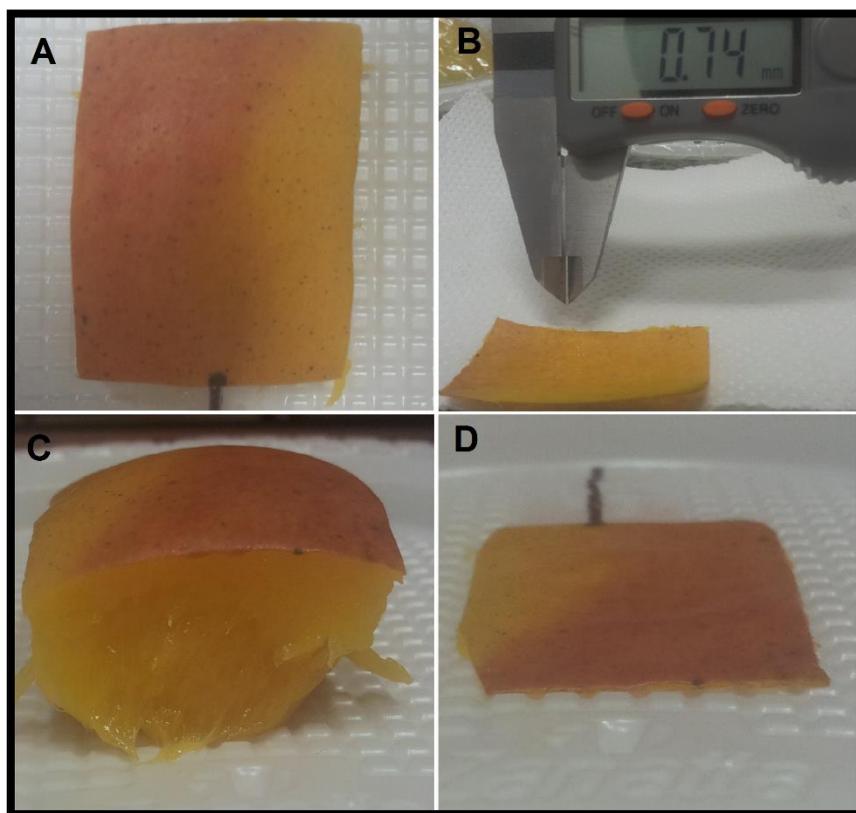
As imagens hiperespectrais foram obtidas no laboratório do Grupo de Instrumentação e Automação da UNICAMP utilizando o sistema de aquisição de imagens SisuCHEMA SWIR (Specim Spectral Imaging Ltd., Oulu, Finlândia) (Figura 18). A aquisição espectral das imagens foi realizada na faixa de 930-2500 nm em intervalos de 6,3 nm. A lente utilizada para a aquisição das imagens foi de 200 mm e o tamanho dos pixels de  $600 \mu\text{m}^2$ . A temperatura foi mantida em  $22^\circ\text{C} (\pm 1^\circ\text{C})$  durante o registro das imagens.



**Figura 18** – Sistema de aquisição de imagens SisuCHEMA SWIR (Specim Spectral Imaging Ltd., Oulu, Finlândia).

3.3.2.3 Penetração da radiação na região do NIR através dos tecidos epidérmicos e subepidérmicos de mangas “Tommy Atkins” utilizando a câmera hiperespectral SisuCHEMA SWIR

Inicialmente os frutos foram cortados ao meio e os caroços descartados, resultando em duas fatias espessas para cada fruto. Cada fatia foi transformada em uma amostra de formato retangular com as dimensões aproximadas de 50 x 30 x 20 mm (comprimento x largura x profundidade), utilizando uma faca para realizar o corte do fruto e a extração de parte da polpa. A amostra foi posicionada sobre um suporte plástico, mantendo-se a região da polpa (mesocarpo do fruto) para baixo e registrando as imagens a partir da casca do fruto (Figura 19-A). Foram obtidas imagens das amostras posicionadas sobre dois suportes distintos: um constituído de poliestireno (PS) e outro constituído de polietileno de alta densidade (PEAD). Após cada aquisição de imagem, utilizou-se uma faca para remover uma fatia do lado oposto ao da casca (mesocarpo do fruto) e, em seguida, uma nova imagem foi registrada (sobre a casca). A espessura das fatias removidas variou de 4,00 a 10,00 mm. Esse procedimento foi repetido várias vezes, resultando em diversas amostras de manga com espessuras variando de aproximadamente 21,00 até 0,60 mm (Figura 19-C e D). A espessura de cada amostra foi determinada por meio de um paquímetro digital (King-Tools, China) antes do registro de cada imagem (Figura 19-B).



**Figura 19** – (A) Amostra de manga com dimensões aproximadas de 50 x 30 x 20 mm (comprimento x largura x profundidade), posicionada sobre um suporte de poliestireno

(PS); (B) determinação da espessura da amostra por meio de um paquímetro digital (King-Tools, China) antes do registro da imagem; (C) amostra de manga com espessura de aproximadamente 21,00 mm; e (D) amostra de manga com espessura de aproximadamente 0,60 mm.

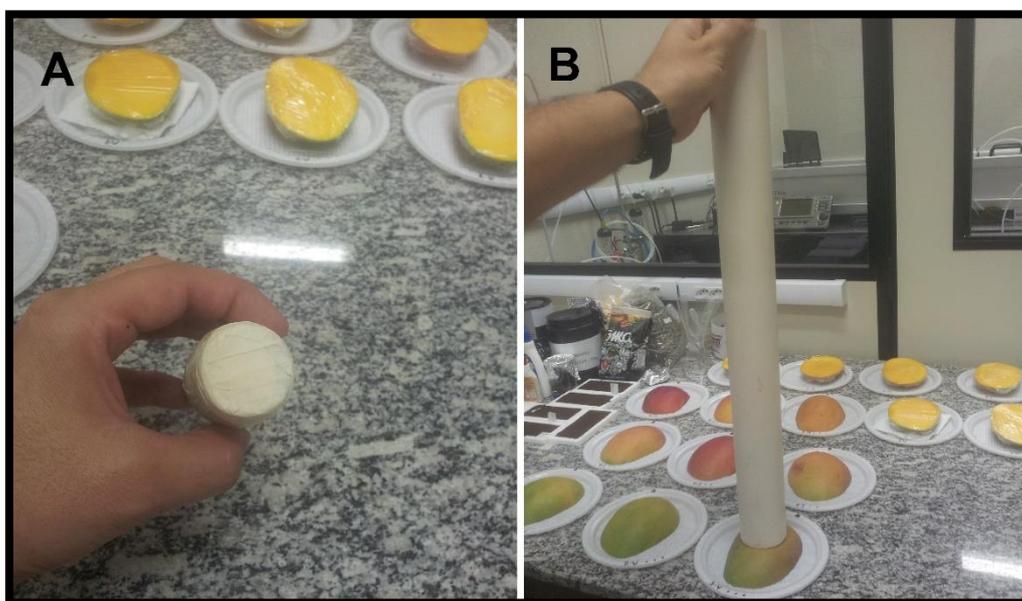
#### 3.3.2.4 Detecção precoce de danos mecânicos induzidos artificialmente em mangas “Tommy Atkins” por meio de imagens hiperespectrais na região do NIR

Neste experimento foram utilizados nove frutos em diferentes estádios de maturação (Figura 20), sendo três frutos em estágio de maturação pouco avançado, que foram chamados de “frutos verdes” (FV); três frutos em estágio de maturação intermediário, que foram chamados de “frutos intermediários” (FI); e três frutos em estágio de maturação avançado, que foram chamados de “frutos maduros” (FM).



**Figura 20** – Frutos em diferentes estádios de maturação utilizados no experimento para detecção precoce de danos mecânicos.

Em função da dimensão dos frutos, não foi possível analisá-los inteiros na câmera, sendo necessário cortá-los na metade antes da obtenção das imagens. Assim, apenas uma metade de cada fruto foi utilizada neste experimento. Os danos mecânicos foram induzidos por meio do impacto de uma peça cilíndrica sobre os frutos. Para tal, uma peça com dimensões de 9,5 x 3,0 cm (comprimento x diâmetro) e massa igual a 130 g (Figura 21-A) foi solta de uma altura de 60 cm. Foi utilizado um cano de PVC com as dimensões de 60 x 4 cm (comprimento x diâmetro) para direcionar a peça cilíndrica durante a queda para o local determinado para o impacto (centro da amostra) (Figura 21-B). As imagens foram registradas em 8 momentos distintos: antes do impacto, imediatamente após o impacto e depois de 1, 2, 4, 24, 48 e 72 h após o impacto.



**Figura 21** – (A) Peça cilíndrica com dimensões de 9,5 x 3,0 cm (comprimento x diâmetro) e massa igual a 130 g utilizada para induzir os danos mecânicos nos frutos; e (B) cano de PVC com dimensões de 60 x 4 cm (comprimento x diâmetro) utilizado para direcionar a peça cilíndrica durante a queda para o local determinado para o impacto (centro da amostra).

### 3.3.2.5 Análise de dados e software

Inicialmente, as imagens registradas foram submetidas a uma etapa prévia de pré-processamento visando a remoção de *dead pixels* e *spikes* e seleção das regiões de interesse, conforme recomendado por Vidal & Amigo (2012). Para correção de efeitos de

espalhamento e desvios de linha de base foi realizado o pré-processamento espectral empregando-se as seguintes técnicas: SNV, primeira derivada por filtro de Savitzky-Golay (polinômio de 2ª ordem e janela 15 pontos), e segunda derivada por filtro de Savitzky-Golay (polinômio de 2ª ordem e janela 15 pontos). Esta etapa de pré-processamento das imagens é essencial para garantir a aplicação adequada dos métodos de calibração e/ou classificação multivariada que serão empregados posteriormente (VIDAL & AMIGO, 2012).

A modelagem dos dados do experimento referente à penetração da radiação através dos tecidos epidérmicos e subepidérmicos de mangas foi baseada na análise direta das imagens e espectros obtidos, além da realização de PCA. A modelagem dos dados referentes ao experimento para detecção precoce de danos mecânicos em mangas foi baseada na realização de PCA e na técnica de agrupamento *K-means*.

Os cálculos relacionados às etapas de pré-processamento, tratamento das imagens e modelagem dos dados foram realizados utilizando-se o software MATLAB versão 2010a (Mathworks, Natick, MA, EUA) e a interface HYPER-Tools toolbox para ambiente MATLAB (disponível em <http://models.life.ku.dk/HYPERTools>).

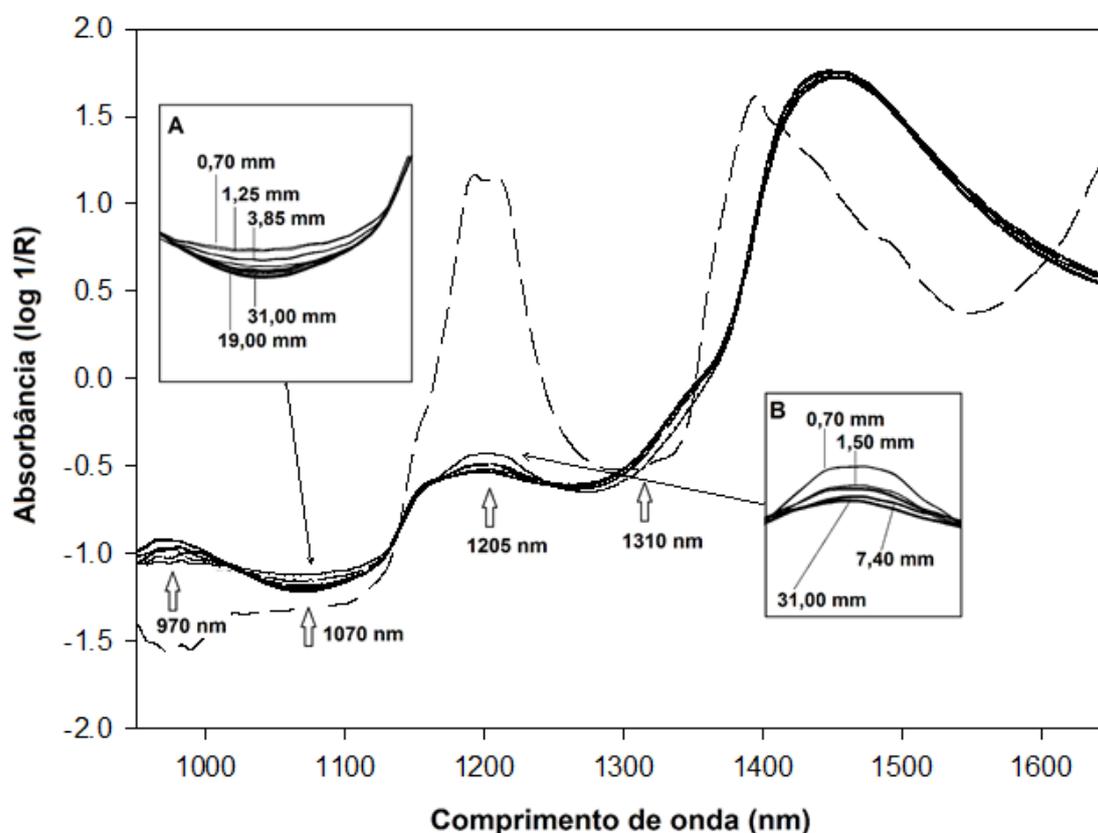
### 3.4 RESULTADOS E DISCUSSÃO

#### **3.4.1 Parte 1: avaliação da qualidade de mangas “Tommy Atkins” utilizando um espectrômetro NIR portátil**

##### 3.4.1.1 Penetração da radiação na região do NIR através dos tecidos epidérmicos e subepidérmicos de mangas “Tommy Atkins” utilizando o espectrômetro NIR portátil MicroNIR 1700

O conhecimento da profundidade de penetração da radiação NIR através dos tecidos epidérmicos e subepidérmicos das mangas é relevante, pois permite estimar o volume do fruto efetivamente amostrado pelo equipamento, que afeta diretamente o espectro de reflectância obtido. Neste sentido, este estudo contribuiu para melhorar o conhecimento a respeito da penetração da radiação NIR em mangas, fornecendo informações relevantes que foram utilizadas na etapa de amostragem do tecido dos frutos para as análises de referência.

Na Figura 22 é possível observar que as regiões espectrais próximas a 970, 1070, 1205 e 1310 nm (destacadas no gráfico) exibem uma variabilidade maior que outras regiões espectrais em função das diferentes espessuras das fatias de manga avaliadas. Nas regiões próximas à 1070 e 1205 nm, pode-se verificar que os espectros provenientes das fatias de menor espessura estão localizados na parte superior do gráfico, enquanto os espectros situados na parte inferior do gráfico representam as fatias mais espessas (Figura 22-A e B). Este fato ocorre, pois, as fatias de menor espessura permitem que uma fração significativa da radiação incidente seja transmitida através do tecido do fruto, resultando em maior quantidade de radiação atingindo o suporte de polietileno situado do lado oposto da fatia, em comparação às fatias de maior espessura. Uma vez que a radiação refletida retorna ao detector do equipamento trazendo informações físicas e químicas dos frutos, a redução da espessura dessas fatias resulta em um incremento das informações espectrais referentes ao suporte de polietileno, o qual apresentou bandas características na região próxima à 970, 1205 e 1310 nm (Figura 22).



**Figura 22** – Espectros pré-processados por SNV, de fatias de manga com diferentes espessuras (linhas sólidas) e do suporte de polietileno (linha tracejada). As regiões

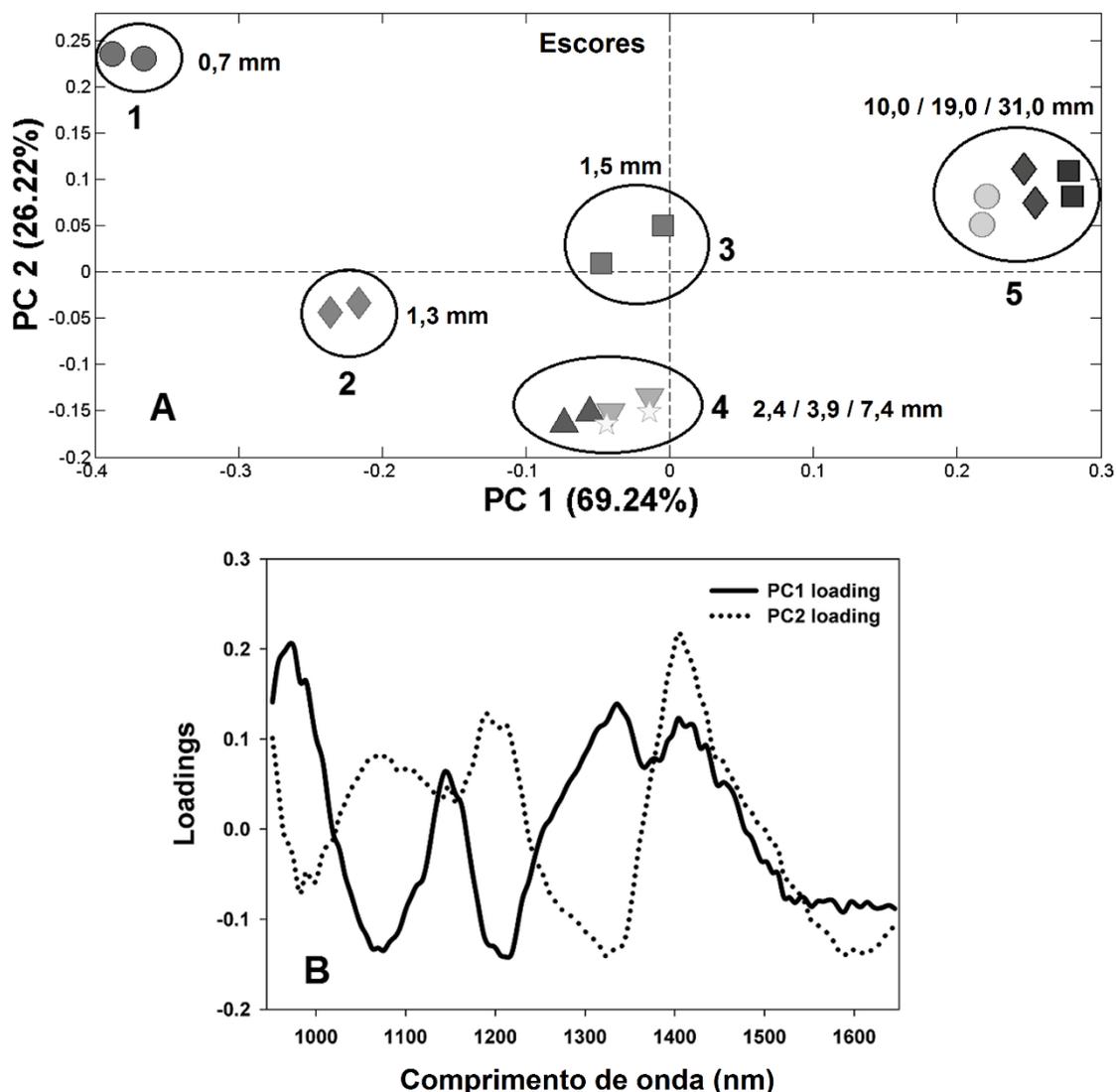
espectrais próximas a 1070 e 1205 nm foram destacadas nos quadros A e B para melhorar a visualização dos espectros.

Os dados espectrais provenientes de todas as fatias de manga foram pré-processados por SNV e submetidos a uma PCA, a fim de se verificar o padrão de distribuição natural destas amostras. A partir do gráfico de escores, pode-se verificar que a primeira e a segunda componente (PC1 e PC2) descrevem juntas 95,5% da variância dos dados (Figura 23-A). A PCA evidenciou a segregação natural das amostras em função de suas espessuras ao longo da PC1, e os *loadings* para esta componente mostraram que as regiões próximas a 970, 1140, 1205, 1330 e 1400 foram as mais importantes para explicar a variância dos dados (Figura 23-B). É possível observar no gráfico de *loadings* para PC2 que as principais regiões espectrais relacionadas à distribuição dos escores nesta componente correspondem a 950, 1070, 1205 e 1440 nm (Figura 23-B), evidenciando a contribuição significativa desta última banda de absorção, que é atribuída à presença água.

A profundidade de penetração da radiação foi definida como a espessura da fatia, cujo espectro foi significativamente diferente daquele obtido a partir da fatia de maior espessura (~ 31 mm), considerando que a fatia mais espessa não permite a transmissão da radiação incidente. Assim, utilizando o gráfico de escores de PC1 *versus* PC2 foi possível inferir que as fatias cujos escores foram agrupados junto com os escores da fatia de maior espessura (~ 31 mm) devem apresentar um comportamento similar em relação à penetração da radiação. Portanto, as fatias com 10,0 e 19,0 mm de espessura não devem permitir que a radiação seja transmitida (Figura 23-A). Em contrapartida, os escores provenientes das fatias com 7,4 mm foram agrupados juntamente com os escores das fatias de menor espessura, evidenciando que as fatias com 7,4 mm de espessura permitem a transmissão parcial da radiação incidente (Figura 23-A). Logo, conclui-se que a radiação NIR pode penetrar até 7,4 mm através do tecido do fruto, mas não além de 10,0 mm.

O método univariado proposto por Lammertyn et al. (2000) foi também utilizado para se calcular a profundidade de penetração da radiação. Este método baseia-se na construção de um gráfico relacionando a espessura das diferentes fatias com a razão entre as absorbâncias da fatia de referência (de maior espessura) e das demais fatias em um comprimento de onda específico, que é usado para calcular a profundidade de penetração da radiação no comprimento de onda em questão. Assim, avaliou-se a profundidade de

penetração da radiação NIR nos comprimentos de onda de 1070, 1205 e 1310 nm. Os resultados indicaram que a radiação pode alcançar até 7,1, 7,7 e 2,2 mm para cada comprimento de onda avaliado, respectivamente (dados não apresentados), os quais são similares aos valores determinados por meio da PCA.

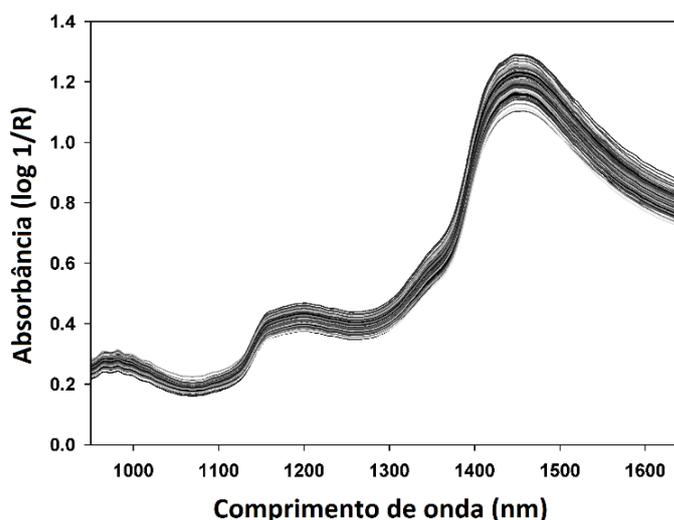


**Figura 23** – (A) Gráfico de escores em PC1 vs PC2 mostrando a separação de grupos de fatias de manga em função de suas espessuras, correspondendo aos números 1, 2, 3, 4 e 5; e (B) gráfico de *loadings* para PC1 vs PC2 mostrando as principais regiões espectrais relacionadas à distribuição dos escores nestas componentes.

3.4.1.2 Desenvolvimento de modelos de calibração multivariada para determinação de parâmetros de qualidade de manga

### 3.4.1.2.1 Atribuição de bandas espectrais

Em geral, frutas e vegetais apresentam em suas composições cerca de 80 a 90% de água, sendo a fração restante constituída por diferentes compostos, tais como carboidratos, proteínas, ácidos orgânicos e outros compostos secundários (GÓMEZ, et al., 2006). Assim, seus espectros NIR apresentam bandas de absorção largas e complexas, resultantes da sobreposição de bandas correspondentes à sobretons e combinações de modos vibracionais fundamentais, principalmente atribuídas às ligações O-H, C-H, N-H e S-H (GOLIC, et al., 2003; NICOLAÏ et al., 2009; MAGWAZA et al., 2012). Estas características podem ser visualizadas na Figura 24, onde são mostrados espectros de absorção (sem pré-processamento) provenientes de 250 mangas (“Tommy Atkins”) intactas, na faixa entre 950 e 1650 nm.



**Figura 24** – Espectros de absorbância (sem pré-processamento) provenientes de 250 mangas (“Tommy Atkins”) intactas, na faixa entre 950 e 1650 nm.

A água apresenta bandas de absorção centradas em aproximadamente 970, 1450 e 1940 (WILLIAMS & NORRIS, 2001). Na Figura 24 é possível visualizar as bandas localizadas em 970 e 1450 nm, que correspondem ao segundo e primeiro sobretons das vibrações de estiramento de ligações O-H, respectivamente.

O amido e os açúcares encontrados nas frutas, tais como sacarose, glicose e frutose, apresentam bandas de absorção de difícil visualização, pois estão muito próximas das regiões onde a água apresenta forte absorção (Figura 24) (DELWICHE et al., 2008). Em geral, estas bandas são provenientes do segundo (920 nm) e terceiro (720 nm)

sobretons de estiramento de ligações O-H, e terceiro (910 nm) e quarto (750 nm) sobretons de estiramento de ligações C-H (GOLIC et al., 2003; SUBEDI et al., 2007). Adicionalmente, a banda de absorção centralizada em 1190 nm corresponde ao segundo sobreton de estiramento de ligações C-H e também está associada à presença de açúcares (OSBO RNE et al., 1993).

Os ácidos orgânicos normalmente apresentam bandas relacionadas ao grupo O-H. Estas bandas estão centralizadas em 1445 nm, 1000 nm e 800 nm e correspondem ao primeiro, segundo e terceiro sobretons, respectivamente (WEYER & LO, 2002). Neste caso, estas bandas também se encontram sobrepostas com as principais bandas da água (Figura 24). Adicionalmente, os ácidos orgânicos exibem uma banda característica centrada em 1890 nm, que está associada à combinação entre o estiramento de ligações O-H e C=O.

#### 3.4.1.2.2 Desenvolvimento dos modelos de calibração

As amostras utilizadas neste estudo apresentaram uma variação significativa nos valores dos parâmetros de qualidade avaliados, que pode ser atribuída aos diferentes estádios de maturação em que os frutos se encontravam durante o período de realização das análises. Os parâmetros estatísticos referentes aos dados de referência são apresentados na Tabela 2 a seguir.

**Tabela 2** – Parâmetros estatísticos relativos aos dados de referência.

Parâmetro de qualidade	<i>n</i> <sup>a</sup>	Faixa	Média	DP <sup>b</sup>	CV <sup>c</sup>
Sólidos Solúveis (°Brix)	250	5,9 – 14,1	10,0	2,1	21%
Matéria Seca (%)	249	9,2 – 16,2	13,3	1,2	9%
Acidez titulável (% ácido cítrico)	250	0,04 – 1,39	0,60	0,40	67%
Firmeza de polpa (N)	250	1,9 – 115,7	39,4	36,9	94%

<sup>a</sup> Número de Amostras utilizadas; <sup>b</sup> Desvio padrão; <sup>c</sup> Coeficiente de variação.

Os modelos de calibração para todos os parâmetros de qualidade foram construídos empregando-se diferentes técnicas de pré-processamento, visando melhorar suas capacidades preditivas. Neste sentido, os melhores resultados, avaliados por meio dos valores de RMSECV foram obtidos empregando-se SNV. Assim, somente os

resultados obtidos aplicando-se esta técnica de pré-processamento são apresentados neste capítulo. Adicionalmente, os modelos foram construídos utilizando-se duas estratégias distintas: (1) utilizando todas as variáveis contidas na faixa espectral de 950 a 1650 nm (PLS/Full); e (2) utilizando somente as variáveis selecionadas pelo algoritmo *Jack-Knife* (PLS/JK). O desempenho preditivo de todos os modelos desenvolvidos utilizando a estratégia PLS/JK foi equivalente àquele obtido com a estratégia PLS/Full, de acordo com o teste F ao nível de 95% de confiança aplicado aos valores de RMSEP (Tabela 3). Portanto, todos os gráficos apresentados neste capítulo são referentes à estratégia de modelagem PLS/Full. Em relação ao parâmetro bias, não foram observados valores significativos nas etapas de calibração e validação externa, não havendo evidências de erros sistemáticos nos modelos desenvolvidos.

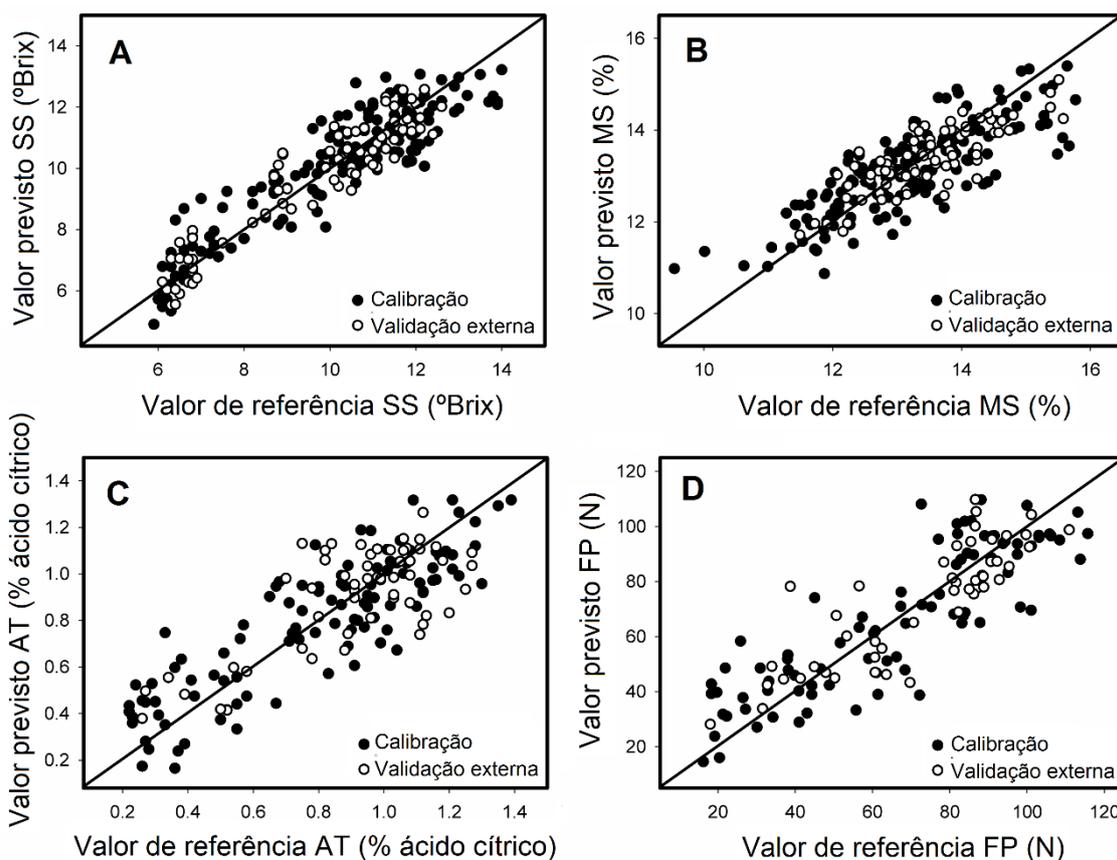
Para SS, o gráfico contendo os valores obtidos pelo método de referência *versus* os valores previstos pelo modelo de calibração é mostrado na Figura 25-A. Os resultados obtidos estão de acordo com outros estudos que reportaram alta exatidão dos modelos de calibração para determinação de SS em manga (SCHMILOVITCH et al., 2000; SARANWONG et al., 2001; DELWICHE et al., 2008; SUBEDI & WALSH, 2011; JHA et al., 2012). Schmilovitch et al. (2000) utilizaram um espectrômetro NIR de bancada, operando na faixa espectral entre 1200 e 2400 nm, para analisar mangas da variedade “Tommy Atkins” contendo SS na faixa entre 7,0 e 24 °Brix. O modelo de calibração desenvolvido empregando-se regressão PLS forneceu valores de 1,45 °Brix e 0,90 para SEP e  $R^2$ , respectivamente, na etapa de validação externa. Resultados um pouco melhores foram obtidos por Delwiche et al. (2008), que utilizaram um espectrômetro de bancada, operando na faixa espectral entre 750 e 1088 nm, para analisar mangas da variedade “Ataúfo” contendo SS na faixa entre 11,7 e 23,1 °Brix. Os resultados obtidos para SEP e  $R^2$  na etapa de validação externa foram 0,94 °Brix e 0,88, respectivamente. Portanto, a exatidão alcançada para a predição de SS no presente estudo, utilizando o espectrômetro MicroNIR 1700, foi equivalente ou superior àquela reportada em estudos prévios que fizeram uso de equipamentos de bancada para determinação de SS em manga.

**Tabela 3** – Parâmetros estatísticos relativos às etapas de calibração e validação externa para determinação de sólidos solúveis (SS), matéria seca (MS), acidez titulável (AT) e firmeza de polpa (FP) em manga “Tommy Atkins”, empregando-se SNV como pré-processamento espectral e duas diferentes faixas espectrais para construção dos modelos de calibração: (1) 950-1650 nm (PLS/Full); e (2) somente os comprimentos de onda selecionados pelo algoritmo *Jack-Knife* (PLS/JK).

Parâmetros	Faixa	Calibração					Validação externa				
		VL <sup>a</sup>	n <sup>b</sup>	R <sup>2</sup> <sub>cv</sub> <sup>c</sup>	RMSECV	Bias <sub>cv</sub> <sup>d</sup>	n <sup>e</sup>	R <sup>2</sup>	RMSEP	Bias	RMSEPr <sup>f</sup>
<b>SS (°Brix)</b>											
PLS/Full	5,9 - 14,1	8	160	0,83	0,88	0,02	78	0,88	0,66	-0,03	6,6
PLS/JK		8	160	0,86	0,78	0,02	79	0,92	0,55	0,00	5,5
<b>MS (%)</b>											
PLS/Full	9,5 - 15,8	7	160	0,61	0,74	-0,07	80	0,67	0,51	-0,07	4,0
PLS/JK		7	157	0,75	0,58	0,00	79	0,57	0,54	0,06	4,3
<b>AT (% ácido cítrico)</b>											
PLS/Full	0,22 - 1,39	6	107	0,75	0,16	0,00	55	0,50	0,17	0,01	21
PLS/JK		6	106	0,77	0,16	0,00	55	0,50	0,17	0,00	21
<b>FP (N)</b>											
PLS/Full	16,2 - 115,7	8	88	0,75	14,3	0,1	46	0,72	12,2	0,5	18
PLS/JK		8	88	0,80	12,7	0,0	46	0,63	13,7	0,7	21

<sup>a</sup> Número de variáveis latentes mantidas no modelo PLS; <sup>b</sup> Número de amostras utilizadas na etapa de calibração excluindo-se os *outliers*; <sup>c</sup> Coeficiente de determinação na etapa de validação cruzada; <sup>d</sup> Bias na etapa de validação cruzada; <sup>e</sup> Número de amostras utilizadas na etapa de validação externa excluindo-se os *outliers*; <sup>f</sup> RMSEP relativo, calculado em relação ao valor médio da faixa do parâmetro de qualidade avaliado. Valores expressos em porcentagem.

Para MS, o gráfico contendo os valores obtidos pelo método de referência *versus* os valores previstos pelo modelo de calibração é mostrado na Figura 25-B. O desempenho preditivo para este parâmetro foi similar ao reportado por Saranwong et al. (2003). Estes autores utilizaram um espectrômetro portátil, operando na faixa espectral entre 600 e 1000 nm, e obtiveram valores de SEP e  $R^2$  na etapa de validação externa de 0,52% e 0,88, respectivamente. Em outro estudo, Saranwong et al. (2004) utilizaram um equipamento de bancada, operando na faixa espectral entre 700 e 1100 nm, para analisar mangas da variedade “Mahajanaka” contendo MS na faixa entre 15,7 e 23,5%. O modelo de calibração desenvolvido empregando-se regressão MLR resultou em valores de 0,41% e 0,92 para SEP e  $R^2$ , respectivamente, na etapa de validação externa. Em relação ao baixo valor de  $R^2$  obtido no presente estudo em relação aos obtidos pelos autores citados anteriormente, é possível que a faixa mais restrita de MS encontrada nos frutos utilizados neste estudo tenha contribuído negativamente para o valor de  $R^2$ .



**Figura 25** – Valores obtidos pelo método de referência *versus* os valores previstos pelo modelo de calibração nas etapas de calibração e validação externa para os parâmetros: (A) sólidos solúveis (SS), (B) matéria seca (MS), (C) acidez titulável (AT) e (D) firmeza

de polpa (FP). A linha sólida representa a bissetriz. Os resultados das etapas de calibração e validação externa são representados por círculos preenchidos e círculos vazios, respectivamente.

Em geral, a AT é um dos parâmetros mais difíceis de serem modelados, sendo este fato atribuído ao seu baixo teor nos frutos, que normalmente é menor que 1% (FLORES et al., 2009; DE OLIVEIRA et al., 2014). De fato, a AT foi o parâmetro mais difícil de ser modelado neste estudo, necessitando da eliminação prévia de amostras contendo valores abaixo de 0,20% para este parâmetro. Os modelos de calibração apresentaram desempenho preditivo muito baixos quando valores de referência menores que 0,20% foram utilizados. Após a exclusão destas amostras foi possível obter valores de 0,17% (ácido cítrico) e 0,50 para RMSEP e  $R^2$ , respectivamente, na etapa de validação externa (Tabela 3). O gráfico contendo os valores obtidos pelo método de referência *versus* os valores previstos pelo modelo de calibração para determinação de AT é mostrado na Figura 25-C. Os resultados obtidos foram similares aos reportados em estudos prévios que abordaram a determinação de AT em manga. Schmilovitch et al. (2000) analisaram mangas da variedade “Tommy Atkins” contendo AT na faixa entre 0,15 e 0,85% e obtiveram valores de SEP e  $R^2$  iguais a 0,22% e 0,40, respectivamente. No estudo desenvolvido por Mahayothee et al. (2004) foram desenvolvidos modelos de calibração com dados provenientes de um espectrômetro Vis/NIR com transformada de Fourier, operando na faixa espectral entre 650 e 2500 nm. Para mangas da variedade “Chok Anan”, contendo AT na faixa entre 0,29 e 1,09%, os autores obtiveram valores de SEP e  $R^2$  iguais a 0,20% e 0,76, respectivamente.

A modelagem de FP também requereu a eliminação prévia de amostras que apresentaram valores abaixo de 15,0 N para este parâmetro, de maneira análoga ao que foi reportado previamente para a modelagem de AT. Somente após a eliminação destas amostras foi possível construir modelos de calibração com capacidade preditiva satisfatória. O gráfico contendo os valores obtidos pelo método de referência *versus* os valores previstos pelo modelo de calibração para determinação de FP é mostrado na Figura 25-D. Os modelos de calibração desenvolvidos para determinação de FP neste estudo apresentaram desempenho preditivo superior ao encontrado no trabalho desenvolvido por Schmilovitch et al. (2000), onde os autores utilizaram mangas com valores de FP na faixa entre 12,0 e 125,0 N e obtiveram valores de SEP e  $R^2$  iguais a 23,3 N e 0,69, respectivamente. No estudo reportado por Mahayothee et al. (2004) foram

desenvolvidos modelos de calibração para mangas da variedade “Chok Anan”, apresentando FP na faixa entre 43,0 e 250,0 N. Neste caso, os autores obtiveram valores de SEP e  $R^2$  iguais a 41,1 N e 0,85, respectivamente.

Inicialmente, os modelos de calibração desenvolvidos para determinação de FP foram construídos utilizando-se os espectros brutos, sem aplicação de qualquer técnica de pré-processamento espectral. Esta estratégia foi empregada, pois a FP é uma propriedade física da amostra. Portanto, esperava-se que a aplicação de pré-processamento espectral seria prejudicial para o desempenho preditivo dos modelos de calibração, já que poderia ocorrer a remoção de informações de natureza física contidas nos espectros. Entretanto, o desempenho preditivo dos modelos foi melhor quando SNV foi aplicada. Assim, é provável que determinação de FP tenha sido realizada indiretamente por meio de informações relativas à composição química das amostras. Embora relevante, uma investigação detalhada a respeito deste fato está fora do escopo deste capítulo e deverá ser abordada em estudos futuros.

#### 3.4.1.3 Monitoramento não destrutivo da qualidade dos frutos durante o amadurecimento.

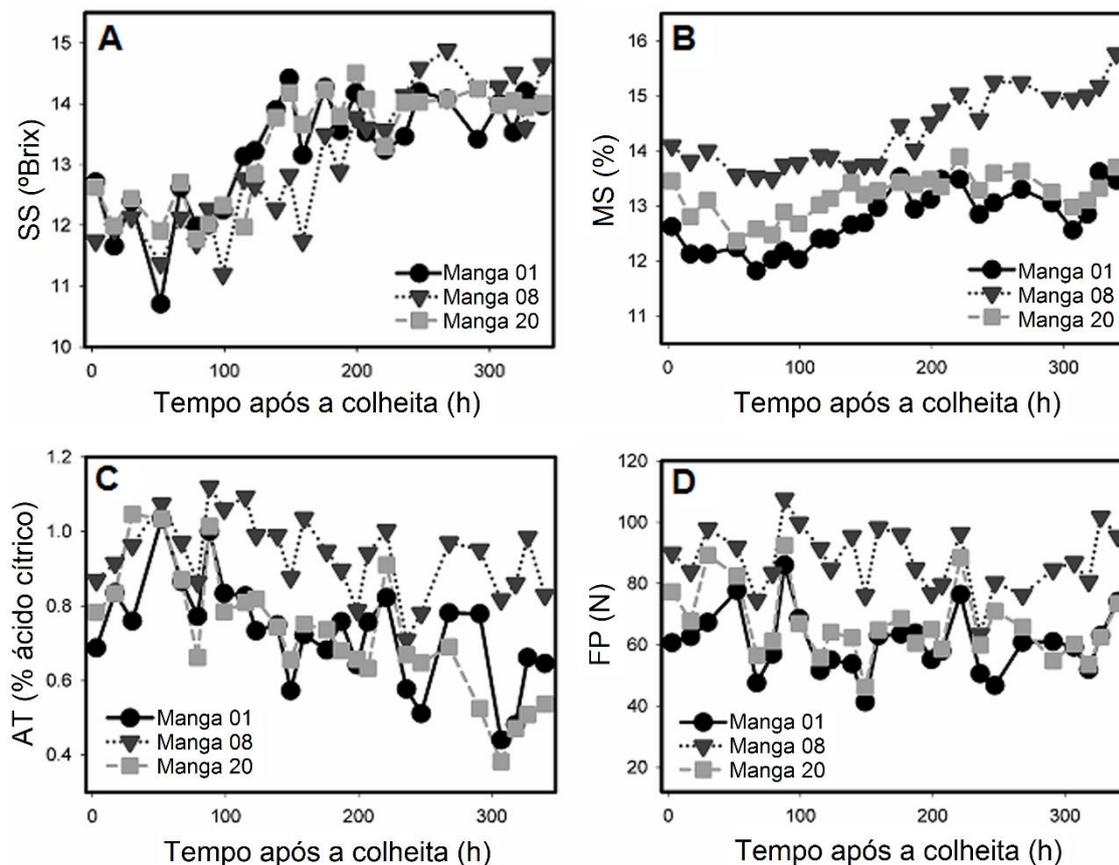
Os modelos de calibração desenvolvidos para determinação de parâmetros de qualidade de manga podem ser utilizados para monitorar as mudanças físico-químicas que ocorrem durante o amadurecimento dos frutos. Neste caso, a técnica de espectroscopia NIR mostra-se como uma ferramenta interessante, pois permite o monitoramento individual dos frutos. Este tipo de abordagem é possível em função da natureza não destrutiva desta técnica analítica.

Durante o processo de amadurecimento de frutos, como é o caso da manga, o teor de MS se mantém praticamente constante, enquanto o conteúdo de SS aumenta de forma significativa. Este aumento ocorre devido à conversão do amido contido no mesocarpo do fruto em açúcares solúveis (SUBEDI et al., 2007). Em contrapartida, os parâmetros AT e FP sofrem uma redução significativa durante o período de amadurecimento. A redução da AT ocorre, principalmente, devido ao uso dos ácidos orgânicos como substratos para a respiração dos frutos (MEDLICOTT & THOMPSON, 1985). No caso da FP, a redução é atribuída, principalmente, ao aumento da atividade de enzimas capazes de degradar a parede celular do fruto (MITCHAM & MCDONALD, 1992).

Neste estudo foi realizado o acompanhamento dos frutos durante o período de amadurecimento (14 dias em temperatura de  $24^{\circ}\text{C} \pm 1^{\circ}\text{C}$ ) empregando-se o espectrômetro

MicroNIR 1700. Os resultados obtidos foram condizentes com o comportamento esperado para SS e MS durante o amadurecimento. Neste caso, o teor de SS apresentou um aumento significativo a partir de 6 dias (150 h) após a colheita (Figura 26-A), enquanto o teor de MS permaneceu relativamente constante ao longo do amadurecimento, conforme pode ser observado para as mangas 1 e 20 (Figura 26-B). Entretanto, alguns frutos apresentaram um aumento no valor deste parâmetro após cerca de 7 ou 8 dias (180 ou 200 h) após a colheita, que foi atribuído à perda de água durante o período de armazenamento, conforme pode ser observado para a manga 8 (Figura 26-B).

Em contraste ao que foi observado para SS e MS, os valores previstos pelos modelos para AT e FP não exibiram o comportamento esperado durante o amadurecimento dos frutos (Figura 26-C e D). Esperava-se que ambos os parâmetros apresentassem um declínio ao longo do tempo, mas esta tendência não foi observada para a maioria dos frutos avaliados. Em geral, os valores previstos para estes parâmetros parecem ter sido afetados de forma significativa por erros aleatórios. De fato, os valores de RMSEP dos modelos construídos para a previsão destes parâmetros foram relativamente altos em comparação à faixa utilizada para a calibração destes parâmetros (Tabela 3). Os valores de RMSEP relativo (RMSEPr) exibidos pelos modelos de calibração para AT e FP foram de aproximadamente 20%, enquanto os modelos para SS e MS apresentaram valores de 6% e 4%, respectivamente. Assim, este fato pode justificar o baixo desempenho preditivo apresentado pelos modelos durante o monitoramento de um conjunto independente de frutos, onde a faixa de variação de AT e FP foi pequena em comparação aos valores de RMSEP dos obtidos.



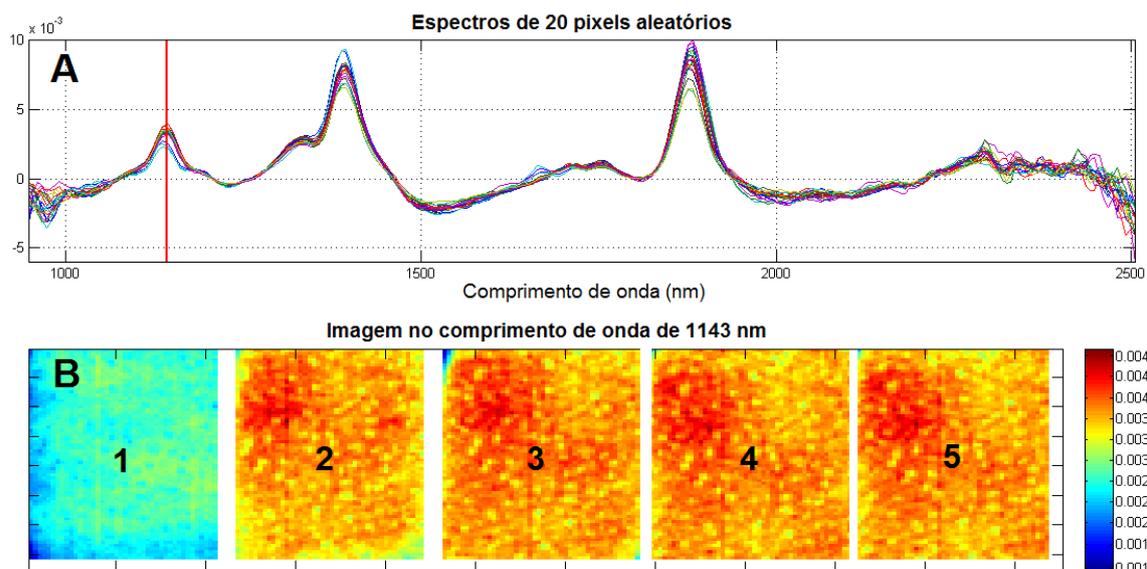
**Figura 26** – Valores previstos pelos modelos de calibração para: (A) sólidos solúveis (SS), (B) matéria seca (MS), (C) acidez titulável (AT) e (D) firmeza de polpa (FP). Estes parâmetros foram determinados em mangas “Tommy Atkins” ao longo do período de armazenamento, que correspondeu a 14 dias (360 h) em temperatura de 24°C ( $\pm$  1°C). Apenas três mangas representativas do conjunto são mostradas na figura, a fim de permitir melhor visualização dos resultados.

Apesar da capacidade preditiva insatisfatória para essa aplicação dos modelos para a determinação de AT e FP, os resultados obtidos no estudo mostraram-se promissores para a avaliação da qualidade de manga. Isto, pois os teores de SS e MS são os parâmetros mais importantes para a avaliação da qualidade dos frutos. Diversos estudos têm demonstrado que a alta qualidade de consumo de mangas está diretamente relacionada com altos teores de SS nos frutos maduros, sendo estes definidos por altos teores de amido e matéria seca nos frutos no momento da colheita (NASSUR, 2013). Adicionalmente, o teor de SS na manga madura pode ser previsto a partir dos valores de MS e amido determinados no momento da colheita por meio da espectroscopia NIR (SARANWONG et al., 2003, SARANWONG et al., 2004; SUBEDI et al., 2007; NASSUR, 2013).

### **3.4.2 Parte 2: Identificação de danos mecânicos em mangas “Tommy Atkins” utilizando imagens hiperespectrais na região do NIR**

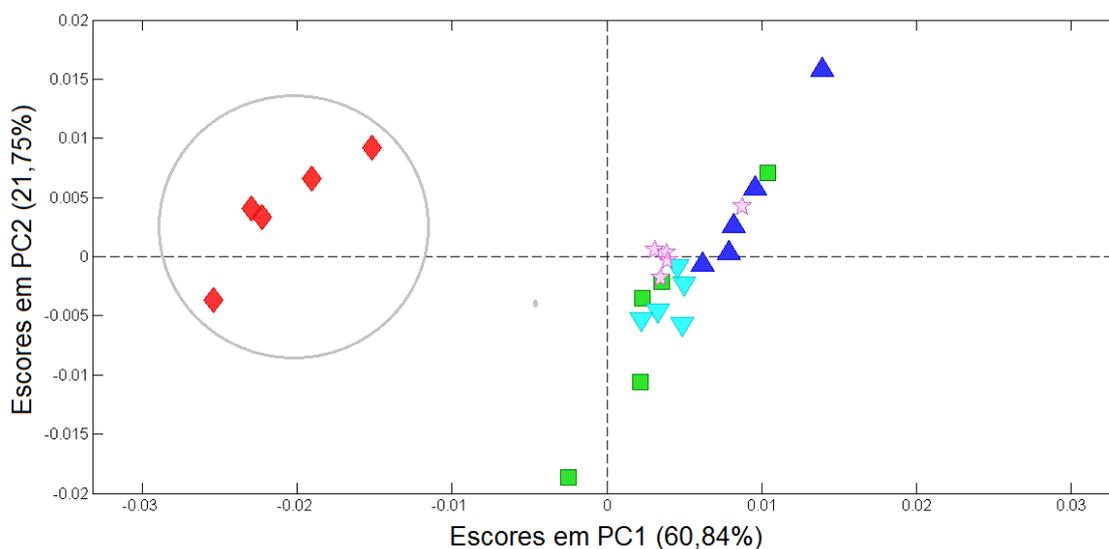
#### **3.4.2.1 Penetração da radiação na região do NIR através dos tecidos epidérmicos e subepidérmicos de mangas “Tommy Atkins” utilizando a câmera hiperespectral SisuCHEMA SWIR**

Inicialmente, a profundidade de penetração da radiação foi avaliada utilizando-se o suporte de PS. Foram obtidas imagens de amostras com as seguintes espessuras: 0,74, 4,13, 7,90, 10,50 e 20,70 mm. Na Figura 27-A são mostrados os espectros pré-processados, utilizando primeira derivada com filtro de Savitzky-Golay (polinômio de 2ª ordem e janela 15 pontos), de 20 pixels aleatórios (um número reduzido de pixels a fim de permitir uma melhor visualização dos espectros) provenientes das imagens das 5 amostras de manga com diferentes espessuras. O pré-processamento espectral utilizando a primeira derivada forneceu melhores resultados em comparação às outras técnicas de pré-processamento e, portanto, apenas os resultados utilizando esta técnica são apresentados. Na Figura 27-B são mostradas as imagens destas 5 amostras em pseudocores no comprimento de onda de 1143 nm (B-1 correspondendo à amostra menos espessa e B-5 correspondendo à amostra mais espessa). O comprimento de onda de 1143 nm foi escolhido, pois nesta região espectral os espectros provenientes de diferentes amostras exibiram variações significativas na intensidade de absorbância, permitindo diferenciar claramente a imagem correspondente à amostra de menor espessura das demais imagens. As imagens em pseudocores estão relacionadas com a absorbância das amostras em um comprimento de onda específico. A partir destas imagens pode se verificar que apenas a amostra de menor espessura (imagem B-1) apresentou pixels com distribuição de pseudocores distinta das demais amostras, em geral, exibindo valores de absorbância menores no comprimento de onda avaliado (1143 nm). Este fato sugere que a espessura reduzida desta amostra (0,74 mm) permite que uma fração significativa da radiação incidente seja transmitida através do tecido do fruto e atinja o suporte de PS situado do lado oposto da amostra. Conseqüentemente, além de informações relativas à composição química e física do fruto, a radiação refletida irá retornar ao detector do equipamento levando também informações espectrais referentes à composição do suporte de PS.



**Figura 27** – (A) Espectros pré-processados, utilizando primeira derivada com filtro de Savitzky-Golay (polinômio de 2<sup>a</sup> ordem e janela 15 pontos), de 20 pixels aleatórios provenientes das imagens de 5 amostras de manga de diferentes espessuras; e (B) imagens destas 5 amostras (B-1 correspondendo à amostra menos espessa e B-5 correspondendo à amostra mais espessa) em pseudocores no comprimento de onda de 1143 nm. As cores azul e vermelha correspondem aos valores mínimos e máximos de absorvância em cada pixel no comprimento de onda de 1143 nm, respectivamente.

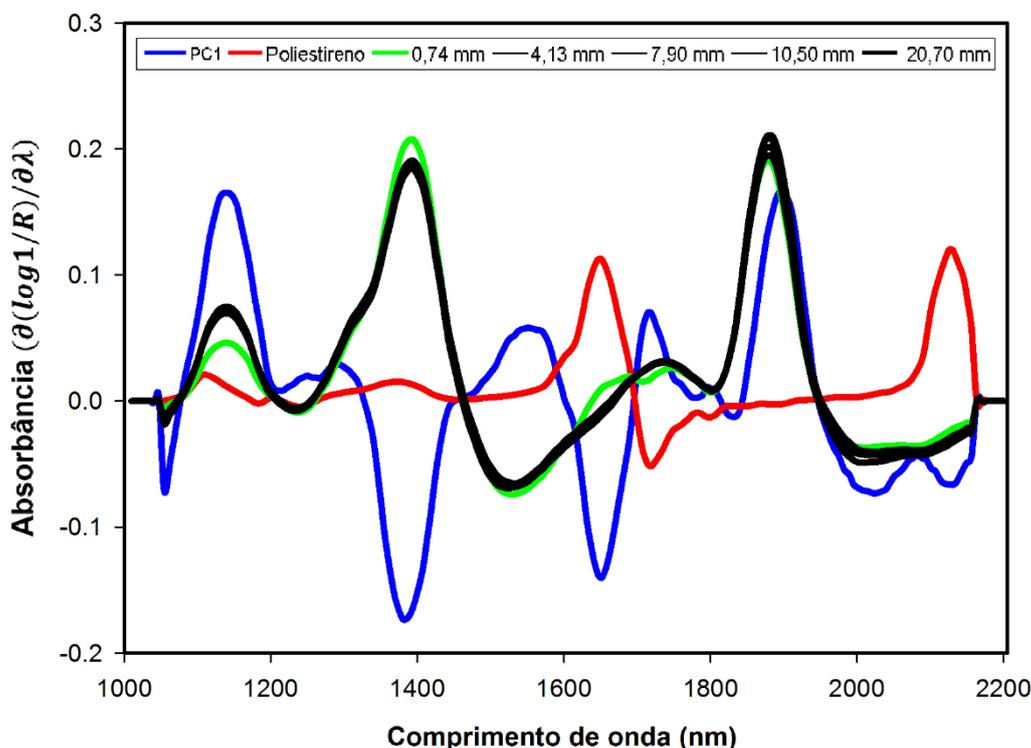
Em seguida, foram selecionados 20 pixels da parte central de cada imagem, os quais foram reduzidos a 5 espectros médios, por meio do cálculo do espectro médio de 4 pixels por vez. Foi realizada uma PCA com o conjunto dos 25 espectros médios resultantes, provenientes de pixels pertencentes às 5 amostras de manga de diferentes espessuras mantidas sobre o suporte de PS. O gráfico de escores em PC1 *versus* PC2 pode ser visualizado na Figura 28, onde PC1 e PC2 descreveram juntas 82,6% da variância dos dados. É possível observar uma separação clara ao longo da PC1 entre os escores referentes à amostra de manga de menor espessura (símbolos em vermelho que apresentam escores negativos em PC1) e os escores referentes às demais amostras.



**Figura 28** – Gráfico de escores (PC1 *versus* PC2) resultante da PCA realizada com pixels provenientes das imagens das amostras de manga de diferentes espessuras mantidas sobre o suporte de PS. Os símbolos em vermelho referem-se à amostra com menor espessura, enquanto os outros símbolos referem-se às demais amostras.

Após a aplicação da PCA foi realizada uma comparação entre os perfis espectrais dos pixels correspondentes às imagens das diferentes amostras de manga, dos pixels correspondentes ao suporte de PS e dos *loadings* em PC1. Esses perfis corresponderam aos espectros médios, obtidos pela média dos espectros de 25 pixels (conforme descrito anteriormente) provenientes de uma mesma amostra. O gráfico obtido é mostrado na Figura 29, onde os *loadings* em PC1 são representados na cor azul, o espectro médio do suporte de PS é representado na cor vermelha, o espectro médio da amostra de manga de menor espessura é representado na cor verde, e os espectros médios das demais amostras de manga são todos representados na cor preta. Como mencionado, as amostras correspondentes a menor espessura tem valores de escores mais negativos em PC1. Avaliando-se os *loadings* (Figura 29), observa-se que essas amostras são influenciadas por regiões espectrais características do polímero, particularmente em torno 1650 nm (Figura 29). Uma outra comparação foi realizada utilizando-se o espectro médio do suporte de PS e os espectros médios de todas as amostras de manga. Neste caso, pode-se observar que apenas o espectro médio da amostra de manga de menor espessura apresenta valores de absorvância mais elevados nos comprimentos de onda próximos à 1390 e 1650 nm (Figura 29). Este fato sugere que a diminuição da espessura das amostras permite registrar informações espectrais relativas ao suporte de PS, mantido sob as amostras

durante o registro das imagens, o qual também apresentou bandas características nas regiões espectrais próximas à 1390 e 1650 nm (Figura 29). Portanto, há novamente indícios de que a radiação proveniente do equipamento é capaz de penetrar somente através dos tecidos epidérmicos e subepidérmicos da amostra de menor espessura.

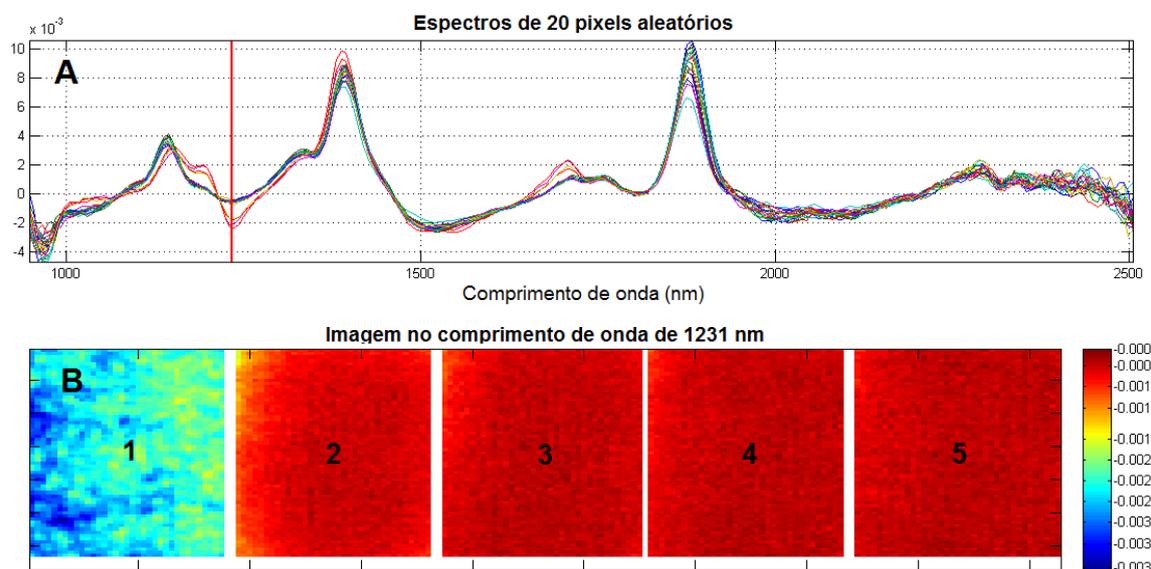


**Figura 29** – Gráfico exibindo os *loadings* em PC1 (em azul), o espectro médio do suporte de PS (em vermelho) e os espectros médios das amostras de manga de diferentes espessuras (amostra com menor espessura em verde e demais amostras em preto).

Paralelamente à aquisição das imagens utilizando o suporte de PS foram também registradas imagens das mesmas amostras de manga sobre um suporte composto por PEAD. A aquisição e tratamento dos dados provenientes destas novas imagens foi realizado com o objetivo de evidenciar os resultados apresentados anteriormente. O tratamento das imagens e procedimentos adotados para avaliar os resultados foram idênticos aos utilizados com as imagens obtidas sobre o suporte de PS e são apresentados a seguir.

Na Figura 30-A são mostrados os espectros pré-processados, utilizando primeira derivada com filtro de Savitzky-Golay (polinômio de 2ª ordem e janela 15 pontos), de 20 pixels aleatórios provenientes das imagens das 5 amostras de manga com diferentes espessuras. O pré-processamento espectral utilizando a primeira derivada forneceu

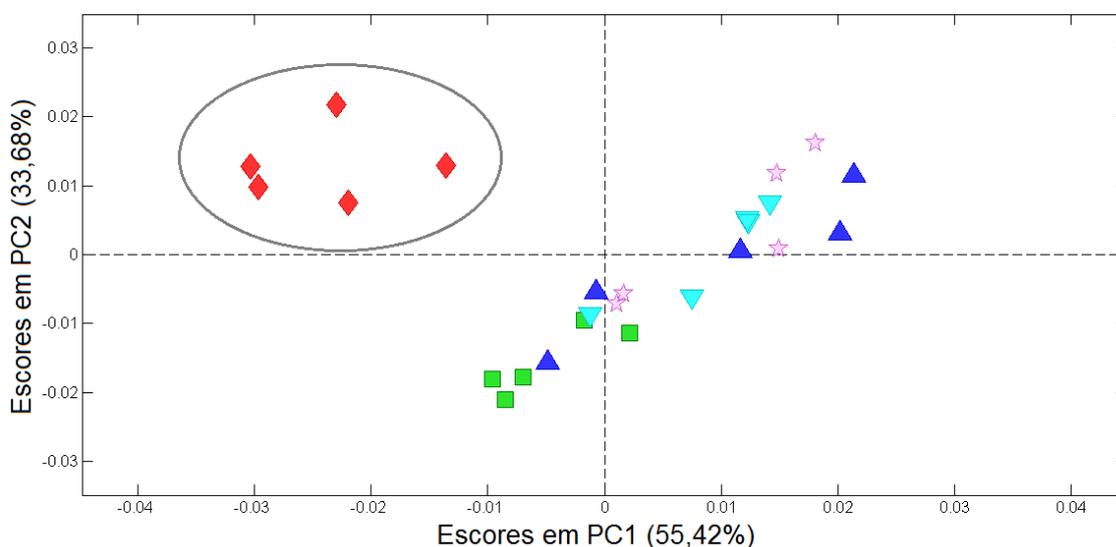
melhores resultados em comparação às outras técnicas de pré-processamento e, portanto, apenas os resultados utilizando esta técnica são apresentados. Na Figura 30-B são mostradas as imagens destas 5 amostras em pseudocores no comprimento de onda de 1231 nm (B-1 correspondendo à amostra menos espessa e B-5 correspondendo à amostra mais espessa). O comprimento de onda de 1231 nm foi escolhido, pois nesta região espectral os espectros provenientes de diferentes amostras exibiram variações significativas na intensidade de absorbância, permitindo diferenciar claramente a imagem correspondente à amostra de menor espessura das demais imagens. A partir destas imagens pode se verificar que apenas a amostra de menor espessura (imagem B-1) apresentou pixels com distribuição de pseudocores distinta das demais amostras, em geral, exibindo valores de absorbância menores no comprimento de onda avaliado (1231 nm). De maneira análoga ao que foi observado com as amostras sobre o suporte de PS, este fato sugere que a amostra de menor espessura (0,74 mm) permite que uma fração significativa da radiação incidente seja transmitida através do tecido do fruto e atinja o suporte de PEAD situado do lado oposto da amostra, retornando ao detector do equipamento com informações espectrais referentes à composição da amostra de manga e do suporte de PEAD.



**Figura 30** – (A) Espectros pré-processados, utilizando primeira derivada com filtro de Savitzky-Golay (polinômio de 2<sup>a</sup> ordem e janela 15 pontos), de 20 pixels aleatórios provenientes das imagens de 5 amostras de manga de diferentes espessuras; e (B) imagens destas 5 amostras (B-1 correspondendo à amostra com menor espessura e B-5 correspondendo à amostra com maior espessura) em pseudocores no comprimento de

onda de 1231 nm. As cores azul e vermelha correspondem aos valores mínimos e máximos de absorbância em cada pixel no comprimento de onda de 1231 nm, respectivamente.

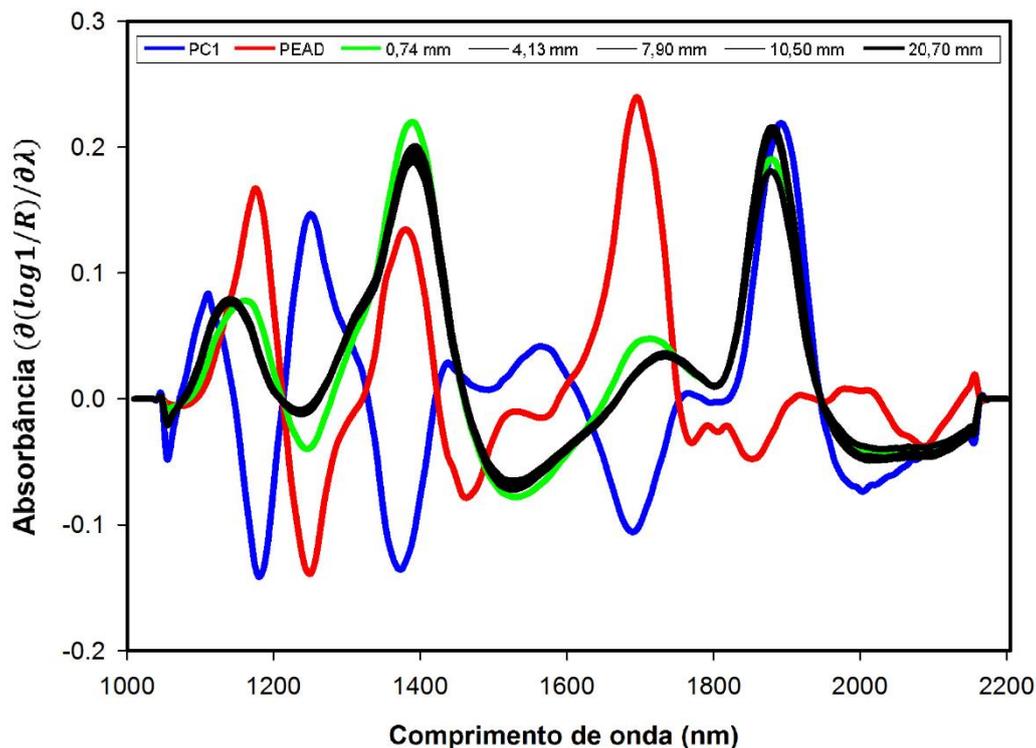
De maneira análoga ao que foi descrito anteriormente, foram selecionados 20 pixels da parte central de cada imagem, os quais foram reduzidos a 5 espectros médios, por meio do cálculo do espectro médio de 4 pixels por vez. Em seguida, foi realizada uma PCA com o conjunto dos 25 espectros médios resultantes, provenientes de pixels pertencentes às 5 amostras de manga de diferentes espessuras mantidas sobre o suporte de PEAD. O gráfico de escores em PC1 *versus* PC2 pode ser visualizado na Figura 31, onde PC1 e PC2 descreveram juntas 89,1% da variância dos dados. É possível observar novamente uma separação clara ao longo da PC1 entre os escores referentes à amostra de manga de menor espessura (símbolos em vermelho) e os escores referentes às demais amostras.



**Figura 31** – Gráfico de escores (PC1 *versus* PC2) resultante da PCA realizada com pixels provenientes das imagens das amostras de manga de diferentes espessuras mantidas sobre o suporte de PEAD. Os símbolos em vermelho referem-se à amostra menos espessa, enquanto os outros símbolos referem-se às demais amostras.

Após a obtenção dos resultados por meio da PCA foi realizada a comparação entre os perfis espectrais dos pixels correspondentes às imagens das diferentes amostras de manga, dos pixels correspondentes ao suporte de PEAD e dos *loadings* em PC1. Os perfis

corresponderam aos espectros médios, obtidos pela média dos espectros de 25 pixels (conforme descrito anteriormente) provenientes de uma mesma amostra. Foram utilizados os espectros médios de cada amostra, obtidos pela média dos espectros de todos os pixels contidos em uma mesma amostra. O gráfico obtido é mostrado na Figura 32, onde os *loadings* em PC1 são representados na cor azul, o espectro médio do suporte de PEAD é representado na cor vermelha, o espectro médio da amostra de manga de menor espessura é representado na cor verde, e os espectros médios das demais amostras de manga são todos representados na cor preta. Comparando-se os valores dos *loadings* em PC1 com o espectro médio do suporte de PEAD pode-se observar que estes apresentam perfis praticamente inversos em algumas regiões espectrais, notadamente ao redor de 1150, 1250, 1390 e 1650 nm (Figura 32), de modo que nas regiões espectrais em que PC1 possui valores negativos, o suporte de PEAD exibe valores positivos; e quando que PC1 possui valores positivos, o suporte de PEAD exibe valores negativos. Assim como observado anteriormente com o suporte de PS, este fato evidencia que a PC1 deve descrever algumas características espectrais associadas à composição do suporte de PEAD. A comparação realizada utilizando-se o espectro médio do suporte de PEAD e os espectros médios de todas as amostras de manga evidenciou que apenas o espectro médio da amostra de manga de menor espessura apresenta um perfil espectral com maior similaridade ao perfil espectral exibido pelo suporte de PEAD, especificamente nas regiões próximas à 1150, 1250, 1390 e 1650 nm (Figura 32). De maneira análoga ao que foi observado a partir das imagens das amostras registradas sobre o suporte de PS, este fato sugere que a diminuição da espessura das amostras permite o registro de informações espectrais relativas ao suporte de PEAD, o qual também apresentou bandas características nas regiões espectrais próximas à 1150, 1250, 1390 e 1650 nm (Figura 32). Portanto, estes resultados reforçam as constatações anteriores de que a radiação proveniente do equipamento é capaz de penetrar somente através dos tecidos epidérmicos e subepidérmicos da amostra de menor espessura.

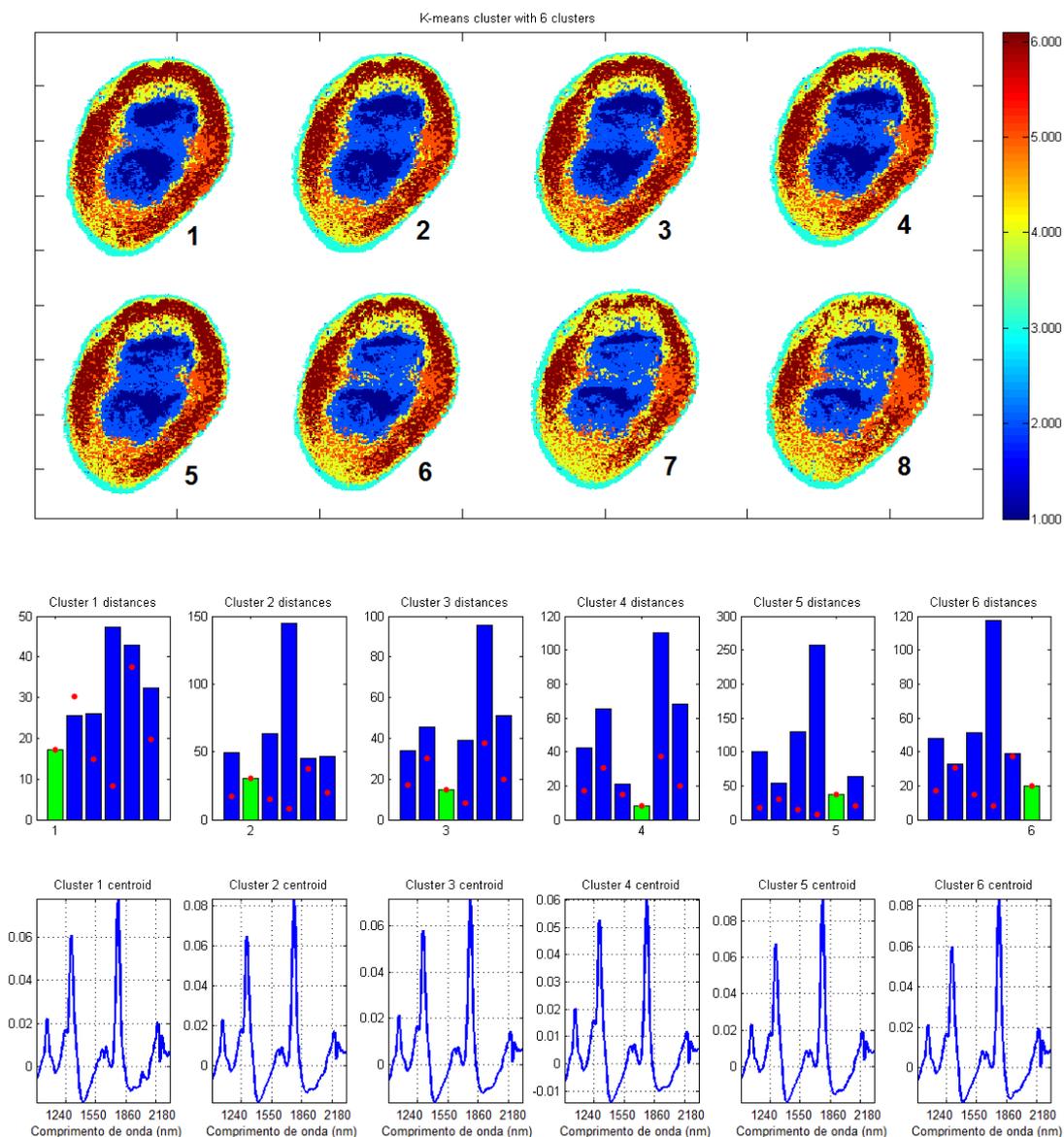


**Figura 32** – Gráfico exibindo os *loadings* em PC1 (em azul), o espectro médio do suporte de PEAD (em vermelho) e os espectros médios das amostras de manga de diferentes espessuras (amostra com menor espessura em verde e demais amostras em preto).

Adicionalmente, as avaliações descritas anteriormente foram repetidas utilizando-se um outro fruto, do qual foram obtidas imagens de amostras com as seguintes espessuras: 0,63, 3,70, 6,30, 10,60 e 20,10 mm (posicionados sobre os suportes de PS e PEAD). Os resultados foram similares aos obtidos com o primeiro fruto (dados não apresentados), evidenciando que a radiação proveniente do equipamento foi capaz de penetrar somente através dos tecidos epidérmicos e subepidérmicos das amostras de menor espessura (0,63 e 0,74 mm). Portanto, os resultados indicam que a radiação, na faixa espectral avaliada, pode penetrar até 0,74 mm através do tecido dos frutos, mas não é capaz de penetrar além de 3,70 mm. O conhecimento da profundidade de penetração da radiação NIR através dos tecidos epidérmicos e subepidérmicos dos frutos é relevante, pois permite estimar o volume do fruto e a região que efetivamente pode ser amostrada pelo equipamento, que afeta diretamente o espectro de reflectância obtido.

### 3.4.2.2 Detecção precoce de danos mecânicos induzidos artificialmente em mangas “Tommy Atkins” por meio de imagens hiperespectrais na região do NIR

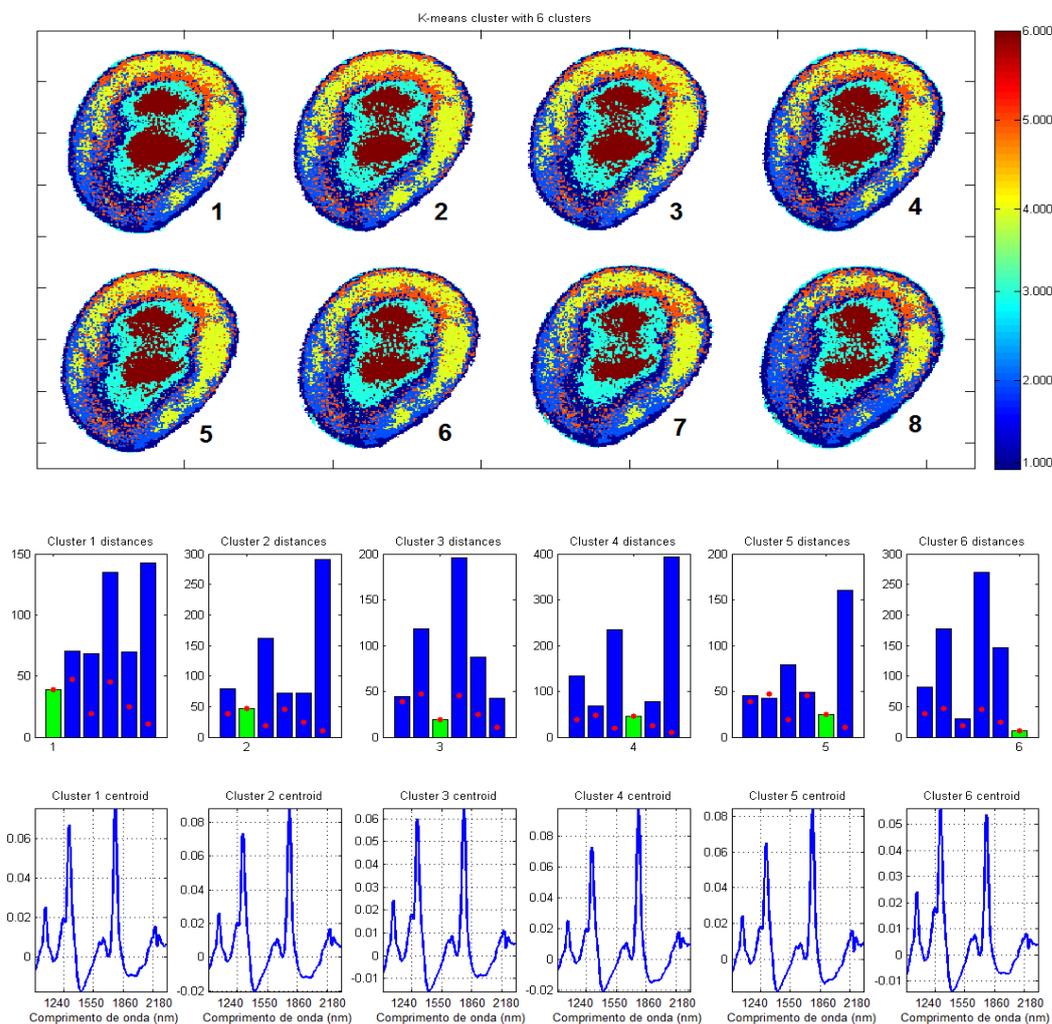
A possibilidade de detecção precoce de danos mecânicos induzidos artificialmente nas mangas “Tommy Atkins” foi avaliada utilizando-se o método de agrupamento *K-means*, empregando diferentes números de clusters (de 3 até 9). Este método foi utilizado com o objetivo de agrupar os pixels de cada imagem em função da similaridade entre seus espectros. Assim, seria possível separar e identificar os pixels relacionados com a área do fruto que sofreu danos (por meio de impacto mecânico). Na Figura 33 são mostrados os resultados obtidos após a aplicação da técnica de agrupamento *K-means* (utilizando 6 clusters, como exemplo) nas imagens provenientes de um fruto em estágio de maturação pouco avançado (FV), registradas em 8 momentos distintos (antes do impacto, imediatamente após o impacto e depois de 1, 2, 4, 24, 48 e 72 h após o impacto). A partir das imagens apresentadas em pseudocores (Figura 33) pode-se verificar que não é possível detectar qualquer agrupamento de pixels especificamente na região do fruto que recebeu o impacto (no ponto central das imagens). Há apenas um grande agrupamento de pixels (em azul) ocupando uma área bem maior que a área relativa ao impacto. Os outros 5 agrupamentos de pixels estão relacionados com outras áreas do fruto, que apresentam perfis espectrais levemente distintos. Pode-se observar que, apesar dos pixels estarem divididos em 6 conjuntos distintos, os espectros médios apresentados por cada *cluster* (denominados de “centroides”) apresentam perfis ainda muito similares. De fato, analisando o gráfico de barras mostrado na Figura 33, pode-se verificar a existência de *clusters* com características muito similares. É possível observar as seguintes relações: (1) *cluster* 1 não apresenta similaridade com os demais *clusters*; (2) *cluster* 2 é muito similar aos *clusters* 5 e 6; (3) *cluster* 3 não apresenta similaridade com os demais *clusters*; (4) *cluster* 4 é similar ao *cluster* 3; (5) *cluster* 5 é similar aos *clusters* 2 e 6; e (6) *cluster* 6 é similar aos *clusters* 2 e 5. Estas similaridades indicam que, possivelmente, há um sobre-ajuste na modelagem dos dados em função da escolha de um número excessivo de clusters ( $n = 6$ ). Entretanto, refazendo a análise com um número menor de *clusters* (dados não apresentados), ainda não foi possível obter os resultados esperados.



**Figura 33** – Resultados obtidos após a aplicação da técnica de agrupamento *K-means* (utilizando 6 clusters) nas imagens provenientes de um fruto em estágio de maturação pouco avançado (FV), registradas em 8 momentos distintos: antes do impacto (1), imediatamente após o impacto (2) e depois de 1 h (3), 2 h (4), 4 h (5), 24 h (6), 48 h (7) e 72 h (8) após o impacto.

Os resultados obtidos após a aplicação da técnica de agrupamento *K-means* (utilizando 6 clusters) nas imagens provenientes de um fruto em estágio de maturação intermediário (FI) e de um fruto em estágio de maturação avançado (FM) são mostrados nas Figuras 34 e 35, respectivamente. Nestes casos, as imagens também foram registradas em 8 momentos distintos (antes do impacto, imediatamente após o impacto e depois de 1, 2, 4, 24, 48 e 72 h após o impacto). Similarmente ao que foi observado anteriormente

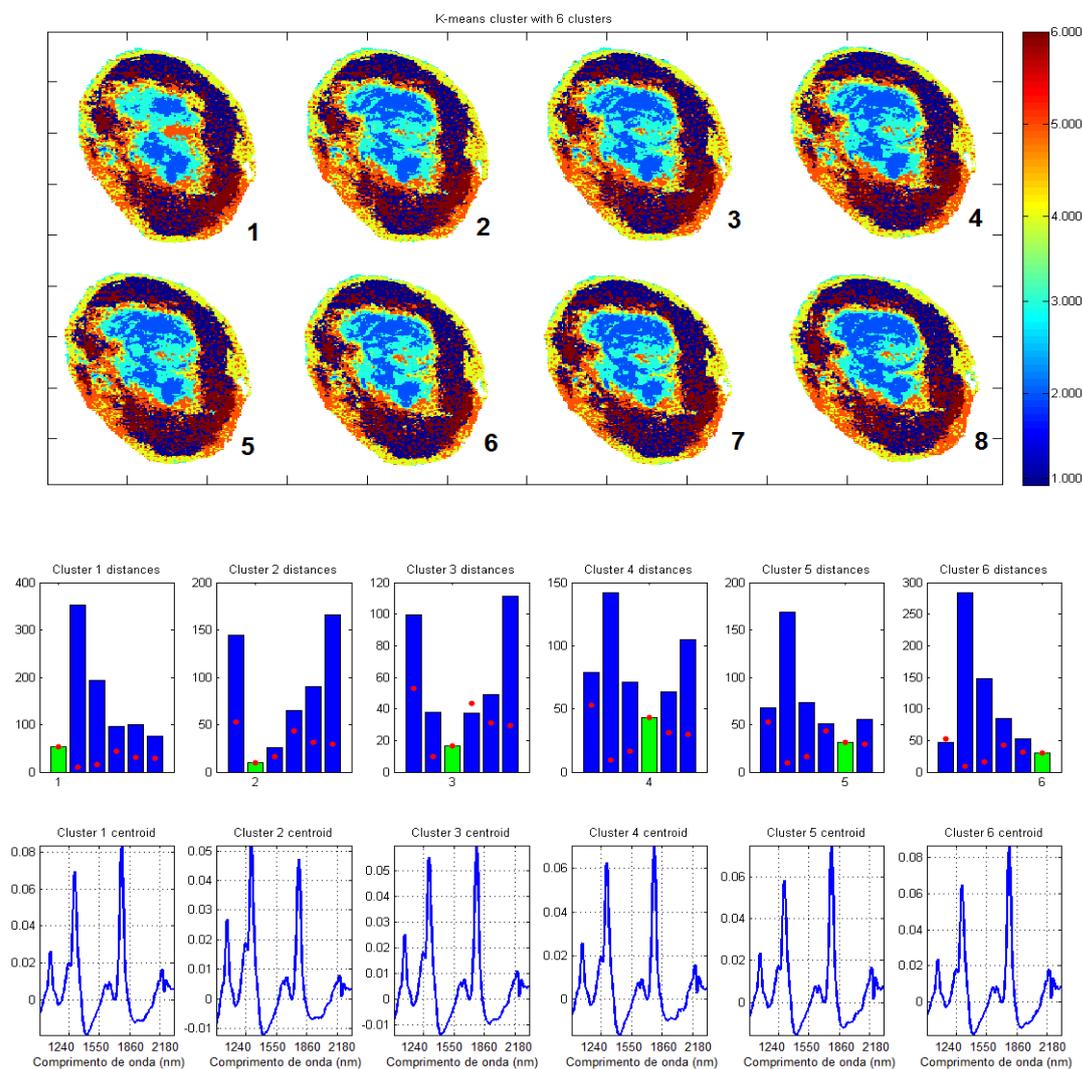
com o FV, não foi possível detectar qualquer agrupamento de pixels que estivesse relacionado com danos nos frutos ocasionados pelo choque mecânico.



**Figura 34** – Resultados obtidos após a aplicação da técnica de agrupamento *K-means* (utilizando 6 clusters) nas imagens provenientes de um fruto em estágio de maturação intermediário (FI), registradas em 8 momentos distintos: antes do impacto (1), imediatamente após o impacto (2) e depois de 1 h (3), 2 h (4), 4 h (5), 24 h (6), 48 h (7) e 72 h (8) após o impacto.

Analisando os gráficos de barras mostrados nas Figura 34 e 35, pode-se verificar a existência de *clusters* com características muito similares. De maneira análoga ao que foi observado anteriormente para o FV, estas similaridades indicam que, possivelmente, há um sobre-ajuste na modelagem dos dados em função da escolha de um número excessivo de clusters ( $n = 6$ ). Entretanto, refazendo a análise com um número menor de

*clusters* (dados não apresentados), ainda não foi possível detectar conjuntos de pixels que pudessem estar associados às regiões do fruto danificadas pelo choque mecânico.



**Figura 35** – Resultados obtidos após a aplicação da técnica de agrupamento *K-means* (utilizando 6 clusters) nas imagens provenientes de um fruto em estágio de maturação avançado (FM), registradas em 8 momentos distintos: antes do impacto (1), imediatamente após o impacto (2) e depois de 1 h (3), 2 h (4), 4 h (5), 24 h (6), 48 h (7) e 72 h (8) após o impacto.

A técnica *K-means*, empregando-se diferentes números de clusters (de 3 até 9), foi aplicada em todas as imagens provenientes dos 9 frutos avaliados e os resultados obtidos foram similares, não permitindo a identificação de um conjunto de pixels associado aos danos provocados pelo impacto nos frutos. Adicionalmente, foram realizadas PCA em cada uma das imagens, no intuito de verificar a existência de

agrupamentos naturais de pixels com características espectrais similares. Entretanto, também não foi possível identificar qualquer conjunto de pixels localizados especificamente nas regiões dos frutos que haviam recebido o impacto. Uma característica observada em todas as imagens obtidas foi a presença de um agrupamento de pixels ocupando uma área relativamente grande, localizado na região central das amostras. Acredita-se que este efeito esteja associado à iluminação direta dos frutos proporcionada pelo sistema de aquisição de imagens, não estando relacionado com informações químicas e/ou físicas relevantes das amostras. Este fato representou um problema significativo para o tratamento das imagens e impediu a obtenção de informações relevantes a respeito de quaisquer alterações na constituição das amostras após estas terem sido submetidas ao choque mecânico. Infelizmente, não foi possível remover este efeito nas etapas de pré-processamento das imagens e este fato prejudicou de forma significativa a obtenção dos resultados esperados.

Outro problema que ocorreu durante a realização deste experimento foi a impossibilidade de manter os frutos inteiros. É possível que o desenvolvimento destes e, conseqüentemente, o desenvolvimento das lesões provocadas pelos choques mecânicos, tenha sido afetado de forma significativa devido ao fato dos frutos terem sido cortados na metade antes do início do experimento. Este fato pode ter influenciado negativamente o desenvolvimento deste estudo. Idealmente, os frutos deveriam ter sido mantidos inteiros durante todo o experimento, mas não foi possível em função das dimensões do equipamento.

### 3.5 CONCLUSÕES

Na Parte 1 deste capítulo, foram desenvolvidos modelos de calibração multivariada empregando o espectrômetro portátil MicroNIR 1700 para a determinação rápida e não destrutiva de sólidos solúveis, matéria seca, acidez titulável e firmeza de polpa em mangas intactas da variedade “Tommy Atkins”. Inicialmente, verificou-se que a radiação NIR, na faixa utilizada neste estudo, foi capaz de penetrar até 7,4 mm através dos tecidos epidérmicos e subepidérmicos dos frutos avaliados. Esta informação permitiu a realização de uma amostragem adequada do tecido dos frutos, o qual foi utilizado para as análises de referência. Na parte do estudo referente à construção dos modelos de calibração, foi possível verificar que o espectrômetro avaliado pode ser utilizado para a

determinação de sólidos solúveis e matéria seca em mangas com exatidão aceitável para atestar a qualidade dos frutos.

Adicionalmente, os modelos de calibração foram utilizados com sucesso para monitorar os teores de sólidos solúveis e matéria seca de um conjunto de frutos independentes ao longo do período de amadurecimento. Entretanto, os modelos falharam na predição de acidez titulável e firmeza de polpa nessa aplicação, e este fato foi atribuído aos altos valores dos erros de predição (RMSEP relativo de aproximadamente 20%) em relação à faixa utilizada para a calibração destes parâmetros. Portanto, conclui-se que o espectrômetro MicroNIR 1700 é um equipamento promissor para ser utilizado na cadeia produtiva de manga, permitindo a avaliação e monitoramento da qualidade dos frutos comercializados e oferecidos aos consumidores.

Na Parte 2 deste capítulo, verificou-se que a radiação NIR proveniente do sistema de aquisição de imagens SisuCHEMA SWIR (Specim Spectral Imaging Ltd., Oulu, Finlândia) foi capaz de penetrar até 0,74 mm através dos tecidos epidérmicos e subepidérmicos dos frutos avaliados, mostrando que o sistema é capaz de registrar informações espectrais em um nível mais superficial das amostras analisadas.

Em relação à detecção precoce dos danos mecânicos nos frutos, não foi possível diferenciar o tecido sadio do tecido danificado por meio da técnica de agrupamento *K-means*. É possível que o desenvolvimento dos frutos e, conseqüentemente, o desenvolvimento das lesões provocadas pelos choques mecânicos, tenha sido afetado de forma significativa devido ao fato dos frutos terem sido cortados na metade antes do início do experimento. Adicionalmente, efeitos atribuídos à iluminação direta dos frutos prejudicaram o tratamento das imagens e, conseqüentemente, tiveram um impacto negativo em relação aos resultados obtidos.

### 3.6 PERSPECTIVAS FUTURAS

Ao longo do desenvolvimento dos estudos realizados neste capítulo foi possível identificar alguns aspectos que podem ser melhorados e novos estudos que podem ser realizados em propostas futuras, dentre os quais podem ser destacados os seguintes:

Em relação aos estudos utilizando espectrômetros portáteis:

- Avaliar o desempenho de novos espectrômetros portáteis de baixo custo (entre R\$ 1.000,00 e 15.000,00) para serem utilizados na avaliação da qualidade de mangas;
- Incluir maior variabilidade nos modelos de calibração desenvolvidos, visando aumentar a robustez dos mesmos (ex.: incluir mangas de diferentes variedades, de diferentes pomares, colhidas em diferentes safras, entre outras variáveis);
- Aplicar métodos adequados para efetuar a compensação da temperatura durante a etapa de construção dos modelos de calibração, permitindo que os equipamentos portáteis possam ser utilizados adequadamente sob diferentes temperaturas (situação comum para aplicações em campo);
- Desenvolver modelos de classificação.

Em relação aos estudos utilizando imagens hiperespectrais:

- Minimizar efeitos indesejados resultantes da iluminação direta das amostras por meio de adequações físicas no sistema de aquisição de imagens (ex.: utilizar iluminação indireta e/ou alterar o ângulo de iluminação);
- Aplicar métodos matemáticos que permitam a eliminação de artefatos existentes nas imagens, resultantes da iluminação direta das amostras e de variação de superfície (variação espacial da altura da amostra);
- Realizar um novo estudo visando a determinação precoce de danos mecânicos em manga utilizando os frutos inteiros;
- Testar outras técnicas quimiométricas para tratamento dos dados;
- Utilizar o sistema de aquisição de imagens para realizar outros estudos em mangas, tais como: detecção precoce de doenças (utilizando folhas da planta), determinação de parâmetros de qualidade como o teor de sólidos solúveis e/ou matéria seca, entre outros.

## 4 TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE ESPECTRÔMETROS NIR DE BANCADA E PORTÁTIL PARA DETERMINAÇÃO DA QUALIDADE DE MAÇÃS

### 4.1 INTRODUÇÃO

A maçã é a terceira fruta mais cultivada e consumida no mundo, ficando atrás somente das bananas e melancias. Em 2014 foram produzidas quase 85 milhões de toneladas de maçãs, sendo a China responsável por 50% da produção mundial, seguida dos Estados Unidos (6%), Itália (3%), Turquia (3%) e Argentina (1%) (FAOSTAT, 2016).

O mercado consumidor de maçãs é altamente exigente quanto à qualidade, o que demanda um sistema de beneficiamento que permita realizar uma seleção criteriosa das frutas com potencial de mercado *in natura*. Assim, antes de serem comercializadas, as frutas devem atender a requisitos mínimos de qualidade, que incluem atributos relacionados com características externas, tais como: tamanho, formato e cor; bem como internas, tais como: teor de açúcares, acidez e firmeza (PENG & LU, 2008). A avaliação destes atributos é importante, pois afetam diretamente o sabor e a textura das frutas e, conseqüentemente, a aceitação dos consumidores.

O uso da espectroscopia NIR para determinação da qualidade de produtos agrícolas, tais como hortaliças e frutas, tem sido investigada durante a última década (BOBELYN et al., 2010). Neste contexto, diversos autores têm reportado o uso da espectroscopia NIR para avaliação de parâmetros de qualidade de maçãs, dentre os quais pode se destacar a determinação de sólidos solúveis, acidez, firmeza de polpa, índices de maturação, polifenóis e vitamina C (LAMMERTYN et al., 1998; VENTURA et al., 1999; LU et al., 2000; MCGLONE et al., 2002; MEHINAGIC et al., 2003; PARK et al., 2003; PEIRS et al., 2003; SANCHEZ et al., 2003; ZUDE et al., 2006; LIU et al., 2007; PENG & LU, 2008; ANGRA et al., 2009; PAZ et al., 2009; BOBELYN et al., 2010; SHI et al., 2011; OUYANG et al., 2012; PISSARD et al., 2012; LIU & ZHOU, 2013; GIOVANELLI et al., 2014; KUMAR et al., 2015; FAN et al., 2016).

Como já mencionado, a espectroscopia NIR apresenta a vantagem de ser uma técnica analítica com características para aplicações rápidas e não destrutivas, não necessitando de preparo de amostras ou uso de reagentes químicos, estando, portanto, alinhada aos conceitos da Química Verde. Entretanto, seu uso em condições de campo foi

limitado durante muitos anos, em função de restrições relacionadas ao tamanho e à pouca robustez dos equipamentos disponíveis no mercado. Apenas recentemente, o desenvolvimento de novas tecnologias utilizadas na construção dos espectrômetros NIR têm permitindo a redução significativa do tamanho e do custo destes equipamentos. Assim, os espectrômetros NIR portáteis representam atualmente uma alternativa atrativa e viável para as agroindústrias e pequenos produtores realizarem o monitoramento da qualidade dos frutos na cadeia produtiva de maçãs.

Neste novo cenário, onde os espectrômetros NIR portáteis estão disponíveis comercialmente, surge um desafio para os laboratórios e usuários da técnica: como utilizar as bases de dados históricas e modelos de calibração desenvolvidos previamente com os espectrômetros de bancada, sem que seja necessária a aquisição de uma nova base de dados utilizando os equipamentos portáteis? Uma recalibração completa iria demandar tempo e investimentos significativos, já que haveria a necessidade de aquisição de novas amostras, realização de análises de referência, registro de espectros e tratamento matemático adequado dos dados. Portanto, a melhor opção seria utilizar um procedimento de transferência de calibração, preferencialmente simples, para correção das diferenças instrumentais e/ou ambientais, tornando os dados provenientes do equipamento portátil compatíveis com os modelos de calibração desenvolvidos com o equipamento de bancada.

Neste contexto, são apresentados neste capítulo estudos sobre a avaliação da possibilidade de transferência de calibração entre espectrômetros NIR de bancada e portátil, visando à determinação de parâmetros de qualidade de maçãs; além da comparação prévia entre o desempenho destes espectrômetros quando utilizados na avaliação da qualidade destas frutas. Os estudos foram realizados em parceria com o *Centre Walloon de Recherches Agronomiques (CRA-W)*, localizado na cidade de Gembloux na Bélgica, no âmbito de um acordo de cooperação científica entre Brasil (CAPES) e Bélgica (Wallonie-Bruxelles International – WBI). As atividades desenvolvidas estão diretamente relacionadas ao projeto intitulado de “*PhotonFruit: Emergent Spectroscopic Techniques for the Quality Control and Traceability of Fruits and Fruits Based Products*”, que tem como objetivo principal o desenvolvimento de novas técnicas espectroscópicas para análise de frutas e produtos derivados de frutas.

## 4.2 OBJETIVOS

### 4.2.1 Objetivo geral

O objetivo geral do trabalho apresentado neste capítulo foi realizar a transferência de calibração entre um espectrômetro NIR de bancada (XDS Rapid Content Analyzer, FOSS NIRSystems, MD, EUA) e um espectrômetro NIR portátil (MicroNIR 1700, VIAVI, CA, EUA), visando a determinação de parâmetros de qualidade de maçãs.

### 4.2.2 Objetivos específicos

Os objetivos específicos foram os seguintes: (1) comparar o desempenho dos espectrômetros XDS e MicroNIR para realizar a determinação de parâmetros de qualidade de maçãs; (2) construir modelos de calibração empregando diferentes técnicas de calibração multivariada para determinação de parâmetros de qualidade de maçãs utilizando espectros provenientes do espectrômetro XDS; e (3) desenvolver uma estratégia satisfatória para realizar a transferência de calibração entre os espectrômetros XDS e MicroNIR, considerando a existência de algumas limitações práticas características do conjunto de dados em estudo.

## 4.3 MATERIAL E MÉTODOS

O trabalho apresentado neste capítulo é constituído por dois estudos distintos. Em relação à descrição do presente tópico (“4.3 – Material e Métodos”), é importante ressaltar que os itens referentes às amostras (item 4.3.1), à instrumentação e aquisição espectral (item 4.3.2) e às análises de referência (item 4.3.3) são comuns aos dois estudos. Estes se diferenciam, principalmente, em relação aos métodos quimiométricos e estratégias adotadas para resolver seus problemas específicos, que são descritos separadamente nos itens relativos aos estudos 1 e 2 (itens 4.3.4 e 4.3.5).

### 4.3.1 Amostras

As amostras utilizadas foram provenientes de 5 projetos desenvolvidos no CRA-W durante o período de 2008 até 2014. Cada projeto continha uma base de dados

constituída por espectros de maçãs de diferentes variedades (exceto o projeto 1) em diferentes estádios de maturação registrados na região do NIR, e análises de referência para vários parâmetros de qualidade da fruta. No total, foram analisadas 1861 maçãs considerando todos os 5 projetos. Na Tabela 4 são mostradas algumas informações relativas às bases de dados provenientes de cada projeto.

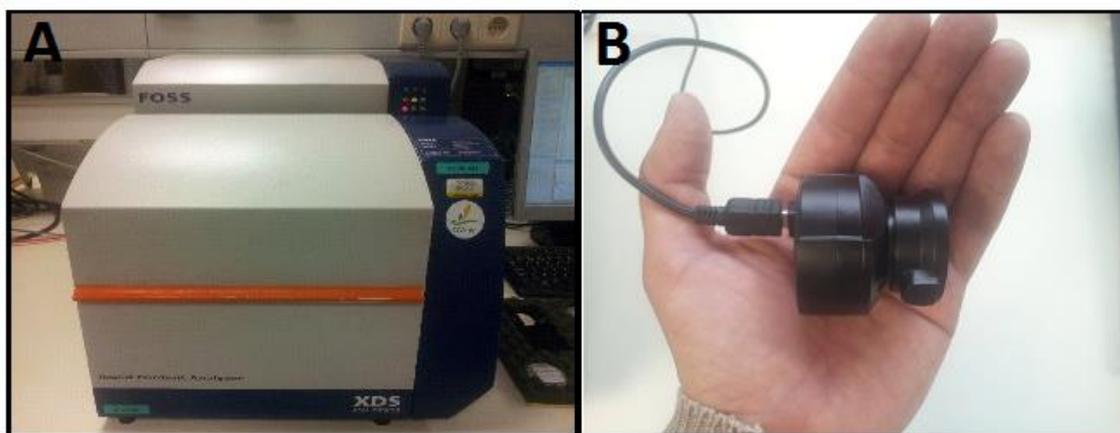
**Tabela 4** – Informações referentes às bases de dados utilizadas neste trabalho, descrevendo o período de realização dos projetos, o número de variedades de maçãs, o número de amostras analisadas e os parâmetros de qualidade determinados por meio de análises de referência.

Projeto	Período	Número de variedades de maçãs	Número de amostras	Parâmetros avaliados
1	2008 a 2010	1	559	Sólidos solúveis, acidez titulável, firmeza de polpa e índice de amido.
2	2009 a 2010	31	244	Sólidos solúveis, acidez titulável, firmeza de polpa e índice de amido.
3	2011	48	239	Sólidos solúveis, acidez titulável e firmeza de polpa.
4	2012 a 2014	2	699	Sólidos solúveis, acidez titulável, firmeza de polpa, índice de amido e índice Streif.
5	2013	12	120	Sólidos solúveis, acidez titulável, firmeza de polpa e índice de amido.

#### 4.3.2 Instrumentação e aquisição espectral

Os espectros foram obtidos em temperatura de  $22 \pm 1$  °C utilizando-se um espectrômetro NIR de bancada (XDS Rapid Content Analyzer, FOSS NIRSystems, MD,

EUA) (Figura 36-A) e um espectrômetro NIR portátil (MicroNIR 1700, VIAVI, CA, EUA) (Figura 36-B). Os dois equipamentos são dispersivos, mas possuem tecnologias diferentes para a decomposição do feixe luminoso: no XDS uma rede de difração é usada como elemento dispersivo, enquanto no MicroNIR um filtro linear variável desempenha esta função.



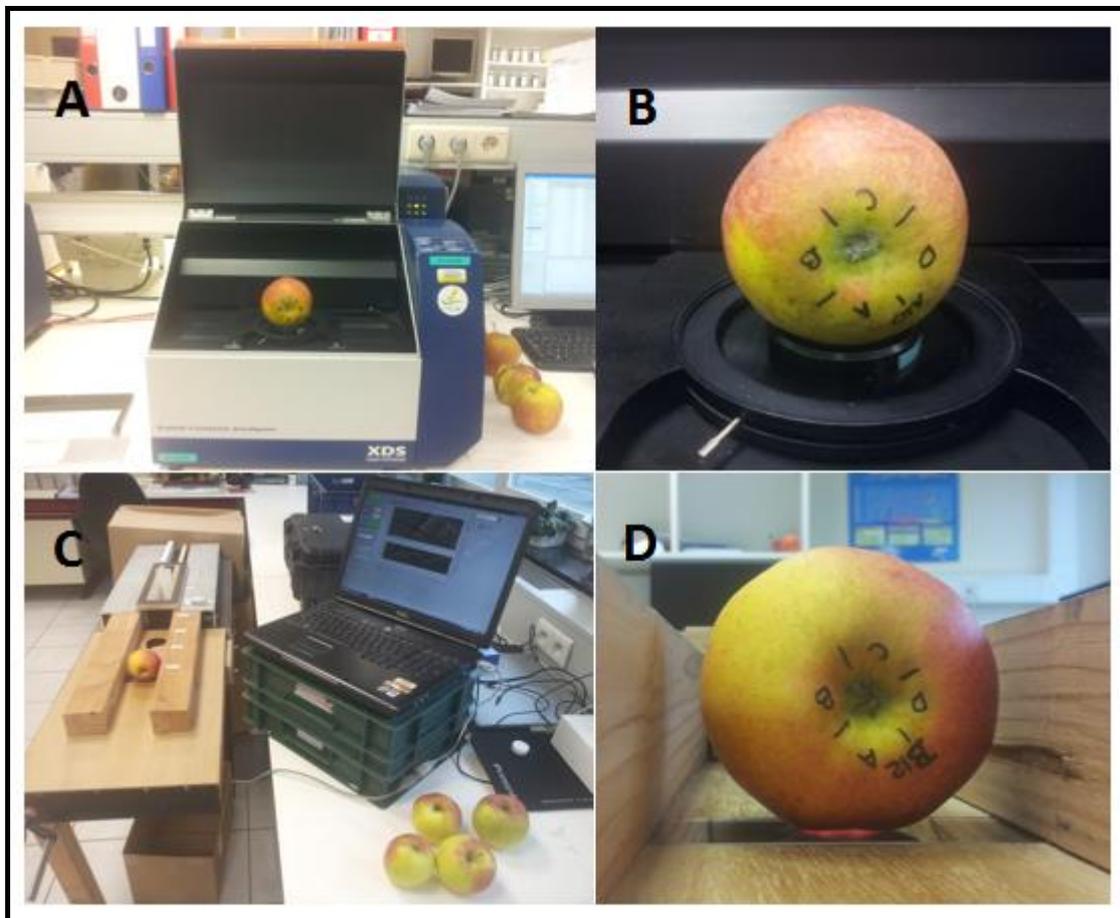
**Figura 36** – (A) espectrômetro NIR de bancada modelo XDS Rapid Content Analyzer (FOSS NIRSystems, MD, EUA); e (B) espectrômetro NIR portátil modelo MicroNIR 1700 (VIAVI, CA, EUA).

A aquisição dos espectros, utilizando-se o espectrômetro XDS, foi realizada no modo de refletância ( $\log 1/R$ ), na faixa espectral entre 400 e 2500 nm em intervalos de 2 nm. O tempo de integração foi de aproximadamente 1 s, sendo que cada espectro registrado correspondeu a média de 32 varreduras, resultando em um tempo de medida de aproximadamente 32 s.

A aquisição dos espectros, utilizando-se o espectrômetro MicroNIR, foi realizada no modo de refletância ( $\log 1/R$ ), na faixa espectral entre 1100 e 1600 nm em intervalos de 6,25 nm. O tempo de integração foi de 12 ms, sendo que cada espectro registrado correspondeu a média de 800 varreduras, resultando em um tempo de medida de 9,6 s.

Os espectros foram obtidos a partir de leituras realizadas em 4 pontos equidistantes distribuídos ao longo da região equatorial de cada fruto (Figuras 37- B e D). Nas análises realizadas com o espectrômetro XDS as amostras foram posicionadas diretamente sobre a janela de leitura do equipamento, utilizando-se apenas um adaptador “tipo íris” (Figuras 37-A e B). Nas análises realizadas com o espectrômetro MicroNIR as amostras foram posicionadas sobre a janela de leitura de uma mesa desenvolvida

especialmente para acomodar este instrumento (construída no próprio Laboratório de Espectroscopia NIR do CRA-W) (Figuras 37-C e D). Os modelos de calibração foram construídos utilizando-se os espectros médios dos frutos.



**Figura 37** – (A) Análise de maçãs utilizando-se o espectrômetro XDS; (B) detalhe do posicionamento da amostra diretamente sobre a janela de leitura deste equipamento, utilizando-se um adaptador “tipo íris”; (C) análise de maçãs utilizando-se o espectrômetro MicroNIR; e (D) detalhe do posicionamento da amostra sobre a janela de leitura da mesa desenvolvida especialmente para acomodar este equipamento. Em todos os casos, os espectros foram obtidos a partir de leituras realizadas em 4 pontos equidistantes distribuídos ao longo da região equatorial de cada fruto.

#### 4.3.3 Análises de referência

Inicialmente foi realizada a determinação da firmeza de polpa (FP) utilizando-se os frutos intactos. Em seguida, cada maçã foi cortada em pequenos pedaços, descartando-

se a parte interna onde estavam localizadas as sementes. Uma fração destes pedaços foi escolhida aleatoriamente e, em seguida, prensada por meio de um extrator de sucos manual. A fração líquida resultante da prensagem foi utilizada para a determinação de sólidos solúveis (SS) e acidez titulável (AT).

Algumas maçãs foram escolhidas para serem utilizadas na determinação do índice de amido (IA) separadamente. Estes frutos foram escolhidos sistematicamente de acordo com a coloração da casca, obtendo-se uma amostragem representativa de diferentes estádios de maturação. Neste caso, foram utilizados frutos cortados na metade através da região equatorial. A seguir são detalhados os procedimentos utilizados para a realização das análises de referência (item 4.3.3.1 ao 4.3.3.5).

#### 4.3.3.1 Determinação de sólidos solúveis (SS)

A determinação de SS foi realizada por meio de um refratômetro digital portátil (Atago, Tóquio, Japão) com compensação automática de temperatura. As leituras foram realizadas utilizando-se aproximadamente 1 mL de suco. Os resultados obtidos foram expressos em °Brix.

#### 4.3.3.2 Determinação de acidez titulável (AT)

A determinação da AT foi realizada por meio da titulação de 1 mL de suco com uma solução de NaOH 0,1 mol L<sup>-1</sup>, utilizando-se uma bureta digital. Uma solução de fenolftaleína foi utilizada como indicador do ponto de viragem (AOAC, 1995). As determinações foram realizadas em duplicata e os valores médios foram utilizados para a construção dos modelos de calibração. Os resultados obtidos foram expressos em miliequivalentes de ácido málico por litro de suco (mEq ácido málico L<sup>-1</sup>).

#### 4.3.3.3 Determinação de firmeza de polpa (FP)

A determinação da FP foi realizada utilizando-se um texturômetro digital modelo TA-XT2i (Stable Micro Systems, Godalming, Inglaterra) com ponteira de 11 mm de diâmetro. Este parâmetro foi determinado nos frutos inteiros e os resultados representam a força necessária para efetuar 10 mm de penetração da ponteira através da polpa do fruto. Foram realizadas duas medidas por fruto, em lados opostos na região equatorial, após a

remoção de uma pequena área da epiderme ( $\sim 1 \text{ cm}^2$ ), e os valores médios foram utilizados para a construção dos modelos de calibração. Os resultados obtidos foram expressos em quilogramas por centímetro quadrado ( $\text{kg cm}^{-2}$ ).

#### 4.3.3.4 Determinação de índice de amido (IA)

O parâmetro IA, também conhecido como “índice de iodo-amido”, foi determinado por meio do “teste de iodo-amido”. Este teste é baseado na reação entre uma solução de iodo e o amido presente no fruto, resultando no aparecimento de uma coloração azul em sua superfície. Para tal, os frutos foram cortados transversalmente na zona equatorial e uma das metades foi imersa durante 1 minuto em uma solução preparada pela adição de 2,2 g de iodo e 8,8 g de iodeto de potássio em 1 L de água destilada (BLANPIED & KENNETH, 1992). Em seguida, a metade do fruto foi retirada da solução e mantida por 10 minutos em repouso. Após este período, realizou-se a comparação da coloração da superfície do fruto com as cores de uma tabela apropriada para esta finalidade. Os resultados foram expressos utilizando-se uma escala com valores de 1 a 10, com intervalos de uma unidade. Neste caso, o valor 1 na escala corresponde a frutas totalmente verdes, nas quais não ocorreu degradação do amido, e o valor 10 corresponde a frutas com maturação avançada, sem presença de amido.

#### 4.3.3.5 Determinação de índice Streif (IS)

A determinação do IS foi realizada por meio da relação entre FP, SS e IA, de acordo com a seguinte equação (STREIF & BUFFER, 1990):  $IS = FP / (SS \times IA)$ . Os resultados obtidos foram expressos em quilogramas por centímetro quadrado por  $^{\circ}\text{Brix}$  ( $\text{kg cm}^{-2} ^{\circ}\text{Brix}^{-1}$ ).

### **4.3.4 Estudo 1: Comparação entre o desempenho dos espectrômetros XDS e MicroNIR para a determinação de parâmetros de qualidade de maçãs**

Neste estudo foi realizada a comparação entre o desempenho dos espectrômetros XDS e MicroNIR para a determinação de parâmetros de qualidade de maçãs. Os espectros e as análises de referência foram obtidos a partir de um conjunto de 93 amostras pertencentes ao projeto 4 (descrito na Tabela 4), correspondendo a duas cultivares de

maças: Pinova, que é uma cultivar de origem alemã amplamente comercializada no mundo todo; e AG90, que é uma nova cultivar desenvolvida pelo próprio CRA-W.

#### 4.3.4.1 Análise de dados e software

Inicialmente, foram avaliadas diferentes técnicas de pré-processamento espectral: SNV, primeira e segunda derivadas com filtro de Savitzky-Golay utilizando polinômio de 2ª ordem e janela de 15 pontos. Os modelos de calibração foram construídos utilizando-se regressão PLS e duas faixas espectrais (400-2500 nm e 1100-1600 nm) para os dados provenientes do espectrômetro XDS e uma faixa espectral (1100-1600 nm) para os dados provenientes do espectrômetro MicroNIR.

O número de variáveis latentes mantidas nos modelos de calibração foi determinado de acordo com o de RMSECV, utilizando-se validação cruzada completa. Em função do número reduzido de amostras ( $n = 93$ ), o desempenho preditivo dos modelos foi avaliado na etapa de validação interna por meio dos valores de RMSECV,  $R^2$  e *bias*. Empregou-se o teste F ao nível de 95% de confiança para comparação entre os valores de RMSECV dos modelos obtidos com os dois equipamentos, bem como as diferentes técnicas de pré-processamento aplicadas.

As amostras anômalas (*outliers*) foram identificadas e removidas empregando-se o gráfico dos valores de  $T^2$  de Hotelling (que permite a identificação de valores extremos) *versus* a soma quadrática dos resíduos (Q) (que permite a identificação de falta de ajuste à equação de regressão). Os limiares adotados para a exclusão das amostras corresponderam aos valores determinados automaticamente pelo software utilizado para a construção dos modelos de calibração.

Os cálculos relacionados às etapas de pré-processamento e construção dos modelos de calibração multivariada foram realizados utilizando-se o software The Unscrambler versão 9.8 (CAMO, Oslo, Noruega).

#### **4.3.5 Estudo 2: Transferência de calibração entre os espectrômetros NIR de bancada (XDS) e portátil (MicroNIR)**

Este estudo foi realizado em duas etapas. Inicialmente, foram construídos modelos de calibração empregando-se diferentes técnicas de calibração multivariada visando a determinação de parâmetros de qualidade de maçãs utilizando um conjunto abrangente

de amostras ( $n = 1861$ ) analisadas com o espectrômetro XDS; em seguida, avaliou-se a possibilidade de transferência de calibração entre os espectrômetros XDS e MicroNIR, considerando a existência de algumas limitações práticas características do conjunto de dados em estudo.

A princípio, poderiam ser utilizados para este estudo todos os espectros e análises de referência provenientes do conjunto de 1861 amostras pertencentes aos projetos 1, 2, 3, 4 e 5 (descritos na Tabela 4). Entretanto, apenas um conjunto de 93 amostras pertencentes ao “projeto 4” foram analisadas nos dois equipamentos. Logo, apenas estas amostras poderiam ser empregadas para obtenção das funções de transferência de calibração utilizadas neste estudo. Este fato gerou um problema de representatividade, já que as 93 amostras disponíveis para obtenção das funções de transferência não representavam adequadamente a variabilidade do conjunto total das amostras ( $n = 1861$ ). Para que o processo de transferência de calibração fosse realizado com êxito, seria necessário que o conjunto de amostras (conjunto de transferência) utilizado para obtenção das funções de transferência fosse representativo da variabilidade do conjunto de dados em estudo.

A aquisição de espectros de novas amostras de maçãs (representativas do conjunto em estudo) nos dois equipamentos, poderia ser uma forma de resolver o problema referente à representatividade. Entretanto, tratando-se de um estudo de caso real, não havia possibilidade de obtenção de novas amostras de maçãs para serem incluídas no estudo e, portanto, foi necessário buscar outra solução para este problema. Assim, optou-se por utilizar apenas as amostras com maior similaridade espectral em relação às amostras do conjunto de transferência ( $n = 93$ ) para a construção dos modelos de calibração multivariada.

#### 4.3.5.1 Análise de dados e software

O procedimento empregado para efetuar a seleção das amostras que seriam eliminadas ou mantidas para a construção dos modelos de calibração baseou-se na aplicação da técnica de Análise de Componentes Principais (PCA). Foram realizadas 5 PCA, de modo que em cada avaliação foram incluídas as amostras do conjunto de transferência e as amostras de um projeto específico (projetos 1, 2, 3, 4 e 5 mostrados na Tabela 4). Em cada PCA realizada, efetuou-se a exclusão das amostras que apresentaram menor similaridade espectral em relação às amostras de transferência por meio do gráfico

de escores. Assim, as amostras excluídas corresponderam a valores de escores mais distantes dos escores correspondentes às amostras de transferência no espaço das componentes principais (PC). Desta forma, pode-se considerar que o conjunto de amostras de transferência foi transformado “artificialmente” em um conjunto representativo por meio da exclusão de amostras específicas. Em seguida, as amostras mantidas foram divididas entre os conjuntos de calibração e validação externa por meio do algoritmo SPXY (GALVÃO et al., 2005) para cada parâmetro de qualidade avaliado.

Os modelos de calibração foram construídos empregando-se os métodos de regressão MLR, PLS e LS-SVM, utilizando-se a faixa espectral de 1100 a 1600 nm. No caso dos modelos construídos empregando-se MLR foi utilizado o algoritmo SPA (ARAÚJO et al, 2001) para realizar a seleção de variáveis. Em todos os casos, o pré-processamento dos espectros foi realizado aplicando-se a técnica SNV.

No caso dos modelos obtidos por meio de regressão PLS, o número de variáveis latentes mantidas foi determinado de acordo com o valor de RMSECV na etapa de validação interna, empregando-se o conjunto de amostras de calibração. Nesta etapa, foi utilizado o método sistemático “*venetian blinds*” para a seleção dos subconjuntos de validação interna, utilizando 10% das amostras de cada subconjunto para realizar a predição.

No caso dos modelos obtidos por meio de LS-SVM, os valores dos parâmetros gama ( $\gamma$ ) e sigma quadrático ( $\sigma^2$ ) foram otimizados visando a obtenção do valor mínimo de RMSECV. Foi realizada a compressão dos dados espectrais originais utilizando-se regressão PLS com diferentes números de variáveis latentes. A SVM utilizada foi do tipo “Epsilon-SVR”, empregando-se uma função núcleo do tipo gaussiana de base radial (RBF).

As amostras anômalas (*outliers*) foram identificadas e removidas empregando-se o gráfico dos valores de  $T^2$  de Hotelling (que permite a identificação de valores extremos) *versus* a soma quadrática dos resíduos (Q) (que permite a identificação de falta de ajuste à equação de regressão). Os limiares adotados para a exclusão das amostras corresponderam aos valores determinados automaticamente pelo software utilizado para a construção dos modelos de calibração.

O desempenho preditivo dos modelos de calibração, construídos utilizando-se MLR, PLS e LS-SVM, foi avaliado na etapa de validação externa (predição), empregando-se um conjunto de amostras independentes, considerando os valores de RMSEP,  $R^2$  e *bias*. Empregou-se o teste F ao nível de 95% de confiança para comparação

entre os valores de RMSEP obtidos pelos diferentes métodos utilizados para a construção dos modelos de calibração.

A etapa de transferência de calibração foi realizada empregando-se a técnica de padronização direta (DS). O espectrômetro de bancada (XDS) foi tratado como equipamento primário e o espectrômetro portátil (MicroNIR) como secundário. Portanto, os espectros do equipamento portátil foram transformados (padronizados) e utilizados nos modelos desenvolvidos previamente com os espectros provenientes do equipamento de bancada.

O conjunto de amostras utilizado nesta etapa foi constituído por 93 maçãs pertencentes ao projeto 4 (Tabela 4), cujos espectros foram registrados nos dois espectrômetros. Estas amostras não foram incluídas nos modelos de calibração construídos previamente. O algoritmo SPXY (GALVÃO et al., 2005) foi utilizado para selecionar as amostras de transferência ( $n = 30$ ) (utilizadas para obtenção das funções de transferência) e as amostras de predição ( $n = 63$  para os parâmetros SS, AT e IA; e  $n = 43$  para os parâmetros FP e IS).

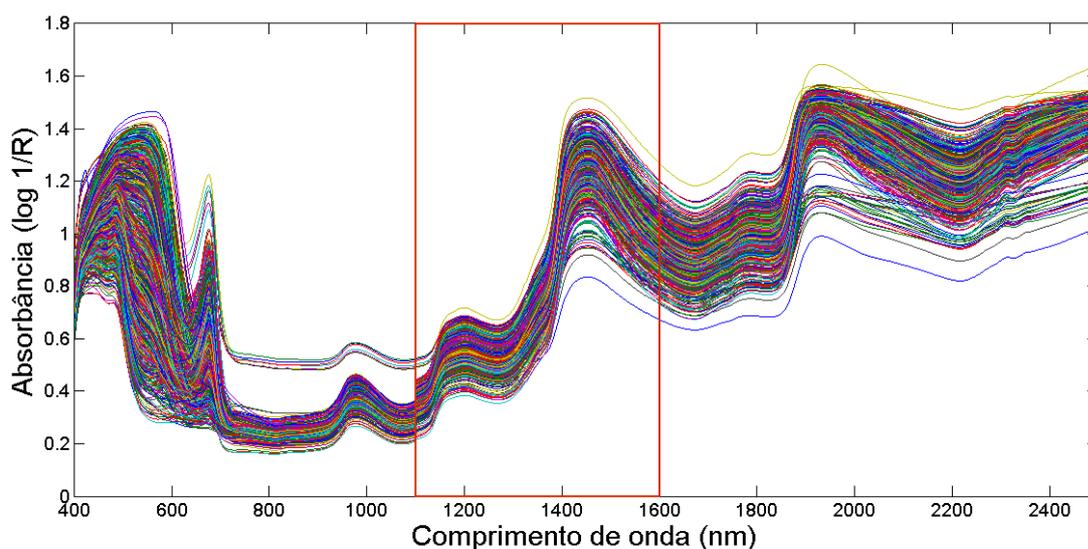
Os cálculos relacionados às etapas de pré-processamento, seleção de amostras, seleção de variáveis, construção dos modelos de calibração multivariada e transferência de calibração foram realizados utilizando-se os softwares MATLAB versão 2010a (Mathworks, Natick, MA, EUA) e PLS\_Toolbox versão 8.2 (Eigenvector Research, Manson, WA, EUA). A interface gráfica para MATLAB “SPA\_GUI” (disponível em <http://www.ele.ita.br/~kawakami/spa/>), desenvolvida pelo professor Dr. Roberto Kawakami (Departamento de Engenharia Eletrônica - ITA) e equipe, foi utilizada para efetuar a seleção de variáveis pelo algoritmo SPA.

## 4.4 RESULTADOS E DISCUSSÃO

### 4.4.1 Atribuição de bandas espectrais

Conforme discutido anteriormente no Capítulo 3 (item 3.4.1.2.1), as frutas apresentam em sua composição cerca de 80 a 90% de água, sendo a fração restante constituída por diferentes compostos, tais como carboidratos, proteínas, ácidos orgânicos e outros compostos secundários (GÓMEZ, et al., 2006). Assim, seus espectros na região do NIR apresentam bandas de absorção largas e complexas, resultantes da sobreposição de bandas correspondentes a sobretons e combinações de modos vibracionais

fundamentais, principalmente atribuídas às ligações O-H, C-H, N-H e S-H (GOLIC, et al., 2003; NICOLAÏ et al., 2009; MAGWAZA et al., 2012). Estas características podem ser visualizadas na Figura 38, onde são mostrados os espectros de absorvância (sem pré-processamento) de 1861 maçãs intactas (provenientes dos 5 projetos utilizados), registrados por meio do espectrômetro XDS na faixa entre 400 e 2500 nm. A região destacada corresponde à faixa espectral registrada pelo espectrômetro MicroNIR, entre 1100 e 1600 nm.



**Figura 38** – Espectros de absorvância (sem pré-processamento) de 1861 maçãs intactas (provenientes dos 5 projetos), registrados por meio do espectrômetro XDS na faixa entre 400 e 2500 nm. A região destacada na figura corresponde à faixa espectral registrada pelo espectrômetro MicroNIR, entre 1100 e 1600 nm.

A descrição completa da atribuição de bandas espectrais foi realizada anteriormente no item “3.4.1.2.1 – Atribuição de bandas espectrais”, localizado no Capítulo 3 desta tese (p.79). A interpretação utilizada para o caso das maçãs é a mesma empregada previamente para as mangas.

#### 4.4.2 Análises de referência

Os valores obtidos para os parâmetros de qualidade determinados pelos métodos de referência são mostrados na Tabela 5, assim como alguns parâmetros estatísticos considerados relevantes para o desenvolvimento dos modelos de calibração. É possível

verificar, que as amostras analisadas apresentaram uma variação significativa nos valores dos parâmetros de qualidade avaliados, que pode ser atribuída aos diferentes estádios de maturação e do número de variedades de maçãs utilizadas. Neste caso, o parâmetro SS apresentou a menor variação, exibindo um coeficiente de variação (CV) de 16%, enquanto o parâmetro IS apresentou a maior variação, exibindo um CV de 56%. É importante ressaltar que nem todas as 1861 amostras foram utilizadas para a determinação de cada parâmetro de qualidade, conforme pode ser visualizado na Tabela 5.

Ao longo do amadurecimento dos frutos o valor de SS aumenta, enquanto o valor de IA diminui, sendo que estes parâmetros estão intrinsecamente relacionados ao longo do processo. Durante a fotossíntese, os açúcares sintetizados nas folhas são polimerizados e armazenados nas frutas na forma de amido. Ao longo da maturação, o amido armazenado na polpa da fruta é hidrolisado, transformando-se em açúcares solúveis. Quanto mais madura a fruta, menor o conteúdo de amido e maior o de açúcares (GIRARDI et al., 2004). É importante ressaltar, que a concentração de açúcares em diversas frutas é avaliada pelo teor de SS, já que constituem a maior fração dos sólidos encontrados em solução.

**Tabela 5** – Parâmetros estatísticos relativos aos parâmetros de qualidade determinados pelos métodos de referência.

Parâmetros de qualidade	Projetos <sup>a</sup>	n <sup>b</sup>	Min. <sup>c</sup>	Máx. <sup>d</sup>	Média <sup>e</sup>	DP <sup>f</sup>	CV <sup>g</sup>
Sólidos solúveis (SS) (°Brix)	Todos	1484	7,8	21,4	12,3	2,0	16%
Acidez titulável (AT) (mEq ácido málico L <sup>-1</sup> )	Todos	1484	0,86	28,35	6,67	2,94	44%
Firmeza de Polpa (FP) (kg cm <sup>-2</sup> )	Todos	1440	1,5	12,7	7,3	2,1	29%
Índice de amido (IA) (adimensional)	1, 2, 4 e 5	1222	2	10	8	2	25%
Índice Streif (IS) (kg cm <sup>-2</sup> °Brix <sup>-1</sup> )	4	604	0,04	0,47	0,09	0,05	56%

<sup>a</sup> As amostras analisadas foram provenientes destes projetos; <sup>b</sup> número de amostras utilizadas; <sup>c</sup> valor mínimo para o parâmetro; <sup>d</sup> valor máximo para o parâmetro; <sup>e</sup> valor médio para o parâmetro; <sup>f</sup> desvio padrão; <sup>g</sup> coeficiente de variação.

Os valores dos parâmetros AT e FP tendem a diminuir ao longo do processo de amadurecimento. No caso da AT, a redução ocorre, principalmente, devido ao uso dos ácidos orgânicos como substratos para a respiração do fruto (MEDLICOTT & THOMPSON, 1985). Para a FP, a redução ocorre em função do aumento da atividade de enzimas responsáveis pela degradação da parede celular dos frutos (MITCHAM & MCDONALD, 1992).

O valor do parâmetro IS também diminui ao longo do amadurecimento, já que se trata de uma relação matemática entre os parâmetros SS, FP e IA  $\{IS = [FP/(SS \times IA)]\}$ . O uso do IS é considerado mais seguro do que o uso individual dos parâmetros SS, FP e IA para se determinar o estágio de maturação e o início da colheita de maçãs, sendo utilizado com êxito há mais de duas décadas por produtores do Brasil e de diversos países da Europa (STREIF, 1994; ARGENTA et al., 1995).

Um fator importante que deve ser avaliado previamente à construção dos modelos de calibração é a existência de correlação entre os parâmetros de qualidade determinados por meio das análises de referência. A existência de correlação pode resultar na determinação indireta de um ou mais parâmetros por meio de outro, levando a possíveis interpretações errôneas a respeito dos modelos construídos. Neste caso, é possível construir um modelo de calibração para um parâmetro, que na realidade não está sendo detectado pelo espectrômetro, mas possui correlação com outro parâmetro que é detectado pelo instrumento.

Em geral, não foi verificada correlação entre os parâmetros de qualidade avaliados neste trabalho, exceto entre IA e IS, conforme pode ser visualizado na matriz de correlação mostrada na Tabela 6.

**Tabela 6** – Matriz de correlação entre os parâmetros de qualidade determinados pelos métodos de referência.

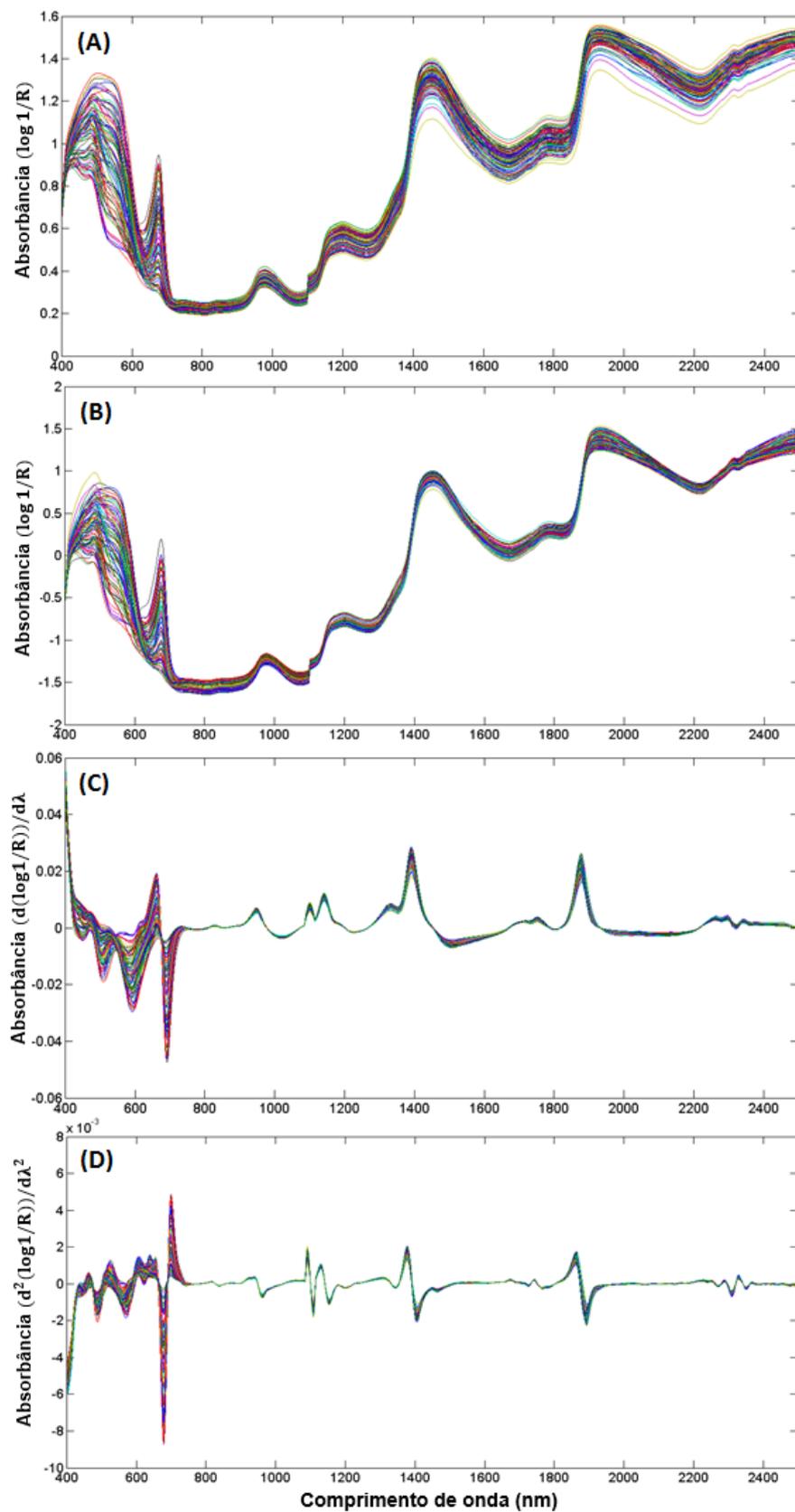
	SS <sup>a</sup>	AT <sup>b</sup>	FP <sup>c</sup>	IA <sup>d</sup>	IS <sup>e</sup>
SS	1.00				
AT	0.15	1.00			
FP	-0.08	0.29	1.00		
IA	0.17	-0.10	-0.29	1.00	
IS	-0.35	0.14	0.27	-0.81	1.00

<sup>a</sup>Sólidos solúveis; <sup>b</sup>acidez titulável; <sup>c</sup> firmeza de polpa; <sup>d</sup> índice de amido; <sup>e</sup> índice Streif.

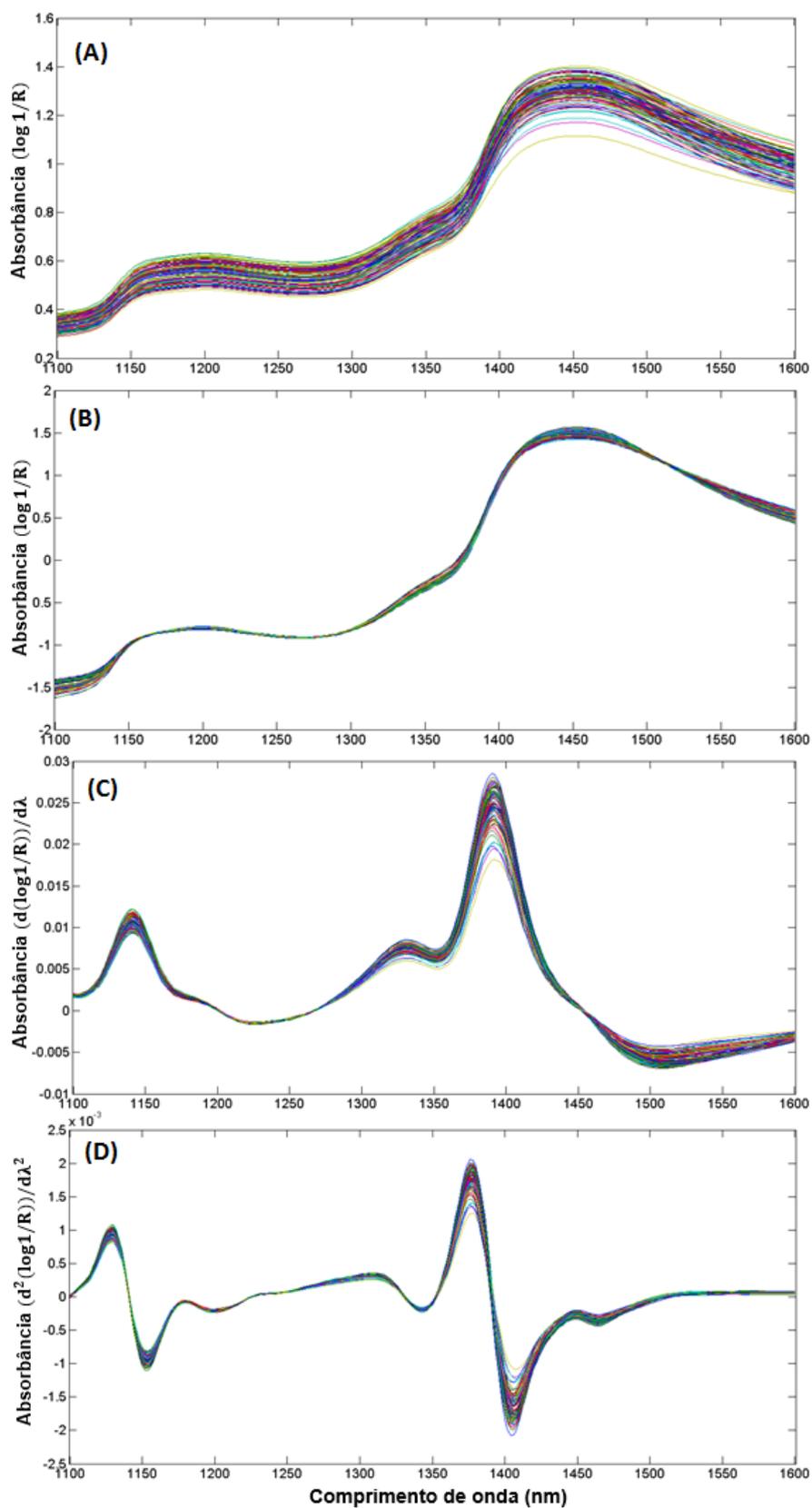
No caso dos parâmetros IA e IS a correlação é esperada, pois ambos se referem ao amadurecimento ou estágio de maturação dos frutos e, de algum modo, estão relacionados com a concentração de amido. Para o IA, a concentração de amido é determinada indiretamente por meio de uma mudança de coloração induzida por uma reação química entre amido e iodo; e para o IS, utiliza-se uma equação para a obtenção do resultado, sendo que um dos termos desta equação é o próprio valor do parâmetro IA. Logo, a avaliação dos modelos de calibração construídos para estimar estes parâmetros deve ser realizada de forma criteriosa, a fim de evitar interpretações equivocadas.

#### 4.4.3 Estudo 1: Comparação entre o desempenho dos espectrômetros XDS e MicroNIR para a determinação de parâmetros de qualidade de maçãs

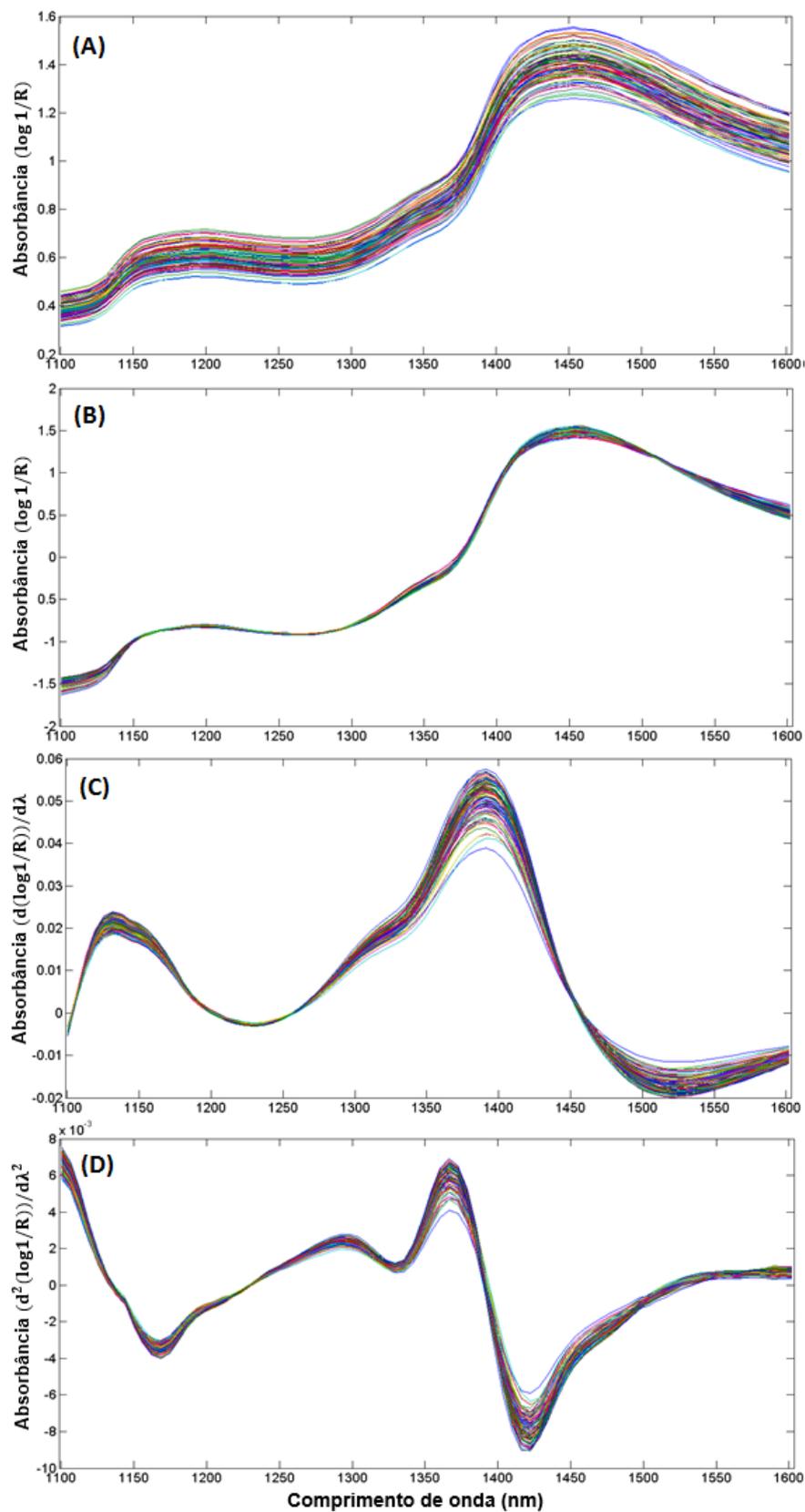
Neste estudo foi utilizado um conjunto de amostras pertencentes ao projeto 4 ( $n = 93$ ), que foram analisadas com os espectrômetros XDS e MicroNIR. Os parâmetros de qualidade avaliados foram SS, AT, FP, IA e IS. Na Figura 39-A, B, C e D são mostrados os espectros registrados com o espectrômetro XDS na faixa espectral de 400 a 2500 nm e submetidos a diferentes técnicas de pré-processamento. De maneira análoga, nas Figuras 40-A, B, C, D e Figura 41-A, B, C, D são mostrados os espectros registrados com os espectrômetros XDS e MicroNIR, respectivamente, na faixa espectral de 1100 a 1600 nm e submetidos a diferentes técnicas de pré-processamento. Em todos os casos foram utilizadas as seguintes técnicas para pré-processamento espectral: SNV, primeira derivada com filtro de Savitzky-Golay utilizando polinômio de 2ª ordem e janela de 15 pontos (1DSG\_P2\_J15), e segunda derivada com filtro de Savitzky-Golay utilizando polinômio de 2ª ordem e janela de 15 pontos (2DSG\_P2\_J15).



**Figura 39** – Espectros registrados com o espectrômetro XDS na faixa espectral de 400 a 2500 nm e submetidos a diferentes técnicas de pré-processamento: (A) sem pré-processamento; (B) SNV; (C) 1DSG\_P2\_J15; e (D) 2DSG\_P2\_J15.



**Figura 40** – Espectros registrados com o espectrômetro XDS na faixa espectral de 1100 a 1600 nm e submetidos a diferentes técnicas de pré-processamento: (A) sem pré-processamento; (B) SNV; (C) 1DSG\_P2\_J15; e (D) 2DSG\_P2\_J15.



**Figura 41** – Espectros registrados com o espectrômetro MicroNIR na faixa espectral de 1100 a 1600 nm e submetidos a diferentes técnicas de pré-processamento: (A) sem pré-processamento; (B) SNV; (C) 1DSG\_P2\_J15; e (D) 2DSG\_P2\_J15.

**Tabela 7** – Resultados obtidos na etapa de validação interna para os parâmetros de qualidade de maçãs avaliados utilizando o espectrômetro XDS (em duas faixas espectrais distintas) e o espectrômetro MicroNIR. Foram utilizadas as seguintes técnicas de pré-processamento espectral: SNV; primeira derivada com filtro de Savitzky-Golay utilizando polinômio de 2ª ordem e janela de 15 pontos (1DSG\_P2\_J15); e segunda derivada com filtro de Savitzky-Golay utilizando polinômio de 2ª ordem e janela de 15 pontos (2DSG\_P2\_J15).

Parâmetro	Instrumento / Faixa espectral	Pré-processamento	Parâmetros estatísticos (etapa de validação interna)							
			Min. <sub>a</sub>	Max. <sub>b</sub>	n <sup>c</sup>	Outliers <sup>d</sup>	VL <sub>e</sub>	RMSE CV	R <sup>2</sup>	Bias
SS	XDS espectro completo	SNV	9,4	16,2	91	2 (2,2%)	5	0,56	0,91	-0,005
		1DSG_P2_J15	9,4	16,2	90	3 (3,3%)	5	0,56	0,91	-0,005
		2DSG_P2_J15	9,4	16,2	90	3 (3,3%)	5	0,57	0,91	-0,010
	XDS 1100-1600 nm	SNV	9,4	16,2	90	3 (3,3%)	7	0,62	0,89	0,010
		1DSG_P2_J15	9,4	16,2	90	3 (3,3%)	7	0,64	0,88	-0,011
		2DSG_P2_J15	9,4	16,2	90	3 (3,3%)	7	0,58	0,90	-0,005
	MicroNIR 1100-1600 nm	SNV	9,4	17,1	90	3 (3,3%)	6	0,57	0,91	0,016
		1DSG_P2_J15	9,4	16,2	91	2 (2,2%)	7	0,64	0,88	0,014
		2DSG_P2_J15	9,4	16,2	91	2 (2,2%)	7	0,61	0,89	0,015
AT	XDS espectro completo	SNV	4,5	12,1	87	6 (6,9%)	10	0,83	0,76	0,012
		1DSG_P2_J15	4,5	12,1	88	5 (5,7%)	11	0,77	0,79	0,000
		2DSG_P2_J15	4,5	11,2	87	6 (6,9%)	13	0,73	0,80	-0,011
	XDS 1100-1600 nm	SNV	4,5	12,1	88	5 (5,7%)	9	0,88	0,71	0,016
		1DSG_P2_J15	4,5	11,2	88	5 (5,7%)	8	0,90	0,69	0,005
		2DSG_P2_J15	4,5	12,1	87	6 (6,9%)	6	0,85	0,76	-0,030
	MicroNIR 1100-1600 nm	SNV	4,5	11,1	87	6 (6,9%)	9	0,80	0,76	0,025
		1DSG_P2_J15	4,5	11,2	88	5 (5,7%)	14	0,83	0,74	0,009
		2DSG_P2_J15	5,1	12,5	88	5 (5,7%)	7	0,95	0,67	0,005
FP	XDS espectro completo	Nenhum	5,80	10,10	71	2 (2,8%)	12	0,60	0,63	0,007
		SNV	5,80	10,45	69	4 (5,8%)	10	0,58	0,69	-0,002
		1DSG_P2_J15	5,80	10,45	71	2 (2,8%)	11	0,53	0,71	-0,012
		2DSG_P2_J15	5,80	10,45	72	1 (1,4%)	14	0,58	0,68	-0,016
	XDS 1100-1600 nm	Nenhum	5,80	10,45	71	2 (2,8%)	6	0,52	0,76	-0,001
		SNV	5,80	10,10	72	1 (1,4%)	9	0,49	0,74	0,010
		1DSG_P2_J15	5,80	10,10	72	1 (1,4%)	7	0,51	0,73	0,003
		2DSG_P2_J15	5,80	10,10	72	1 (1,4%)	8	0,48	0,76	0,004
	MicroNIR 1100-1600 nm	Nenhum	5,80	10,45	71	2 (2,8%)	7	0,55	0,72	0,008
SNV		5,80	10,45	70	3 (4,3%)	9	0,50	0,75	-0,020	
1DSG_P2_J15		5,80	9,65	69	4 (5,5%)	6	0,49	0,71	0,003	
		2DSG_P2_J15	5,80	10,10	70	3 (4,3%)	8	0,51	0,73	0,006

<sup>a</sup> Valor mínimo para o parâmetro; <sup>b</sup> valor máximo para o parâmetro; <sup>c</sup> número de amostras utilizadas; <sup>d</sup> número de amostras anômalas e porcentagem em relação ao número total de amostras; <sup>e</sup> número de variáveis latentes utilizadas para construção dos modelos.

**Tabela 7 (continuação)** – Resultados obtidos na etapa de validação interna para os parâmetros de qualidade de maçãs avaliados utilizando o espectrômetro XDS (em duas faixas espectrais distintas) e o espectrômetro MicroNIR. Foram utilizadas as seguintes técnicas de pré-processamento espectral: SNV; primeira derivada com filtro de Savitzky-Golay utilizando polinômio de 2ª ordem e janela de 15 pontos (1DSG\_P2\_J15); e segunda derivada com filtro de Savitzky-Golay utilizando polinômio de 2ª ordem e janela de 15 pontos (2DSG\_P2\_J15).

Parâmetro	Instrumento / Faixa espectral	Pré-processamento	Parâmetros estatísticos (etapa de validação interna)							
			Min. a	Max. <sup>b</sup>	n <sup>c</sup>	Outliers <sup>d</sup>	VL <sup>e</sup>	RMSE CV	R <sup>2</sup>	Bias
IA	XDS espectro completo	SNV	2	10	91	2 (2,2%)	8	0,98	0,86	-0,034
		1DSG_P2_J15	2	10	91	2 (2,2%)	7	0,80	0,91	0,003
		2DSG_P2_J15	2	10	91	2 (2,2%)	5	0,94	0,87	0,014
	XDS 1100-1600 nm	SNV	2	10	88	5 (5,7%)	10	0,94	0,76	-0,004
		1DSG_P2_J15	2	10	88	5 (5,7%)	9	0,93	0,88	-0,007
		2DSG_P2_J15	2	10	91	2 (2,2%)	8	0,92	0,88	-0,011
	MicroNIR 1100-1600 nm	SNV	2	10	88	5 (5,7%)	7	1,03	0,85	0,018
		1DSG_P2_J15	2	10	89	4 (4,3%)	8	0,99	0,86	-0,002
		2DSG_P2_J15	2	10	90	3 (3,2%)	8	0,95	0,87	-0,013
IS	XDS espectro completo	SNV	0,04	0,17	73	0 (0,0%)	3	0,02	0,55	0,000
		1DSG_P2_J15	0,04	0,14	70	3 (4,1%)	4	0,02	0,66	0,000
		2DSG_P2_J15	0,04	0,14	70	3 (4,1%)	3	0,02	0,66	-0,001
	XDS 1100-1600 nm	SNV	0,04	0,17	73	0 (0,0%)	7	0,02	0,47	0,000
		1DSG_P2_J15	0,04	0,17	72	1 (1,4%)	9	0,02	0,53	0,000
		2DSG_P2_J15	0,04	0,14	71	2 (2,7%)	9	0,02	0,60	0,000
	MicroNIR 1100-1600 nm	SNV	0,04	0,16	72	1 (1,4%)	10	0,02	0,57	0,000
		1DSG_P2_J15	0,04	0,16	72	1 (1,4%)	12	0,02	0,64	0,000
		2DSG_P2_J15	0,04	0,16	72	1 (1,4%)	9	0,02	0,65	0,000

<sup>a</sup> Valor mínimo para o parâmetro; <sup>b</sup> valor máximo para o parâmetro; <sup>c</sup> número de amostras utilizadas; <sup>d</sup> número de amostras anômalas e porcentagem em relação ao número total de amostras; <sup>e</sup> número de variáveis latentes utilizadas para construção dos modelos.

Foi realizada uma comparação entre os valores de RMSECV fornecidos pelos modelos para cada parâmetro de qualidade avaliado. A comparação foi realizada considerando os dois equipamentos e suas respectivas faixas espectrais (XDS espectro completo; XDS 1100-1600 nm; MicroNIR 1100-1600 nm), bem como as diferentes técnicas de pré-processamento utilizadas (sem pré-processamento (somente para o parâmetro FP); SNV; 1DSG\_P2\_J15; 2DSG\_P2\_J15). Assim, foram realizadas 36 comparações para os parâmetros SS, AT, IA e IS, e 66 comparações para o parâmetro FP.

Após aplicação de teste F (ao nível de 95% de confiança), verificou-se que não há diferença significativa entre os valores de RMSECV provenientes dos modelos obtidos para cada parâmetro de qualidade em função do equipamento/faixa espectral ou tipo de pré-processamento utilizado. Logo, pode-se concluir que o espectrômetro MicroNIR oferece desempenho similar ao espectrômetro de bancada para a determinação dos parâmetros de qualidade de maçãs avaliados neste estudo.

#### **4.4.4 Estudo 2: Transferência de calibração entre os espectrômetros NIR de bancada (XDS) e portátil (MicroNIR)**

Os resultados obtidos neste estudo foram divididos em duas partes e são apresentados a seguir nos itens “4.4.4.1 – Modelos de calibração multivariada” e “4.4.4.2 – Transferência de calibração”.

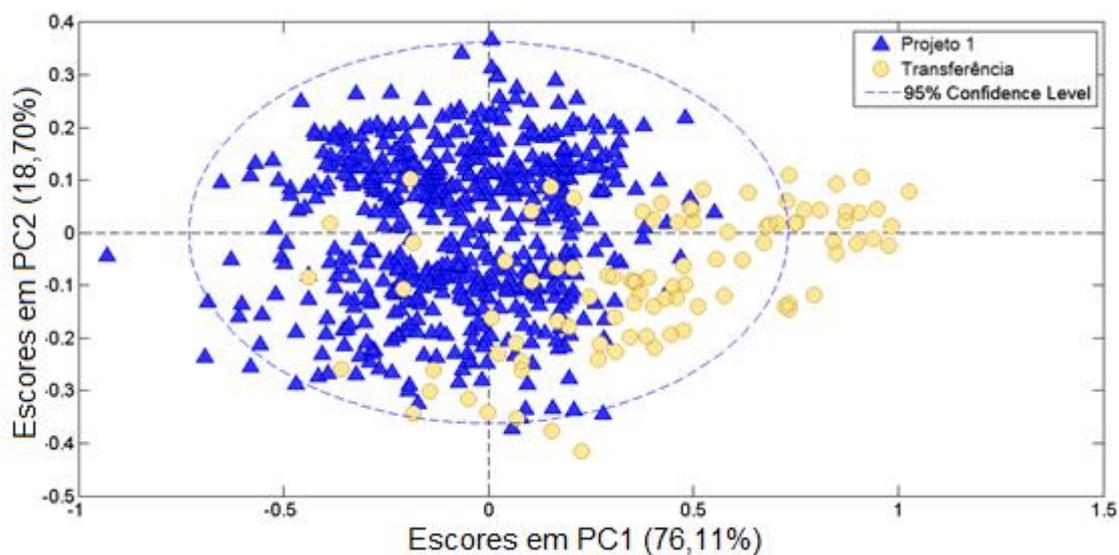
##### **4.4.4.1 Modelos de calibração multivariada**

Nesta parte do estudo foi utilizado um conjunto abrangente de amostras ( $n = 1861$ ) pertencentes aos projetos 1, 2, 3, 4 e 5 que foram analisadas com o espectrômetro XDS. Inicialmente, efetuou-se a exclusão das amostras que apresentaram menor similaridade espectral em relação ao conjunto de amostras utilizadas no estudo de transferência de calibração (93 maçãs analisadas nos 2 espectrômetros). Este procedimento foi realizado, pois os modelos de calibração desenvolvidos nesta etapa do trabalho seriam utilizados posteriormente no estudo sobre transferência de calibração (uma explicação mais detalhada sobre a necessidade de realização deste procedimento foi apresentada anteriormente no item “4.3.5 – Estudo 2: Transferência de calibração entre os espectrômetros NIR de bancada (XDS) e portátil (MicroNIR)”).

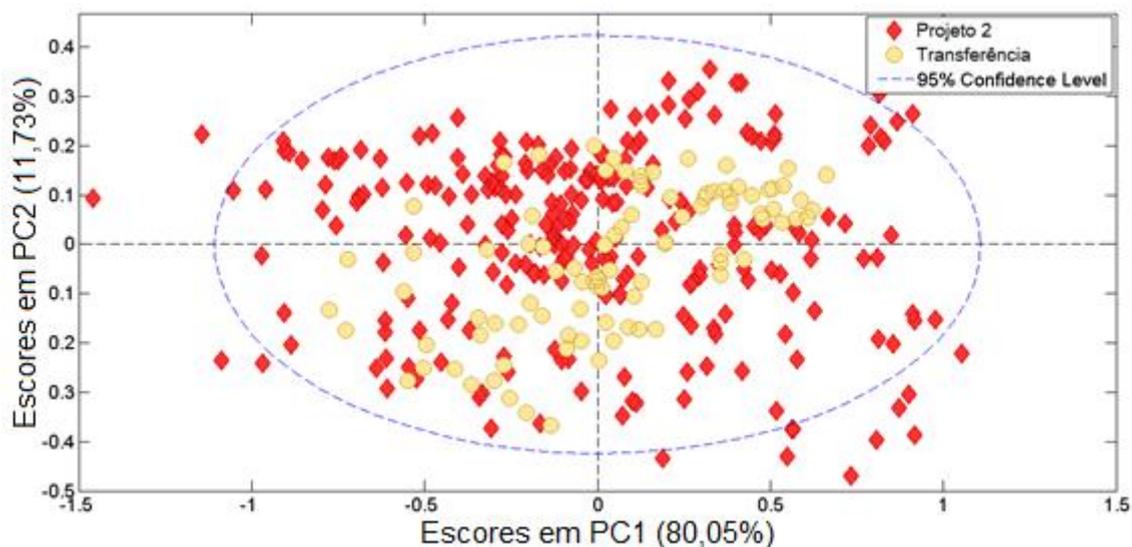
O procedimento de seleção e exclusão das amostras foi realizado por meio de 5 análises de componentes principais utilizando os espectros pré-processados por SNV, de modo que em cada análise foram avaliadas as amostras do conjunto de transferência e as amostras de um projeto específico (projetos 1, 2, 3, 4 e 5 mostrados na Tabela 4). Não foi realizada uma única PCA incluindo todas as 1861 amostras, pois problemas na visualização dos dados dificultaram a aplicação adequada do procedimento.

Nas Figuras 42, 43, 44, 45 e 46 são mostrados os gráficos de escores em PC1 e PC2 relativos às PCA realizadas para seleção e exclusão de amostras que apresentaram

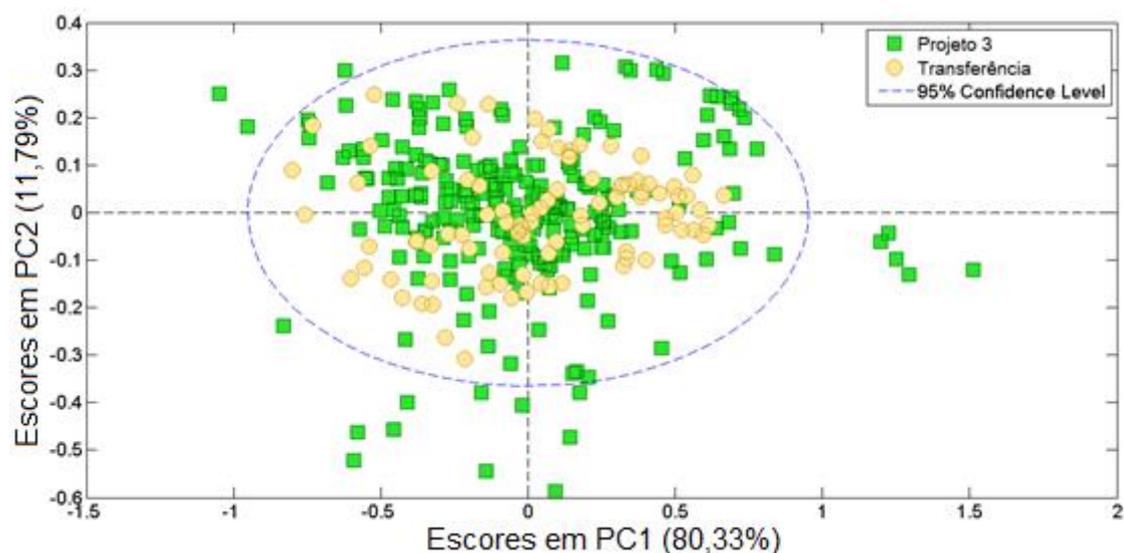
menor similaridade espectral em relação ao conjunto de amostras de transferência para os projetos 1, 2, 3, 4 e 5, respectivamente. Nestes gráficos, a PC3 não foi incluída, pois em todos os casos esta componente explicou menos de 5% da variância dos dados, não sendo utilizada para a realização deste procedimento. Os gráficos dos *loadings* relativos a cada PC também não foram apresentados, pois o procedimento de exclusão das amostras foi baseado unicamente na distribuição relativa entre os escores provenientes das amostras pertencentes aos projetos 1, 2, 3, 4 e 5 e os escores provenientes das amostras do conjunto de transferência.



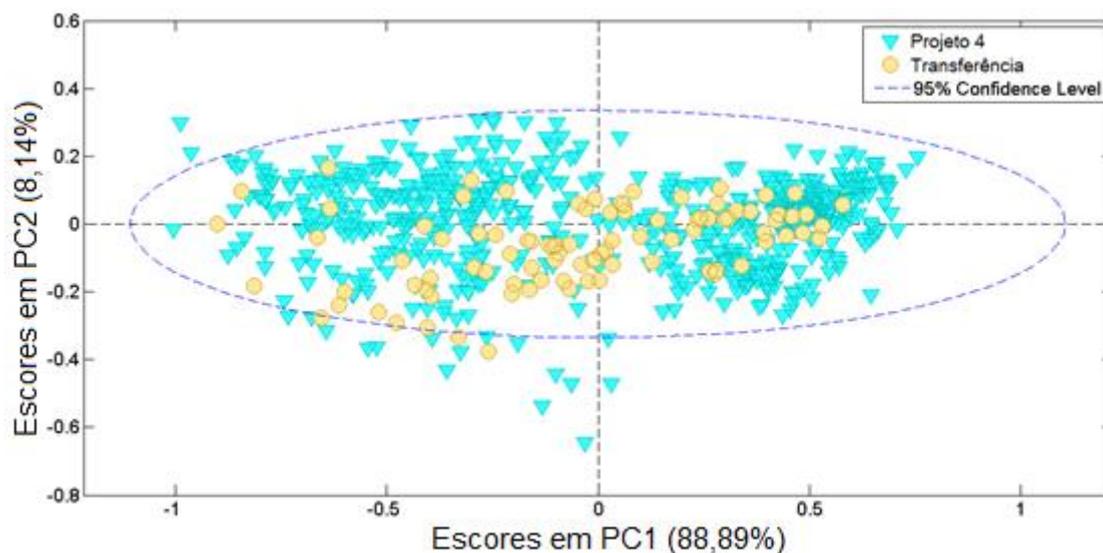
**Figura 42** – Gráficos de escores em PC1 e PC2 relativos à PCA realizada para seleção e exclusão de amostras que apresentaram menor similaridade espectral em relação ao conjunto de amostras de transferência. Esta PCA foi realizada com as amostras provenientes do projeto 1 (símbolos na cor azul) e do conjunto de transferência (símbolos na cor bege).



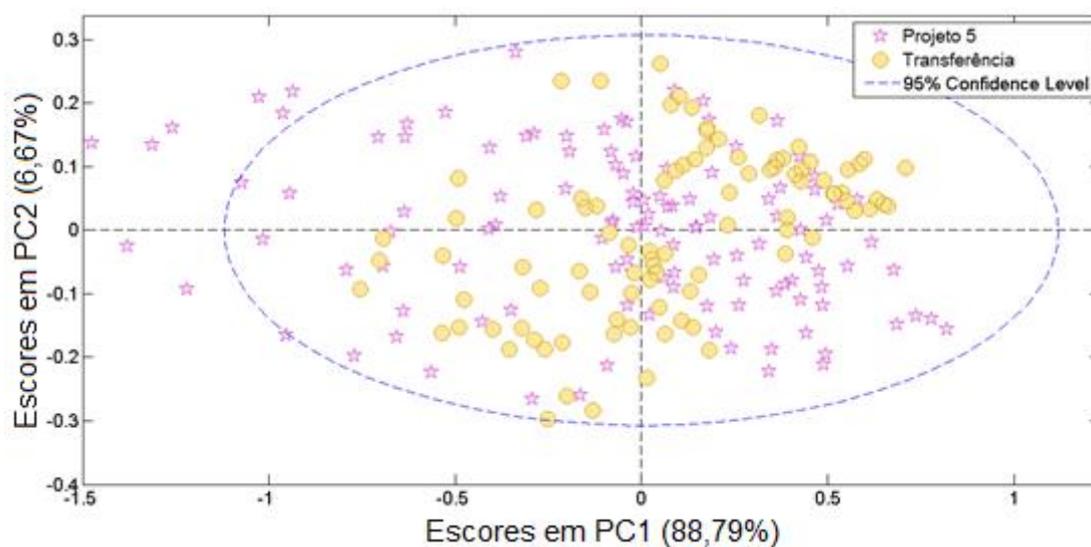
**Figura 43** – Gráficos de escores em PC1 e PC2 relativos à PCA realizada para seleção e exclusão de amostras que apresentaram menor similaridade espectral em relação ao conjunto de amostras de transferência. Esta PCA foi realizada com as amostras provenientes do projeto 2 (símbolos na cor vermelha) e do conjunto de transferência (símbolos na cor bege).



**Figura 44** – Gráficos de escores em PC1 e PC2 relativos à PCA realizada para seleção e exclusão de amostras que apresentaram menor similaridade espectral em relação ao conjunto de amostras de transferência. Esta PCA foi realizada com as amostras provenientes do projeto 3 (símbolos na cor verde) e do conjunto de transferência (símbolos na cor bege).



**Figura 45** – Gráficos de escores em PC1 e PC2 relativos à PCA realizada para seleção e exclusão de amostras que apresentaram menor similaridade espectral em relação ao conjunto de amostras de transferência. Esta PCA foi realizada com as amostras provenientes do projeto 4 (símbolos na cor azul) e do conjunto de transferência (símbolos na cor bege).

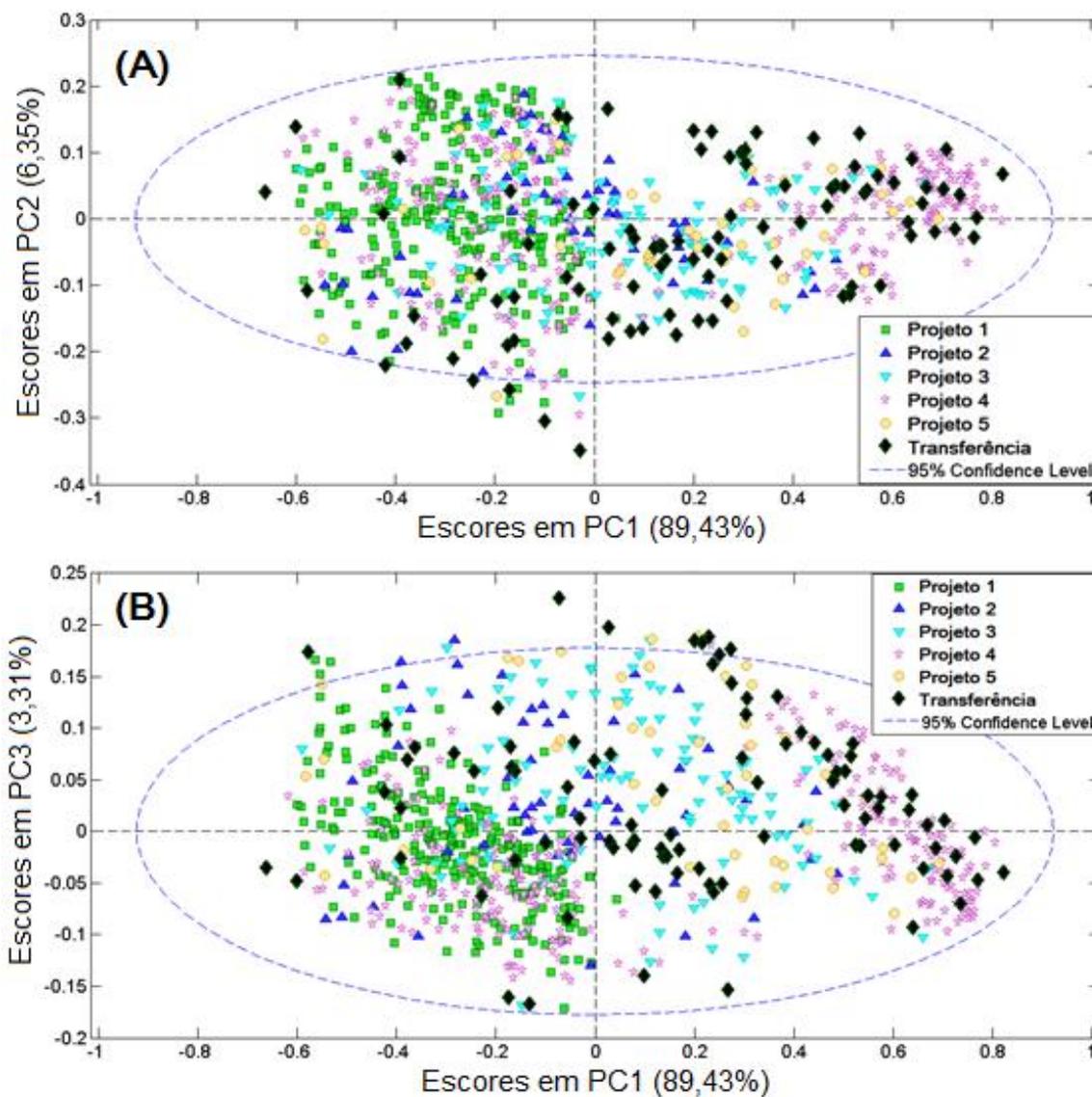


**Figura 46** – Gráficos de escores em PC1 e PC2 relativos à PCA realizada para seleção e exclusão de amostras que apresentaram menor similaridade espectral em relação ao conjunto de amostras de transferência. Esta PCA foi realizada com as amostras provenientes do projeto 5 (símbolos na cor lilás) e do conjunto de transferência (símbolos na cor bege).

Ao final do processo de exclusão o número de amostras foi reduzido de 1861 para 773. Uma nova PCA foi realizada incluindo todas as amostras remanescentes e aquelas pertencentes ao conjunto de transferência. Assim, foi possível visualizar a distribuição das amostras remanescentes em relação às amostras de transferência, por meio dos gráficos de escores em PC1, PC2 e PC3, e avaliar se o procedimento forneceu o resultado esperado. Na Figura 47-A e B é possível observar, a partir da distribuição dos escores em PC1, PC2 e PC3, que as amostras do conjunto de transferência passaram a representar adequadamente a variabilidade do novo conjunto de amostras ( $n = 773$ ), confirmando a adequação do procedimento adotado.

As amostras remanescentes ( $n = 773$ ) foram utilizadas para construção dos modelos de calibração empregando-se diferentes técnicas de calibração multivariada (PLS, SPA-MLR e LS-SVM) visando à determinação dos parâmetros de qualidade SS, AT, FP, IA e IS. É importante ressaltar que as amostras do conjunto de transferência ( $n = 93$ ) não foram utilizadas na construção dos modelos de calibração e não foram incluídas no conjunto de validação externa, sendo utilizadas novamente apenas na parte do estudo sobre transferência de calibração.

Nas Tabelas 8, 9, 10, 11 e 12 são mostrados parâmetros estatísticos relativos às etapas de calibração e validação externa, empregando-se SNV como pré-processamento espectral e os métodos de regressão PLS, SPA-MLR e LS-SVM para construção dos modelos de calibração multivariada para determinação dos parâmetros SS, AT, FP, IA e IS. Na Figura 48-A, B e C são mostrados gráficos contendo valores obtidos por meio das análises de referência *versus* valores obtidos por meio dos modelos de calibração multivariada (PLS, SPA-MLR e LS-SVM) para o parâmetro SS. De maneira análoga, nas Figuras 49-A, B e C; 50-A, B e C; 51-A, B e C; e 52-A, B e C são mostrados os resultados obtidos para os parâmetros AT, FP, IA e IS, respectivamente.

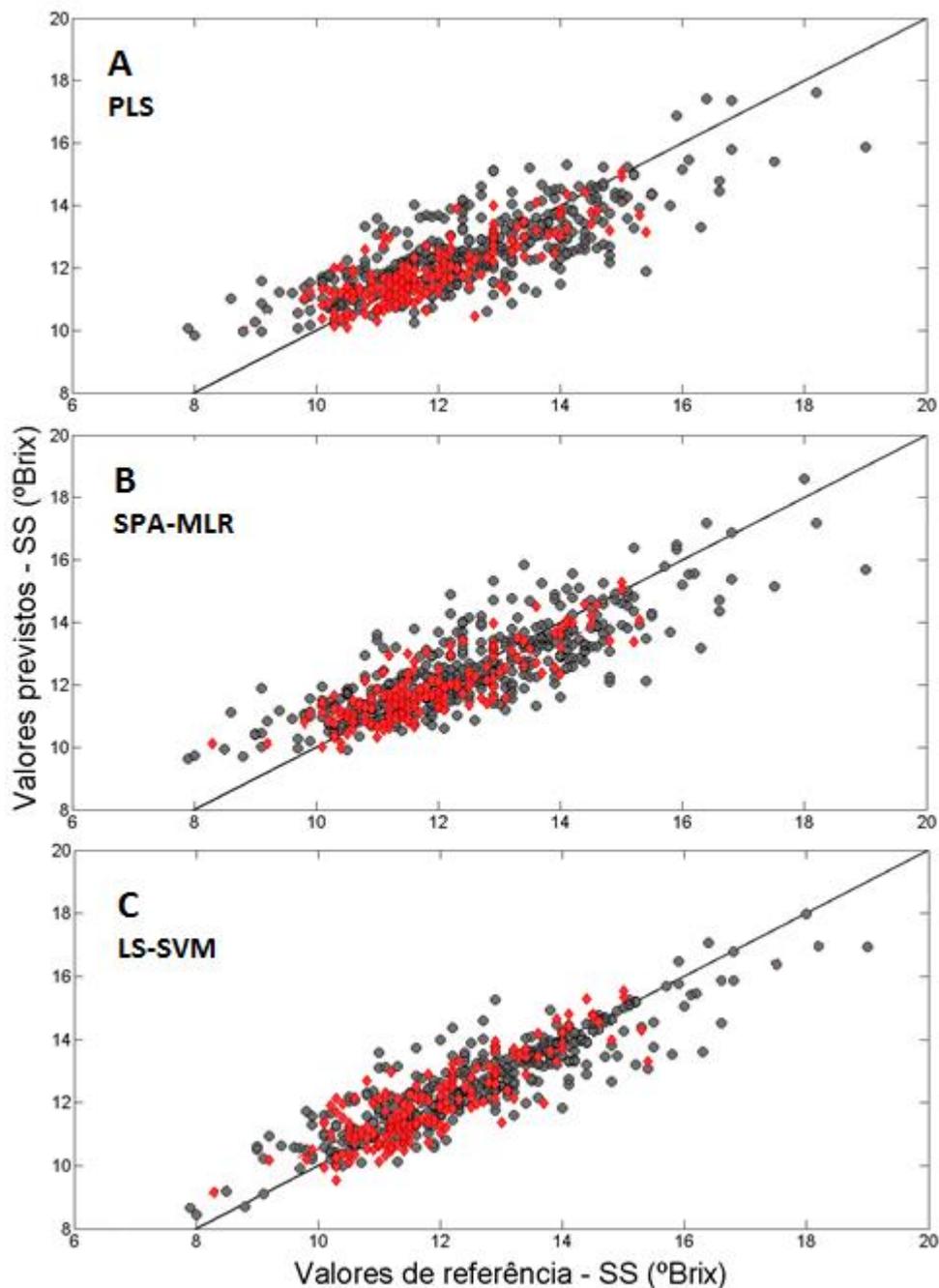


**Figura 47** – Gráficos de escores em PC1 e PC2 (A) e PC1 e PC3 (B) obtidos incluindo-se todas as amostras remanescentes (símbolos de diferentes cores) e as amostras de transferência (símbolos na cor preta), após a exclusão das amostras que apresentaram menor similaridade espectral em relação ao conjunto de amostras de transferência.

**Tabela 8** – Parâmetros estatísticos relativos às etapas de calibração e validação externa para determinação de sólidos solúveis (SS), empregando-se SNV como pré-processamento espectral e os métodos de regressão PLS, SPA-MLR e LS-SVM para construção dos modelos de calibração multivariada.

Método	Faixa (°Brix)	Calibração								Validação externa					
		<i>n</i> <sup>a</sup>	<i>Outliers</i>	<i>V</i> <sup>b</sup>	$\Gamma$	$\sigma^2$	$R^2$	RMSEC	Bias	<i>n</i> <sup>c</sup>	<i>Outliers</i>	$R^2$	RMSEP <sup>d</sup>	Bias	RMSEPr <sup>e</sup>
PLS		396	16 (3,9%)	6	-	-	0,61	1,0	0,000	176	1 (0,6%)	0,65	1,0 a	0,006	6,3%
SPA-MLR	7,8 - 24.1	411	1 (0,2%)	12	-	-	0,64	0,8	0,000	176	1 (0,6%)	0,70	0,9 a	0,003	5,6%
LS-SVM		408	4 (1,0%)	15	10	0,01	0,81	0,7	0,040	176	1 (0,6%)	0,73	0,7 b	0,098	4,4%

<sup>a</sup> Número de amostras utilizadas na etapa de calibração excluindo-se os *outliers*; <sup>b</sup> Número de variáveis utilizadas: para PLS representa o número de variáveis latentes, para SPA-MLR representa o número de variáveis espectrais, e para LS-SVM representa o número de variáveis latentes mantidas no modelo PLS utilizado na etapa de compressão dos dados; <sup>c</sup> Número de amostras utilizadas na etapa de validação externa excluindo-se os *outliers*; <sup>d</sup> Valores de RMSEP com letras iguais não diferem estatisticamente ao nível de 95% de confiança; <sup>e</sup> RMSEP relativo, calculado em relação ao valor médio da faixa do parâmetro de qualidade avaliado.



**Figura 48** – Valores obtidos pelo método de referência *versus* valores previstos pelos modelos de calibração, nas etapas de calibração (símbolos em preto) e validação externa (símbolos em vermelho) para o parâmetro sólidos solúveis (SS), empregando-se SNV como pré-processamento espectral e diferentes técnicas de calibração multivariada: (A) PLS, (B) SPA-MLR, e (C) LS-SVM. A linha preta representa a bissetriz.

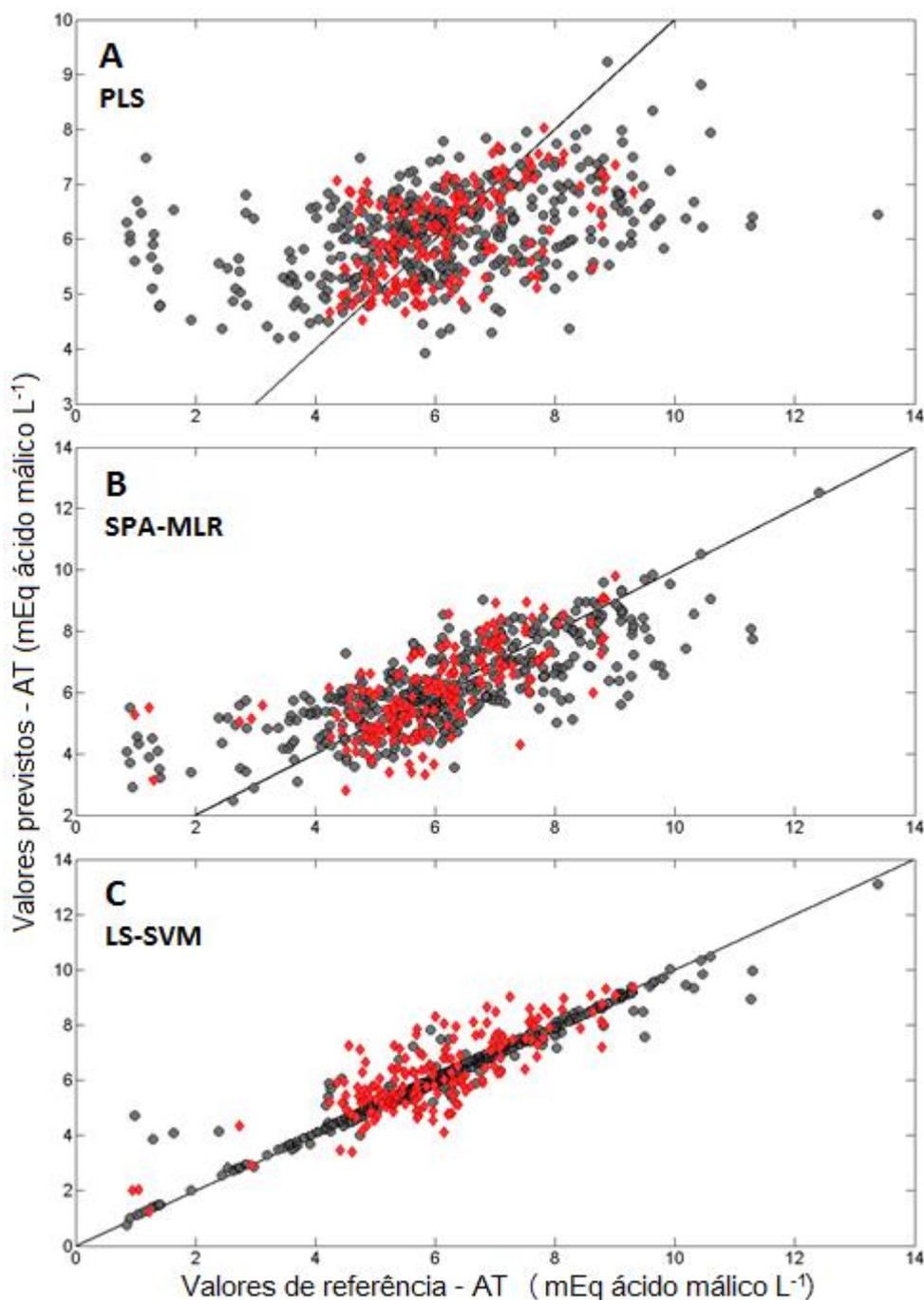
Os modelos de calibração desenvolvidos por PLS, SPA-MLR e LS-SVM para determinação de SS apresentaram desempenho preditivo satisfatório, com valores de RMSEPr (RMSEP relativo, calculado em relação ao valor médio da faixa do parâmetro de qualidade avaliado) variando de 4,4% para LS-SVM até 6,3% para PLS. Os valores de  $R^2$  também foram satisfatórios, variando de 0,73 para LS-SVM até 0,65 para PLS, indicando a adequação dos modelos para explicar a variância total dos dados.

Aplicando-se o teste F ao nível de 95% de confiança para comparação entre os valores de RMSEP obtidos pelos diferentes métodos utilizados para a construção dos modelos de calibração verificou-se que: PLS e SPA-MLR não apresentam diferença estatística significativa; PLS e LS-SVM apresentam diferença estatística significativa; e SPA-MLR e LS-SVM apresentam diferença estatística significativa. Portanto, de modo geral, o modelo desenvolvido por LS-SVM mostrou melhor desempenho preditivo em comparação aos modelos desenvolvidos por PLS e SPA-MLR, exibindo valor de RMSEP significativamente menor.

**Tabela 9** – Parâmetros estatísticos relativos às etapas de calibração e validação externa para determinação de acidez titulável (AT), empregando-se SNV como pré-processamento espectral e os métodos de regressão PLS, SPA-MLR e LS-SVM para construção dos modelos de calibração multivariada.

Método	Faixa (mEq ácido málico L <sup>-1</sup> )	Calibração								Validação externa					
		<i>n</i> <sup>a</sup>	<i>Outliers</i>	<i>V</i> <sup>b</sup>	$\Gamma$	$\sigma^2$	$R^2$	RMSEC	Bias	<i>n</i> <sup>c</sup>	<i>Outliers</i>	$R^2$	RMSEP <sup>d</sup>	Bias	RMSEPr <sup>e</sup>
PLS		407	13 (3,1%)	6	-	-	0,18	1,79	0,000	169	8 (4,5%)	0,24	1,03 a	-0,055	14,4%
SPA-MLR	0,90-13,40	405	15 (3,6%)	30	-	-	0,58	1,24	0,000	172	5 (2,8%)	0,53	1,00 a	-0,042	14,0%
LS-SVM		405	15 (3,6%)	12	10	0,1	0,95	0,44	0,026	175	1 (0,6%)	0,68	0,89 a	0,126	12,4%

<sup>a</sup> Número de amostras utilizadas na etapa de calibração excluindo-se os *outliers*; <sup>b</sup> Número de variáveis utilizadas: para PLS representa o número de variáveis latentes, para SPA-MLR representa o número de variáveis espectrais, e para LS-SVM representa o número de variáveis latentes mantidas no modelo PLS utilizado na etapa de compressão dos dados; <sup>c</sup> Número de amostras utilizadas na etapa de validação externa excluindo-se os *outliers*; <sup>d</sup> Valores de RMSEP com letras iguais não diferem estatisticamente ao nível de 95% de confiança; <sup>e</sup> RMSEP relativo, calculado em relação ao valor médio da faixa do parâmetro de qualidade avaliado.



**Figura 49** – Valores obtidos pelo método de referência *versus* valores previstos pelos modelos de calibração, nas etapas de calibração (símbolos em preto) e validação externa (símbolos em vermelho) para o parâmetro acidez titulável (AT), empregando-se SNV como pré-processamento espectral e diferentes técnicas de calibração multivariada: (A) PLS, (B) SPA-MLR, e (C) LS-SVM. A linha preta representa a bissetriz.

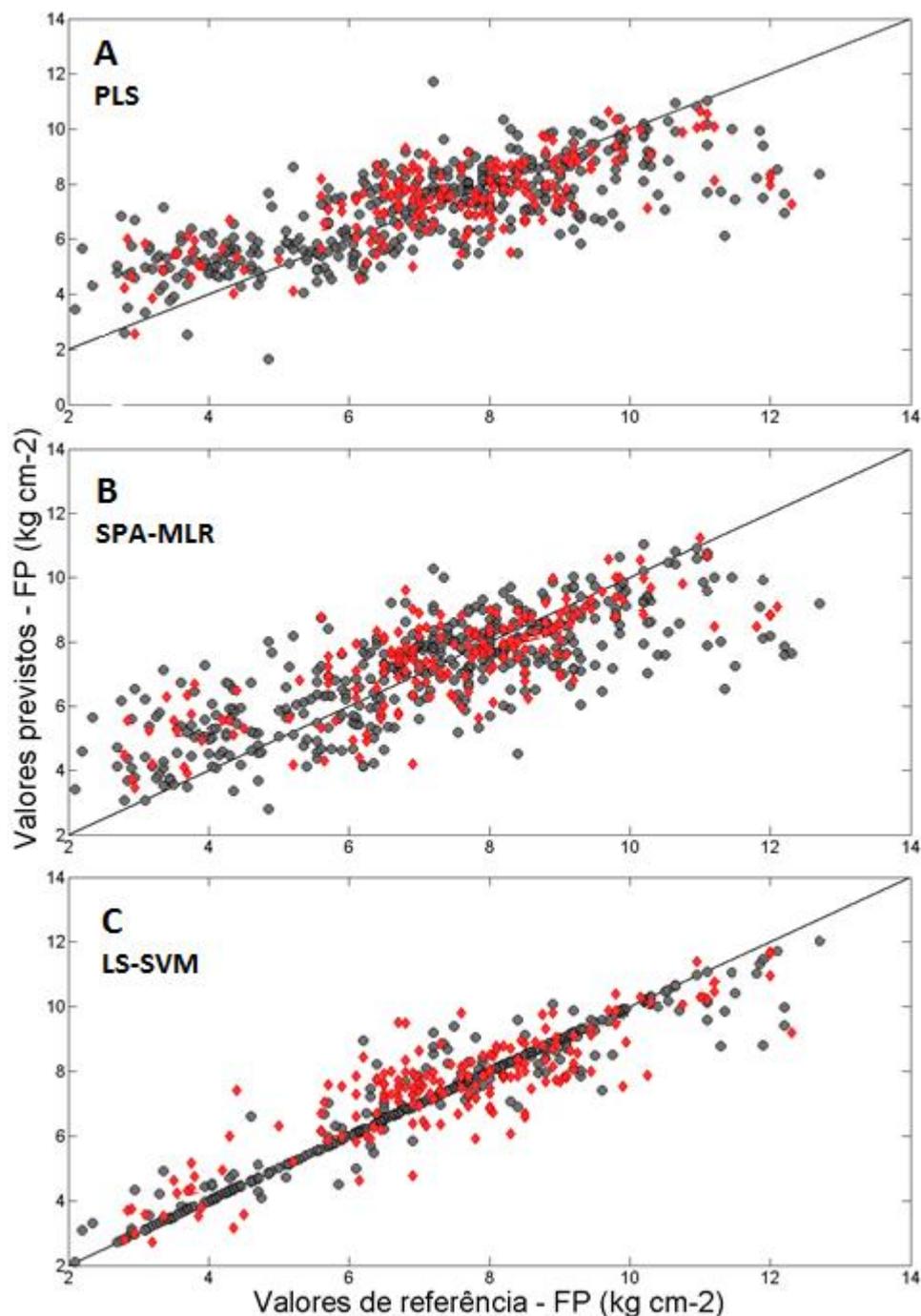
Os modelos de calibração desenvolvidos pelos três métodos para determinação de AT apresentaram desempenho preditivo similares, com valores de RMSEPr variando de 12,4% para LS-SVM até 14,4% para PLS. Os valores de  $R^2$  também foram similares para LS-SVM (0,68) e SPA-MLR (0,53), e mais baixo para o modelo desenvolvido por PLS (0,24).

Aplicando-se o teste F ao nível de 95% de confiança, verificou-se que os valores de RMSEP obtidos pelos diferentes métodos utilizados para a construção dos modelos de calibração para determinação de AT não apresentam diferença estatística significativa. Entretanto, é importante ressaltar, que este resultado deve ser interpretado com cautela, pois os valores obtidos para RMSEP para os modelos PLS e SPA-MLR foram menores que os valores de RMSEC. Este fato pode ter sido causado em função da remoção excessiva de amostras consideradas anômalas do conjunto de validação externa, que pode ter levado a uma melhoria 'artificial' no desempenho preditivo dos modelos construídos por estes dois métodos. Por outro lado, sem a remoção destes *outliers* o desempenho preditivo destes modelos foi insatisfatório, além de resultar em valores de  $R^2$  muito baixos. Assim, optou-se pela eliminação destas amostras. Em geral, numa situação ideal seria esperado que os valores de RMSEP fossem similares aos valores de RMSEC.

**Tabela 10** – Parâmetros estatísticos relativos às etapas de calibração e validação externa para determinação de firmeza de polpa (FP), empregando-se SNV como pré-processamento espectral e os métodos de regressão PLS, SPA-MLR e LS-SVM para construção dos modelos de calibração multivariada.

Método	Faixa (kg cm <sup>-2</sup> )	Calibração								Validação externa					
		<i>n</i> <sup>a</sup>	<i>Outliers</i>	<i>V</i> <sup>b</sup>	$\Gamma$	$\sigma^2$	$R^2$	RMSEC	Bias	<i>n</i> <sup>c</sup>	<i>Outliers</i>	$R^2$	RMSEP <sup>d</sup>	Bias	RMSEPr <sup>e</sup>
PLS		402	13 (3,2%)	10	-	-	0,56	1,47	0,000	177	2 (1,1%)	0,52	1,38 a	0,046	18,6%
SPA-MLR	2,1 - 12,7	415	0 (0,0%)	21	-	-	0,59	1,43	0,000	179	0 (0,0%)	0,55	1,23 a,b	-0,080	16,6%
LS-SVM		413	2 (0,5%)	12	3,16	0,01	0,94	0,54	-0,005	179	0 (0,0%)	0,74	0,99 b	0,084	13,4%

<sup>a</sup> Número de amostras utilizadas na etapa de calibração excluindo-se os *outliers*; <sup>b</sup> Número de variáveis utilizadas: para PLS representa o número de variáveis latentes, para SPA-MLR representa o número de variáveis espectrais, e para LS-SVM representa o número de variáveis latentes mantidas no modelo PLS utilizado na etapa de compressão dos dados; <sup>c</sup> Número de amostras utilizadas na etapa de validação externa excluindo-se os *outliers*; <sup>d</sup> Valores de RMSEP com letras iguais não diferem estatisticamente ao nível de 95% de confiança; <sup>e</sup> RMSEP relativo, calculado em relação ao valor médio da faixa do parâmetro de qualidade avaliado.



**Figura 50** – Valores obtidos pelo método de referência *versus* valores previstos pelos modelos de calibração, nas etapas de calibração (símbolos em preto) e validação externa (símbolos em vermelho) para o parâmetro firmeza de polpa (FP), empregando-se SNV como pré-processamento espectral e diferentes técnicas de calibração multivariada: (A) PLS, (B) SPA-MLR, e (C) LS-SVM. A linha preta representa a bissetriz.

O modelo de calibração desenvolvido para determinação de FP por LS-SVM apresentou desempenho preditivo satisfatório, exibindo RMSEPr com valor de 13,4%, enquanto este parâmetro apresentou valores mais altos para os modelos desenvolvidos por PLS (18,6%) e SPA-MLR (16,6%). Os valores de  $R^2$  foram similares para PLS (0,52) e SPA-MLR (0,55), e superiores para LS-SVM (0,74), indicando a maior capacidade deste método de regressão para explicar de forma satisfatória a variância total dos dados.

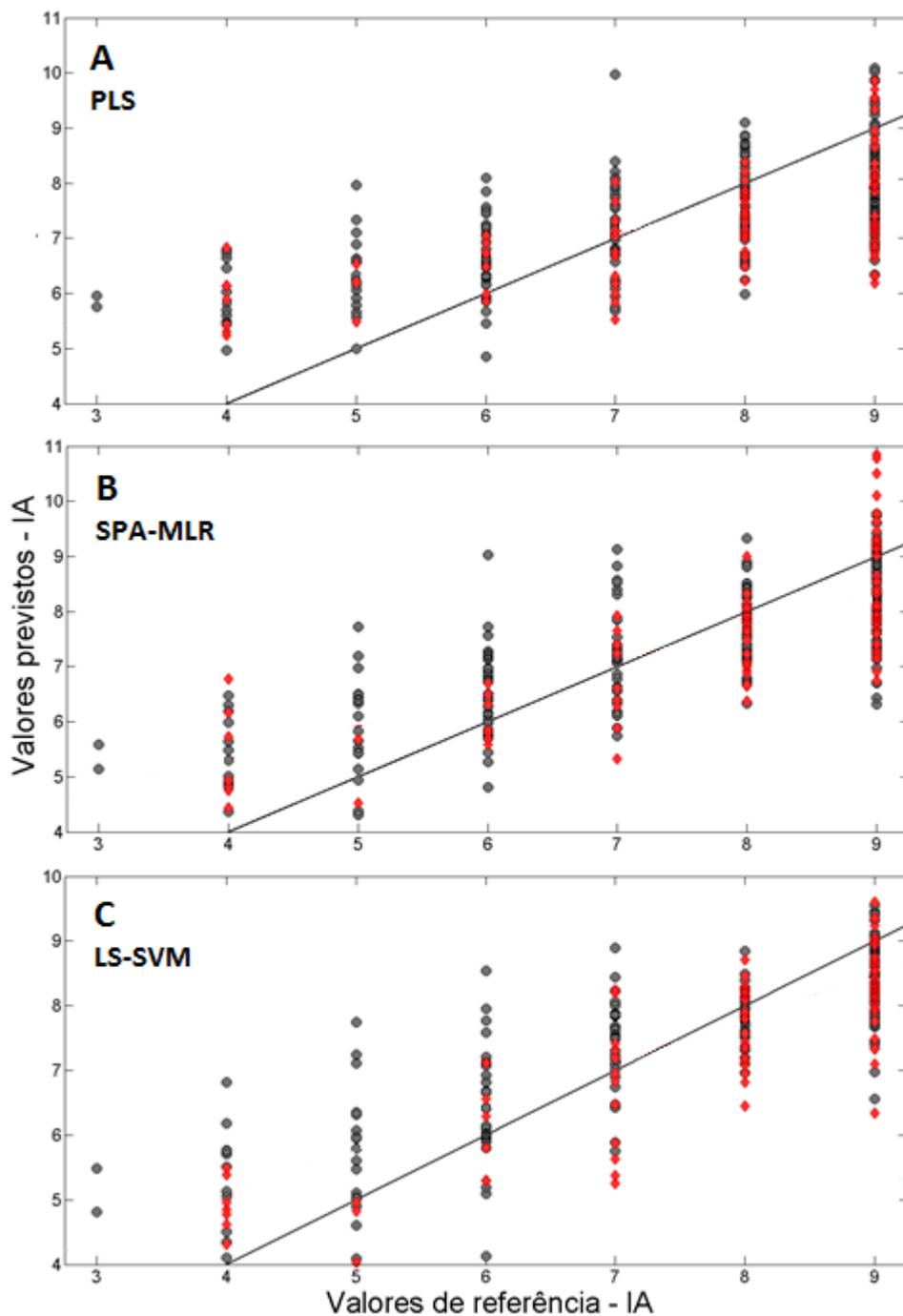
Aplicando-se o teste F ao nível de 95% de confiança para comparação entre os valores de RMSEP obtidos pelos diferentes métodos utilizados para a construção dos modelos de calibração para determinação de FP verificou-se que: PLS e SPA-MLR não apresentam diferença estatística significativa; PLS e LS-SVM apresentam diferença estatística significativa; e SPA-MLR e LS-SVM não apresentam diferença estatística significativa. Contudo, de modo similar ao que foi observado para o parâmetro AT, os valores de RMSEP obtidos para os modelos PLS e SPA-MLR foram menores que os valores de RMSEC. Entretanto, diferentemente do que foi observado para o modelo PLS, não houve remoção de qualquer amostra do conjunto de validação externa para o modelo SPA-MLR. É possível, que neste caso, as amostras selecionadas pelo algoritmo para compor o conjunto de validação externa possam de alguma forma ter favorecido o desempenho do modelo na etapa de predição destes dados.

Outro fato que deve ser destacado refere-se ao uso de SNV para realizar o pré-processamento dos dados espectrais utilizados para construção de modelos de calibração para predição de FP, já que este parâmetro corresponde a uma propriedade física da amostra. Assim, a princípio, os espectros brutos deveriam ser utilizados para a modelagem, evitando a remoção de informações de natureza física contidas nos espectros. Entretanto, o desempenho preditivo dos modelos foi melhor quando SNV foi utilizada com pré-processamento. Este fato sugere, portanto, que a determinação de FP foi realizada indiretamente por meio de informações relativas, predominantemente, à composição química das amostras. Embora relevante, a investigação deste fato está fora do escopo deste capítulo e poderá ser abordada em estudos futuros.

**Tabela 11** – Parâmetros estatísticos relativos às etapas de calibração e validação externa para determinação de índice de amido (IA), empregando-se SNV como pré-processamento espectral e os métodos de regressão PLS, SPA-MLR e LS-SVM para construção dos modelos de calibração multivariada.

Método	Faixa (Adimensional)	Calibração								Validação externa					
		<i>n</i> <sup>a</sup>	<i>Outliers</i>	<i>V</i> <sup>b</sup>	$\Gamma$	$\sigma^2$	$R^2$	RMSEC	Bias	<i>n</i> <sup>c</sup>	<i>Outliers</i>	$R^2$	RMSEP <sup>d</sup>	Bias	RMSEPr <sup>e</sup>
PLS		213	12 (5,3%)	8	-	-	0,44	1,19	0,000	99	1 (1,0%)	0,37	1,30 a	-0,543	21,7%
SPA-MLR	3-9	225	0 (0,0%)	22	-	-	0,57	1,03	0,000	100	0 (0,0%)	0,56	1,18 a	-0,327	19,7%
LS-SVM		225	0 (0,0%)	16	10	0,01	0,72	0,83	0,075	100	0 (0,0%)	0,73	0,84 b	-0,332	14,0%

<sup>a</sup> Número de amostras utilizadas na etapa de calibração excluindo-se os *outliers*; <sup>b</sup> Número de variáveis utilizadas: para PLS representa o número de variáveis latentes, para SPA-MLR representa o número de variáveis espectrais, e para LS-SVM representa o número de variáveis latentes mantidas no modelo PLS utilizado na etapa de compressão dos dados; <sup>c</sup> Número de amostras utilizadas na etapa de validação externa excluindo-se os *outliers*; <sup>d</sup> Valores de RMSEP com letras iguais não diferem estatisticamente ao nível de 95% de confiança; <sup>e</sup> RMSEP relativo, calculado em relação ao valor médio da faixa do parâmetro de qualidade avaliado.



**Figura 51** – Valores obtidos pelo método de referência *versus* valores previstos pelos modelos de calibração, nas etapas de calibração (símbolos em preto) e validação externa (símbolos em vermelho) para o parâmetro índice de amido (IA), empregando-se SNV como pré-processamento espectral e diferentes técnicas de calibração multivariada: (A) PLS, (B) SPA-MLR, e (C) LS-SVM. A linha preta representa a bissetriz.

O modelo de calibração desenvolvido para determinação de IA por LS-SVM apresentou desempenho preditivo satisfatório, exibindo RMSEPr com valor de 14,0%, enquanto este parâmetro apresentou valores mais altos para os modelos desenvolvidos por PLS (21,7%) e SPA-MLR (19,7%). Os valores de  $R^2$  foram crescentes: 0,37 para PLS, 0,56 para SPA-MLR e 0,73 para LS-SVM.

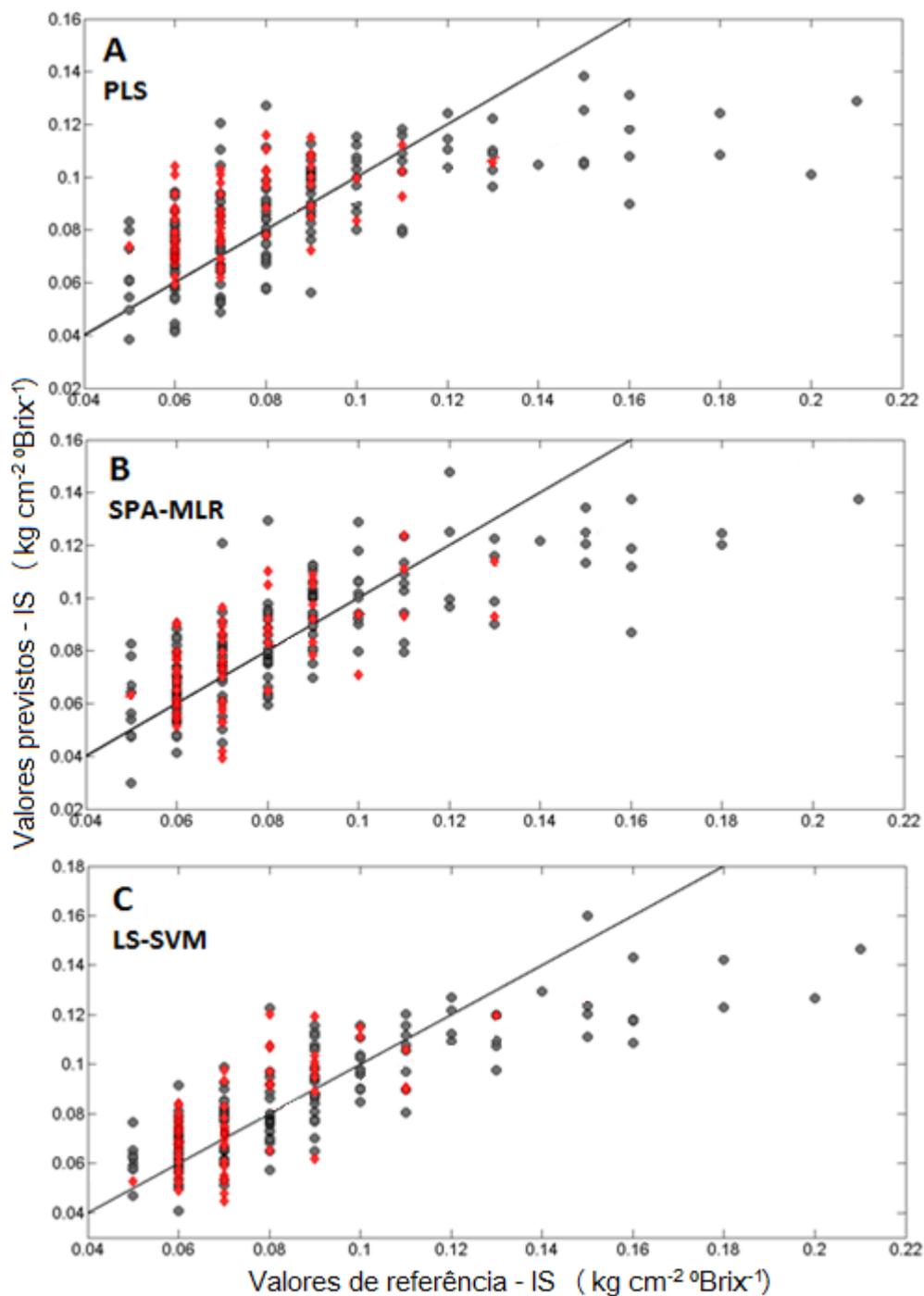
Os altos valores de RMSEPr obtidos para os modelos PLS e SPA-MLR possivelmente inviabilizarão suas aplicações práticas. Uma situação semelhante foi verificada no capítulo 3 (item 3.4.3) desta tese, onde modelos de calibração desenvolvidos para determinação de AT e FP em manga falharam na predição destes parâmetros durante o acompanhamento do amadurecimento dos frutos, sendo este fato atribuído aos altos valores dos erros de predição (RMSEPr de aproximadamente 20%).

Aplicando-se o teste F ao nível de 95% de confiança para comparação entre os valores de RMSEP obtidos pelos diferentes métodos utilizados para a construção dos modelos de calibração para determinação de IA verificou-se que: PLS e SPA-MLR não apresentam diferença estatística significativa; PLS e LS-SVM apresentam diferença estatística significativa; e SPA-MLR e LS-SVM apresentam diferença estatística significativa. Logo, o modelo desenvolvido por LS-SVM mostrou melhor desempenho preditivo em comparação aos modelos desenvolvidos por PLS e SPA-MLR, apresentando valor de RMSEP significativamente menor.

**Tabela 12** – Parâmetros estatísticos relativos às etapas de calibração e validação externa para determinação de índice Streif (IS), empregando-se SNV como pré-processamento espectral e os métodos de regressão PLS, SPA-MLR e LS-SVM para construção dos modelos de calibração multivariada.

Método	Faixa (kg cm <sup>-2</sup> °Brix <sup>-1</sup> )	Calibração								Validação externa					
		<i>n</i> <sup>a</sup>	<i>Outliers</i>	<i>V</i> <sup>b</sup>	$\Gamma$	$\sigma^2$	<i>R</i> <sup>2</sup>	RMSEC	Bias	<i>n</i> <sup>c</sup>	<i>Outliers</i>	<i>R</i> <sup>2</sup>	RMSEP <sup>d</sup>	Bias	RMSEPr <sup>e</sup>
PLS		166	4 (2,4%)	9	-	-	0,46	0,019	0,0000	75	3 (3,8%)	0,35	0,018 a	0,0125	13,8%
SPA-MLR	0,05-0,21	168	2 (1,2%)	15	-	-	0,59	0,013	0,0000	77	1 (1,3%)	0,40	0,015 a	0,0041	11,5%
LS-SVM		166	4 (2,4%)	15	100	0,0001	0,70	0,012	-0,0014	77	1 (1,3%)	0,48	0,014 a	0,0043	10,8%

<sup>a</sup> Número de amostras utilizadas na etapa de calibração excluindo-se os *outliers*; <sup>b</sup> Número de variáveis utilizadas: para PLS representa o número de variáveis latentes, para SPA-MLR representa o número de variáveis espectrais, e para LS-SVM representa o número de variáveis latentes mantidas no modelo PLS utilizado na etapa de compressão dos dados; <sup>c</sup> Número de amostras utilizadas na etapa de validação externa excluindo-se os *outliers*; <sup>d</sup> Valores de RMSEP com letras iguais não diferem estatisticamente ao nível de 95% de confiança; <sup>e</sup> RMSEP relativo, calculado em relação ao valor médio da faixa do parâmetro de qualidade avaliado.



**Figura 52** – Valores obtidos pelo método de referência *versus* valores previstos pelos modelos de calibração, nas etapas de calibração (símbolos em preto) e validação externa (símbolos em vermelho) para o parâmetro índice Streif (IS), empregando-se SNV como pré-processamento espectral e diferentes técnicas de calibração multivariada: (A) PLS, (B) SPA-MLR, e (C) LS-SVM. A linha preta representa a bissetriz.

Os modelos de calibração desenvolvidos por PLS, SPA-MLR e LS-SVM para determinação de IS apresentaram desempenho preditivo satisfatório, com valores de

RMSEPr variando de 10,8% para LS-SVM até 13,8% para PLS. Entretanto, os valores de  $R^2$  foram baixos, variando de 0,35 para PLS até 0,48 para LS-SVM, evidenciando a baixa capacidade dos modelos para explicar a variância total dos dados.

Aplicando-se o teste F ao nível de 95% de confiança, verificou-se que os valores de RMSEP obtidos pelos diferentes métodos utilizados para a construção dos modelos de calibração para determinação de IS não apresentam diferença estatística significativa.

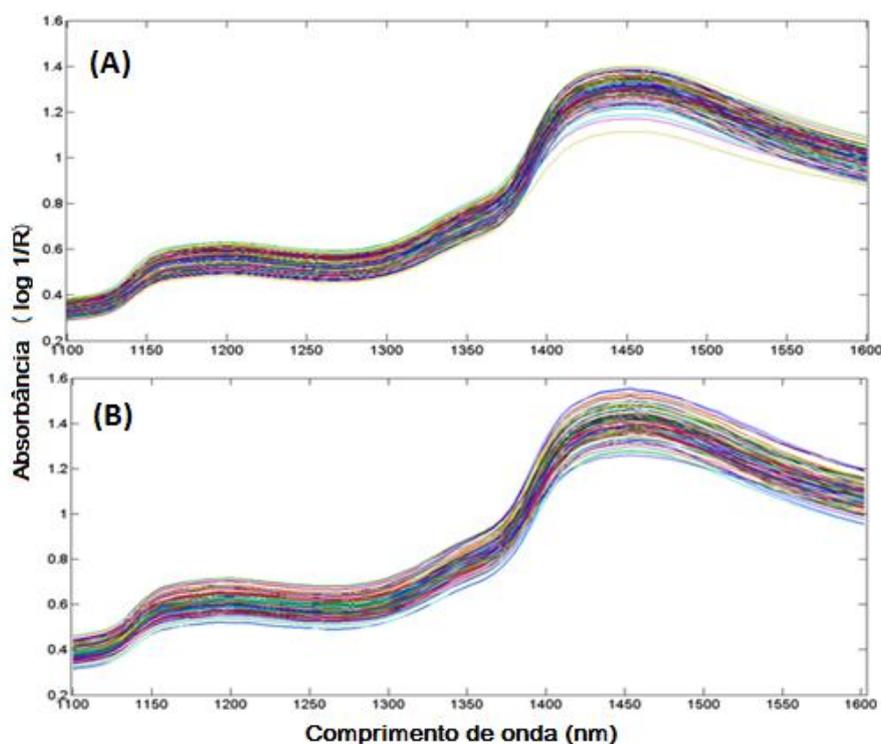
Em relação ao parâmetro estatístico bias, as amostras, em geral, apresentaram-se bem distribuídas ao longo da reta de regressão não havendo indícios de erros sistemáticos significativos para os modelos construídos para determinação de SS, AT e FP, considerando as três técnicas de regressão empregadas. Entretanto, para os parâmetros IA e IS foram observados valores de bias mais elevados, chegando a atingir um valor correspondente a 69% do valor de RMSEP do modelo desenvolvido para determinação de IS por PLS.

De modo geral, pode-se verificar que os modelos desenvolvidos por LS-SVM mostraram melhor desempenho preditivo em comparação aos modelos desenvolvidos por PLS e SPA-MLR, exibindo valores de RMSEP significativamente menores para a maioria dos parâmetros de qualidade avaliados. Adicionalmente, nos casos em que os valores de RMSEP foram considerados estatisticamente equivalentes aos obtidos por PLS e SPA-MLR, os valores de  $R^2$  obtidos por LS-SVM foram sempre mais altos, evidenciando a maior capacidade destes modelos para explicar a variância total dos dados.

O desempenho, em geral, superior dos modelos de calibração desenvolvidos por LS-SVM em relação aos desenvolvidos por PLS e SPA-MLR pode ser atribuída à capacidade desta técnica em descrever relações não-lineares existentes entre os dados espectrais e os valores de referência dos parâmetros avaliados. É provável, que o número elevado de amostras e de variedades de maçãs (caracterizando um conjunto de dados complexo) utilizado neste trabalho tenha resultado em desvios consideráveis da linearidade na correlação entre as variáveis independentes ( $\mathbf{X}$ ) e dependentes ( $\mathbf{Y}$ ). Este fato justifica a proposição e avaliação do desempenho de modelos desenvolvidos por meio de técnicas não-lineares, como é o caso das LS-SVM, em relação aos métodos clássicos lineares (PLS e MLR). Em geral, desvios da linearidade são esperados em análises espectroscópicas de matrizes complexas, tais como alimentos, tecidos vegetais e solos (SOUZA et al., 2012).

#### 4.4.4.2 Transferência de calibração

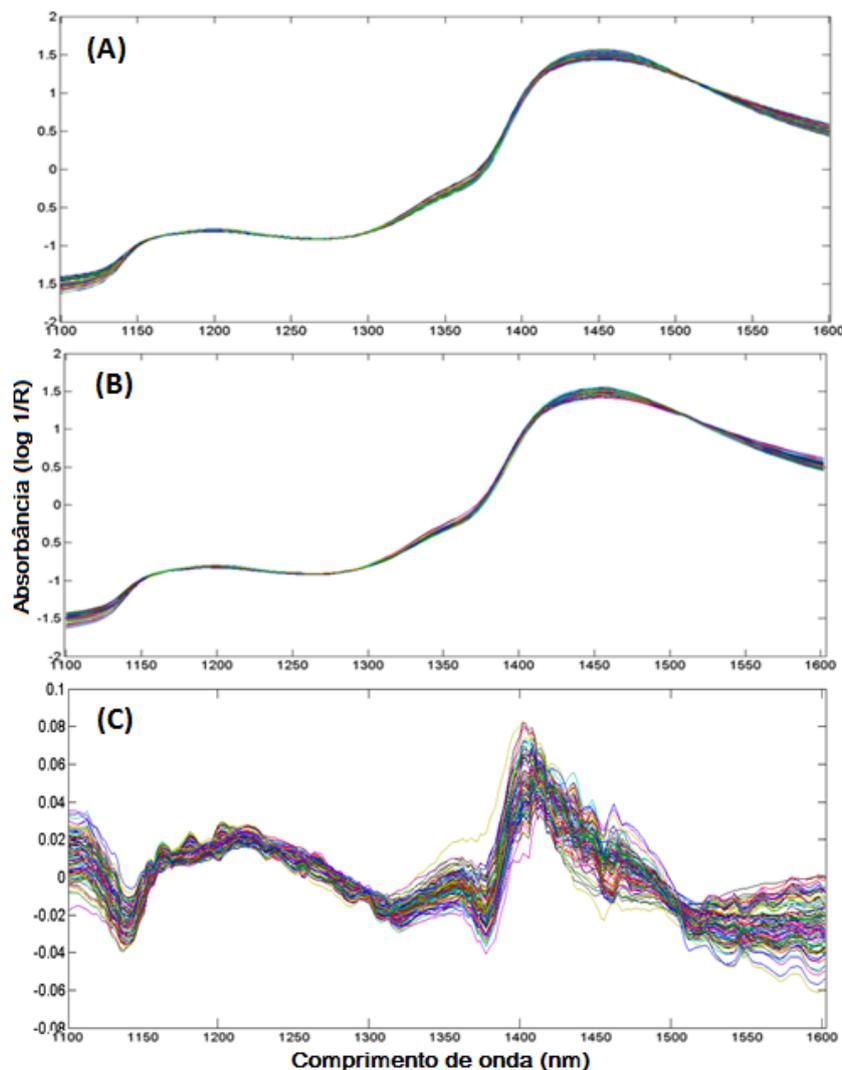
Nesta etapa do trabalho, foram utilizados os modelos de calibração desenvolvidos previamente na etapa anterior (item “4.4.4.1 – Modelos de calibração multivariada”) e um conjunto de amostras independentes ( $n = 93$ , não utilizadas para a construção destes modelos), cujos espectros foram registados nos equipamentos primário e secundário. Este conjunto foi dividido em amostras de transferência ( $n = 30$ ) e amostras de predição ( $n = 63$  ou  $n = 43$ , dependendo do parâmetro de qualidade avaliado) por meio do algoritmo SPXY. Os perfis dos espectros originais (sem pré-processamento), registados nos equipamentos primário e secundário são apresentados na Figura 53-A e B, respectivamente.



**Figura 53** – Conjunto de espectros originais (sem pré-processamento), provenientes de 93 maçãs registados nos equipamentos primário (A) e secundário (B).

Os dados espectrais provenientes das amostras em estudo foram pré-processados aplicando-se SNV. Na Figura 54-A, B e C são mostrados os espectros pré-processados registados nos equipamentos primário e secundário e as diferenças espectrais existentes entre os dois espectrômetros, respectivamente. Aparentemente, os espectros registados com os dois espectrômetros exibem perfis muito similares. Entretanto, será mostrado a

seguir neste estudo, como estas pequenas diferenças espectrais são suficientes para inviabilizar que os espectros registrados no equipamento secundário sejam utilizados diretamente, sem a aplicação prévia de procedimentos de transferência de calibração, nos modelos desenvolvidos com espectros provenientes do equipamento primário.



**Figura 54** – Conjunto de espectros, pré-processados por SNV, provenientes de 93 maçãs registrados nos equipamentos primário (A) e secundário (B), e diferenças espectrais existentes entre os dois espectrômetros (C).

O desempenho preditivo dos modelos foi avaliado utilizando-se o conjunto de amostras de predição, cujos espectros foram registrados nos equipamentos primário e secundário. Na Tabela 13 são mostrados os valores de RMSEP obtidos pelos métodos PLS, SPA-MLR e LS-SVM para os diferentes parâmetros de qualidade avaliados. Os

valores de  $RMSEP^S_P$  (obtido pelo conjunto de predição do equipamento secundário utilizando o modelo calculado no primário) evidenciam a impossibilidade de se utilizar diretamente os espectros registrados no equipamento secundário nos modelos desenvolvidos com dados provenientes do equipamento primário. Este fato é observado para todos os métodos de calibração empregados e parâmetros avaliados (excluindo-se o modelo LS-SVM para determinação de FP). É importante ressaltar, que os valores de  $RMSEP^S_P$  foram calculados aplicando-se previamente um procedimento de interpolação à matriz de dados espectrais do equipamento primário, a fim de reduzir o número de variáveis. Após a aplicação desse procedimento, as matrizes espectrais obtidas com ambos os equipamentos passaram a apresentar o mesmo número de colunas e, portanto, a mesma dimensionalidade em relação às variáveis (FERNÁNDEZ-PIERNA et al., 2014).

Os modelos desenvolvidos por SPA-MLR apresentaram-se mais sensíveis a mudanças nas características das respostas instrumentais, quando comparados aos modelos desenvolvidos por PLS e LS-SVM. Este fato pode ser constatado em função dos altos valores de  $RMSEP^S_P$ , quando comparados aos valores de  $RMSEP^P_P$ . Em contrapartida, os modelos desenvolvidos por LS-SVM apresentaram menor sensibilidade, de modo que  $RMSEP^P_P$  e  $RMSEP^S_P$  exibiram valores próximos. Este comportamento pode ser atribuído à elevada capacidade de generalização, que é uma característica intrínseca das máquinas de vetores de suporte (FERRÃO et. al., 2007). Assim, modelos de calibração desenvolvidos por LS-SVM geralmente apresentam desempenho razoável na previsão de parâmetros de interesse de amostras não pertencentes aos conjuntos de calibração e/ou validação. Isto pode ser observado para o parâmetro FP, para o qual os valores de  $RMSEP^P_P$  e  $RMSEP^S_P$  não apresentaram diferença significativa após a aplicação de teste F ao nível de 95% de confiança. Entretanto, mesmo com valores de  $RMSEP^S_P$  menores, os modelos desenvolvidos por LS-SVM para os outros 4 parâmetros ainda não apresentaram desempenho preditivo satisfatório, justificando a necessidade de aplicação de técnicas de transferência de calibração.

**Tabela 13** – Valores de RMSEP obtidos pelos métodos PLS, SPA-MLR e LS-SVM para os diferentes parâmetros de qualidade avaliados.

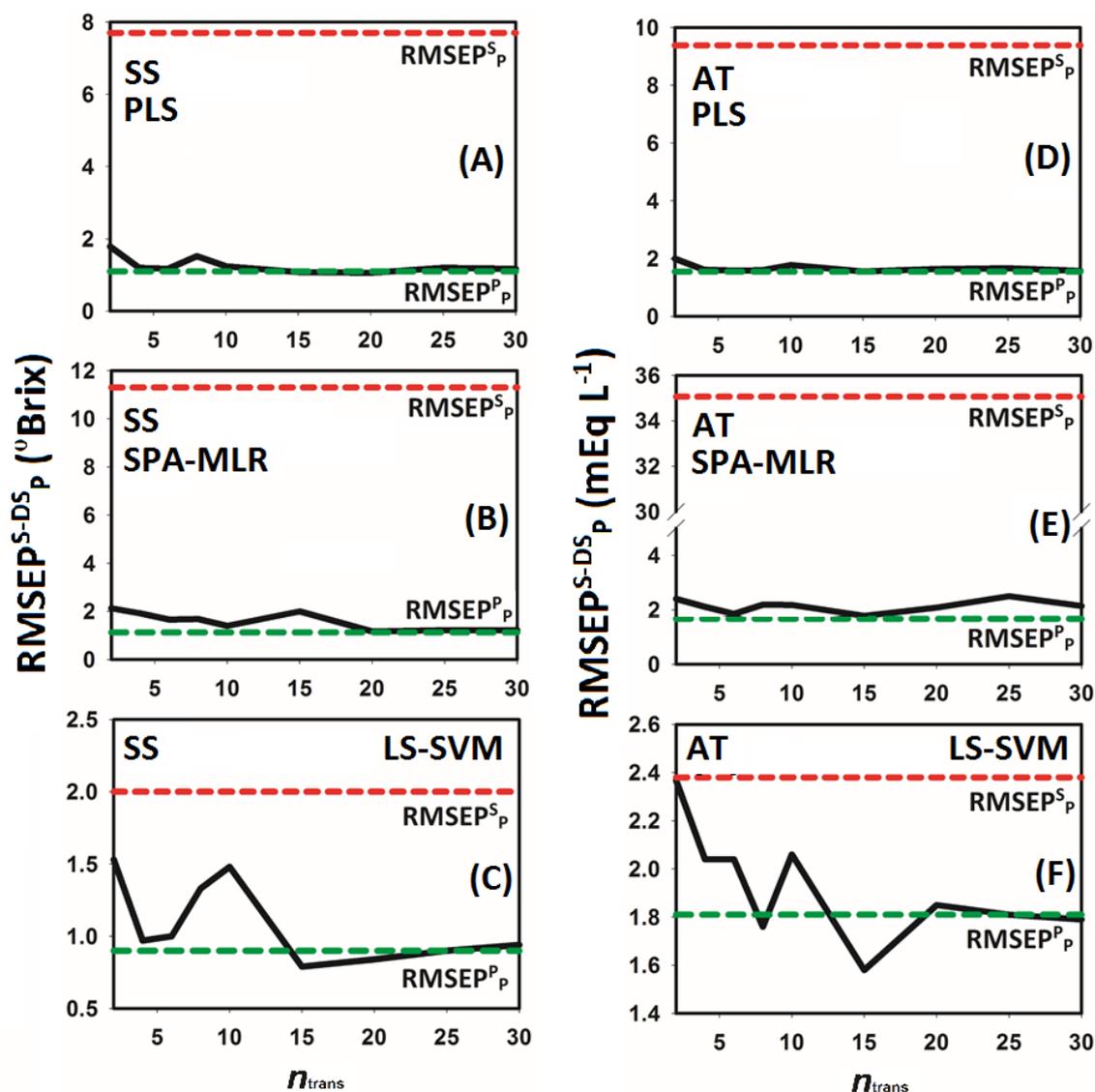
RMSEP	PLS	SPA-MLR	LS-SVM
<i>Sólidos Solúveis (°Brix)</i>			
RMSEP <sup>P</sup> <sub>a</sub>	1,1 (6) <sup>c</sup>	1,1 (12) <sup>d</sup>	0,9 (15) <sup>e</sup>
RMSEP <sup>S</sup> <sub>b</sub>	7,7 (6)	11,3 (12)	2,0 (15)
<i>Acidez titulável (mEq ácido málico L<sup>-1</sup>)</i>			
RMSEP <sup>P</sup>	1,55 (6)	1,67 (30)	1,75 (12)
RMSEP <sup>S</sup>	9,38 (6)	35,06 (30)	2,41 (12)
<i>Firmeza de polpa (kg cm<sup>-2</sup>)</i>			
RMSEP <sup>P</sup>	1,25 (10)	1,52 (21)	1,24 (12)
RMSEP <sup>S</sup>	22,73 (10)	12,20 (21)	1,38 (12)
<i>Índice de amido (adimensional)</i>			
RMSEP <sup>P</sup>	1,88 (8)	1,73 (22)	1,42 (16)
RMSEP <sup>S</sup>	9,20 (8)	16,71 (22)	2,63 (16)
<i>Índice Streif (kg cm<sup>-2</sup> °Brix<sup>-1</sup>)</i>			
RMSEP <sup>P</sup>	0,039 (9)	0,020 (15)	0,046 (15)
RMSEP <sup>S</sup>	0,128 (9)	0,112 (15)	0,114 (15)

<sup>a</sup> Valor de RMSEP obtido pelo conjunto de predição do equipamento primário utilizando o modelo calculado neste equipamento; <sup>b</sup> Valor de RMSEP obtido pelo conjunto de predição do equipamento secundário utilizando o modelo calculado no primário; <sup>c</sup> Número de variáveis latentes mantidas no modelo PLS; <sup>d</sup> Número de variáveis originais utilizadas no modelo SPA-MLR; <sup>e</sup> Número de variáveis latentes mantidas no modelo PLS utilizado na etapa de compressão de dados para o modelo LS-SVM.

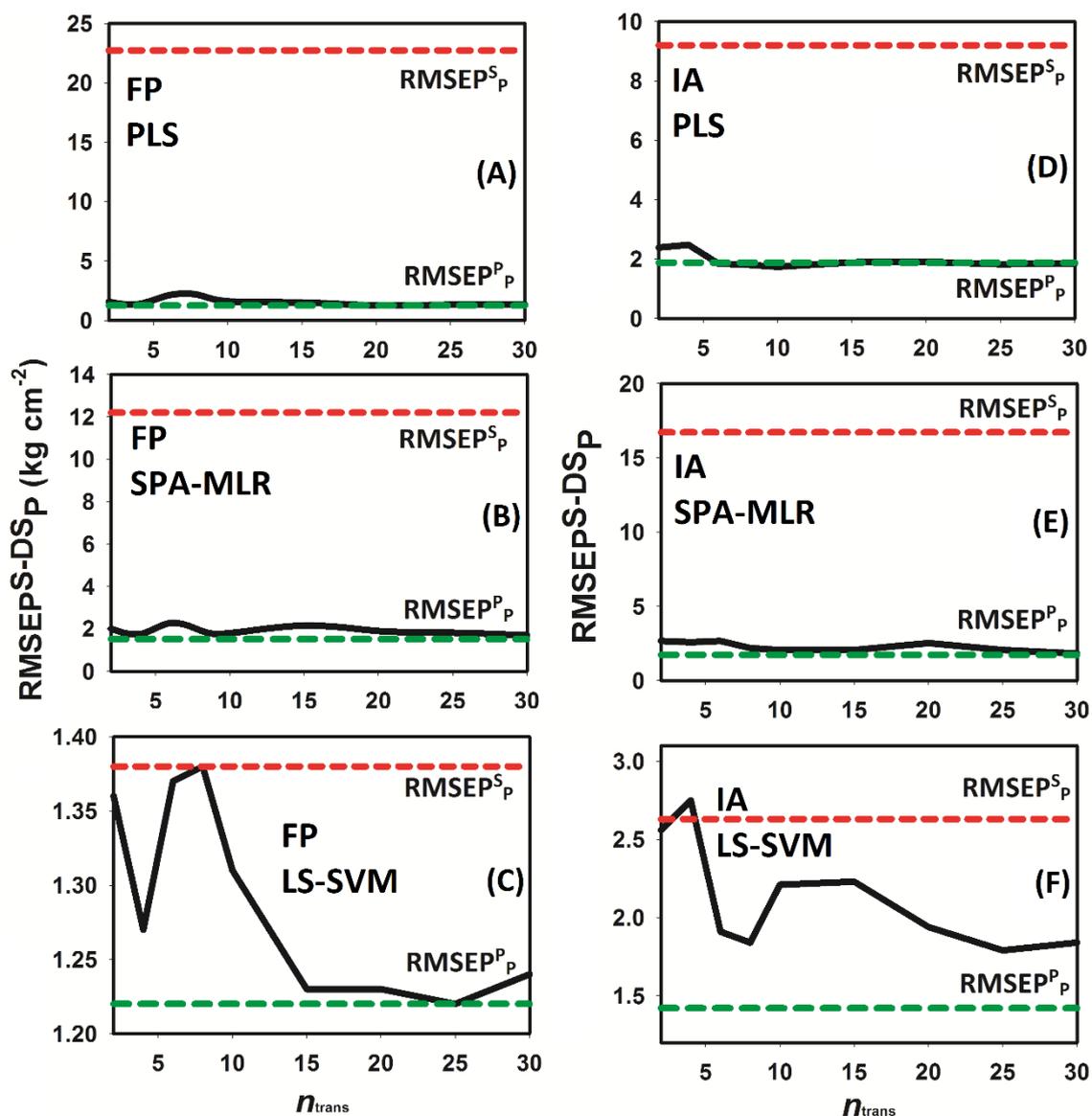
Os modelos desenvolvidos por SPA-MLR apresentaram-se mais sensíveis a mudanças nas características das respostas instrumentais, quando comparados aos modelos desenvolvidos por PLS e LS-SVM. Este fato pode ser constatado em função dos altos valores de RMSEP<sup>S</sup><sub>b</sub>, quando comparados aos valores de RMSEP<sup>P</sup><sub>a</sub>. Em

contrapartida, os modelos desenvolvidos por LS-SVM apresentaram menor sensibilidade, de modo que  $RMSEP^P_P$  e  $RMSEP^S_P$  exibiram valores próximos. Este comportamento pode ser atribuído à elevada capacidade de generalização, que é uma característica intrínseca das máquinas de vetores de suporte (FERRÃO et. al., 2007). Assim, modelos de calibração desenvolvidos por LS-SVM geralmente apresentam desempenho razoável na previsão de parâmetros de interesse de amostras não pertencentes aos conjuntos de calibração e/ou validação. Isto pode ser observado para o parâmetro FP, para o qual os valores de  $RMSEP^P_P$  e  $RMSEP^S_P$  não apresentaram diferença significativa após a aplicação de teste F ao nível de 95% de confiança. Entretanto, mesmo com valores de  $RMSEP^S_P$  menores, os modelos desenvolvidos por LS-SVM para os outros 4 parâmetros ainda não apresentaram desempenho preditivo satisfatório, justificando a necessidade de aplicação de técnicas de transferência de calibração.

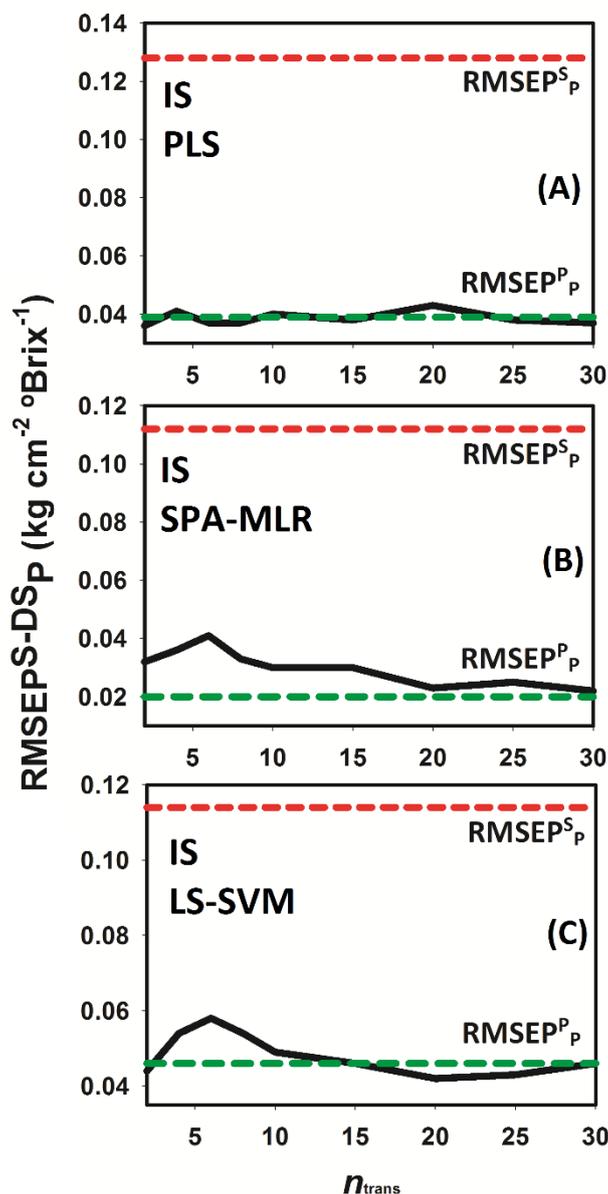
Os resultados do procedimento de transferência de calibração pelo método DS para os modelos construídos por PLS, SPA-MLR e LS-SVM para os diferentes parâmetros de qualidade avaliados são apresentados nas Figuras 55-A, B, C, D, E e F; 56-A, B, C, D, E e F; e 57-A, B e C, em termos dos valores de RMSEP. Nesses gráficos, os valores de  $RMSEP^{S-DS}_P$  (conjunto de predição do equipamento secundário padronizado por DS, aplicado ao modelo calculado no primário – representados pela linha preta contínua) são apresentados em função do número de amostras de transferência ( $n = 2$  até  $n = 30$ ) utilizadas para obtenção das funções de transferência. A linha verde tracejada representa o valor de  $RMSEP^P_P$  (RMSEP obtido pelo conjunto de predição do equipamento primário utilizando o modelo calculado neste equipamento). A linha vermelha tracejada representa o valor de  $RMSEP^S_P$  (RMSEP obtido pelo conjunto de predição do equipamento secundário utilizando o modelo calculado no primário).



**Figura 55** – Valores de  $RMSEP^{S-DS}_p$  (linha preta) obtidos em função do número de amostras de transferência ( $n = 2$  até  $n = 30$ ). Os gráficos A, B e C referem-se aos modelos de calibração para determinação de sólidos solúveis, construídos por PLS, SPA-MLR e LS-SVM, respectivamente; enquanto os gráficos D, E e F referem-se aos modelos de calibração para determinação de acidez titulável, construídos por PLS, SPA-MLR e LS-SVM, respectivamente. As linhas verde e vermelha tracejadas representam os valores de  $RMSEP^P_p$  e  $RMSEP^S_p$ , respectivamente.



**Figura 56** – Valores de  $RMSEP^{S-DS}_P$  (linha preta) obtidos em função do número de amostras de transferência ( $n = 2$  até  $n = 30$ ). Os gráficos A, B e C referem-se aos modelos de calibração para determinação de firmeza de polpa, construídos por PLS, SPA-MLR e LS-SVM, respectivamente; enquanto os gráficos D, E e F referem-se aos modelos de calibração para determinação de índice de amido, construídos por PLS, SPA-MLR e LS-SVM, respectivamente. As linhas verde e vermelha tracejadas representam os valores de  $RMSEP^{P}_P$  e  $RMSEP^{S}_P$ , respectivamente.



**Figura 57** – Valores de  $RMSEP^{S-DS}_P$  (linha preta) obtidos em função do número de amostras de transferência ( $n = 2$  até  $n = 30$ ). Os gráficos A, B e C referem-se aos modelos de calibração para determinação de índice Streif, construídos por PLS, SPA-MLR e LS-SVM, respectivamente. As linhas verde e vermelha tracejadas representam os valores de  $RMSEP^P_P$  e  $RMSEP^S_P$ , respectivamente.

É possível verificar para os modelos construídos por PLS e SPA-MLR, que o uso de duas amostras para obtenção das funções de transferência foi suficiente para reduzir de forma significativa os valores de RMSEP em comparação aos valores obtidos sem aplicação do procedimento de transferência de calibração. Neste caso, as reduções relativas entre os dois procedimentos variaram de 72% para determinação de IS por PLS até 93% para determinação de AT por SPA-MLR. Em geral, os valores de  $RMSEP^{S-DS}_P$

para os modelos construídos por PLS e SPA-MLR se estabilizaram a partir do uso de cinco amostras de transferência, mantendo-se muito próximos dos valores de  $RMSEP^P$ .

No caso dos modelos de calibração desenvolvidos por LS-SVM, os valores de  $RMSEP^{S-DS}_P$  também apresentaram menos oscilações após o uso de cerca de 20 amostras de transferência. Entretanto, as reduções nos valores de RMSEP não foram muito significativas em comparação aos valores obtidos sem aplicação do procedimento de transferência de calibração. As reduções relativas entre os dois procedimentos variaram de 10% para determinação de FP até 60% para determinação de SS. Este fato era esperado, uma vez que os modelos de calibração desenvolvidos por LS-SVM apresentaram desempenho preditivo muito superiores aos modelos desenvolvidos por PLS e SPA-MLR quando utilizados com o conjunto de predição sem aplicação prévia dos procedimentos de transferência de calibração. Isto foi evidenciado pelos baixos valores de  $RMSEP^S_P$  obtidos, que foram muito similares aos valores de  $RMSEP^P_P$ .

#### 4.5 CONCLUSÕES

No primeiro estudo apresentado neste capítulo foi realizada a comparação entre o desempenho dos espectrômetros de bancada (XDS) e portátil (MicroNIR) para a determinação de parâmetros de qualidade de maçãs. Verificou-se que o desempenho preditivo dos modelos de calibração desenvolvidos para determinação de sólidos solúveis, acidez titulável, firmeza de polpa, índice de amido e índice Streif foram similares utilizando-se ambos os equipamentos, já que os valores de RMSECV obtidos não apresentaram diferença estatística significativa. Assim, pode-se concluir que o espectrômetro MicroNIR ofereceu desempenho comparável ao espectrômetro de bancada para a determinação dos parâmetros de qualidade de maçãs avaliados.

O segundo estudo apresentado neste capítulo foi dividido em duas etapas. Na primeira, foram desenvolvidos modelos de calibração empregando-se diferentes técnicas de regressão (PLS, SPA-MLR e LS-SVM) para determinação de parâmetros de qualidade de maçãs utilizando uma base de dados histórica (obtida entre 2008 e 2014), cujos espectros foram registrados com o espectrômetro de bancada (XDS). Nesta etapa, verificou-se que os modelos desenvolvidos por LS-SVM apresentaram melhor desempenho preditivo em comparação aos modelos desenvolvidos por PLS e SPA-MLR, exibindo valores de RMSEP significativamente menores para a maioria dos parâmetros de qualidade avaliados.

Na segunda etapa, avaliou-se a possibilidade de transferência de calibração entre os equipamentos de bancada (XDS) e portátil (MicroNIR), visando a determinação de parâmetros de qualidade de maçãs. Neste caso, o método de padronização direta (DS) foi empregado com êxito para realizar o procedimento de transferência de calibração. Este método permitiu reduzir de forma muito significativa os valores de RMSEP dos modelos desenvolvidos por PLS e SPA-MLR. Entretanto, para os modelos desenvolvidos por LS-SVM a redução nos valores de RMSEP não foi tão evidente. Este fato foi atribuído à elevada capacidade de generalização deste método de regressão, que permitiu a obtenção de valores de  $RMSEP^S_P$  e  $RMSEP^P_P$  mais próximos, mesmo sem a aplicação do procedimento de transferência de calibração. No entanto, recomenda-se que a transferência de calibração seja realizada para todos os métodos de calibração empregados neste estudo, já que na maioria dos casos foi inviável a utilização direta dos espectros registrados no equipamento secundário (portátil) nos modelos desenvolvidos com dados provenientes do equipamento primário (bancada).

De forma geral, verificou-se que um método simples de transferência de calibração permitiu que os dados obtidos com o equipamento portátil fossem utilizados nos modelos desenvolvidos com o equipamento de bancada. Desta forma, a base de dados histórica registrada ao longo de vários anos com o equipamento de bancada pôde ser utilizada, eliminando a necessidade de aquisição de uma nova base de dados utilizando o equipamento portátil e reconstrução dos modelos de calibração.

#### 4.6 PERSPECTIVAS FUTURAS

Ao longo do desenvolvimento dos estudos realizados neste capítulo foi possível identificar alguns aspectos que podem ser melhorados e novos estudos que podem ser realizados em propostas futuras, dentre os quais podem ser destacados os seguintes:

Em relação ao estudo utilizando espectrômetro portátil:

- Avaliar o desempenho de novos espectrômetros portáteis de baixo custo (entre R\$ 1.000,00 e 15.000,00) para serem utilizados na avaliação da qualidade de maçãs;

- Incluir maior variabilidade nos modelos de calibração desenvolvidos, visando aumentar a robustez dos mesmos (ex.: incluir maçãs de diferentes variedades, de diferentes pomares, colhidas em diferentes safras, entre outras variáveis);
- Aplicar métodos adequados para efetuar a compensação da temperatura durante a etapa de construção dos modelos de calibração, permitindo que os equipamentos portáteis possam ser utilizados adequadamente sob diferentes temperaturas (situação comum para aplicações em campo).

Em relação ao estudo sobre transferência de calibração:

- Avaliar o desempenho de outros métodos de transferência de calibração para o mesmo conjunto de dados utilizados neste estudo;
- Utilizar métodos de transferência de calibração que não necessitem de amostras de transferência (ex.: *Linear Model Correction*, Filtros de Resposta a Impulso Finita, entre outros).

**REFERÊNCIAS**

ALAMAR, M. C.; BOBELYN, E.; NICOLAÏ, B. M.; MOLTÓ, E. Calibration transfer between NIR diode array and FT-NIR spectrophotometers for measuring the soluble solids contents of apple. **Postharvest Biology and Technology**, v.45 (1), p.38-45, 2007.

AMIGO, J. M.; MARTÍ, I.; GOWEN, A. **Hyperspectral Imaging and Chemometrics: A Perfect Combination for the Analysis of Food Structure, Composition and Quality**. In: FEDERICO, M. (Ed.). *Data Handling in Science and Technology*: Elsevier, v.28, p.343-370, 2013.

ANGRA, S. K.; DIMRI, A. K.; KAPUR, P. Nondestructive brix evaluation of apples of different origin using near infrared (NIR) filter based reflectance spectroscopy. **Instrumentation Science & Technology**, v.37, p.241–253, 2009.

ANUÁRIO BRASILEIRO DA FRUTICULTURA. Santa Cruz do Sul: Ed. Gazeta Santa Cruz, 2017. 90p.

ANUÁRIO BRASILEIRO DA FRUTICULTURA. Santa Cruz do Sul: Ed. Gazeta Santa Cruz, 2013. 140p.

AOAC. Association of Official Analytical Chemistry. *Official methods of analysis* (16<sup>th</sup> ed.), Arlington: AOAC, 1995. 1141 p.

ARAÚJO, M. C. U.; SALDANHA, T. C. B.; GALVÃO, R. K. H.; YONEYAMA, T.; CHAME, H. C.; VISANI, V. The successive projections algorithm for variable selection in spectroscopy multicomponent analysis. **Chemometrics and Intelligent Laboratory Systems**, v.57, p.65, 2001.

ARFKEN, G. B.; WEBER, H. J. **Mathematical Methods for Physicists** (4<sup>th</sup> edition). Academic Press: New York, 1995. 1029p.

ARGENTA, L. C.; BENDER, R. J.; KREUZ, C. L.; MONDARDO, M. Padrões de maturação e índices de colheita de maçãs cvs. Gala, Golden Delicious e Fuji. **Pesquisa Agropecuária Brasileira**, v.30 (10), p.1.259-1.266, 1995.

BARANOWSKI, P.; MAZUREK, W.; WOZNIAK, J.; MAJEWSKA, U. Detection of early bruises in apples using hyperspectral data and thermal imaging. **Journal of Food Engineering**, v.110, p.345–355, 2012.

BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. 25 Anos de Quimiometria no Brasil. **Química Nova**, v.29, p.1401, 2006.

BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. **Chemometrics: a practical guide**. New York: A Wiley - Interscience Publication, 1998.

BETEMPS, D. L.; FACHINELLO, J.C.; GALARÇA, S.P. Espectroscopia do visível e infravermelho próximo (VIS/NIR) na avaliação da qualidade de mangas 'Tommy Atkins'. **Revista Brasileira de Fruticultura**, v.33, p.306-313, 2011.

BLANCO, M.; VILLARROYA, I. NIR spectroscopy: a rapid-response analytical tool. **Trends in analytical chemistry**, v.21, p.240, 2002.

BLANPIED, G. D.; SILSBY, K. J. Predicting Harvest Date Windows for Apples. **Cornell University Cooperative Extension Bulletin 221**, 1992. 12 p.

BOBELYN, E.; SERBAN, A-S.; NICU, M.; LAMMERTYN, J.; NICOLAÏ, B. M.; SAEYS, W. Postharvest quality of apple predicted by NIR spectroscopy: study of the effect of biological variability on spectra and model performance. **Postharvest Biology and Technology**, v.55 p.133–143, 2010.

BORIN, A. **Aplicações de Máquinas de vetores de suporte por mínimos quadrados (LS-SVM) na quantificação de parâmetros de qualidade de matrizes lácteas**. Universidade Estadual de Campinas. 2007. Tese de doutorado, 122p.

BRERETON, R. G. Introduction to multivariate calibration in analytical chemistry. **Analyst**, v.125, p.2125-2154, 2000.

CERQUEIRA, E. O.; ANDRADE, J. C.; POPPI, R. J. Redes Neurais e suas aplicações em calibração multivariada. **Química Nova**, v.24, p.864, 2001.

CERQUEIRA, E. O.; POPPI, R. J.; KUBOTA, L. T.; MELO, C. Utilização de filtro de transformada de Fourier para minimização de ruídos em sinais analíticos. **Química Nova**, v.23, p.690, 2000.

CHENG, J.-H.; SUN, D.-W. Hyperspectral imaging as an effective tool for quality analysis and control of fish and other seafoods: Current research and potential applications. **Trends in Food Science & Technology**, v. 37, p.78-91, 2014.

CODGILL, R. P.; DARDENNE, P. Least-square support vector machines for chemometrics: an introduction and evaluation. **Journal of Near Infrared Spectroscopy**, v.12, p.93-100, 2004.

COSTA FILHO, P. A.; POPPI, R. J. Aplicação de algoritmos genéticos na seleção de variáveis em espectroscopia no infravermelho médio. Determinação simultânea de glicose, maltose e frutose. **Química Nova**, v.25, p.46, 2002.

CRISTIANINI, N.; SHAW-TAYLOR J. **An introduction to support vector machines**. Cambridge University Press: UK, 2000. 189p.  
DE LIRA, L. F. B. **Desenvolvimento de métodos analíticos para monitoramento da qualidade do biodiesel e suas misturas**. Universidade Federal de Pernambuco, 2010. Tese de doutorado, 145p.

DELWICHE, S. R.; MEKWATANAKARN, W.; WANG, C. Y. Soluble solids and simple sugars measurement in intact mango using near infrared spectroscopy. **HortTechnology**, v.18 (3), p.325-544, 2008.

DI NEZIO, M.S.; PISTONESI, M.F.; FRAGOSO, W.D.; PONTES, M.J.C.; GOICOECHEA, H.C; ARAUJO, M.C.U.; FERNÁNDEZ BAND, B.S. Successive projections algorithm improving the multivariate simultaneous direct spectrophotometric determination of five phenolic compounds in sea water. **Microchemical Journal**, v.85, p.194, 2007.

DUBEY, S. R.; DIXIT, P.; SINGH, N.; GUPTA, J. P. Infected Fruit Part Detection using K-Means Clustering Segmentation Technique. **International Journal of Artificial Intelligence and Interactive Multimedia**, v. 2, p.65-72, 2013.

ELMASRY, G.; WANG, N.; ELSAYED, A.; NGADI, M. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. **Journal of Food Engineering**, v.81, p.98–107, 2007.

FACHINELLO, J. C.; NACHTIGAL, J. C.; KERSTEN, E. **Fruticultura: fundamentos e práticas**. Pelotas: Embrapa Clima Temperado (Série Livro, online). 2009.

FAN, S.; ZHANG, B.; LI, J.; HUANG, W.; WANG, C. 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.

FAOSTAT. Food and Agriculture Organization of the United Nations, Rome: Italy,

2016. Disponível em: <<http://faostat.fao.org/>>. Acessado em maio de 2017.

FERRÃO, M. F.; MELLO, C.; BORIN, A.; MARETTO, D. A.; POPPI, R. LS-SVM: uma nova ferramenta quimiométrica para regressão multivariada. Comparação de modelos de regressão LS-SVM e PLS na quantificação de adulterantes em leite em pó empregando NIR. **Química Nova**, v.30 (4), p.852-859, 2007.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S., VOLPE, P. L. O. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, v.22, p.724, 1999.

FONSECA, A.; RAIMUNDO Jr., I. M. A simple method for water discrimination based on an light emitting diode (LED) photometer. **Analytica Chimica Acta**, v.596 (1), p.66-72, 2007.

FORINA, M.; LANTERI, S.; CASALE, M.; OLIVEROS, M. C. C. Stepwise orthogonalization of predictors in classification and regression techniques: an “old” technique revisited. **Chemometrics and Intelligent Laboratory Systems**, v.87, p.252, 2007.

GALVÃO, R. K. H., ARAÚJO, M. C. U., FRAGOSO, W. D., SILVA, E. C., JOSÉ, G. E., SOARES, S. F. C., PAIVA, H. M. A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm. **Chemometrics and Intelligent Laboratory Systems**, v.92 p.83-91, 2008.

GALVÃO, R. K. H.; ARAÚJO, M. C. U. The Successive Projections Algorithm for Variable Selection in NIR Spectrometry. Curso ministrado na 17ª Conferência Internacional sobre Espectroscopia no Infravermelho Próximo, Foz do Iguaçu, Brasil, 18 de outubro de 2015.

GALVÃO, R. K. H.; ARAUJO, M. C. U.; JOSÉ, G. E.; PONTES, M. J. C.; SILVA, E. C.; SALDANHA, T. C. B. A method for calibration and validation subset partitioning. **Talanta**, v.67, p.36, 2005.

GALVÃO, R. K. H.; ARAÚJO, M. C. U.; SALDANHA, T. C. B., VISANI, V.; PIMENTEL, M. F. Estudo comparativo sobre filtragem de sinais instrumentais usando transformadas de Fourier e Wavelet. **Química Nova**, v.24, p.874, 2001.

GALVÃO, R. K. H.; ARAÚJO, M. C. U.; SILVA, E. C.; JOSÉ, G. E.; SOARES, S. F. C.; PAIVA, H. M. Cross-validation for the selection of spectral variables using the Successive projections algorithm. **Journal of Brazilian Chemical Society**, v.18, p.1580, 2007.

GALVÃO, R. K. H.; JOSÉ, G. E.; DANTAS FILHO, H. A.; ARAUJO, M. C. U.; DA SILVA, E. C.; PAIVA, H. M.; SALDANHA, T. C. B.; DE SOUZA, E. S. O. N. Optimal wavelet filter construction using X and Y data. **Chemometrics and Intelligent Laboratory Systems**, v.70 (1), p.1-10, 2004.

GELADI, P. L. M.; GRAHN, H. F.; BURGER, J. E. **Multivariate Images, Hyperspectral Imaging: Background and Equipment**. In: HANS F. GRAHN & PAUL GELADI (Ed.). *Techniques and Applications of Hyperspectral Image Analysis*. John Wiley & Sons, Ltd, p.1-15, 2007.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, v.185, p.1-17, 1986.

GENÚ, P. J. C.; PINTO, A. C. Q. **A cultura da mangueira**. Brasília: EMBRAPA, p.452, 2002.

GIOVANELLI, G.; SINELLI, N.; BEGHI, R.; GUIDETTI, R.; CASIRAGHI, E. NIR spectroscopy for the optimization of postharvest apple management. **Postharvest Biology and Technology**, v.87, p.13-20, 2014.

GIOVANELLI, G.; SINELLI, N.; BEGHI, R.; GUIDETTI, R.; CASIRAGHI, E. NIR spectroscopy for the optimization of postharvest apple management. **Postharvest Biology and Technology**, v.87, p.13-20, 2014.

GIRARDI, L. C.; NACHTIGALL, G. R.; PARUSSOLO, A. **Fatores pré-colheita que interferem na qualidade da fruta**. In: GIRARDI, L. C. (Ed.). *Maçã: Pós-colheita* (Coleção Frutas do Brasil). Bento Gonçalves: Embrapa Uva e Vinho; Brasília: Embrapa Informação Tecnológica, 2004. 109 p.

GOLIC, M., WALSH, K. B., & LAWSON, P. Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature. **Applied Spectroscopy**, v.57, p.139–145, 2003.

GÓMEZ, A. H., HE, Y., & PEREIRA, A. G. Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using vis–NIR spectroscopy techniques. **Journal of Food Engineering**, v.77, p.313–319, 2006.

GÓMEZ, A. H., HE, Y., & PEREIRA, A. G. Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using vis-NIR spectroscopy techniques. *Journal of Food Engineering*, v.77, p.313–319, 2006.

GOWEN, A. A.; O'DONNELL, C. P.; CULLEN, P. J.; DOWNEY, G.; FRIAS, J. M. Hyperspectral Imaging – an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, v.18, p.590–598, 2007.

GUTHRIE, J. A., & WALSH, K. B. Non-invasive assessment of pineapple and mango fruit quality using near infrared spectroscopy. *Australian Journal of Experimental Agriculture*, v.37, p.253-263, 1997.

HAHN, F. Mango firmness sorter. *Biosystems Engineering*, v.89, p.309-319, 2004.

HAMZAÇEBI, C.; KUTAY, F.; A heuristic approach for finding the global minimum: Adaptive random search technique. *Applied Mathematics and Computation*, v.173, p.1323, 2006.

HONORATO, F. A. **Previsão das propriedades das gasolinas do Nordeste empregando espectroscopia NIR/MID e transferência de calibração**. Universidade Federal de Pernambuco, 2006. Tese de doutorado, 106p.

HONORATO, F. A.; BARROS NETO, B.; MARTINS, M. N.; GALVÃO, R. K. H.; PIMENTEL, M. F. Transferência de calibração em métodos multivariados. *Química Nova*, v.30, p.1301, 2007.

IGNE, B. Intra and inter-brand calibration transfer for near infrared spectrometers (PhD Thesis). Iowa State University: Iowa, EUA, 2009. 158p.

INSTITUTO BRASILEIRO DE FRUTAS - IBRAF. Disponível em: <<http://www.ibraf.org.br>>. Acesso em: 15 de setembro de 2016.

JHA, S. N.; CHOPRA, S.; KINGSLEY, A. R. P. Modeling of color values for nondestructive evaluation of maturity of mango. *Journal of Food Engineering*, v.78, p.22–26, 2007.

JHA, S. N.; JAISWAL, P.; NARSAIAH, K.; GUPTA, M.; BHARDWAJ, R.; SINGH, A. K. Nondestructive prediction of sweetness of intact mango using near infrared spectroscopy. *Scientia Horticulturae*, v.138, p.171-175, 2012.

JHA, S. N.; NARSAIAH, K.; JAISWAL, P.; BHARDWAJ, R.; GUPTA, M.; KUMAR, R.; SHARMA, R. Nondestructive prediction of maturity of mango using near infrared spectroscopy. **Journal of Food Engineering**, v.124, p.152–157, 2014.

JHA, S.N.; MATSUOKA, T. Non-destructive techniques for quality evaluation of intact fruit and vegetables. **Food Science and Technology Research**, v.6, p. 248-251, 2000.

JOUAN-RIMBAUD, D.; MASSART, D. L.; LEARDI, R.; NOORD, O. E., Genetic algorithms as a tool for wavelength selection in multivariate calibration. **Analytical Chemistry**, v.67, p.4295, 1995.

JOUAN-RIMBAUD, D.; WALCZACK, B.; POPPI, R. J.; DENOORD, O. E.; MASSART, D. L.; Application of Wavelet Transform To Extract the Relevant Component from Spectral Data for Multivariate Calibration. **Analytical Chemistry**, v.69, p.4317, 1997.

KARAMAN, İ.; QANNARI, E. M.; MARTENS, H.; HEDEMANN, M. S.; KNUDSEN, K. E. B.; KOHLER, A. Comparison of Sparse and Jack-knife partial least squares regression methods for variable selection. **Chemometrics and Intelligent Laboratory Systems**, v.122, p.65-77, 2013.

KENNARD, R. W.; STONE, L. A. Computer-aided design of experiments. **Technometrics**, v.11, p.137, 1969.

KIENZLE, S.; SRUAMSIRI, P.; CARLE, R.; SIRISAKULWAT, S.; SPREER, W.; & NEIDHART, S. Harvest maturity specification for mango fruit (*Mangifera indica* L. ‘Chok Anan’) in regard to long supply chains. **Postharvest Biology and Technology**, v.61, p.41–55, 2011.

KUBELKA, P.; MUNK, F. Ein Beitrag zur Optik der Farbanstriche. **Zeitschrift für technische Physik**, v.12, p.593–601, 1931.

KUMAR, S.; MCGLONE, A.; WHITWORTH, C.; VOLZ, R. Postharvest performance of apple phenotypes predicted by near-infrared (NIR) spectral analysis. **Postharvest Biology and Technology**, v.100, p.16-22, 2015.

LACERDA, M. A. D.; LACERDA, R. D.; ASSIS, P. C.O. A participação da fruticultura no agronegócio brasileiro. **Revista de Biologia e Ciências da Terra**, v.4 (1), 2004.

LAMMERTYN, J.; NICOLAÏ, B. M.; OOMS, K.; DE SMEDT, V.; DE BRAEDEMÄEKER, J. Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR spectroscopy. **Transactions of the ASAE**, v.41, p.1089–1094, 1998.

LAMMERTYN, J.; PEIRS, A.; DE BAERDEMAEKER, J.; NICOLAÏ, B. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. **Postharvest Biology Technology**, v.18, p.121–132, 2000.

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

LI, Z.; THOMAS, C. Quantitative evaluation of mechanical damage to fresh fruits. **Trends in Food Science & Technology**, v.35, p.138-150, 2014.

LINDEN, R. Técnicas de agrupamento - Tutorial. **Revista de Sistemas de Informação da FSMA**, n.4, p.18-36, 2009.

LIU, Y.; ZHOU, Y. Quantification of the soluble solids content of intact apples by vis-NIR transmittance spectroscopy and the LS-SVM method. **Spectroscopy**, v.28, 32–43, 2013.

LIU, Y.D.; YING, Y. B.; FU, X. P.; LU, H. S. Experiments on prediction sugar content in apples by FT-NIR technique. **Journal of Food Engineering**, v.80, p.986–989, 2007.

LU, R. GUYER, D. E.; BEAUDRY, R. M. Determination of firmness and sugar content of apples using near-infrared diffuse reflectance. **Journal of texture studies**, v.31, p.615–630, 2000.

LU, R. **Quality Evaluation of Fruit by Hyperspectral Imaging**. In: DA-WEN SUN (Ed.). *Computer Vision Technology for Food Quality Evaluation* (1<sup>st</sup> ed., p.319–348). Burlington (USA): Academic Press. 2008.

LUCASIU, C. B.; BECKERS, M. L. M.; KATEMAN, G.; Genetic algorithms in wavelength selection: a comparative study. **Analytica Chimica Acta**, v.286, p.135, 1994.

MAGWAZA, L. S., OPARA, U. L., NIEUWOUDT, H., CRONJE, P. J. R., SAEYS, W., & NICOLAÏ, B. NIR spectroscopy applications for internal and external quality analysis of citrus fruit – A review. **Food Bioprocess Technology**, v.5, p.425–444, 2012.

MAGWAZA, L. S., OPARA, U. L., NIEUWOUDT, H., CRONJE, P. J. R., SAEYS, W., & NICOLAÏ, B. NIR spectroscopy applications for internal and external quality analysis of citrus fruit – A review. **Food Bioprocess Technology**, v.5, p.425–444, 2012.

MAHAYOTHEE, B., LEITENBERGER, M., NEIDHART, S., MÜHLBAUER, W., & CARLE, R. Nondestructive determination of maturity of Thai mangoes by near infrared spectroscopy. **Proceedings of the VII international mango symposium**. Acta Horticulture (ISHS), v.645, p.581–588, 2004.

MARQUES, E. J. N.; FREITAS, S. T.; PIMENTEL, M. F.; PASQUINI, C. Rapid and non-destructive determination of quality parameters in the ‘Tommy Atkins’ mango using a novel handheld near infrared spectrometer. **Food Chemistry**, v.197, p.1207–1214, 2016.

MARTÍNEZ-ROMERO, D.; SERRANO, M.; CARBONELL, A.; CASTILLO, S.; RIQUELME, F.; VALERO, D. **Mechanical damage during fruit post-harvest handling: Technical and physiological implications**. In P. DRIS, & M. S. JAIN (Eds.). *Quality handling and evaluation: Vol. 3. Quality handling and evaluation* (p.233–252). Dordrecht (The Netherlands): Kluwer Academic Publishers. 2004.

MCGLONE, V.A.; JORDAN, R. B.; MARTINSEN, P. J. Vis/NIR estimation at harvest of pre- and post-storage quality indices for ‘Royal Gala’ apple. **Postharvest Biology and Technology**, 25:135–144 (2002).

MEDLICOTT, A. P., & THOMPSON, A. K. Analysis of sugars and organic acids in ripening of mango fruit (*Mangifera indica* L. var. Keitt) by high-performance liquid chromatography. **Journal of the Science of Food and Agriculture**, v.36, p.561–566, 1985.

MEHINAGIC, E.; ROYER, G.; BERTRAND, D.; SYMONEAUX, R.; LAURENS, F.; JOURJON, F. Relationship between sensory analysis, penetrometry and visible–NIR spectroscopy of apples belonging to different cultivars. **Food Quality and Preference**, v.14, p.473–484, 2003.

MENDOZA, F.; LU, R.; ARIANA, D.; CEN, H.; BAILEY, B. Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content. **Postharvest Biology and Technology**, v.62, p.149–160, 2011.

MILLER, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. **Analytical Chemistry**, v.31, p.426-428, 1959.

MITCHAM, E. J., & MCDONALD, R. E. Cell wall modification during ripening of 'Keitt' and 'Tommy Atkins' mango fruit. **Journal of the American Society for Horticultural Science**, v.117, p.919-924, 1992.

NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. **A user-friendly guide to Multivariate Calibration and Classification**. NIR Publications: Chichester-UK, 2002. 354p.

NAGATA, M.; TALLADA, J. G.; KOBAYASHI, T. Bruise Detection using NIR Hyperspectral Imaging for Strawberry (*Fragaria×ananassa* Duch.). **Environmental Control in Biology**, v.44, p.133-142, 2006.

NASSUR, R. C. M. R. **Indicadores de qualidade em mangas durante o amadurecimento**. Universidade Federal de Lavras, 2013. Tese de doutorado, 86p.

NICOLAÏ, B. M., BULENS, I., DE BAERDEMAKER, J., DE KETELAERE, B., HERTOOG, M. L. A. T. M., VERBOVEN, P., & LAMMERTYN, J. **Non-destructive evaluation: detection of external and internal attributes frequently associated with quality and damage**. In: W. J. Florkowski, R. L. Shewfelt, B. Brueckner, & S. E. Prussia (Eds.), *Postharvest handling: A systems approach*. Amsterdam: Academic Press, Elsevier, 2009.

NØRGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L.; ENGELSEN, S. B. Interval partial least-squares regression (i-PLS): a comparative chemometric study with an example from near infrared spectroscopy. **Applied Spectroscopy**, v.54, p.413, 2000.

OSBORNE, B. G., FEARN, T., & HINDLE, P. T. **Practical NIR spectroscopy with applications in food and beverage analysis** (2<sup>nd</sup> ed.). Harlow, UK: Addison-Wesley Longman Ltd., 1993.

OUYANG, A.; XIE, X.; ZHOU, Y.; LIU, Y. Partial least squares regression variable screening studies on apple soluble solids NIR spectral detection. **Spectroscopy and Spectral Analysis**, v.32, p.2680-2684, 2012.

PARK, B.; ABBOTT, J.A.; LEE, K.J.; CHOI, C.H.; CHOI, K.H. Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of delicious and gala apples. **Transactions of the ASAE**, v.46, p.1721–1731, 2003.

PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. **Journal Brazilian Chemical Society**, v.14, p.198, 2003.

PAZ, P.; SANCHEZ, M. T.; PEREZ-MARIN, D.; GUERRERO, J. E.; GARRIDO-VARO, A. Evaluating NIR instruments for quantitative and qualitative assessment of intact apple quality. **Journal of the Science of Food and Agriculture**, v.89, p.781–790, 2009.

PEDERSON, C. G.; FRIEDRICH, D. M.; HSIUNG, C.; GUNTEN, M. V.; O'BRIEN, N. A.; RAMAKER, H. J.; SPRANG, E. V.; DREISCHOR, M. Pocket-size near-infrared spectrometer for narcotic materials identification. *Proceedings of the SPIE*, v.9101, p.1-11, 2014.

PEIRS, A. TIRRY, J.; VERLINDEN, B.; DARIUS, P.; NICOLAÏ, B. M. Effect of biological variability on the robustness of NIR models for soluble solids content of apples. **Postharvest Biology and Technology**, v.28, p.269–280, 2003.

PENG, Y.; LU, R. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. **Postharvest Biology and Technology**, v.48, p.52–62, 2008.

PERALTA-ZAMORA, P.; CORNEJO-PONCE, L., NAGATA, N., POPPI, R. J. Alternativas quimiométricas para a resolução de problemas analíticos clássicos. Determinação espectrofotométrica de misturas de zircônio e háfnio. **Química Nova**, v.20, p.469, 1997.

PIERNA, J. A. F.; VERMUELEN, P.; LECLER, B.; BAETEN, V.; DARDENNE, P. Calibration transfer from dispersive instruments to handheld spectrometers. **Applied Spectroscopy**, v.64, p.644-648, 2010.

PINTO, A. C. Q. **O agronegócio da manga**. In: ALBUQUERQUE, A. C. S.; DA SILVA, A. G. (Ed.). *Agricultura Tropical: quatro décadas de inovações tecnológicas, institucionais e políticas*. Brasília, DF: Embrapa Informação Tecnológica, p. 402, 2008. PIO-CORRÊA, M. **Dicionário das plantas úteis do Brasil**. Rio de Janeiro: Ministério da Agricultura, v. 5, p. 87-88, 1974.

PISSARD, A.; PIERNA, J. A. F.; BAETEN, V.; SINNAEVE, G.; LOGNAY, G.; MOUTEAU, A.; DUPONT, P.; RONDIA, A.; LATEUR, M. 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, p.238–244, 2012.

PISTONESI, M. F.; DI NEZIO, M.; CENTURIÓN, M. E.; LISTA, A. G.; FRAGOSO, W. D.; PONTES, M. J. C.; ARAÚJO, M. C. U.; FERNÁNDEZ BAND. Simultaneous determination of hydroquinone, resorcinol, phenol, m-cresol and p-cresol in untreated air samples using spectrofluorimetry and a custom multiple linear regression-successive projection algorithm. **Talanta**, v.83, p.320, 2010.

PLANCHON, V.; LATEUR, M.; DUPONT, P.; LOGNAY, G. Ascorbic acid level of Belgian apple genetic resources. **Scientia Horticulturae**, v.100, p.51–61, 2004.

PONTES, M. J. C.; CORTEZ, J.; GALVÃO, R. K. H.; PASQUINI, C.; COELHO, R. M.; CHIBA, M. K.; ABREU, M. F.; MADARI, B. E. Classification of Brazilian soils by using LIBS and variable selection in the wavelet domain. **Analytica Chimica Acta**, v.12, p.642, 2009.

PONTES, M. J. C.; ROCHA, A. M. J.; PIMENTEL, M. F.; PEREIRA, C. F. Determining the quality of insulating oils using near infrared spectroscopy and wavelength selection. **Microchemical Journal**, v.98, p.254-259, 2011.

QIN, J.; LU, R. Measurement of the optical properties of fruits and vegetables using spatially resolved hyperspectral diffuse reflectance imaging technique. **Postharvest Biology and Technology**, v.49, p. 355-365, 2008.

RIVERA, N. V.; GÓMEZ-SANCHIS, J.; CHANONA-PÉREZ, J.; CARRASCO, J. J.; MILLÁN-GIRALDO, M.; LORENTE, D.; CUBERO, S.; BLASCO, J. Early detection of mechanical damage in mango using NIR hyperspectral images and machine learning. **Biosystems Engineering**, v.122, p.91-98, 2014.

SALGUERO-CHAPARRO, L.; PALAGOS, B.; PEÑA-RODRIGUES, F.; ROGER, J. M. Calibration transfer of intact olive NIR spectra between a pre-dispersive instrument and a portable spectrometer. **Computers and Electronics in Agriculture**, v.96, p.202-208, 2013.

SANCHEZ, N. H.; LUROL, S.; ROGER, J. M.; BELLON-MAUREL, W. Robustness of models based on NIR spectra for sugar content prediction in apples. **Journal of Near Infrared Spectroscopy**, v.11, p.97–107, 2003.

SANTOS, D. B.; PEREIRA, M. E. C.; VIEIRA, E. L.; LIMA, M. A. C. Caracterização físico-química dos estádios de maturação da manga 'Tommy Atkins' produzida no município de Iaçu-BA. **Magistra**, v.20, p.342-348, 2008.

SARANWONG, S.; SORNSRIVICHAI, J.; KAWANO, S. Improvement of PLS calibration for Brix value and dry matter of mango using information from MLR. **Journal of Near Infrared Spectroscopy**, v.9, p.287–295, 2001.

SARANWONG, S.; SORNSRIVICHAI, J.; KAWANO, S. On-tree evaluation of harvesting quality of mango fruit using a hand-held NIR instrument. **Journal of Near Infrared Spectroscopy**, v.11, p.283-293, 2003.

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, p.137–145, 2004.

SCHMILOVITCH, Z.; MIZRACH, A.; HOFFMAN, A.; EGOZI, A.; FUCHS, Y. Determination of mango physiological indices by near-infrared spectrometry. **Postharvest Biology and Technology**, v.19, p.245–252, 2000.

SELIM, S. Z.; ISMAIL, M. A. K -Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.6, p.81-86, 1984.

SENA, M. M.; POPPI, R. J.; FRIGHETTO, R. T. S.; VALARINI, P. J. Avaliação do uso de métodos quimiométricos em análise de solos. **Química Nova**, v.23, p.547, 2000.

SHI, B.; ZHAO, L.; WANG, H.; ZHU, D. Signal optimization approaches on the prediction of apples firmness by near infrared spectroscopy. **Sensor Letters**, v.9, p.1062–1068, 2011.

SIGRIST, J. M. M. **Tecnologia pós-colheita para a comercialização de manga *in natura***. In: ROZANE, D. E.; DAREZZO, R. J.; AGUIAR, R. L.; AGUILERA, G. H. A.; ZAMBOLIM, L. (Ed.). *Manga - Produção integrada, industrialização e comercialização*. Viçosa: UFV, 2004. 604 p.

SIMÕES, S. S. **Desenvolvimento de métodos validados para a determinação de captopril usando espectrometria NIRR e calibração multivariada**. Universidade Federal da Paraíba, 2008. Tese de doutorado, 98p.

SKOOG, D. A.; HOLLER, F. J.; NEIMAN, T. A. **Princípios de Análise Instrumental**. 5ª ed. Porto Alegre: Bookman, 2006.

SLAUGHTER, D. C. Non-destructive determination of internal quality in peaches and nectarines. **Transactions of the ASAE**, v.38, p.617–623, 1995.

SLINKARD, K.; SINGLETON, V. L. Total phenol analysis: automation and comparison with manual methods. **American Journal of Enology and Viticulture**, v.16, p.144–158, 1977.

SNEE, R. D. Validation of regression models: methods and examples. **Technometrics**, v.19, p.415–428, 1977.

SOUZA, A. M.; BREITKREITZ, M. C.; FILGUEIRAS, P. R.; ROHWEDDER, J. J. R.; POPPI, R. J. Experimento didático de quimiometria para calibração multivariada na determinação de paracetamol em comprimidos comerciais utilizando espectroscopia no infravermelho próximo: um tutorial, parte II. **Química Nova**, v.36 (7), p.1057-1065, 2013.

SOUZA, D. M.; MADARI, B. E.; GUIMARÃES, F. F. Aplicação de técnicas multivariadas e inteligência artificial na análise de espectros de infravermelho para determinação de matéria orgânica em amostras de solo. **Química Nova**, v.35 (9), 1738-1745, 2012.

STATISTA. Apple production in Belgium from 2005/06 to 2016/17 (in tons). 2017. Disponível em <<https://www.statista.com/statistics/628823/apple-production-in-belgium/>>. Acessado em julho de 2017.

STREIF, J. **Optimum harvest date for different apple cultivars in the 'Bodensee' area**. In: DE JAGER, A., D. JOHNSON AND E. HOHN (Editors). European Commission COST 94: The postharvest treatment of fruit and vegetables - Determination and prediction of optimum harvest date of apples and pears. Proceedings of June, workshop, Loftus, Norway, 1994.

STREIF, J.; BUFFER, G. **Physiological ripening parameters at optimum picking date of different apple varieties**. Proceedings of the 23<sup>rd</sup> International Horticultural Congress, p.636, Florence, Italy, 1990.

SUBEDI, P. P.; WALSH, K. B. Assessment of sugar and starch in intact banana and mango fruit by SWNIRS spectroscopy. **Postharvest Biology and Technology**, v.62, p.238-45, 2011.

SUBEDI, P. P.; WALSH, K. B.; & OWENS, G. Prediction of mango eating quality at harvest using short-wave near infrared spectrometry. **Postharvest Biology and Technology**, 43, 326–334, 2007.

SUNU, J. Applications of K-means and spectral clustering to hyperspectral video sequences. California State University, 2014 (Thesis - Master of Science). 59p.

SUYKENS, J. A. K.; VAN GESTEL, T.; DE BRABANTER, J.; DE MOOR, B.; VANDEWALLE, J. **Least-Squares Support Vector Machines**, World Scientific: Singapore, 2002.

TALLADA, J. G.; NAGATA, M.; KOBAYASHI, T. Non-destructive estimation of firmness of strawberries (fragaria x ananassa duch.) using nir hyperspectral imaging. **Environmental Control in Biology**, v.44, p.245–255, 2006.

TEÓFILO, R. F. **Métodos Quimiométricos: Uma Visão Geral - Conceitos básicos de quimiometria**. Universidade Federal de Viçosa, Viçosa, vol. 1, 2013.

THANGARAJ, T.; IRULAPPAN, I. Studies on the maturity standards for mango fruit. **South Indian Horticulture**, v.37, p.341, 1989.

THISSEN, U.; USTUN, B.; MELSSSEN, W. J.; BUYDENS, L. M. C. Multivariate calibration with Least-squares support vector machine. **Analytical Chemistry**, 76, p.3099-3105, 2004.

VALENTE, M.; LEARDI, R.; SELF, G.; LUCIANO, G.; PAIN, J. P. Multivariate calibration of mango firmness using vis/NIRS spectroscopy and acoustic impulse method. **Journal of Food Engineering**, v.94, p.7-13, 2009.

VANDEGINSTE, B. G. M., MASSART, D. L., BUYDENS, L. M. C., JING, S., LEWI, P. J., SMEYERS-VERBEKE, J., **Handbook of Chemometrics and Qualimetrics: Part B**. Amsterdam: Elsevier, 1998.

VAPNIK, V. **Statistical Learning Theory**, John Wiley & Sons: New York, 1998.

VENTURA, M.; DE JAGER, A.; DE PUTTER, H.; ROELOFS, P. M. Non-destructive determination of soluble solids in apple fruits by near infrared spectroscopy (NIRS). **Postharvest Biology and Technology**, v.14, p.21–27, 1998.

VIDAL, M.; AMIGO, J. M. Pre-processing of hyperspectral images. Essential steps before image analysis. **Chemometrics and Intelligent Laboratory Systems**, v.117, p.138–148, 2012.

WANG, H.; PENG, J.; XIE, C.; BAO, Y.; HE, Y. Fruit Quality Evaluation Using Spectroscopy Technology: A Review. **Sensors**, v.15, p.11889-11927, 2015.

WANG, J.; NAKANO, K.; OHASHI, S.; KUBOTA, Y.; TAKIZAWA, K.; SASAKI, Y. Detection of external insect infestations in jujube fruit using hyperspectral reflectance imaging. **Biosystems Engineering**, v.108, p.345–351, 2011.

WANG, Y.; VELTKAMP, D. J.; KOWALSKI, R. Multivariate Instrument Standardization. **Analytical Chemistry**, v.63, p.2750-2756, 1991.

WENTZELL, P. D.; MONTOTO, L. V. Comparison of Principal Components Regression and Partial Least Squares Regression through Complex Mixture Simulation. **Chemometrics and Intelligent Laboratory System**, v.65, p.257-279, 2003.

WESTAD, F.; MARTENS H. Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression. **Journal of Near Infrared Spectroscopy**, v.8 (2), p.117-124, 2000.

WEYER, L., & LO, S.-C. **Spectra-structure correlations in the near-infrared**. In: J. M. CHALMERS & P. R. GRIFFITHS (Eds.). *Handbook of vibrational spectroscopy* (p.171–185). UK: Wiley and Sons, 2002.

WILLIAMS, P., & NORRIS, K. H. (2001). **Variable affecting near infrared spectroscopic analysis**. In: P. Williams & K. H. Norris (Eds.). *Near infrared technology in the agriculture and food industries* (2<sup>nd</sup> ed., p.171–185). St Paul: The American Association of Cereal Chemists.

WU, D.; SUN, D.-W. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review – Part I: Fundamentals. **Innovative Food Science & Emerging Technologies**, v.19, 2013, p.1-14, 2013.

ZUDE, M.; HEROLD, B.; ROGER, J. M.; BELLON-MAUREL, V.; LANDAHL, S. Nondestructive tests on the prediction of apple fruit firmness and soluble solids content on tree and in shelf life. **Journal of food engineering**, v.77, p.254–260, 2006.

## APÊNDICE A – PRODUÇÃO CIENTÍFICA

Ao longo do período de doutoramento foram publicados artigos científicos e um capítulo de livro com conteúdo diretamente relacionado ao desenvolvimento da presente tese. Os principais estudos são apresentados a seguir, enquanto outras produções científicas realizadas neste período (2013-2017) podem ser visualizadas na Plataforma Lattes: <http://lattes.cnpq.br/2477680651316334>.

- MARQUES, E. J. N.; PISSARD, A.; DARDENNE, P.; LATEUR, M.; PASQUINI, C.; PIMENTEL, M. F.; BAETEN, V. Evaluation of a handheld ultra-compact NIR spectrometer for rapid and non-destructive determination of apple fruit quality. **Artigo em processo de submissão para a revista Food Chemistry.**
  
- MARQUES, E. J. N.; FREITAS, S. T.; CAVALCANTE, I. H. L.; FIGUEIREDO NETO, A. Espectroscopia na região do infravermelho próximo (NIR): técnica analítica não destrutiva para determinação da qualidade de manga. In: Acácio Figueiredo Neto; Francisco de Assis Cardoso Almeida; Ítalo Herbert Lucena Cavalcante. (Org.). Manga - Maturação, Colheita e Conservação (1ed.). Petrolina: Fundação UNIVASF, 2017.
  
- MARQUES, E. J. N.; FREITAS, S. T.; PIMENTEL, M. F.; PASQUINI, C. Rapid and non-destructive determination of quality parameters in the ‘Tommy Atkins’ mango using a novel handheld near infrared spectrometer. **Food Chemistry**, v.197, p.1207-1214, 2016.



A manga (*Mangifera indica* L.) tem conquistado um espaço de destaque na fruticultura nacional. Em resposta a uma demanda crescente e dinâmica, o cultivo da mangueira vem se destacando pelo consistente emprego de tecnologia nos pomares, pelas altas produtividades e pela excelente qualidade dos frutos. Essa realidade é observada, principalmente, nos perímetros irrigados do Semiárido nordestino, onde são determinantes o uso da irrigação, a prática da indução floral e o emprego de variedades melhoradas. Conseqüentemente, o Nordeste já responde por 77% da produção de mangas do Brasil e por mais de 90% das exportações brasileiras desta fruta tropical. É necessário produzir, com o menor custo possível, frutos de alta qualidade e boa aparência, colhê-los no ponto de maturação adequado ao mercado de destino e mantê-los nessas condições pelo máximo de tempo possível. Além disso, é imprescindível estar atento às tendências dos mercados, pois estes estão em constante transformação, exigindo dos produtores habilidade para ajustar sua capacidade de produção e comercialização. A presente obra surge como uma opção de leitura para aqueles que se dedicam ao cultivo da mangueira, sejam eles técnicos, professores ou pesquisadores (Acácio Figueiredo Neto – UNIVASF).



## Analytical Methods

## Rapid and non-destructive determination of quality parameters in the 'Tommy Atkins' mango using a novel handheld near infrared spectrometer



Emanuel José Nascimento Marques<sup>a,\*</sup>, Sérgio Tonetto de Freitas<sup>b</sup>, Maria Fernanda Pimentel<sup>c</sup>, Celso Pasquini<sup>d</sup>

<sup>a</sup> Department of Fundamental Chemistry, Federal University of Pernambuco (UFPE), 50670-901 Recife, PE, Brazil

<sup>b</sup> Brasília Tropical Semi-Arid, Brazilian Agricultural Research Corporation, 55302-970 Petrolina, PE, Brazil

<sup>c</sup> Department of Chemical Engineering, Federal University of Pernambuco (UFPE), 50670-901 Recife, PE, Brazil

<sup>d</sup> Institute of Chemistry, University of Campinas (UNICAMP), 13084-971 Campinas, SP, Brazil

## ARTICLE INFO

## Article history:

Received 31 August 2015

Received in revised form 9 November 2015

Accepted 14 November 2015

Available online 24 November 2015

## Keywords:

Fruit quality

Non-destructive analysis

NIR spectroscopy

PLS regression

Mango

## ABSTRACT

The objective of this study was to evaluate the potential of a new handheld ultra-compact near infrared (NIR) spectrometer, based on the linear-variable filter (LVF) technology for rapid and non-destructive quality control analysis of the 'Tommy Atkins' mango. Multivariate calibration models were built using the Partial Least Squares (PLS) regression method to determine soluble solids (SS), dry matter (DM), titratable acidity (TA) and pulp firmness (PF). Different spectral pre-processing techniques were tested. Coefficient of determination and root mean square errors of prediction (RMSEP) values were, respectively: 0.92 and 0.55 % for SS, 0.67 and 0.51% for DM, 0.50 and 0.17% citric acid for TA, 0.72 and 12.2 N for PF. The predictive models allowed monitoring physico-chemical changes in each fruit during ripening. The results show the feasibility of using the new NIR handheld spectrometer to determine quality parameters in the 'Tommy Atkins' mango.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Mango (*Mangifera indica* L.) is a climacteric fruit cultivated in tropical and subtropical regions in the world and its production is increasing each year following consumer demand (Jahurul et al., 2015; Schmilovitch, Mizrahi, Hoffman, Egozi, & Fuchs, 2000). In 2013, mango production worldwide reached 40 million tons, moving more than 25 billion dollars in the producing countries (FAOSTAT, 2013). India, China, Thailand, Indonesia, Mexico, Pakistan and Brazil are the largest mango producers, accounting for about 75% of world production (FAOSTAT, 2013).

Although mango production is an important activity worldwide, problems related to fruit quality limit the consumption of this fruit. One of the main problems is marketing fruits with different maturity stages and consumer quality in the same batch (Kienzie et al., 2011). This problem is due to low precision of the visual parameters used to determine the maturity stage at harvest and to monitor fruit consumer quality during storage and marketing (Kienzie et al., 2011; Thangaraj & Irulappan, 1988). In addition, destructive

methods often used to estimate quality of the fruit sent to the market are labor-intensive, expensive, due to fruit sampling, and may not be representative of the commercial batch. Therefore, the development of non-destructive, reliable, accurate, fast and robust methods is essential to ensure a better quality of fruit delivered to consumers.

In the past decades, near infrared (NIR) spectroscopy has proven to be a reliable analytical technique for qualitative and quantitative analyses of a great variety of compounds present in different agricultural and food products (Cen & He, 2007; Nicolai et al., 2007; Norris, 1984). The technique has the advantages of being rapid and non-destructive, precise, with no sample preparation requirements. These characteristics make NIR spectroscopy an important tool for fruit quality control.

Previous studies have demonstrated the potential of NIR spectroscopy for nondestructive determination of quality parameters in mango, such as soluble solids, dry matter, titratable acidity, pulp firmness, starch content and other physico-chemical properties (Betemps, Fachinello & Galarça, 2011; Delwiche, Mekwatanakam, & Wang, 2008; Guthrie & Walsh, 1997; Jha et al., 2012, 2014; Mahayothee, Leitberger, Neidhart, Mühlbauer, & Carle, 2004; Sarawong, Somsriwichai, & Kawano, 2001, 2003, 2004;

\* Corresponding author.

E-mail address: [emanueljn@gmail.com](mailto:emanueljn@gmail.com) (E.J.N. Marques).