



**UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO
CURSO DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA**

MEIRIANE DIAS DE SOUZA

**DETERMINAÇÃO DE COMPOSTOS BIOQUÍMICOS EM FOLHAS DE
MANGUEIRA E MARACUJAZEIRO AMARELO COM USO DA
ESPECTROSCOPIA VIS-NIR**

JUAZEIRO-BA

2022

MEIRIANE DIAS DE SOUZA

**DETERMINAÇÃO DE COMPOSTOS BIOQUÍMICOS EM FOLHAS DE
MANGUEIRA E MARACUJAZEIRO AMARELO COM USO DA
ESPECTROSCOPIA VIS-NIR**

Trabalho de Dissertação apresentado à Universidade Federal do Vale do São Francisco – UNIVASF, Campus Juazeiro, como requisito para obtenção do título de Mestre em Engenharia Agrícola.

Orientador: Prof. Dr. Daniel dos Santos Costa

Coorientadores: Prof. Dr. Ítalo Herbert Lucena Cavalcante e Prof. Dr. José Aliçandro Bezerra da Silva

JUAZEIRO-BA

2022

Souza, Meiriane Dias de
729c Determinação de compostos bioquímicos em folhas de mangueira e maracujazeiro amarelo com uso da espectroscopia VIS-NIR / Meiriane Dias de Souza, Juazeiro-Ba, 2022.
xvii, 119 f.: il. ; 29 cm.

Trabalho de Conclusão (Curso de Pós-Graduação em Engenharia Agrícola) -
Universidade Federal do Vale do São Francisco, Campus Juazeiro-BA, 2022.

Orientador: Prof. Dr. Daniel dos Santos Costa.

1. Fruticultura. 2. Métodos ópticos. 3. Análises bioquímicas. 4. Quimiometria I. Título.
II. Costa, Daniel dos Santos. III. Universidade Federal do Vale do São Francisco.

CDD 634.44

Ficha catalográfica elaborada pelo Sistema Integrado de Biblioteca SIBI/UNIVASF
Bibliotecário: Márcio Pataro. CRB-5/1369.

**UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO
CURSO DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA**

**FOLHA DE APROVAÇÃO
MEIRIANE DIAS DE SOUZA**

**DETERMINAÇÃO DE COMPOSTOS BIOQUÍMICOS EM FOLHAS DE
MANGUEIRA E MARACUJAZEIRO AMARELO COM USO DA
ESPECTROSCOPIA VIS-NIR**

Trabalho de conclusão de curso apresentado como requisito parcial para obtenção do título de Mestre em Engenharia Agrícola, pela Universidade Federal do Vale do São Francisco.

Aprovado em: 25 de fevereiro de 2022.

Banca Examinadora



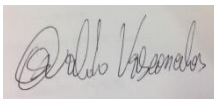
Prof. Dr. Daniel dos Santos Costa (PPGEA/UNIVASF)
(Orientador)



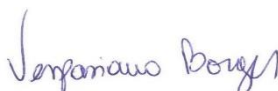
Prof. Dr. Ítalo Herbert Lucena Cavalcante (PPGPV/UNIVASF)
(Coorientador)



Prof. Dr. José Aliçandro Bezerra da Silva (PPGEA/UNIVASF)
(Coorientador)



Prof. Dr. Osvaldo Campelo de Mello Vasconcelos (IFAP)



Prof. Dr. Vespasiano Borges de Paiva Neto (PPGPV/UNIVASF)

À minha família e ao meu companheiro.

AGRADECIMENTOS

À Universidade Federal do Vale do São Francisco (UNIVASF) e ao Programa de Pós-graduação em Engenharia Agrícola, pela oportunidade e todo aprendizado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 pela concessão da bolsa de estudos.

Ao professor DSc. Daniel dos Santos Costa, pela orientação, ensinamentos e confiança, e a todos os meus colegas do Laboratório de Energia na Agricultura (LENA).

Aos professores DSc. Ítalo Herbert Lucena Cavalcante e DSc. José Aliçandro Bezerra da Silva, ao Grupo de Pesquisa em Fruticultura do Vale do São Francisco, FRUTVASF, e a outros que colaboraram diretamente para a realização desse trabalho.

Aos meus pais, Rosimary e Edimicio, por todo apoio, incentivo, força e por tudo que fizeram e ainda fazem por mim e por meus irmãos, Maria Ana e Daniel, que também contribuíram muito para eu pudesse me dedicar a esse trabalho.

Ao meu companheiro, parceiro, amigo e noivo, mein Lieber Folkert, pela paciência, companheirismo, força e apoio incondicional.

A todos os meus colegas de curso que de alguma forma contribuíram para a realização desse trabalho.

E a todos os meus amigos e familiares que sempre acreditaram em mim e me apoiaram em todos os momentos.

RESUMO

Alguns compostos bioquímicos podem fornecer informações importantes sobre o estado fisiológico das plantas e acerca de processos morfológicos, fisiológicos e bioquímicos que ocorrem durante o seu desenvolvimento. Conseqüentemente, o monitoramento desses compostos pode contribuir para o processo de tomada de decisões no campo e aumento de produtividade da cultura e para que não haja comprometimento de produtividade, é imprescindível o monitoramento constante da planta. Entretanto, embora existam diversos métodos para determinação de compostos bioquímicos em amostras vegetais, normalmente trata-se de métodos analíticos laboratoriais, que apesar de confiáveis, podem ser demorados, complexos e de alto custo. Com isso, o desenvolvimento de métodos de análises mais rápidos, simples e não destrutivos é crucial para otimizar os processos produtivos. Diante disso, o principal objetivo do trabalho foi determinar os atributos bioquímicos em folhas de mangueira (*Mangifera indica* L.) e maracujazeiro amarelo (*Passiflora edulis*) com uso da espectroscopia de refletância Vis-NIR. O trabalho foi desenvolvido na região do Vale do São Francisco, seguindo as seguintes etapas: (1) amostragem das folhas; (2) aquisição dos espectros de refletância; (3) determinação dos teores dos compostos bioquímicos com uso dos métodos de referência; e (4) desenvolvimento dos modelos preditivos. Os melhores modelos de predição dos diferentes atributos bioquímicos foram selecionados com base no desempenho dos modelos, sendo os parâmetros estatísticos avaliados: coeficiente de determinação (R^2); erro quadrático médio de calibração (RMSEC); erro padrão de calibração (SEC); erro quadrático médio de validação cruzada (RMSECV); erro padrão de validação cruzada (SECV). Os métodos de regressão utilizados foram: *Partial Least Squares Regression* (PLSR), *Multiple Linear Regression* (MLR), para ambas as culturas; *Gaussian Processes* (GP), *Multilayer Perceptron* (MPL), *Model Tree* (M5P), *Random Forest* (RF) e *Support Vector Machine* (SVM), para as folhas de mangueira. Para folhas de mangueira, os melhores modelos foram encontrados utilizando o método MLR, com R^2_{cv} iguais a 0,92, 0,71 e 0,55 para amido, carboidratos e sacarose, respectivamente. Para folhas de maracujazeiro, os melhores modelos foram encontrados utilizando o método PLSR para amido e carboidratos, com R^2_{cv} iguais a 0,83 e 0,39; e MLR para sacarose, aminoácido, proteína, nitrato e nitrato redutase, com R^2_{cv} iguais a 0,98, 0,99, 0,86, 0,99 e 0,84, respectivamente. Deste modo, a espectroscopia é uma ferramenta com alto potencial para determinação de compostos bioquímicos em folhas de mangueira e maracujazeiro.

Palavras-chave: Fruticultura. Métodos ópticos. Análises bioquímicas. Quimiometria.

ABSTRACT

Some biochemical compounds can provide important information about the physiological state of plants and about morphological, physiological and biochemical processes that occur during their development. Consequently, the monitoring of these compounds can contribute to the decision making process in the field and increase crop productivity. However, although there are several methods for determination of biochemical compounds in plant samples, usually these are laboratory analytical methods, which despite being reliable, can be time consuming, complex and high cost. Therefore, the development of faster, simple and non-destructive analysis methods is crucial to optimize the production processes. Therefore, the main objective of this work was to determine the biochemical attributes in mango (*Mangifera indica* L.) and yellow passion fruit (*Passiflora edulis*) leaves using Vis-NIR reflectance spectroscopy. The work was developed in the Vale do São Francisco region, following the following steps: (1) sampling of the leaves; (2) acquisition of the reflectance spectra; (3) determination of the contents of the biochemical compounds using the reference methods; and (4) development of the predictive models. The best predictive models for the different biochemical attributes were selected based on the performance of the models, and the following statistical parameters were evaluated: coefficient of determination (R^2); root mean square error of calibration (RMSEC); standard error of calibration (SEC); root mean square error of cross-validation (RMSECV); standard error of cross-validation (SECV). The regression methods used were Partial Least Squares Regression (PLSR), Multiple Linear Regression (MLR) for both crops; Gaussian Processes (GP), Multilayer Perceptron (MPL), Model Tree (M5P), Random Forest (RF) and Support Vector Machine (SVM) for mango leaves. For mango leaves, the best models were found using the MLR method, with R^2_{cv} equal to 0.92, 0.71 and 0.55 for starch, carbohydrates and sucrose, respectively. For passion fruit leaves, the best models were found using the PLSR method for starch and carbohydrates, with R^2_{cv} equal to 0.83 and 0.39; and MLR for sucrose, amino acid, protein, nitrate and nitrate reductase, with R^2_{cv} equal to 0.98, 0.99, 0.86, 0.99 and 0.84, respectively. Thus, spectroscopy is a tool with high potential for determination of biochemical compounds in mango and passion fruit leaves.

Keywords: Fruticulture. Optical methods. Biochemical analysis. Chemometrics.

LISTA DE FIGURAS

- Figura 1 - A natureza ondulatória de um feixe de radiação com uma única 23
frequência. Em (a), uma onda plano-polarizada é apresentada
propagando-se ao longo do eixo x, onde o campo elétrico oscila em um
plano perpendicular ao campo magnético. Em (b), somente as oscilações
do campo elétrico são mostradas.
- Figura 2 - Espectro eletromagnético. 25
- Figura 3 - Métodos de absorção de radiação. A radiação com potência incidente 26
igual a P_0 pode ser absorvida pelo analito, resultando em um feixe
transmitido de menor potência P. O espectro de absorção resultante é
exposto em (b).
- Figura 4 - Diagrama do processo de desenvolvimento dos modelos de predição 34
teores de carboidratos solúveis, amido e sacarose em folhas de
mangueira.
- Figura 5 - Sistema óptico eletrônico constituído por um espectrômetro *FieldSpec3*, 35
equipado com *probe plant*; e computador com o software RS3 em (a).
Em (b), momento de análise da folha de mangueira utilizando a *probe
plant*.
- Figura 6 - Espectro bruto de absorbância média de amostras foliares de mangueira 41
das variedades Keitt, Omer e Shelly, e Palmer.
- Figura 7 - Gráficos de carregamento de peso das variáveis independentes: (a) 45
amido com pré-tratamento MA 11 pts, OSC, conjunto geral; (b)
carboidrato com pré-tratamento MA 29 pts, OSC, conjunto geral; (c)
sacarose com pré-tratamento MA 23 pts, OSC, conjunto Keitt.
- Figura 8 - Gráficos de dispersão de calibração e validação cruzada dos modelos 49
MLR: (a) amido com pré-tratamento MA 11 pts, OSC, conjunto geral;
(b) carboidrato com pré-tratamento MA 29 pts, OSC, conjunto geral; (c)
sacarose com pré-tratamento MA 23 pts, OSC, conjunto Keitt.
- Figura 9 - Diagrama do processo de desenvolvimento dos modelos de predição 70
para compostos bioquímicos em folhas de maracujazeiro.

- Figura 10 - Sistema óptico eletrônico constituído por um espectrômetro *FieldSpec3*, 71
equipado com *probe plant*; e computador com o software RS3 em (a).
Em (b), momento de análise da folha de mangueira utilizando a *probe
plant*.
- Figura 11 - Espectro bruto de absorbância média das amostras foliares de 76
maracujazeiro amarelo.
- Figura 12 - Gráficos de carregamento de peso das variáveis independentes: (a) 80
amido com pré-tratamento MA 17 pts, OSC; (b) carboidratos solúveis
com pré-tratamento MA 5 pts, OSC; (c) sacarose com pré-tratamento
MA 11 pts, OSC; (d) aminoácido com pré-tratamento MA 5 pts, OSC;
(e) proteína com pré-tratamento MA 23 pts, OSC; (f) nitrato com pré-
tratamento MA 5 pts, OSC; (g) nitrato redutase com pré-tratamento MA
23 pts, OSC.
- Figura 13 - Gráficos de dispersão de calibração e validação cruzada dos melhores 85
modelos de predição: (a) amido com pré-tratamento MA 17 pts, OSC
(MLR); (b) carboidratos solúveis com pré-tratamento MA 5 pts (MLR),
OSC; (c) sacarose com pré-tratamento MA 11 pts, OSC (PLSR); (d)
aminoácidos totais com pré-tratamento MA 5 pts, OSC (PLSR); (e)
proteína com pré-tratamento MA 23 pts, OSC (PLSR); (f) nitrato com
pré-tratamento MA 5 pts, OSC (PLSR); (g) nitrato redutase com pré-
tratamento MA 23 pts, OSC (PLSR).

LISTA DE TABELAS

- Tabela 1 - Estatística descritiva dos teores de amido ($\mu\text{g g}^{-1}$ de MF), carboidratos solúveis (mmol g^{-1} de MF) e sacarose (mg g^{-1} de MF) em folhas de mangueira para os conjuntos de dados "Keitt", "Omer", "Shelly, Palmer" e "Geral". 40
- Tabela 2 - Parâmetros estatísticos dos melhores modelos PLSR para predição de amido, carboidratos solúveis e sacarose em folhas de mangueira para os conjuntos de dados "Keitt", "Omer, Shelly", "Palmer" e "Geral". 42
- Tabela 3 - Parâmetros estatísticos dos melhores modelos MLR com variáveis selecionadas para os atributos amido, carboidratos solúveis (conjunto geral) e sacarose (conjunto Keitt) em folhas de mangueira para os conjuntos de dados "Keitt", "Omer, Shelly", "Palmer" e "Geral". 44
- Tabela 4 - Parâmetros estatísticos dos melhores modelos PLSR com variáveis selecionadas para predição de amido, carboidratos solúveis e sacarose em folhas de mangueira para os conjuntos de dados "Keitt", "Omer, Shelly", "Palmer" e "Geral". 47
- Tabela 5 - Parâmetros estatísticos dos modelos matemáticos *Gaussian Processes* (GP), *Multilayer Perceptron* (MPL), *Model Tree* (M5P), *Random Forest* (RF) e *Support Vector Machine* (SVM) para predição de amido, carboidratos solúveis e sacarose em folhas de mangueira para os conjuntos de dados "Keitt", "Omer, Shelly", "Palmer" e "Geral". 48
- Tabela 6 - Estatística descritiva dos teores de amido (mmol g^{-1} de MF), carboidratos solúveis (mg g^{-1} de MF), sacarose (mg g^{-1} de MF), aminoácidos totais ($\mu\text{g g}^{-1}$ de MF), proteína (mg g^{-1} de MF), nitrato (mg g^{-1} de MF) e nitrato redutase ($\mu\text{g g}^{-1}$ de MF) em folhas de maracujazeiro. 75
- Tabela 7 - Parâmetros estatísticos dos melhores modelos PLSR para predição de amido, carboidratos solúveis, sacarose, aminoácidos totais, proteína, nitrato e nitrato redutase em folhas de maracujazeiro. 77
- Tabela 8 - Parâmetros estatísticos dos melhores modelos MLR com variáveis selecionadas para predição de amido, carboidratos solúveis, sacarose, aminoácidos totais, proteína, nitrato e nitrato redutase em folhas de maracujazeiro. 79

Tabela 9 - Parâmetros estatísticos dos melhores modelos PLSR com variáveis 84 selecionadas para predição de amido, carboidrato, sacarose, aminoácido, proteína, nitrato e nitrato redutase em folhas de maracujazeiro.

LISTA DE ABREVIATURAS E SIGLAS

ANOVA	Análise de Variância
AA totais	Aminoácidos totais
BSh	Clima Semiárido quente
cal	Calibração
cv	Validação cruzada
CHOsol	Carboidratos solúveis
GP	<i>Gaussian Processes</i>
KCN	Cianeto de potássio
MAE	<i>Mean Absolute Error</i>
Máx.	Máximo
MF	Material foliar fresco
Mín.	Mínimo
MLR	<i>Multiple Linear Regression</i>
MPL	<i>Multilayer Perceptron</i>
MSC	<i>Multiplicative Scatter Correction</i>
M5P	<i>Model Tree</i>
MCW	Metanol, clorofórmio e água
NIR	<i>Near Infrared</i>
OSC	<i>Orthogonal Signal Correction</i>
PCA	<i>Principal Component Analysis</i>
PLSR	<i>Partial Least Squares Regression</i>
PVPP	Polivinilpolipirrolidona
R ²	<i>Coefficient of determination</i>

RAE	Erro relativo absoluto
RER	<i>Range Error Ratio</i>
RF	<i>Random Forest</i>
RMSEC	<i>Root Mean Squared Error of Calibration</i>
RMSECV	<i>Root Mean Squared Error of Cross-Validation</i>
RPD	<i>Residual Predictive Desviation</i>
RRSE	<i>Root Relative Squared Error</i>
SD	<i>Standard deviation</i>
SEC	<i>Standard Error of Calibration</i>
SECV	<i>Standard Error of Cross-Validation</i>
SNV	<i>Standard Normal Variate</i>
SVM	<i>Support Vector Machine</i>
SWIR 1	<i>Short Wavelength Infrared (1001nm – 1830 nm)</i>
SWIR 2	<i>Short Wavelength Infrared (1831 nm – 2500 nm)</i>
Vis	<i>Visible</i>

LISTA DE SÍMBOLOS

ATP	Adenosina trifosfato
%	Porcentagem
°C	Graus Celsius
λ	Comprimento de onda
μl	Microlitro
μg	Micrograma
cm	Centímetro
EDTA	Etilenodiamino tetra acético
f	Frequência
g	Gramma
h	Constante de Planck
H_2SO_4	Ácido sulfúrico
Hz	Hertz
GOGAT	Glutamina sintetase
GS	Glutamato sintase
KCN	Cianeto de potássio
M	Molar
mmol	Milimol
mg	Miligrama
ml	Mililitro
P	Potência
P_o	Radiação incidente
pH	Potencial Hidrogeniônico

N	Nitrogênio
nm	Nanômetro
NH_4^+	Nitrato
NO_3^-	Amônio
R	Refletância
<i>T</i>	Período

SUMÁRIO

1 INTRODUÇÃO GERAL	15
1.1 OBJETIVO GERAL.....	17
1.2 OBJETIVOS ESPECÍFICOS	17
1.3 HIPÓTESE GERAL	17
2 REFERENCIAL TEÓRICO.....	18
2.1 A CULTURA DA MANGUEIRA	18
2.2 A CULTURA DO MARACUJAZEIRO.....	19
2.3 PARÂMETROS BIOQUÍMICOS EM FOLHAS VEGETAIS	19
2.3.1 Carboidratos	20
2.3.2 Aminoácidos e proteínas	20
2.3.3 Nitrato e nitrato redutase	21
2.4 ESPECTROSCOPIA	22
2.5 QUIMIOMETRIA	26
3 ARTIGOS.....	30
3.1 DETERMINAÇÃO DO TEOR DE CARBOIDRATOS SOLÚVEIS, AMIDO E SACAROSE EM FOLHAS DE MANGUEIRA COM USO DA ESPECTROSCOPIA VIS-NIR	30
3.1.1 INTRODUÇÃO.....	32
3.1.2 MATERIAL E MÉTODOS.....	33
3.1.3 RESULTADOS	39
3.1.4 DISCUSSÃO	51
3.1.5 CONCLUSÕES	57
3.1.6 REFERÊNCIAS	59
3.2 DETERMINAÇÃO DE COMPOSTOS BIOQUÍMICOS EM FOLHAS DE MARACUJAZEIRO AMARELO COM USO DA ESPECTROSCOPIA VIS-NIR.....	66

3.2.1 INTRODUÇÃO.....	68
3.2.2 MATERIAL E MÉTODOS.....	69
3.2.3 RESULTADOS	75
3.2.4 DISCUSSÃO.....	89
3.2.5 CONCLUSÕES.....	93
3.2.6 REFERÊNCIAS	95
4 DISCUSSÃO GERAL.....	99
5 CONSIDERAÇÕES FINAIS	101
6 REFERÊNCIAS	102
ANEXO A – VARIÁVEIS SELECIONADAS DO ARTIGO 1.....	113
ANEXO B – VARIÁVEIS SELECIONADAS DO ARTIGO 2.....	114

1 INTRODUÇÃO GERAL

O principal papel da agricultura é garantir a segurança alimentar da população mundial através da produção de alimentos de forma mais sustentável possível, além de fornecer matéria-prima para produtos indispensáveis para a sociedade. Entretanto, o cenário global não é necessariamente favorável: a população mundial tem a perspectiva de aumentar de 7,7 bilhões para 9,7 bilhões até 2050 (ONU, [s.d.]); supostas mudanças climáticas, que podem comprometer a produção de alimentos (YASSITEPE *et al.*, 2020); expansão de terras agrícolas com uso de queimada (SILVA *et al.*, 2021); exploração intensiva dos recursos naturais não renováveis (LANNA *et al.*, 2021); consumidores cada vez mais exigentes quanto a qualidade e funcionalidade dos alimentos (LUCAS *et al.*, 2021).

Deste modo, o grande desafio é otimizar as cadeias produtivas do setor agrícola com o propósito de garantir melhores valores de produtividade em menores áreas, com produtos de maior qualidade e produzidos com menor impacto ambiental. E para que isso seja possível, é imprescindível o desenvolvimento e adoção de novas tecnologias (COLEZEA *et al.*, 2018).

Nesse contexto surge então a agricultura digital, que consiste na inserção de tecnologias digitais na cadeia produtiva de produtos agrícolas, promovendo vantagens competitivas e benefícios socioambientais. Este novo modelo de agricultura baseia-se no processamento de dados coletados em todas as etapas produtivas, de modo que esses dados podem ser coletados por meio de diversas ferramentas como drones, sensores, satélites e veículos, conectados entre si e com uma central de coleta de dados (MASSARUHÁ *et al.*, 2020).

A análise desses dados combinados com estratégias de aprendizagem de máquina e algoritmos computacionais possibilita uma análise inovadora desses dados (SOUZA *et al.*, 2019), em diferentes níveis de complexidade, incluindo a previsão de aptidão de terras para cultivo (GANESAN *et al.*, 2021), mapeamento de atributos do solo e práticas agrícolas (MENDES *et al.*, 2020), classificação de áreas em cultivo agrícola a partir de imagens (CARDOSO, 2020), predição de evapotranspiração de referência (ALTHOFF *et al.*, 2018), dentre outros.

Uma ferramenta com alto potencial tecnológico e que tem se mostrado muito promissora é a espectroscopia do infravermelho, proporcionando soluções fantásticas para os propósitos da nova revolução da agricultura moderna (DALE *et al.*, 2013). Por décadas tem sido utilizada para identificar, compreender e quantificar materiais sólidos, líquidos e gasosos

em várias áreas do conhecimento. Essas medições espectroscópicas detectam absorção de luz devido às bandas químicas específicas, e análises detalhadas são utilizadas para determinar a abundância e estado físico das espécies de absorbância detectadas (PABÓN, 2015).

Os métodos químicos analíticos convencionais para análises de atributos bioquímicos em folhas vegetais são complexos, demorados e normalmente requerem insumos sofisticados, muitos deles de uso restrito. A determinação de carboidratos, no método proposto por Dubois *et al.* (1956) por exemplo, requer o preparo das amostras, usos de reagentes como ácido sulfúrico, cuja compra e utilização é controlada pela polícia federal (BRASIL, 2019), utilização de materiais e equipamentos como vidrarias, tubos de ensaio, balança analítica e espectrofotômetro. Essa complexidade de análise limita o número de amostras analisadas, limitando conseqüentemente a detecção da variabilidade no campo, bem como a contribuição das análises para o processo de tomada de decisão, em virtude do tempo necessário para a obtenção dos resultados das análises.

A espectroscopia está sendo utilizada em diversas áreas do setor agrícola tais como: análise de fertilidade, atributos e qualidade biológica do solo (COMINO *et al.*, 2018; MUNAWAR *et al.*, 2020; REDA *et al.*, 2021); avaliação da umidade do solo e planta para tomada de decisão quanto à irrigação (FILIPPUCCI *et al.*, 2020; TORRES *et al.*, 2019); avaliação da maturação e qualidade de pós-colheita de frutas e vegetais (SHAH *et al.*, 2020; WALSH *et al.*, 2020); diferenciação entre produtos de origem convencional e orgânico (SAKANE; MATSUURA; SIMIONI, 2019) e avaliação de alterações bioquímicas em plantas doentes (SKOLIK *et al.*, 2020).

Considerando a importância e a necessidade da medição de atributos bioquímicos em folhas vegetais de plantas frutíferas, proporcionando informações fisiológicas importantes com maior rapidez para a tomada de decisão no campo, o desenvolvimento de modelos para predição de atributos bioquímicos em folhas, com uso da espectroscopia Vis-NIR, pode contribuir para o desenvolvimento tecnológico da agricultura, além de um potencial objeto de estudo para desenvolvimento de equipamentos eletrônicos que realizem essas leituras bioquímicas ainda no pomar.

Diante do exposto, existe uma demanda por métodos não destrutivos e rápidos de análise bioquímica em folhas de plantas frutíferas, como mangueira e maracujazeiro, para a tomada de decisão no campo, e por meio da espectroscopia Vis-NIR pode ser uma ótima ferramenta para essa finalidade.

1.1 OBJETIVO GERAL

Determinar compostos bioquímicos em folhas de mangueira (*Mangifera indica* L.), variedades Keitt, Omer, Shelly e Palmer, e maracujazeiro amarelo (*Passiflora edulis*) com uso da espectroscopia de refletância Vis-NIR.

1.2 OBJETIVOS ESPECÍFICOS

- Desenvolver modelos preditivos para a determinação de carboidratos solúveis, amido e sacarose com uso da espectroscopia Vis-NIR em folhas de mangueira;
- Desenvolver modelos preditivos para a determinação de carboidratos solúveis, amido, sacarose, aminoácidos totais, proteínas, nitrato e nitrato redutase com uso da espectroscopia Vis-NIR em folhas de maracujazeiro amarelo;
- Identificar as assinaturas espectrais em relação a carboidratos solúveis, amido e sacarose em folhas de mangueira;
- Identificar as assinaturas espectrais em relação a carboidratos solúveis, amido, sacarose, aminoácidos totais, proteínas, nitrato e nitrato redutase em folhas de maracujazeiro.

1.3 HIPÓTESE GERAL

É possível determinar os teores de carboidratos solúveis, amido e sacarose em folhas de mangueira e maracujazeiro, além dos teores de proteínas, aminoácidos, nitrato e nitrato redutase em folhas de maracujazeiro utilizando a espectroscopia de refletância Vis-NIR.

2 REFERENCIAL TEÓRICO

2.1 A CULTURA DA MANGUEIRA

No Brasil, a manga é uma das frutas mais produzidas e consumidas, e a mais exportada, com receita de 223 milhões de dólares no período de janeiro a novembro de 2021 e volume de quase 245 mil toneladas, 16% a mais que o volume exportado no mesmo período do ano anterior (CEPEA, 2021). Desse volume total, quase 90% são produzidos na região do Vale do São Francisco, que teve aumento significativo de área produtiva devido à alta rentabilidade dos anos anteriores, chegando a 51 mil hectares em 2021 com perspectiva de continuar em expansão no ano de 2022 (EMBRAPA, 2021).

Independente da região, os maiores desafios da produção de manga são a variação do florescimento, em função da desuniformidade da maturação dos ramos, e a tomada da decisão quanto ao momento ideal para indução floral (OLIVEIRA *et al.*, 2019). O florescimento da mangueira ocorre no Brasil de forma natural entre os meses de junho e julho, devido às temperaturas mais baixas, ou de forma artificial durante todo o ano, através da aplicação de reguladores de crescimento, técnicas de manejo, de acordo com as condições da planta e demanda comercial. Além disso, é uma das espécies frutíferas mais favoráveis para a manipulação floral, característica de grande importância para o gerenciamento da época e qualidade do florescimento, que resulta na colheita de produtos de qualidade, com maior uniformidade, em épocas de maior demanda e maior proveito das oportunidades de mercado (DAVENPORT, 2007).

As técnicas de indução floral podem ser baseadas no uso de retardantes de crescimento vegetal, como o PBZ (paclobutrazol), na aplicação de fertilizantes como os nitratos de cálcio e potássio, além de alguns tratamentos culturais como podas e despontes, para evitar o gasto de energia para as partes vegetativas da planta, estresse hídrico, entre outros (LIMA *et al.*, 2016).

O uso dessas técnicas de indução de florescimento permite que os produtores possam planejar as colheitas dos anos seguintes, a partir da última colheita, entretanto, ainda não há parâmetros técnicos bem definidos para determinar o melhor momento para a indução floral. Nesse contexto, a determinação do teor de carboidratos e enzimas associadas apresenta um papel promissor no estabelecimento de características fisiológicas para delimitação do momento ideal para indução floral.

2.2 A CULTURA DO MARACUJAZEIRO

O gênero *Passiflora*, o mais comum da família Passifloraceae, é caracterizado pela diversidade morfológica e compreende pouco mais de 500 espécies, das quais cerca de um quarto são nativas do Brasil. Entre as espécies nativas, cerca de 70 produzem frutos comestíveis, sendo o maracujazeiro amarelo (*Passiflora edulis*), também chamado de maracujazeiro azedo, o mais cultivado no país (CERVI, 2006; ULMER; MACDUGAL, 2004).

Atualmente o Brasil é o maior produtor e exportador mundial de maracujá, com quase 700 mil toneladas produzidas em 2020, dos quais cerca de 65% é produzido na região nordeste (ABRAFRUTAS, 2021; IBGE, 2021). Consumida em diversas formas em todas as regiões do Brasil, a fruta ganhou importância comercial na segunda metade da década de 1970, quando a produção nacional começou a crescer em função da demanda nacional e para exportação. Até a década de 1990, o estado do Pará foi o maior produtor de maracujá do país. Entretanto, devido a desafios impostos por problemas fitopatológicos, custos de produção, além das limitações de transporte, houve o declínio da produção no estado, resultando na concentração de plantio da fruta na região nordeste, principalmente nos estados da Bahia e Ceará (LANDAU *et al.*, 2020; PIRES *et al.*, 2011).

Apesar do destaque, como produtor e exportador, a produtividade média da fruta no Brasil ainda é baixa. Estima-se que os produtores consigam atingir produtividade média entre 14 a 30 toneladas por hectare, enquanto em cultivos mais tecnológicos, a produtividade pode chegar a 60 toneladas por hectare (ABRAFRUTAS, 2021).

2.3 PARÂMETROS BIOQUÍMICOS EM FOLHAS VEGETAIS

Para se alcançar bons valores de produtividade em espécies frutíferas, é imprescindível um conhecimento consolidado acerca do ciclo da cultura, bem como das alterações fisiológicas e bioquímicas no decorrer de seu desenvolvimento, inclusive em resposta às determinadas condições ambientais ou diante de estresses bióticos e abióticos. Isto é, compreender e analisar os aspectos fisiológicos e bioquímicos de uma planta é um marco essencial para o aumento da produtividade (ELY *et al.*, 2019; SERBIN *et al.*, 2012; WHITE *et al.*, 2016).

Além disso, alguns compostos bioquímicos fornecem informações importantes sobre o estado fisiológico da planta, exemplo: a diminuição do teor de amido nas folhas pode indicar indiretamente uma redução da atividade fotossintética (EARLES *et al.*, 2018); baixa atividade da enzima nitrato redutase indica baixa assimilação de nitrogênio

(MARTUSCELLO *et al.*, 2016). A partir do monitoramento constante de compostos bioquímicos na planta, é possível tornar o processo de tomada de decisão no campo mais assertivo, evitando assim danos à cultura, além de promover aumento da produtividade.

2.3.1 Carboidratos

Os principais carboidratos não estruturais acumulados na parte aérea de plantas são o amido e os açúcares solúveis redutores, como glicose e frutose, e não redutores, como a sacarose. Esses carboidratos são provenientes da etapa bioquímica da fotossíntese e possuem diversas atribuições no metabolismo vegetal. Durante esse processo são produzidas seis moléculas chamadas trioses fosfatos, que são a base para obtenção dos carboidratos como amido, principal carboidrato de reserva das plantas, inclusive de armazenamento foliar, e a sacarose, forma preferencial de transporte de carboidrato nos vegetais (TAIZ *et al.*, 2017).

A magnitude do acúmulo de carboidratos depende muito da espécie vegetal, do momento do ciclo e das condições ambientais e de manejo. Entretanto, de modo geral, o amido funciona como um mecanismo de armazenamento de energia quando a síntese e o transporte de sacarose são limitados durante o dia ou durante a noite, quando não há atividade fotossintética. Além disso, há mecanismos de regulação do metabolismo de carbono. Maiores teores de carboidratos em tecidos drenos estimulam o crescimento e o desenvolvimento da planta, enquanto baixos teores estimulam a atividade fotossintética. O transporte de sacarose depende da disponibilidade de carboidratos nas folhas-fonte e da necessidade para uso de energia e a formação de polissacarídeos de reserva nos tecidos-dreno (TAIZ *et al.*, 2017).

Em síntese, os carboidratos são importantes fatores relacionados ao crescimento, desenvolvimento e adaptabilidade da cultura a diferentes ambientes (AMARANTE *et al.*, 2007), além de serem fundamentais para o fornecimento de energia para a formação e desenvolvimento de flores e frutos, bem como para a biossíntese de aminoácidos (BRONDANI *et al.*, 2012; PRASAD *et al.*, 2014; SILVA *et al.*, 2020; LOBO *et al.*, 2019).

2.3.2 Aminoácidos e proteínas

As proteínas são polímeros de aminoácidos e controlam praticamente todos os processamentos metabólicos que acontecem dentro de uma célula, assumindo assim uma infinidade de funções. Os aminoácidos, componentes das proteínas, são monômeros orgânicos compostos basicamente por um grupo carboxila e um grupo amino ligados ao mesmo átomo de carbono. Essas substâncias orgânicas podem atuar na assimilação de nutrientes, ativação

do metabolismo das plantas, bem como no fornecimento de esqueletos de carbono para inúmeras vias metabólicas, incluindo os metabolismos secundário e hormonal (TAIZ *et al.*, 2017).

Podem participar da atividade celular na forma de enzimas, inibidores de enzimas, hormônios, proteínas estruturais, de transporte e reserva. Além disso, também são utilizados como substrato respiratório pelas plantas sob deficiência de carbono, ou seja, durante a redução da taxa fotossintética (CULTIVAR, 2020; VERPOORTE; ALFERMANN, 2000).

O metabolismo do nitrogênio (N) é essencial para a formação de aminoácidos e proteínas, fator determinante para o bom desenvolvimento e produtividade das culturas, de modo que a deficiência desse elemento pode comprometer o metabolismo vegetal. Deste modo, é imprescindível entender o processo de assimilação desse nutriente, cujo objetivo é aumentar a eficiência da sua assimilação pela planta (CARNEIRO *et al.*, 2015; VIANA, 2007).

2.3.3 Nitrato e nitrato redutase

As duas principais formas de N no solo são nitrato (NO_3^-) e amônio (NH_4^+) (MARSCHNER, 2011). A maioria das plantas consegue absorver íons de amônio ou nitrato. A forma de absorção pela planta pode ser controlada pelo genótipo, fase de desenvolvimento da planta, ou até mesmo por propriedades do solo, como textura, estrutura, pH e teor de água (BORGOGNONE *et al.*, 2013).

O nitrato é absorvido de forma ativa via sistema simporte, com transporte simultâneo de dois íons de hidrogênio e um nitrato, com custo de duas moléculas de adenosina trifosfato (ATP), enquanto a absorção do amônio ocorre de forma passiva via sistema antiporte. Quando absorvido na forma nítrica pelas raízes, pode ser reduzido e assimilado ou estocado nos vacúolos dos tecidos radiculares, podendo ainda ser translocado via xilema por meio do fluxo respiratório para parte aérea, para redução ou ainda armazenamento nos vacúolos em tecidos foliares (SANTOS, 2010; TAIZ *et al.*, 2017).

A forma de assimilação do nitrogênio depende da forma que ele é absorvido pela planta. O processo de assimilação do nitrato inicia-se com a redução do nitrato em nitrito, que ocorre no citosol e é catalisado pela enzima nitrato redutase. Posteriormente, as células vegetais transportam o nitrito do citosol para o interior dos cloroplastos nas folhas e dos plastídios nas raízes, para então ocorrer a redução do nitrito a amônio, reação catalisada pela enzima nitrito redutase. A assimilação do amônio pode ocorrer com a sua incorporação à enzima glutamina sintetase (GS), para formação da glutamina, que, em níveis altos, estimula a

atividade da enzima glutamato sintase (GOGAT) e a formação de glutamato. Uma vez assimilado em glutamina e glutamato, o nitrogênio é incorporado a outros aminoácidos (BREDEMEIER; MUNDSTOCK, 2000; TAIZ *et al.*, 2017).

Após absorvido, no nitrato pode ser assimilado na raiz ou ser transportado para as folhas para ser assimilado na parte aérea (BREDEMEIER; MUNDSTOCK, 2000). Entretanto, a redução do nitrato em tecidos fotossintetizantes, teoricamente, é mais adequada, pois há o gasto de oito elétrons no processo completo de redução, que são oriundos da fase fotoquímica da fotossíntese (MARTUSCELLO *et al.*, 2016).

A enzima nitrato redutase é primordial para o metabolismo de N, tornando-a um fator limitante para o crescimento, desenvolvimento e produção de proteínas. Deste modo, sua atividade pode estar indiretamente, relacionada com a produtividade de uma cultura, considerando que plantas com alta atividade dessa enzima apresentem maior capacidade de resposta à adubação nitrogenada (CRAWFORD, 1995; SOLOMONSON; BARBER, 1990; (VIANA; KIEHL, 2010).

Segundo Khouri (2007), os maiores valores de nitrato redutase, e provavelmente maior absorção de nitrato, ocorrem durante os períodos de alta temperatura e demanda evaporativas da planta. Com o aumento do influxo de água, pode aumentar também o influxo de nitrato para as folhas, aumentando a disponibilidade de substrato para a enzima.

2.4 ESPECTROSCOPIA

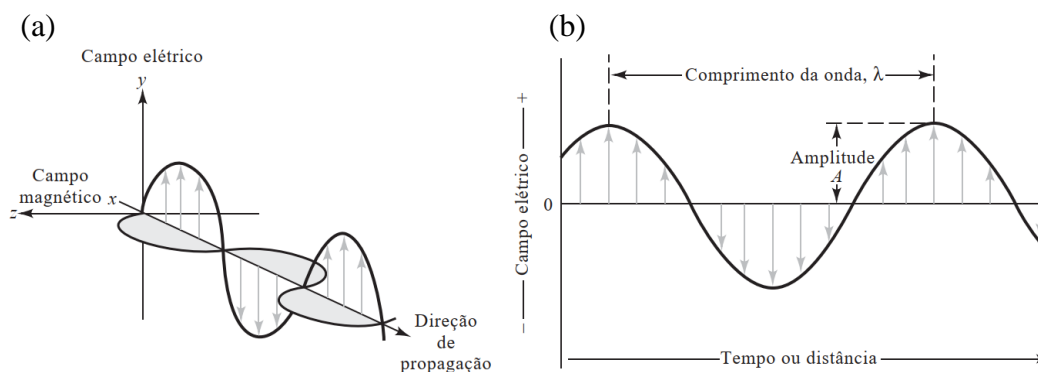
A espectroscopia pode ser definida como o estudo da interação (absorção e emissão de luz e outras radiações) com a matéria, relacionando a dependência destes processos com o comprimento de onda. Recentemente sua definição foi ampliada para incluir o estudo das interações entre partículas, tais como elétrons, prótons e íons, bem como sua interação com outras partículas em função de sua energia de colisão (HURST *et al.*, 2021).

Ondas eletromagnéticas são formadas por campos elétricos e magnéticos que variam com o tempo, e apresentam propriedades como comprimento de onda (λ), frequência, velocidade e amplitude. Entretanto, diferente das ondas sonoras, a luz não precisa de um meio de propagação, ou seja, pode ser propagada no vácuo (HALLIDAY; RESNICK; WALKERT, 2016a).

A radiação eletromagnética é propagada por uma fonte de luz e se move em linha reta se não forem refletidas ou refratadas por um objeto (DA-WEN, 2009). Em fenômenos como reflexão, refração, difração e interferência, essas ondas são modelos de modo que o campo elétrico e campo magnético oscilam e são perpendiculares entre si, conforme a Figura

1a. O campo elétrico (eixo y) oscila de forma senoidal no espaço e no tempo (eixo x), que representa o tempo que a radiação passa por um ponto fixo, conforme a Figura 1b (SKOOG *et al.*, 2006).

Figura 1 - A natureza ondulatória de um feixe de radiação com uma única frequência. Em (a), uma onda plano-polarizada é apresentada propagando-se ao longo do eixo x , onde o campo elétrico oscila em um plano perpendicular ao campo magnético. Em (b), somente as oscilações do campo elétrico são mostradas.



Fonte: Skoog (2006).

A amplitude da onda senoidal (Figura 1b), seja no campo elétrico ou magnético, é o valor máximo das oscilações de uma onda. O tempo necessário para realizar uma oscilação completa, ou seja, passagem de dois máximos ou mínimos sucessivos por um ponto fixo no espaço é denominado período (T). A frequência (f) é o número de oscilações do campo elétrico por unidade de tempo, definido como $1/T$ e expressa em hertz (Hz). O comprimento de uma onda (λ) é a distância linear entre pontos repetidos na forma da onda, como por exemplo, a distância entre dois pontos de máximo conforme a Figura 1b. Outro aspecto interessante é seu comprimento, pois é inversamente proporcional à sua intensidade, de modo que quanto menor o comprimento de onda, maior sua energia (HALLIDAY; RESNICK; WALKER, 2016b; SKOOG *et al.*, 2006).

Já a sua velocidade (c) depende do meio em que atravessa e da frequência, determinada pela razão entre comprimento de onda e período (λ/T), ou ainda pela multiplicação entre frequência e comprimento da onda. Entretanto, todas as ondas eletromagnéticas se propagam na mesma velocidade no vácuo, independentemente de onde ela se situe no espectro eletromagnético (HALLIDAY; RESNICK; WALKER, 2016a).

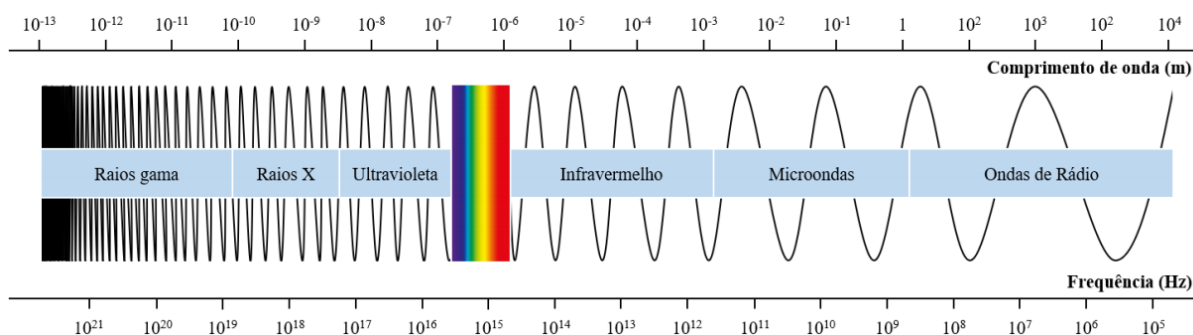
De modo geral, em interações entre a radiação e matéria pode considerar que a luz é constituída por fótons, uma partícula de radiação eletromagnética com massa igual a zero e energia determinada por hf , onde h é a constante de Planck ($6,63 \times 10^{-34}$ J s) e f a frequência da onda (SKOOG *et al.*, 2006). Quando um objeto contém muitos átomos, a tendência é que

predomine o processo de absorção de energia, assim como em objetos escuros ou predominância de emissão em objetos de cores claras. No entanto, independente do processo que ocorra, a variação de energia sempre será igual à energia de um fóton (HALLIDAY; RESNICK; WALKER, 2016a).

Os principais tipos de interação na espectroscopia envolvem transição de diferentes níveis energéticos das espécies químicas. Interações como reflexão, refração, interferência e outros são mais relacionados às alterações das propriedades globais dos materiais do que com os níveis energéticos de átomos e moléculas. Contudo, o tipo de interação depende da energia radiante empregada e do modo de detecção (SKOOG *et al.*, 2006).

O espectro eletromagnético representa a distribuição da radiação eletromagnética em diferentes intervalos, chamados de regiões espectrais, de acordo com o comprimento de onda e frequência, conforme a Figura 2. As ondas que compreendem o espectro, no sentido decrescente de energia, são raios γ , raios-X, ultravioleta, visível, infravermelha, micro-ondas e ondas de rádio. Além disso, é possível também observar na figura o padrão de comportamento das ondas eletromagnéticas, de modo que o comprimento da onda é inversamente proporcional à frequência (FLORENZANO, 2002).

Contudo, as pesquisas em espectroscopia envolvem principalmente as regiões espectrais do visível e infravermelho (DA-WEN, 2009; PASQUINI, 2018; SKOOG *et al.*, 2006). A faixa espectral visível é uma zona em que transições eletrônicas podem acontecer. Isto é, quando uma molécula absorve energia, esse fóton pode excitar o átomo para um nível energético. Moléculas com muitas ligações duplas conjugadas tais como carotenoides, clorofilas e outros pigmentos, absorvem energia nessa região. Já a faixa da região do infravermelho próximo é a primeira faixa espectral com bandas de absorção relacionada com as vibrações de moléculas, e é amplamente utilizada para análise de qualitativa e quantitativa de diversas moléculas e grupos químicos (DA-WEN, 2009).

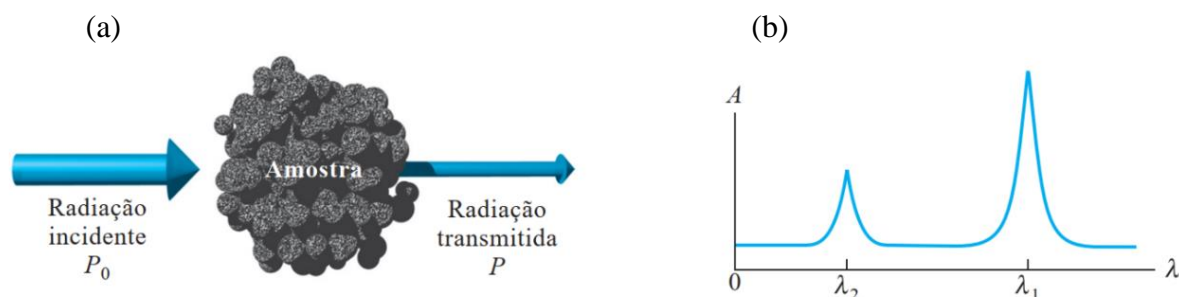
Figura 2 - Espectro eletromagnético.

Fonte: Adaptado de Florenzano, (2007).

Os métodos espectroscópicos de análise, exceto espectroscopia de massa, podem ser classificados de acordo com a região do espectro eletromagnético envolvida na medida (PASQUINI, 2018; SKOOG *et al.*, 2006). A espectroscopia VIS-NIR utiliza a faixa espectral do visível (VIS, 380 - 780 nm) e infravermelha (IR, 780 - 2500 nm), faixa que permite observar os fenômenos de transferência de energia eletromagnética em ligações químicas dos grupos C-H, N-H, O-H e S-H, encontrados facilmente em moléculas orgânicas (CEN; HE, 2007; DA-WEN, 2009; TIBOLA *et al.*, 2018).

Amostras de cada tipo de material apresentam diferentes composições físico-químicas e conseqüentemente tem a capacidade de refletir, absorver, dispersar e emitir energia eletromagnética em padrões e comprimentos de ondas diferentes, chamada de assinatura espectral (ELMASRY *et al.*, 2012; TININI, 2015). Ao ser exposto a uma fonte externa de radiação eletromagnética, parte da radiação incidente (P_o) pode ser absorvida pelo analito, resultando em um feixe transmitido de menor potência (P), conforme a Figura 3a. A partir da medição dessa absorção de luz em função do comprimento de onda, é possível gerar o espectro de absorção de radiação de uma amostra, conforme a Figura 3b, que contém informações qualitativas e quantitativas da amostra (SKOOG *et al.*, 2006).

Figura 3 - Métodos de absorção de radiação. A radiação com potência incidente igual a P_0 pode ser absorvida pelo analito, resultando em um feixe transmitido de menor potência P . O espectro de absorção resultante é exposto em (b).



Fonte: Skoog (2006)

A lei de absorção, também conhecida como lei de Beer-Lambert, é o princípio básico da espectroscopia. À medida que a luz atravessa a amostra, ocorre uma atenuação na intensidade de luz, proporcionalmente à absorção da luz e excitação das moléculas (SKOOG *et al.*, 2006). A absorção de radiação pelo tecido, causando essa atenuação da luz, depende de compostos chamados de cromóforos, qual cada um tem um espectro específico de absorção. Quanto maior a concentração do soluto analisado, maior será a absorção de luz pela amostra e maior atenuação (menor intensidade de luz). A absorbância (A), ou atenuação da luz, pode ser quantificada pela Equação 1 (LI *et al.*, 2020b; PELLICER; BRAVO, 2011; SKOOG *et al.*, 2006).

$$A = \log\left(\frac{P_0}{P}\right) \quad (1)$$

Entretanto, o espectroradiômetro, equipamento utilizado na espectroscopia, nos fornece como resposta após a análise de uma amostra apenas a intensidade de luz absorvida ou refletida pela amostra, sendo necessário o processamento da resposta instrumental para obtenção dos valores de concentração dos atributos de interesse (FERREIRA, 2015).

2.5 QUIMIOMETRIA

Segundo Malinowski (1991), a quimiometria é definida como “uso de métodos matemáticos e estatísticos para o tratamento, a interpretação e previsão de dados químicos”. Ferramenta criada em 1974 pelo químico sueco Svante Wold é utilizada para a extração de informações úteis para predição de uma propriedade de interesse a partir de dados multivariados. A partir disso, é possível a realização de análises quantitativas e qualitativas por meio de métodos de regressão multivariada e métodos de reconhecimento de padrão

(classificação) ou análises exploratórias respectivamente, além da otimização de processos por meio do planejamento de experimentos (FERREIRA *et al.*, 1999; NESPECA, 2018).

Após a análise de uma amostra, o espectrofotômetro, equipamento utilizado na espectroscopia, nos fornece como resposta após a análise de uma amostra apenas a intensidade de luz absorvida ou refletida pela amostra. Ou seja, o equipamento deve ser “calibrado”. Isso significa que é preciso selecionar um conjunto adequado de amostras de referência com valores conhecidos de concentração dos atributos de interesse para serem correlacionados com a resposta instrumental a partir de cálculos matemáticos, também conhecidos como modelos de regressão ou curva de calibração. O modelo matemático que obtiver o melhor desempenho servirá então para prever as concentrações de interesse de novas amostras a partir de suas leituras espectrais, sem necessidade de novas análises por meio de métodos convencionais (NESPECA, 2018).

O processo de desenvolvimento dos modelos de predição normalmente é constituído pelas etapas: (1) organização e pré-processamento adequado no conjunto de dados; (2) calibração, que é a construção do modelo de regressão; (3) validação do modelo; (4) interpretação dos parâmetros de desempenho do modelo; (5) previsão de novos dados (FERREIRA *et al.*, 1999; MARTENS; NAES, 1989).

Na calibração multivariada de dados, onde o número de variáveis de entrada e saída é alto, é necessária a organização dos dados em forma matricial para facilitar sua manipulação. Os valores relativos às variáveis independentes (espectros das amostras de interesse) e às variáveis dependentes (propriedades das amostras obtidas a partir dos métodos análise de referência) são organizados separadamente nas chamadas matriz absorvância e matriz concentração, respectivamente (LUZ, 2003).

Entretanto, o sinal medido é composto por uma contribuição determinística, que é o sinal relevante para a análise, e por uma contribuição estocástica, que são os erros de natureza aleatória também chamados de ruídos. Para uma boa análise multivariada, é necessária a aplicação de artifícios matemáticos que diminuam a relação entre o sinal verdadeiro (S) e o ruído aleatório (R) (CAPELA *et al.*, 2003).

Com o objetivo de reduzir as variações indesejadas, são aplicados diferentes tratamentos de dados. Entretanto, esse procedimento deve ser feito de forma cautelosa e criteriosa, para que não comprometa os resultados da análise. Os pré-tratamentos podem ser classificados em transformação, aplicados às amostras, ou em pré-processamento, que são aplicados nas variáveis (FERREIRA, 2015). Diversos métodos podem ser aplicados antes dos modelos de calibração. A escolha do método adequado depende do tipo de variações

presentes nos dados espectrais, como ruídos, deslocamento da linha (MARTENS; NAES, 1989).

As principais técnicas de pré-tratamento de dados são: técnicas de alisamento, como os filtros de média, cujo objetivo é a suavização dos ruídos; técnicas de correção de linha base, que são as derivadas; padronização normal do sinal (*Standard Normal Variate*); correções do espalhamento do sinal através da *Multiplicative Scatter Correction* (MSC); e correção ortogonal de sinal, através da *Orthogonal Signal Correction* (OSC) (FERREIRA, 2015).

Outra etapa importante é a análise exploratória dos dados. A partir dessa etapa, é possível analisar os dados previamente, cujo objetivo é compreender os dados obtidos e sua relação com as variáveis analisadas. A *Principal Component Analysis* (PCA) é um algoritmo matemático que reduz a dimensionalidade dos dados multivariados, mantendo a maior parte da variação no conjunto de dados, ou seja, há uma “compressão” dos dados. Essa compressão é feita a partir da identificação de padrões, chamados de componentes principais, abrange as máximas variações dos dados. Com os componentes principais, as amostras podem ser representadas por menos variáveis independentes, ao invés do espectro completo de absorção, a partir de correlações significativas entre as variáveis do conjunto de dados (RINGNÉR, 2008).

Com os dados previamente organizados e pré-tratados, são aplicados os métodos de regressão para construção dos modelos de calibração, que são modelos matemáticos para relacionar os dados espectrais, também chamados de variáveis preditoras ou independentes, e os dados de referência, também chamados de variáveis respostas ou dependentes. Para a análise de dados multivariados, os principais métodos de regressão utilizados são: *Principal Component Regression* (PCR); *Multiple Linear Regression* (MLR); *Partial Least Squares Regression* (PLSR) (FERREIRA, 2015).

Entretanto, para utilização de espectro Vis-NIR na determinação quantitativa de um composto em amostras, devido a sua complexidade e alto número de variáveis preditoras (comprimentos de onda), normalmente recomenda-se o método PLSR, pois esse método consegue resolver o problema de colinearidade entre as amostras e suavizar os ruídos (JIA *et al.*, 2010).

Diferente dos modelos PLSR, os modelos MLR requerem o uso de artifícios matemáticos para reduzir os problemas de colinearidade de dados, como por exemplo, por meio da seleção de variáveis. Esse processo é de extrema importância para melhorar a relação entre o desempenho do modelo e custo computacional, aumentar a previsão das variáveis

selecionadas e melhorar a compreensão e interpretação dos modelos obtidos (YUN *et al.*, 2019). Já os modelos PCR, utilizam em seu modelo as mesmas etapas da análise de componentes principais (PCA), cujo objetivo é reduzir a dimensionalidade dos dados e tornar a análise menos complexa (FERREIRA, 2015).

Embora o PLSR seja o modelo mais comum na espectroscopia, existem diversos outros métodos que podem ser utilizados, tais como *K-Nearest Neighbors* (KNN), *Random Forest* (SHEN *et al.*, 2018), *Model tree* (M5P), *Multilayer Perceptron* (MPL), *Bagging* M5P (BM5P) (WU *et al.*, 2012; DAS *et al.*, 2018; SAMPAIO *et al.*, 2020; PENG *et al.*, 2019), entre outros.

Existem vários fatores que influenciam na escolha do melhor método de regressão: objetivo da análise, tamanho da amostra, suposições inerentes à regressão (linearidade, normalidade) e avaliação dos parâmetros de desempenho gerados pelos modelos (HAIR, 2005; ZHANG *et al.*, 2012; ELHAMDAOUI *et al.*, 2020). Isto é, desde que um modelo se aplique ao contexto do estudo, a escolha do melhor método de regressão é feita a partir do desempenho do modelo gerado. Ou seja, quanto mais métodos de regressão forem testados, maiores as chances de desenvolver um modelo preditivo com melhor desempenho.

Consiste no teste de predição do modelo, que pode ser feito por meio de um conjunto de amostras, para as quais propriedades dependentes e independentes, ou seja, valores de referência e dados espectrais sejam conhecidos, mas que não tenham sido utilizados no processo de calibração, processo conhecido como validação externa. Entretanto, muitos estudos possuem limitações técnicas e de custos, que limitam a quantidade de dados. Uma alternativa é utilização da validação cruzada. Trata-se de um método estatístico de avaliação e comparação de algoritmos de aprendizagem de máquina, que divide o conjunto de dados em dois segmentos: conjunto de calibração e de validação. Os dados dos conjuntos de calibração e validação são cruzados em sucessivas “rodadas” durante a construção do modelo de calibração, de modo que cada ponto de dados tenha uma chance de ser validado (FERREIRA, 2015; REFAEILZADEH *et al.*, 2016).

O desempenho dos modelos de calibração é avaliado a partir da comparação dos parâmetros estatísticos gerados para cada modelo, principalmente o coeficiente de determinação (R^2) e o erro quadrático médio de validação (RMSECV), de modo que os melhores modelos serão aqueles que apresentarem os maiores valores de R^2 e menores erros. A partir da seleção dos melhores modelos desenvolvidos, é possível realizar a predição dos analitos em novas amostras (FERREIRA, 2015).

3 ARTIGOS

3.1 DETERMINAÇÃO DO TEOR DE CARBOIDRATOS SOLÚVEIS, AMIDO E SACAROSE EM FOLHAS DE MANGUEIRA COM USO DA ESPECTROSCOPIA VIS-NIR

RESUMO

A região do Submédio do Vale do São Francisco tem mostrado alto potencial na produção e exportação de frutas, com destaque para a cultura da manga. Os maiores desafios para sua produção são a variação do florescimento, em função da desuniformidade da maturação dos ramos, e a tomada de decisão quanto ao momento ideal para indução floral. Os teores de carboidratos nas folhas estão diretamente envolvidos nesses processos, e determiná-los pode ser um marco para a iniciação da indução floral. Entretanto, métodos convencionais para sua análise são complexos e demorados. Deste modo, a espectroscopia pode otimizar a realização das análises de determinação desses atributos bioquímicos. Diante desses aspectos, o principal objetivo do trabalho foi desenvolver modelos preditivos para determinação dos teores de amido, carboidratos solúveis (como frutose, glicose) e sacarose em folhas de mangueira (*Mangifera indica* L.) com uso da espectroscopia Vis-NIR. O trabalho foi desenvolvido na região do Vale do São Francisco, seguindo as seguintes etapas: (1) a amostragem das folhas de mangueira; (2) aquisição dos espectros de refletância; (3) determinação dos teores de carboidratos (solúveis, amido e sacarose) com uso dos métodos de referência; e (4) desenvolvimento dos modelos preditivos. Os melhores modelos de predição dos diferentes atributos bioquímicos foram selecionados com base no desempenho dos modelos, sendo os parâmetros estatísticos avaliados: coeficiente de determinação (R^2); erro quadrático médio de calibração (RMSEC); erro padrão de calibração (SEC); erro quadrático médio de validação cruzada (RMSECV); erro padrão de validação cruzada (SECV); erro quadrático médio de predição (RMSEP); e erro padrão de previsão (SEP). Os métodos de regressão utilizados foram: *Partial Least Squares Regression* (PLSR), *Multiple Linear Regression* (MLR), *Gaussian Processes* (GP), *Multilayer Perceptron* (MPL), *Model Tree* (M5P), *Random Forest* (RF) e *Support Vector Machine* (SVM). Os melhores modelos preditivos desenvolvidos foram os modelos MLR para amido, carboidratos solúveis e sacarose, com R^2_{cv} de 0,92, 0,71 e 0,55, respectivamente. Deste modo, a espectroscopia se

mostrou uma ferramenta com alto potencial para determinação de atributos bioquímicos em folhas de mangueira.

PALAVRAS-CHAVE: Fruticultura. Indução floral. Métodos ópticos. Quimiometria.

ABSTRACT

The Submedium do Vale do São Francisco region has shown a high potential in fruit production and export, especially mango. The biggest challenges for its production are the variation in flowering, due to the uneven maturation of the branches, and the decision as to the ideal moment for floral induction. The carbohydrate levels in the leaves are directly involved in these processes, and determining them can be a landmark for the initiation of floral induction. However, conventional methods for their analysis are complex and time consuming. In this way, spectroscopy can optimize the performance of the analyses for determining these biochemical attributes. Considering these aspects, the main objective of this work was to develop predictive models for the determination of starch, soluble carbohydrates (as fructose, glucose) and sucrose contents in mango (*Mangifera indica* L.) leaves using Vis-NIR spectroscopy. The work was developed in the Vale do São Francisco region, following the following steps: (1) sampling of mango leaves; (2) acquisition of the reflectance spectra; (3) determination of the carbohydrate contents (soluble, starch and sucrose) using the reference methods; and (4) development of the predictive models. The best predictive models for the different biochemical attributes were selected based on the performance of the models, and the following statistical parameters were evaluated: coefficient of determination (R^2); root mean square error of calibration (RMSEC); standard error of calibration (SEC); root mean square error of cross-validation (RMSECV); standard error of cross-validation (SECV); root mean square error of prediction (RMSEP); and standard error of prediction (SEP). The regression methods used were Partial Least Squares Regression (PLSR), Multiple Linear Regression (MLR), Gaussian Processes (GP), Multilayer Perceptron (MPL), Model Tree (M5P), Random Forest (RF), and Support Vector Machine (SVM). The best predictive models developed were the MLR models for starch, soluble carbohydrates and sucrose, with R^2_{cv} of 0.92, 0.71 and 0.55, respectively. Thus, spectroscopy proved to be a tool with high potential for determining biochemical attributes in mango leaves.

KEYWORDS: Fruticultura. Floral induction. Optical methods. Chemometrics.

3.1.1 INTRODUÇÃO

A região do Submédio do Vale do São Francisco, localizada no Nordeste brasileiro, mais precisamente nos estados de Pernambuco e Bahia, tem mostrado seu alto potencial tanto de produção como em exportação de frutas, destacando-se a cultura da manga. Dentre as frutas produzidas no país, a manga é a mais exportada, com volume de 211 mil toneladas em 2020, dos quais 87% foram produzidas na região do Vale do São Francisco (CEPEA, 2021; EMBRAPA, 2021).

Dois dos maiores desafios da produção de manga são a variação do florescimento, em função da desuniformidade da maturação dos ramos, e a tomada da decisão quanto ao momento ideal para indução floral. Os teores de carboidratos nas folhas, principalmente o amido e os açúcares solúveis, estão envolvidos diretamente no desenvolvimento das flores e na intensidade do florescimento (DAVENPORT, 2007; OLIVEIRA *et al.*, 2018) de modo que brotos com maiores teores de carboidrato têm maiores chances de florescimento (SILVA *et al.*, 2020).

Deste modo, determinar os teores de carboidratos nas folhas pode ser fundamental para determinar o momento ideal para a indução floral. No entanto, os métodos tradicionais de determinação do teor de carboidratos, como por exemplo, o método proposto por Dubois *et al.* (1956), exigem reagentes, insumos e preparação da amostra, tornando os métodos demorados, de alto custo, além do uso restrito de alguns reagentes, que são controlados pelo exército ou polícia federal, como por exemplo, o ácido sulfúrico.

Essa determinação pode ser otimizada por meio da espectroscopia através da relação estabelecida entre os dados espectrais obtidos pela radiação refletida ou absorvida e as propriedades físico-químicas de interesse, por meio do desenvolvimento de modelos preditivos através de métodos de regressão (CANASSA *et al.*, 2018; FERREIRA, 2015).

A espectroscopia Vis-NIR tem se destacado no setor do agronegócio e pode ser utilizada para prever inúmeros atributos de uma planta de interesse comercial, incluindo o teor de clorofila e outros pigmentos fotossintéticos (OLIVEROS *et al.*, 2021; YENDREK *et al.*, 2017), água (LI *et al.*, 2020a; RODRÍGUEZ-PÉREZ *et al.*, 2018), inclusive para determinar o momento ideal para se fazer a irrigação (TORRES *et al.*, 2019), nitrogênio, (DECHANT *et al.*, 2017; PAZ-KAGAN *et al.*, 2020), carboidratos (DAS *et al.*, 2018; DE BEI *et al.*, 2017; LAMMERS *et al.*, 2009), amido (FREY *et al.*, 2020; NAKAJIMA *et al.*, 2019), celulose e lignina (PETISCO *et al.*, 2006), e metabólitos secundários (SERBIN *et al.*, 2012).

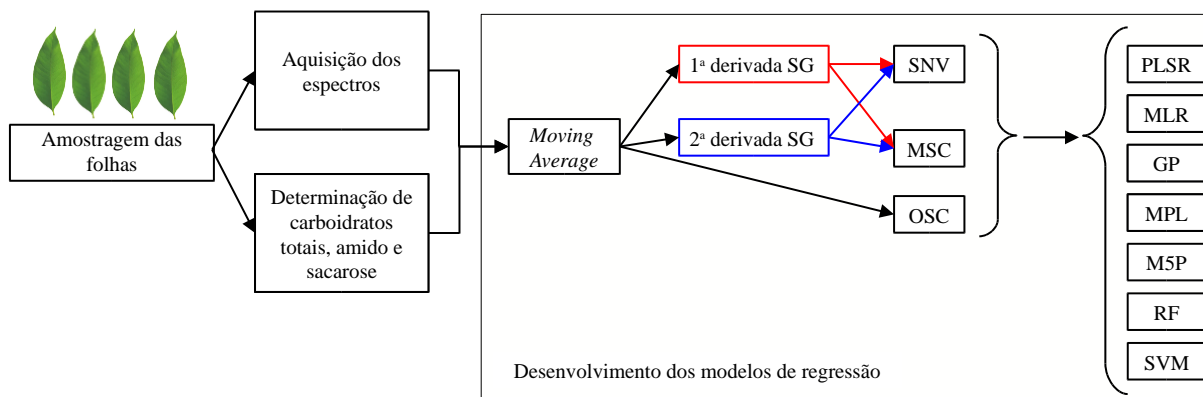
Os modelos preditivos também podem ser utilizados para o desenvolvimento de protótipos e dispositivos óptico eletrônicos que auxiliem diretamente na tomada de decisão no campo, através da predição de atributos químicos em diversas amostras como, por exemplo, os dispositivos: *OPTOfruit*, para predição de sólidos solúveis totais (SST), antocianinas totais e flavonoides amarelos em uvas viníferas (COSTA, 2018); *PHYLIS*, que avalia as condições do solo e da planta através da determinação de teores de nutrientes (AITKENHEAD *et al.*, 2016); *SpecSolo*, desenvolvido pela EMBRAPA Solos (EMBRAPA, 2018); dispositivo para detecção de gás N₂O em solos agrícolas (BANDARA *et al.*, 2021).

Por outro lado, o uso da espectroscopia para a determinação dos teores de carboidratos em plantas ainda é limitado a poucas espécies vegetais e/ou em tecidos específicos (PAZ-KAGAN *et al.*, 2020). Diante disso, esse trabalho teve como objetivo o desenvolvimento de modelos preditivos para determinar os teores de carboidratos solúveis, amido e sacarose em folhas de mangueira das variedades Keitt, Omer, Shelly e Palmer por meio da espectroscopia Vis-NIR.

3.1.2 MATERIAL E MÉTODOS

O presente estudo é dividido em quatro etapas principais: amostragem, análises espectrais, determinação de amido, carboidratos solúveis e sacarose, e desenvolvimento dos modelos de regressão, conforme a Figura 4: após a amostragem, as folhas são submetidas às análises espectrais para aquisição dos espectros de absorbância, e logo em seguida, às análises de referência, que são as análises convencionais de determinação dos compostos em análise; os dados espectrais passam por uma série de pré-tratamentos, cujo objetivo é a remoção de informações irrelevantes, que não contribuem ou diminuem o desempenho dos modelos; após o pré-tratamento dos dados, métodos de regressão são aplicados, para o desenvolvimento dos modelos de predição.

Figura 4 - Diagrama do processo de desenvolvimento dos modelos de predição teores de carboidratos solúveis, amido e sacarose em folhas de mangueira.



Legenda: Derivada SG – derivada de Savitzky-Golay; SNV – Standard Normal Variate; MSC – Multiplicative Scatter Correction; OSC – Orthogonal Signal Correction; PLSR – Partial Least Squares Regression; MLR – Multiple Linear Regression; GP – Gaussian Processes; MPL – Multilayer Perceptron; M5P – Model Tree; RF – Random Forest; SVM – Support Vector Machine.

Amostragem das folhas

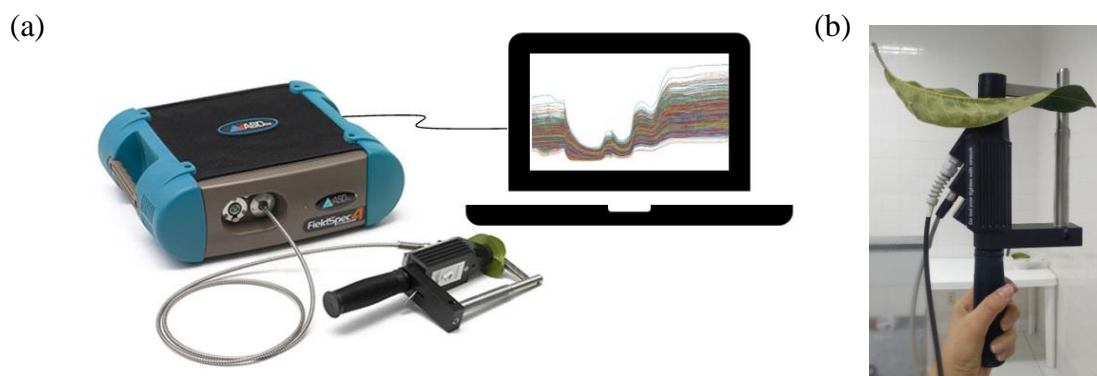
As folhas de mangueira foram coletadas em fazendas localizadas na região do Submédio do Vale do São Francisco (Petrolina, Pernambuco, Brasil: -9.39416; -40.5096; Juazeiro, Bahia, Brasil: -9.43847; -40.5052). Essa região é caracterizada pela ocorrência de clima semiárido quente, classificado como BSh, segundo a classificação climática de Köppen (ALVARES *et al.*, 2013).

As amostras foram compostas por quatro folhas, retiradas na altura mediana da copa da planta, na região mediana do último fluxo vegetativo maduro, em dois lados da planta (voltados para direções leste e oeste), totalizando em quatro conjuntos de dados: “Keitt”, com 96 amostras da variedade Keitt; “Omer, Shelly”, com 94 amostras das variedades Omer e Shelly (47 para cada variedade); “Palmer”, com 112 amostras da variedade Palmer; “Geral”, composto por todas as amostras das analisadas, com 302 amostras das variedades Keitt, Omer, Shelly e Palmer. As coletas foram realizadas na fase de plena florada para a variedade Keitt e na fase vegetativa (pré-poda) para as variedades Omer, Shelly e Palmer, cujo objetivo foi obter maior variabilidade entre as amostras para o desenvolvimento de modelos matemáticos preditivos mais robustos. Ainda no campo, as amostras foram armazenadas em sacos plásticos, devidamente identificados, imersas em caixa de gelo, transportadas até o laboratório e congeladas em freezers verticais (-20 °C) para manutenção de sua integridade e composição até o momento das análises. Destaca-se que as amostras foram descongeladas até atingirem a temperatura ambiente de 25 °C, antes das análises espectrais e convencionais.

Aquisição dos espectros

O sistema de aquisição dos espectros de refletância foi constituído dos seguintes componentes: espectrômetro *FieldSpec3* (*Analytical Spectral Devices*, Boulder, Colorado, USA) que opera na faixa de comprimento de onda de 350 a 2500 nm, com resolução de 3 a 10 nm e tempo por varredura de 100 ms, equipado com *probe plant*; e computador com o software RS3 (*Analytical Spectral Devices*, Boulder, Colorado, USA), conforme a Figura 5. O sistema foi ligado para aquecimento por 30 minutos antes da aquisição dos espectros e as leituras espectrais foram feitas no limbo foliar do lado.

Figura 5 - Sistema óptico eletrônico constituído por um espectrômetro *FieldSpec3*, equipado com *probe plant*; e computador com o software RS3 em (a). Em (b), momento de análise da folha de mangueira utilizando a *probe plant*.



A *probe plant* possui uma referência branca, com a qual as operações de calibração são implementadas antes da aquisição dos dados. Cada espectro é composto por uma média de 30 varreduras executadas pelo equipamento. Os valores de refletância foram transformados em medições de absorbância ($\log(1/R)$), onde R significa refletância, utilizando o software *ViewSpec Pro* (*Analytical Spectral Devices*, Boulder, Colorado, USA).

Determinação de carboidratos totais, amido e sacarose

Após as medições de refletância, as amostras de material foliar fresco foram submetidas às análises bioquímicas de referência para determinação dos teores de amido ($\mu\text{g g}^{-1}$ de MF) e carboidratos solúveis totais (mmol g^{-1} de MF) para todas as variedades, além do teor de sacarose (mg g^{-1} de MF) apenas para a variedade Keitt.

O teor de amido foi determinado segundo Neves e Morais (2005). Para a extração, utilizou-se 0,05 g de material foliar fresco, homogeneizado em 4 ml de H_2SO_4 (ácido sulfúrico). Foram tomadas alíquotas de 0,5 ml de extrato: a cada alíquota foi adicionado

0,95 ml de água e 3 ml de antrona. Após o preparo das amostras, foram realizadas as leituras em um espectrofotômetro a 620 nm.

O teor de carboidratos solúveis foi determinado segundo Dubois *et al.* (1956). Para a extração utilizou-se 1 g de material foliar fresco, homogeneizado em 4 ml de solução tampão fosfato de potássio monobásico, 0,1 M e pH 7,0. Após centrifugação e banho-maria, foram colocados em tubos de ensaio 10 µl do sobrenadante (após teste de diluição), 490 µl do tampão fosfato monobásico, 500 µl de fenol e 2,5 ml de ácido sulfúrico, seguido de repouso e posteriormente agitação, e depois novamente deixados em uma bandeja contendo água à temperatura ambiente (25°C) por 10 minutos. Por fim, foram realizadas as leituras em um espectrofotômetro a 490 nm.

O teor de sacarose foi determinado segundo Van Handel *et al.* (1968). Para a extração, utilizou-se 50 mg de material foliar, homogeneizadas em 1,5 ml de solução de MCW (metanol, clorofórmio e água, na proporção 12:5:3). Após agitação e centrifugação, foi coletado o sobrenadante e o resíduo extraído novamente para nova centrifugação e coleta do sobrenadante. Após a obtenção dos extratos, foi feita a separação de fases por solvente e purificação. A cada 5,0 ml do sobrenadante, foram adicionados 125 ml de clorofórmio e 188 µl de água deionizada, seguido de agitação, centrifugação e banho-maria. Da fase aquosa obtida, foram tomadas alíquotas para as dosagens de sacarose: a cada alíquota de 100 µl de extrato, devidamente diluído, foi adicionado 100 µl de hidróxido de potássio (30%), seguido de agitação, aquecimento, posterior resfriamento e adição de 3 ml de solução de antrona (0,2%) em ácido sulfúrico. Por fim, foram realizadas as leituras em um espectrofotômetro a 620 nm.

Desenvolvimento dos modelos de regressão

Os dados espectrais brutos foram submetidos a uma combinação de diferentes pré-tratamentos matemáticos: filtro de média móvel (*Moving Average*, MA) com tamanho de segmentos de 5, 11, 17, 23, 29 e 35 pontos; primeira e segunda derivada, obtidas pelo método de Savitzky-Golay (GORRY, 1990; SAVITZKY; GOLAY, 1964), com polinômio de segunda ordem e seis segmentos diferentes (5, 11, 17, 23, 29 e 35 pontos). Em seguida, os dados pré-tratados foram corrigidos pelos métodos de correção de dispersão multiplicativa (*Multiplicative Scatter Correction*, MSC) (ISAKSSON; NAES, 1988) e transformação variável normal padrão (*Standard Normal Variate Transformation*, SNV) (BARNES; DHANOA; LISTER, 1989). Por fim, foi aplicada uma correção de sinal ortogonal (*Orthogonal Signal Correction*, OSC) nos dados que foram unicamente filtrados pelo filtro de média móvel.

Assim, foram produzidas 61 matrizes de dados para cada atributo foliar (carboidratos totais, amido e sacarose) de cada variedade, sendo: matriz bruta (sem pré-tratamentos), seis matrizes com filtros de média móvel, doze matrizes com primeira e segunda derivadas (em seis diferentes tamanhos de segmento), dezoito matrizes com SNV, dezoito com MSC e seis com OSC (apenas matrizes com dados filtrados).

Além de testar diferentes pré-tratamentos, a faixa espectral também foi dividida em três segmentos, formando assim quatro bandas espectrais de análise, sendo: FULL, incluindo todos os comprimentos de onda (350-2500 nm); VNIR – *Visible and Near Infrared* (350-1000 nm); SWIR1 - *First Short Wavelength Infrared* (1001-1830 nm); e SWIR2 - *Second Short Wavelength Infrared* (1831-2500 nm).

A partir das matrizes pré-tratadas foram desenvolvidos os modelos de Regressão por Mínimos Quadrados Parciais (*Partial Least Squares Regression*, PLSR) para cada banda espectral e atributo foliar. Para isso, os dados espectrais brutos ou pré-tratados foram utilizados como variáveis de entrada independentes (X) e os atributos de interesse (carboidrato, amido e sacarose) como variáveis de entrada dependentes (Y). Foi aplicada a validação cruzada em todas as amostras dos conjuntos de dados e o número ótimo de fatores foi definido através do menor valor obtido de erro quadrático médio de validação cruzada (RMSECV). O algoritmo utilizado para a construção dos modelos foi Kernel, com 100 interações.

Ao final do processamento dos quatro conjuntos de dados (Keitt, Omer e Shelly, Palmer e Geral), foram gerados 2196 modelos diferentes de PLSR para os atributos amido, carboidrato e sacarose (sendo o último apenas para a variedade Keitt, devido à sua fase reprodutiva). Após processamento, foram selecionados os melhores modelos e combinações de pré-tratamentos para cada atributo foliar, em suas respectivas faixas espectrais e conjunto de dados.

Posteriormente, foi feita a seleção de variáveis através de uma PCA e exclusão das variáveis com menor influência e não significativas (p -valor $> 0,05$) para os modelos através do método Jack-Knife (MARTENS; MARTENS, 2000). Foram selecionados os comprimentos de ondas associados aos picos de absorbâncias em todas as componentes principais selecionadas, seguido da remoção das variáveis com menores coeficientes de regressão. O modelo foi recalculado a cada variável removida, de modo que, o objetivo foi encontrar modelos com os melhores parâmetros de desempenho e menor número de variáveis e que todas fossem significativas para a construção dos modelos. Os comprimentos de onda

selecionados a partir dos picos de absorbâncias e aqueles submetidos à ANOVA e teste de significância a 5% são apresentados no Anexo A.

A partir disso, foi aplicado MLR em todos os atributos em seus respectivos conjuntos de dados, com as variáveis selecionadas. Em seguida, também utilizando apenas as variáveis selecionadas, foram desenvolvidos novos modelos PLSR para todos os conjuntos de dados, bem como a aplicação dos modelos foram submetidos a outros modelos de regressão, sendo eles: *Gaussian Processes* (GP), *Multilayer Perceptron* (MPL), *Model Tree* (M5P), *Random Forest* (RF) e *Support Vector Machine* (SVM) para todos os atributos, utilizando apenas o melhor conjunto de dados.

O desempenho dos modelos foi avaliado a partir da comparação dos parâmetros estatísticos a seguir: coeficiente de determinação (R^2); erro quadrático médio de calibração (RMSEC); erro padrão de calibração (SEC); erro quadrático médio de validação cruzada (RMSECV); erro padrão de validação cruzada (SECV); Estes parâmetros são definidos pelas equações de (2) a (6):

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

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3)$$

$$SEC = \sqrt{\frac{n * RMSEC^2}{n - 1}} \quad (4)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_{i^*})^2}{n}} \quad (5)$$

$$SECV = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i - bias)^2}{n - 1}} \quad (6)$$

Onde \hat{y}_i é o valor previsto pelo modelo de calibração; \hat{y}_{i^*} é o valor previsto pelo modelo na etapa de validação cruzada; y_i é o valor de referência; $\bar{\hat{y}}$ é a média dos valores de predição; n é o número de amostras na calibração ou validação; m é o número de amostras previstas; σ_r é o desvio-padrão dos valores de referência; σ_p é o desvio-padrão dos valores de predição.

A avaliação do parâmetro SECV deve ser feita associando o desvio-padrão (SD) e os valores de amplitude do atributo a ser analisado. Deste modo, o desvio preditivo residual

(RPD) e a relação de erro de intervalo (RER) foram definidos pelas equações 7 e 8, conforme Williams (2001):

$$RPD = \frac{SD}{SECV} \quad (7)$$

$$RER = \frac{Alcance}{SECV} \quad (8)$$

Os melhores modelos são aqueles que apresentam os maiores valores de coeficiente de determinação da validação cruzada (R^2_{cv}), seguido dos menores valores de erro quadrático médio de validação cruzada (RMSECV) e maiores valores de desvio preditivo residual (RPD).

3.1.3 RESULTADOS

Análise química das folhas

A Tabela 1 apresenta a estatística descritiva para os teores de amido, carboidratos e sacarose para cada conjunto de dados, ou seja, o número de amostras para cada conjunto de dados avaliados, valores médios, máximos e mínimos, amplitude, desvio-padrão e variância. É possível observar maior variabilidade para os teores de carboidratos em relação aos outros atributos.

Tabela 1- Estatística descritiva dos teores de amido ($\mu\text{g g}^{-1}$ de MF), carboidratos solúveis (mmol g^{-1} de MF) e sacarose (mg g^{-1} de MF) em folhas de mangueira para os conjuntos de dados "Keitt", "Omer", "Shelly, Palmer" e "Geral".

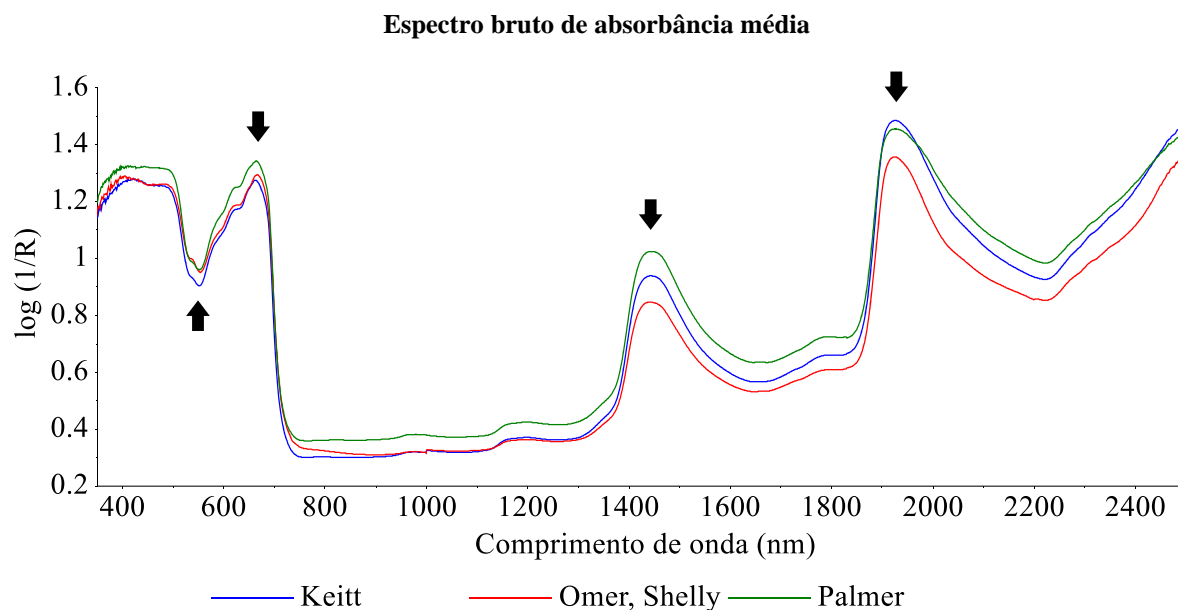
Parâmetro	Amostras	Média	Mín.	Máx.	Amplitude	SD	Variância
Keitt							
Amido	96	0,13	0,06	0,22	0,16	0,04	0,002
CHOsol	96	85,01	41,40	166,80	125,40	23,24	539,987
Sacarose	96	3,22	0,74	11,74	11,00	1,62	2,618
Omer, Shelly							
Amido	94	0,16	0,07	0,52	0,45	0,06	0,003
CHOsol	94	202,88	55,03	303,25	248,22	49,21	2421,575
Palmer							
Amido	112	0,55	0,43	0,70	0,27	0,06	0,004
CHOsol	112	111,53	43,79	216,25	172,45	40,46	1636,831
Geral							
Amido	302	0,29	0,06	0,70	0,64	0,21	0,042
CHOsol	302	131,53	41,40	303,25	261,85	62,86	3950,993

Legenda: MF – material foliar fresco; CHOsol – Carboidratos solúveis; Mín. – valor mínimo; Máx. – valor máximo; SD – desvio-padrão.

Análise exploratória do espectro

Os espectros médios brutos de absorvância de folhas das três variedades de mangueira são mostrados na Figura 6. Esses espectros apresentam quatro principais picos de absorção nas regiões do visível e infravermelho próximo, sendo em torno de 460, 670, 1450 e 1930 nm.

Figura 6 - Espectro bruto de absorvância média de amostras foliares de mangueira das variedades Keitt, Omer e Shelly, e Palmer.



Modelos de regressão PLSR

Foram desenvolvidos modelos PLSR para os conjuntos “Keitt”, “Omer e Shelly”, “Palmer” e “Geral”, que abrange o conjunto de dados das variedades anteriores. A Tabela 2 apresenta os melhores resultados de cada atributo e variedade. O melhor modelo de predição foi para o atributo amido, no conjunto geral, com R^2_{cv} de 0,90 e RMSECV de 0,06. Além disso, todos os modelos obtiveram melhores desempenhos com os pré-tratamentos *Moving Average* e OSC.

Tabela 2 - Parâmetros estatísticos dos melhores modelos PLSR para predição de amido, carboidratos solúveis e sacarose em folhas de mangueira para os conjuntos de dados “Keitt”, “Omer, Shelly”, “Palmer” e “Geral”.

Parâmetro	BoW	Pré-tratamento	F	calibração			validação			RPD
				R^2_{cal}	RMSEC	SEC	R^2_{cv}	RMSECV	SECV	
Keitt										
Amido	VNIR	MA 29 pts, OSC	1	0,73	0,02	0,02	0,73	0,02	0,02	1,89
CHOsol	VNIR	MA 5 pts, OSC	1	0,65	13,75	13,83	0,63	14,04	14,11	1,65
Sacarose	VNIR	MA 23 pts, OSC	7	0,77	0,77	0,77	0,54	1,11	1,12	1,45
Omer, Shelly										
Amido	VNIR	MA 35 pts, OSC	2	0,56	0,04	0,04	0,38	0,05	0,05	1,27
CHOsol	VNIR	MA 23 pts, OSC	1	0,38	38,40	38,60	0,35	39,42	39,63	1,24
Palmer										
Amido	FULL	MA 5 pts, OSC	15	0,96	0,01	0,01	0,52	0,05	0,05	1,37
CHOsol	FULL	MA 5 pts, OSC	2	0,72	21,24	21,33	0,58	26,20	26,31	1,54
Geral										
Amido	FULL	MA 11 pts, OSC	3	0,92	0,06	0,06	0,90	0,06	0,06	3,22
CHOsol	VNIR	MA 29 pts, OSC	4	0,73	32,42	32,47	0,69	35,19	35,25	1,78

Legenda: PLSR – *Partial Least Squares Regression*; CHOsol – Carboidratos solúveis; BoW – banda de comprimento de onda; VNIR – banda espectral dos comprimentos de onda no intervalo de 350 a 1000 nm; FULL – espectro completo, que abrange os comprimentos de onda no intervalo de 350 a 2500 nm; MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*; F – fator; R^2_{cal} – coeficiente de determinação da calibração; RMSEC – erro quadrático médio de calibração; SEC – erro padrão de calibração; R^2_{cv} – coeficiente de determinação da validação cruzada; RMSECV – erro quadrático médio de validação cruzada; SECV – erro padrão de validação cruzada; RPD – desvio preditivo residual.

Para o conjunto de dados da variedade Keitt, a banda espectral que proporcionou os melhores modelos para os atributos foliares foi VNIR, com pré-tratamentos *Moving Average* (29, 5 e 23 pontos) e OSC. Esses obtiveram R^2_{cv} de 0,73, 0,63 e 0,54 e RMSECV de 0,02, 14,04 e 1,11, para amido, carboidratos solúveis e sacarose, respectivamente. Os valores de RPD foram 1,89, 1,65 e 1,45 e os valores de RER foram de 7,18, 8,89 e 9,84, para amido, carboidratos solúveis e sacarose, respectivamente.

Para o conjunto de dados das variedades Omer e Shelly, a banda espectral que proporcionou os melhores modelos para os atributos foliares foi VNIR, com pré-tratamentos *Moving Average* (35 e 23 pontos) e OSC. Esses obtiveram R^2_{cv} de 0,38 e 0,35 e RMSECV de 0,05 e 39,42, para amido e carboidratos solúveis, respectivamente. Os valores de RPD foram de 1,27 e 1,24 e os valores de RER foram de 9,81 e 6,26, para amido e carboidrato, respectivamente.

Para o conjunto de dados da variedade Palmer, a banda espectral que proporcionou os melhores modelos para os atributos foliares foi FULL, com pré-tratamentos *Moving Average* (5 pontos) e OSC. Esses obtiveram R^2_{cv} de 0,52 e 0,58 e RMSECV de 0,05 e

26,20, para amido e carboidratos solúveis, respectivamente. Os valores de RPD foram de 1,37 e 1,54 e os valores de RER foram de 5,98 e 6,55, para amido e carboidratos solúveis, respectivamente.

Para o conjunto geral de dados, a banda espectral que proporcionou os melhores modelos para os atributos amido foi FULL e para carboidratos solúveis foi VNIR, com pré-tratamentos *Moving Average* (11 e 29 pontos) e OSC. Esses obtiveram R^2_{cv} de 0,90 e 0,69 e RMSECV de 0,06 e 35,19, para amido e carboidratos solúveis, respectivamente. Os valores de RPD foram de 3,22 e 1,78 e os valores de RER foram de 10,03 e 7,43, para amido e carboidratos solúveis, respectivamente.

Modelos de regressão com seleção de variáveis

O primeiro método de regressão aplicado foi o MLR, devido à simplicidade de sua modelagem e fácil interpretação dos resultados. Entretanto, em virtude de problemas de colinearidade, foi necessário fazer uma seleção de variáveis nos dados de entrada para aplicação do método MLR. A partir da seleção de variáveis pelo PCA e exclusão das variáveis com menor influência para o modelo através do método Jack-Knife feito anteriormente, foram gerados os modelos MLR para todos os atributos, cujos parâmetros de desempenho são apresentados na Tabela 3.

Tabela 3 - Parâmetros estatísticos dos melhores modelos MLR com variáveis selecionadas para os atributos amido, carboidratos solúveis (conjunto geral) e sacarose (conjunto Keitt) em folhas de mangueira para os conjuntos de dados “Keitt”, “Omer, Shelly”, “Palmer” e “Geral”.

Parâmetro	BoW	Pré-tratamento	Var	calibração			validação			RPD
				R^2_{cal}	RMSEC	SEC	R^2_{cv}	RMSECV	SECV	
Keitt										
Amido	VNIR	MA 29 pts, OSC	3	0,82	0,02	0,02	0,76	0,02	0,02	2,03
CHOsol	VNIR	MA 5 pts, OSC	1	0,65	13,79	13,72	0,64	13,99	14,00	1,66
Sacarose	VNIR	MA 23 pts, OSC	11	0,64	1,04	0,97	0,55	1,10	1,10	1,48
Omer, Shelly										
Amido	VNIR	MA 35 pts, OSC	5	0,58	0,04	0,04	0,49	0,04	0,04	1,40
CHOsol	VNIR	MA 23 pts, OSC	12	0,55	35,38	33,02	0,39	38,87	38,87	1,27
Palmer										
Amido	FULL	MA 5 pts, OSC	7	0,51	0,05	0,04	0,41	0,05	0,05	1,30
CHOsol	FULL	MA 5 pts, OSC	5	0,68	23,37	22,84	0,65	24,08	24,08	1,68
Geral										
Amido	FULL	MA 11 pts, OSC	15	0,93	0,06	0,06	0,92	0,06	0,06	3,46
CHOsol	VNIR	MA 29 pts, OSC	14	0,74	32,92	32,15	0,71	33,81	33,81	1,86

Legenda: MLR – *Multiple Linear Regression*; CHOsol – Carboidratos solúveis; BoW – banda de comprimento de onda; VNIR – banda espectral dos comprimentos de onda no intervalo de 350 a 1000 nm; FULL – espectro completo, que abrange os comprimentos de onda no intervalo de 350 a 2500 nm; MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*; Var – número de variáveis selecionadas (comprimentos de onda) com PCA; F – fator; R^2_{cal} – coeficiente de determinação da calibração; RMSEC – erro quadrático médio de calibração; SEC – erro padrão de calibração; R^2_{cv} – coeficiente de determinação da validação cruzada; RMSECV – erro quadrático médio de validação cruzada; SECV – erro padrão de validação cruzada; RPD – desvio preditivo residual.

Com exceção dos modelos desenvolvidos para os atributos amido no conjunto “Palmer”, todos os modelos MLR gerados obtiveram desempenho semelhante ou ligeiramente superior aos modelos PLSR, devido à seleção dos componentes principais através da PCA e seleção de variáveis.

O atributo amido do conjunto geral, grupo que proporcionou modelos com desempenho superior aos demais conjuntos, obteve R^2_{cv} iguais a 0,92 e 0,90 e valores de RPD iguais a 3,46 e 3,22 em modelos MLR e PLSR, respectivamente.

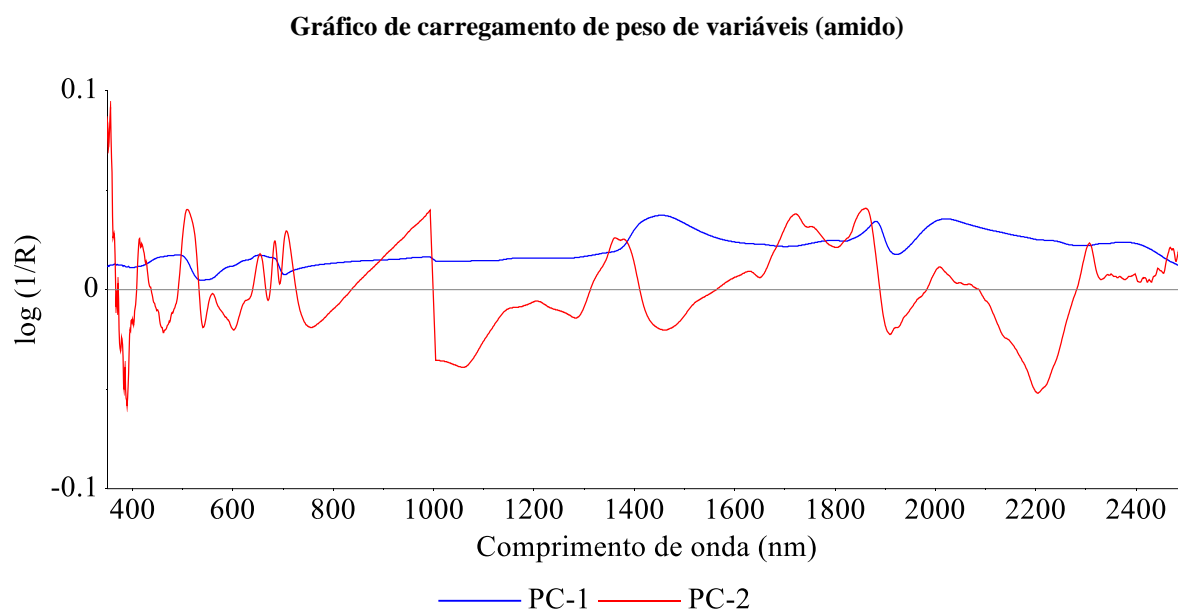
Situação semelhante ocorreu com os atributos carboidratos solúveis (conjunto geral) e sacarose (Keitt), que obtiveram R^2_{cv} iguais a 0,71 e 0,55 em modelos MLR, e 0,69 e 0,54 em modelos PLSR, respectivamente. Em relação aos valores de RPD, os modelos MLR apresentaram valores superiores, sendo 1,86 e 1,48 para MLR e 1,78 e 1,45 para PLSR, para carboidratos solúveis e sacarose, respectivamente.

Vale ressaltar que os modelos MLR utilizaram um número de variáveis independentes (comprimentos de onda) consideravelmente menor em relação aos modelos PLSR. A Figura 7 representa o peso das variáveis independentes para cada atributo,

considerando os melhores conjuntos de dados, e seus respectivos componentes principais utilizados na seleção das variáveis com maior correlação com as variáveis preditoras. A partir deles, é possível visualizar as evidências da possível associação entre os comprimentos de onda e os atributos de interesse.

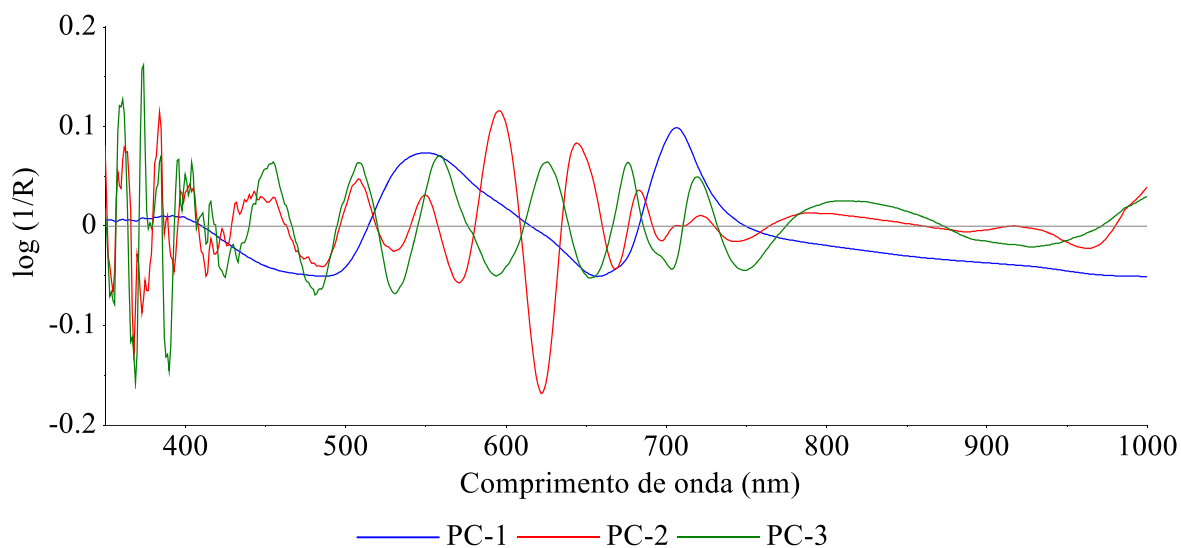
Figura 7 - Gráficos de carregamento de peso das variáveis independentes: (a) amido com pré-tratamento MA 11 pts, OSC, conjunto geral; (b) carboidrato com pré-tratamento MA 29 pts, OSC, conjunto geral; (c) sacarose com pré-tratamento MA 23 pts, OSC, conjunto Keitt.

(a)



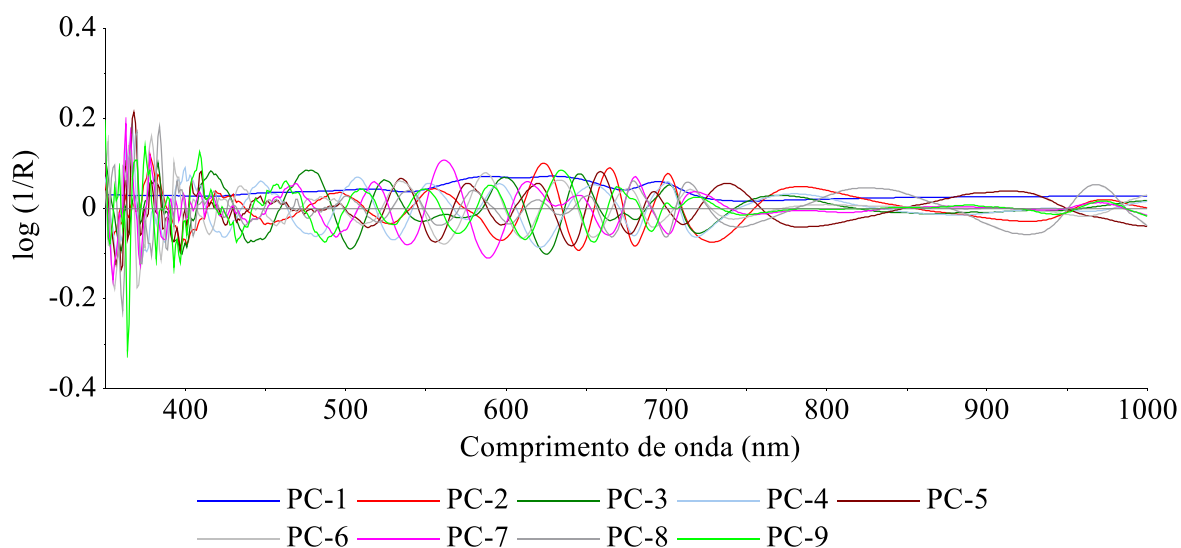
Legenda: PC – *Principal component* (componente principal); MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

(b)

Gráfico de carregamento de peso de variáveis (carboidratos solúveis)

Legenda: PC – *Principal component* (componente principal); MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

(c)

Gráfico de carregamento de peso de variáveis (sacarose)

Legenda: PC – *Principal component* (componente principal); MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

A partir da determinação das assinaturas espectrais, o método de regressão PLSR foi aplicado novamente, desta vez, utilizando apenas os comprimentos de ondas seleccionados como variáveis independentes, cujos parâmetros de desempenho são apresentados na Tabela 4.

Tabela 4 - Parâmetros estatísticos dos melhores modelos PLSR com variáveis selecionadas para predição de amido, carboidratos solúveis e sacarose em folhas de mangueira para os conjuntos de dados “Keitt”, “Omer, Shelly”, “Palmer” e “Geral”.

Parâmetro	BoW	Pré-tratamento	Var	F	calibração			validação				
					R^2_{cal}	RMSEC	SEC	R^2_{cv}	RMSECV	SECV	RPD	
Keitt												
Amido	VNIR	MA 29 pts, OSC	3	3	0,78	0,02	0,02	0,76	0,02	0,02	2,03	
CHOsol	VNIR	MA 5 pts, OSC	1	-	-	-	-	-	-	-	-	
Sacarose	VNIR	MA 23 pts, OSC	11	1	0,46	1,19	1,19	0,42	1,22	1,23	1,31	
Omer, Shelly												
Amido	VNIR	MA 35 pts, OSC	5	4	0,57	0,04	0,04	0,47	0,04	0,04	1,38	
CHOsol	VNIR	MA 23 pts, OSC	12	1	0,38	38,39	38,60	0,35	39,41	39,62	1,24	
Palmer												
Amido	FULL	MA 5 pts, OSC	7	6	0,49	0,04	0,04	0,42	0,05	0,05	1,31	
Carboidrato	FULL	MA 5 pts, OSC	5	4	0,68	22,79	22,90	0,65	23,94	24,05	1,68	
Geral												
Amido	FULL	MA 11 pts, OSC	15	2	0,88	0,07	0,07	0,87	0,07	0,07	2,77	
CHOsol	VNIR	MA 29 pts, OSC	14	1	0,63	38,11	38,17	0,63	38,36	38,42	1,64	

Legenda: PLSR – *Partial Least Squares Regression*; CHOsol – Carboidratos solúveis; BoW – banda de comprimento de onda; VNIR– banda espectral dos comprimentos de onda no intervalo de 350 a 1000 nm; FULL – espectro completo, que abrange os comprimentos de onda no intervalo de 350 a 2500 nm; MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*; Var– número de variáveis selecionadas (comprimentos de onda) com PCA; F – fator; R^2_{cal} – coeficiente de determinação da calibração; R^2_{cv} – coeficiente de determinação da validação cruzada; RMSEC – erro quadrático médio de calibração; SEC – erro padrão de calibração; R^2_{cv} – coeficiente de determinação da validação cruzada; RMSECV – erro quadrático médio de validação cruzada; SECV – erro padrão de validação cruzada; RPD – desvio preditivo residual.

Em relação aos modelos PLSR com seleção de variáveis, o desempenho não apresentou melhora em relação aos modelos PLSR utilizando o espectro completo e modelos MLR. Além disso, é possível observar que não foi possível desenvolver um modelo para o atributo carboidratos solúveis do conjunto Keitt, pois apenas um comprimento de onda foi selecionado como assinatura espectral para esse atributo.

Com o intuito de encontrar um modelo de predição mais robusto que os modelos desenvolvidos (PLSR e MLR), novos métodos de regressão foram aplicados nos três atributos analisados, considerando os melhores pré-tratamentos, bandas espectrais, conjuntos de dados e assinaturas espectrais definidos anteriormente. Foram desenvolvidos modelos *Gaussian Processes* (GP), *Multilayer Perceptron* (MPL), *Model Tree* (M5P), *Random Forest* (RF) e *Support Vector Machine* (SVM). A Tabela 5 apresenta os parâmetros de desempenho encontrados para cada modelo desenvolvido.

Tabela 5 - Parâmetros estatísticos dos modelos matemáticos *Gaussian Processes* (GP), *Multilayer Perceptron* (MPL), *Model Tree* (M5P), *Random Forest* (RF) e *Support Vector Machine* (SVM) para predição de amido, carboidratos solúveis e sacarose em folhas de mangueira para os conjuntos de dados “Keitt”, “Omer, Shelly”, “Palmer” e “Geral”.

Parâmetro	Conjunto	BoW	Pré-tratamento	Modelo	R	R ² _{cv}	MAE	RMSECV	RAE	RRSE
Amido	Geral	FULL	MA 11 pts, OSC	GP	0,85	0,71	0,14	0,16	74,14	76,04
				MPL	0,96	0,92	0,05	0,06	24,25	29,29
				M5P	0,96	0,92	0,04	0,06	22,02	27,54
				RF	0,96	0,92	0,04	0,06	23,20	28,73
				SVM	0,93	0,87	0,06	0,07	30,75	36,14
CHOsol	Geral	VNIR	MA 29 pts, OSC	GP	0,79	0,63	29,82	38,29	54,91	60,87
				MPL	0,73	0,54	34,00	44,27	62,62	70,36
				M5P	0,81	0,65	28,25	37,17	52,03	59,09
				RF	0,78	0,61	29,15	39,63	53,67	63,00
				SVM	0,79	0,63	29,70	38,58	54,71	61,33
Sacarose	Keitt	VNIR	MA 23 pts, OSC	GP	0,50	0,25	1,15	1,58	101,03	96,92
				MPL	0,53	0,28	1,04	1,43	91,02	87,88
				M5P	0,68	0,46	0,79	1,19	69,69	72,71
				RF	0,59	0,35	0,98	1,30	86,38	79,85
				SVM	0,66	0,43	0,80	1,25	69,75	76,36

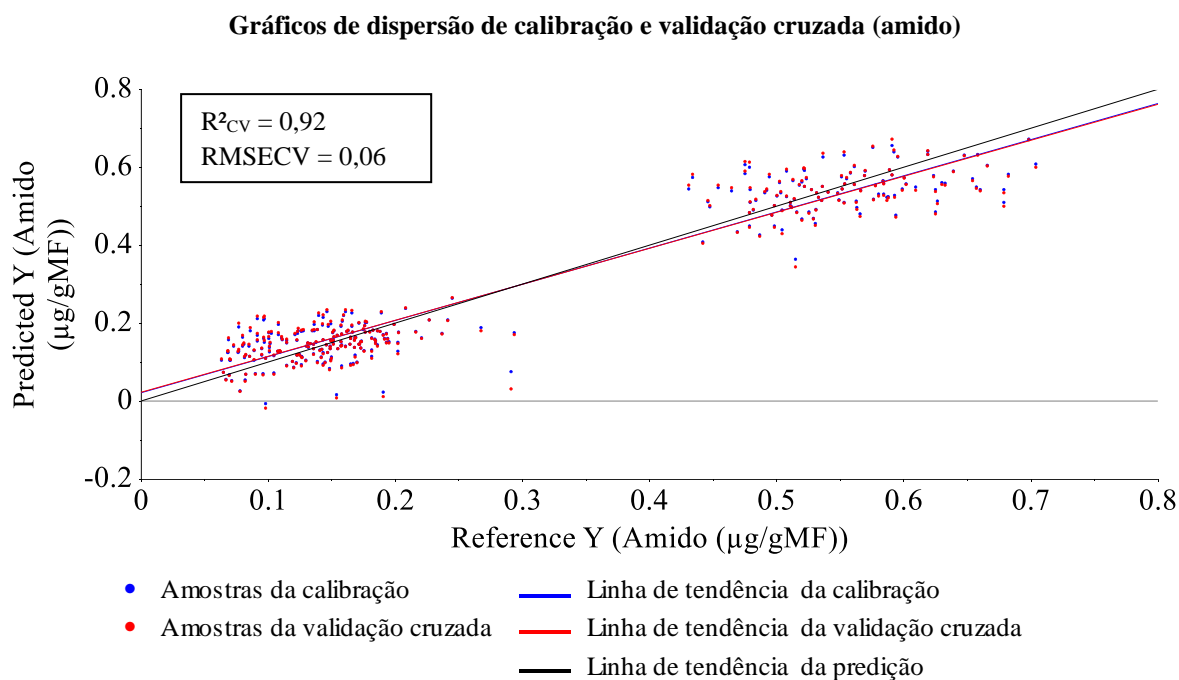
Legenda: CHOsol – Carboidratos solúveis BoW – banda de comprimento de onda; VNIR– banda espectral dos comprimentos de onda no intervalo de 350 a 1000 nm; FULL – espectro completo, que abrange os comprimentos de onda no intervalo de 350 a 2500 nm; MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*; R – coeficiente de correlação; R²_{cv} – coeficiente de determinação da validação cruzada; MAE – erro médio absoluto; RMSECV – erro quadrático médio de validação cruzada; RAE – erro relativo absoluto (%); RRSE – Erro quadrado relativo de raiz (%).

De forma geral, nenhum dos novos modelos gerados superou os modelos MLR e PLSR. Para o atributo amido, os modelos MPL, M5P e RF apresentaram desempenho similar ao modelo MLR, todos com R²_{cv} iguais a 0,92. Entre os modelos gerados para os outros atributos, os melhores modelos foram GP e SVM, ambos com R²_{cv} = 0,63, e M5P com R²_{cv} = 0,46, para carboidratos solúveis e sacarose, respectivamente.

Os gráficos de dispersão de calibração e validação cruzada dos modelos MLR construídos com as assinaturas espectrais são mostrados na Figura 8.

Figura 8 - Gráficos de dispersão de calibração e validação cruzada dos modelos MLR: (a) amido com pré-tratamento MA 11 pts, OSC, conjunto geral; (b) carboidrato com pré-tratamento MA 29 pts, OSC, conjunto geral; (c) sacarose com pré-tratamento MA 23 pts, OSC, conjunto Keitt.

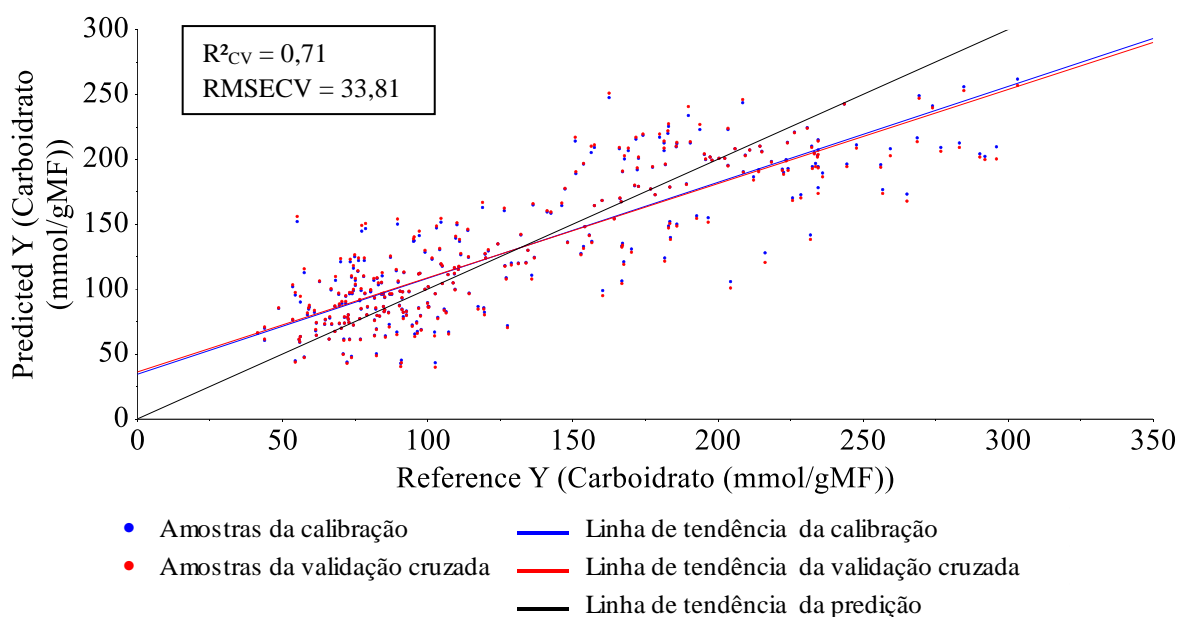
(a)



Legenda: R^2_{cv} – coeficiente de determinação da validação cruzada; RMSECV – erro quadrático médio de validação cruzada MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

(b)

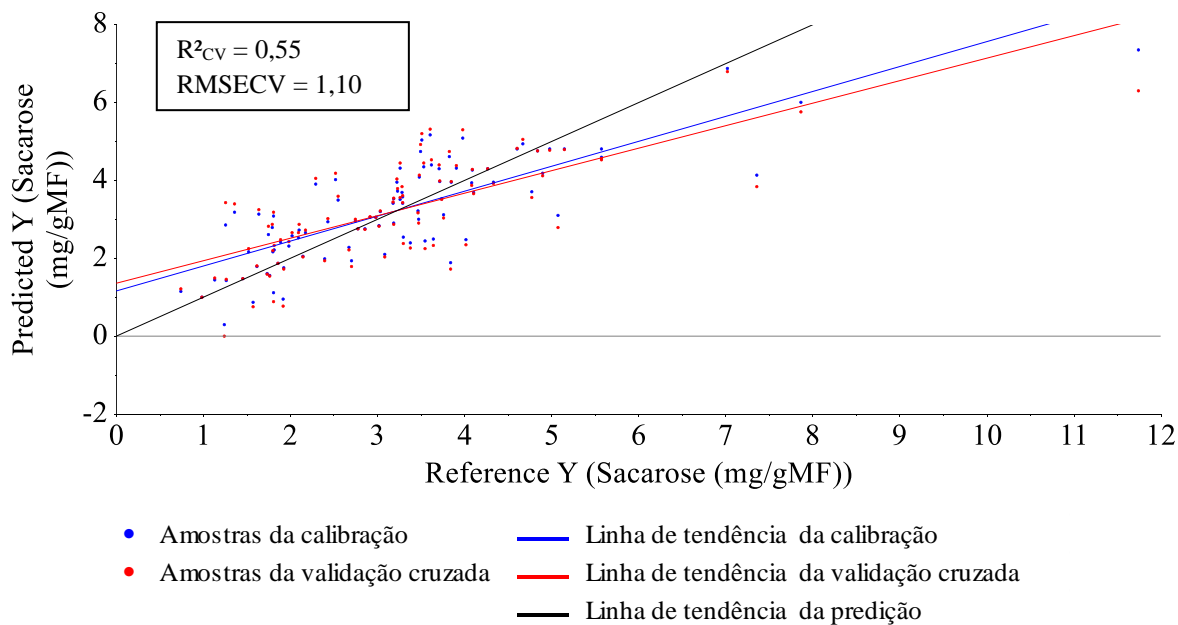
Gráficos de dispersão de calibração e validação cruzada (carboidratos solúveis)



Legenda: R^2_{cv} – coeficiente de determinação da validação cruzada; RMSECV – erro quadrático médio de validação cruzada MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

(c)

Gráficos de dispersão de calibração e validação cruzada (sacarose)



Legenda: R^2_{cv} – coeficiente de determinação da validação cruzada; RMSECV – erro quadrático médio de validação cruzada MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

A partir das assinaturas espectrais identificadas nas bandas espectrais analisadas e dos melhores modelos preditivos desenvolvidos, é possível fazer a predição dos teores de amido, carboidrato e sacarose sem a necessidade de novas análises laboratoriais, considerando a precisão obtida para cada modelo.

3.1.4 DISCUSSÃO

Análise química das folhas

Teores de carboidratos semelhantes aos registrados no presente trabalho foram relatados por Das *et al.* (2019) e Hazi *et al.* (2018). A maior variabilidade para os teores de carboidratos em relação aos outros atributos analisados pode ser atribuída à desuniformidade de maturação dos ramos, fase fenológica da planta, especificidades da variedade ou tratamentos utilizados.

Os teores de carboidratos nas folhas, principalmente amido e açúcares solúveis, estão diretamente relacionados ao crescimento e desenvolvimento das plantas, além de serem fundamentais para o fornecimento de energia para a formação e desenvolvimento de flores e frutos (AMARANTE *et al.*, 2007; LOBO *et al.*, 2019; PRASAD *et al.*, 2014; SILVA *et al.*, 2020).

Entretanto, a necessidade energética de uma mangueira é variável para cada fase de desenvolvimento (crescimento, reprodução e repouso). Quanto maior o período de repouso, maior o acúmulo de reservas e, assim como é possível observar maiores teores de carboidratos nas variedades que estão em repouso (Omer, Shelly e Palmer) em relação a variedade Keitt, que está em pleno florescimento. Além disso, maiores teores de carboidratos proporcionam maiores chances de florescimento, devido à alta demanda de energia na fase reprodutiva (FIERRO e ULLOA, 1991; PRASAD *et al.*, 2014; SILVA *et al.*, 2019).

Análise exploratória do espectro

Os espectros obtidos se mostraram semelhantes para todos os conjuntos de dados. O comportamento espectral de uma folha é função de sua composição, morfologia e estrutura interna e apresenta diferenças desde que sejam de grupos geneticamente diferentes (PONZONI, 2002). Conforme Curran (1989), embora os valores de absorbância sejam variáveis, o comportamento espectral na região de 350 a 2500 nm é notavelmente semelhante entre todas as plantas, principalmente pela interação da luz com a clorofila e com as ligações moleculares O-H, C-H e N-H.

Novo (2010) propôs uma curva de refletância típica de uma folha verde sadia: na região do visível há uma baixa refletância devido à alta absorbância pelas clorofilas, no

infravermelho próximo a absorção diminui, resultado da interação da energia com o mesófilo, de modo que, quanto mais lacunosa for a estrutura interna foliar, menor a absorbância, e no infravermelho médio a absorbância volta a aumentar devido ao conteúdo de água.

Na faixa do visível (400 – 760 nm), ocorre alta absorbância da radiação eletromagnética na folha devido à presença dos pigmentos (clorofila, xantofila e caroteno) nos cloroplastos (GOULD *et al.*, 2018; MENESES *et al.*, 2019). Picos de absorbância semelhantes, em torno de 480 e 680 nm, bem como a queda na curva na região do infravermelho próximo, em torno de 700 a 1300 nm, podem ser observados nos espectros apresentados na Figura 6.

Segundo Meneses *et al.* (2019), a diminuição da absorbância na faixa espectral entre 700 a 1300 nm ocorre devido a estrutura celular, de modo que, quanto mais lacunas entre células paliádicas e mesófilo esponjoso, maior é o espalhamento interno da radiação incidente e menor a absorbância. Essa diminuição de absorbância é importante para impedir o aumento excessivo da temperatura foliar e evitar perdas de proteínas e açúcares.

Os picos de absorbância em torno de 1450 e 1930 nm, que embora sejam normalmente relacionados ao conteúdo de água, podem estar associados à absorbância pelos teores de amido e açúcares nas folhas. De Bei *et al.* (2017), fizeram essa correlação a partir da mensuração do teor de carboidratos não estruturais em amostras de tronco e folhas da videira submetidas à liofilização, ou seja, com conteúdo de água desprezível, reforçando a hipótese da associação desses comprimentos de onda com o conteúdo de amido e açúcares nas folhas.

Modelos de regressão PLSR

Modelos matemáticos com coeficiente de determinação superior a 0,9 ou mais próximo de 1 são considerados excelentes, pois indicam que a proporção da variabilidade dos dados de saída pode ser explicada pela variabilidade das variáveis de entrada (SCHOBER *et al.*, 2018; SHENK; WESTERHAUS, 1996).

Entre os conjuntos e atributos analisados, um ótimo valor de coeficiente de correlação foi obtido para predição do teor de amido para o conjunto geral, pode ser em decorrência das diferentes variedades utilizadas e o maior número de amostras. Segundo Mora e Shimleck (2008), maior variabilidade das amostras propicia o desenvolvimento de modelos matemáticos preditivos mais robustos. Isso ocorre porque os modelos de calibração devem incluir todas as possíveis variações que possam ocorrer em aplicações reais, tendo em vista que o objetivo do modelo é estimar determinada propriedade em amostras desconhecidas.

Assim, quanto maior a variabilidade da propriedade, maior é a probabilidade de obtenção de um maior coeficiente de determinação.

Segundo Williams e Sobering (1996), um bom modelo de predição deve apresentar valor de RPD, que é uma estimativa da predição, próximo a três e RER igual ou superior a dez. Nesses parâmetros, o modelo citado anteriormente obteve um desempenho ligeiramente inferior. De acordo com o RPD obtido, é possível analisar a estimativa de predição de um modelo: entre 3,1 e 4,9, o modelo é adequado para triagem; entre 5,0 e 6,4 para controle de qualidade; entre 6,5 e 8,0 para controle do processo; e por fim, acima de 8,0 é considerado adequado para qualquer aplicação (WILLIAMS, 2001).

Deste modo, para teor de amido (conjunto geral), foram considerados adequados para triagem os modelos preditivos PLSR sem seleção de variáveis e MLR, com valores de RPD iguais a 3,22 e 3,46, respectivamente. De Bei *et al.* (2017) obtiveram resultado semelhante, com RPD de 3,60 em modelos PLSR para predição de amido em folhas de videiras, com um conjunto de dados de 222 amostras.

Os modelos PLSR desenvolvidos para os outros atributos não apresentam os valores mínimos para as aplicações citadas por Williams (2001). Entretanto, Li *et al.* (2015), consideram que, em aplicações agrícolas específicas, os valores de RPD classificam os modelos como: adequados para triagens preliminares e previsões iniciais, para valores superiores a 1,5 (SCHIMLECK *et al.*, 2001); satisfatório para a predição, com valores entre 2,0 e 2,5 (KARLINASARI *et al.*, 2014); adequados para predição, com valores superiores a 3,0 (MOUAZEN *et al.*, 2005).

Diante disso, os modelos desenvolvidos para amido (conjunto geral) foram classificados como adequados para predição. Por outro lado, os modelos para carboidrato (conjunto geral) foram considerados satisfatórios, com valores de RPD iguais a 1,78, 1,64 e 1,86, e os modelos desenvolvidos para sacarose (conjunto Keitt) não obtiveram boa estimativa de predição, com valores de RPD iguais a 1,45, 1,31 e 1,48, para modelos PLSR com a banda espectral completa, PLSR com seleção de variáveis e MLR.

Em relação aos teores de amido, De Bei *et al.* (2017) obtiveram modelos de PLSR em folhas da videira com desempenho semelhante aos modelos para folhas de mangueira desse estudo, obtendo R^2_{cv} 0,87. Entretanto, em um trabalho mais recente com folhas de trevo, os modelos preditivos desenvolvidos por Frey *et al.* (2020), utilizando um maior número de amostras (509) e faixa espectral de 350 a 2500 nm, obtiveram desempenho inferior, com R^2_{cv} de 0,56 e RMSECV de 17,1. Ely *et al.* (2019), analisaram o teor de amido em folhas de diferentes espécies vegetais (folhas de tomateiro, pepino, abobrinha, soja, manjeriço, feijão e

girassol) e alcançaram R^2_{cv} de 0,80, que pode ser considerado satisfatório, mas inferior em relação aos desempenhos dos modelos para folhas de mangueira.

Em relação aos carboidratos solúveis, De Bei *et al.* (2017), obtiveram modelos de PLSR em folhas da videira com desempenho superior aos modelos para folhas de mangueira, obtendo R^2_{cv} 0,86. Segundo o trabalho de Lohr *et al.* (2017), o modelo PLSR para predição de carboidratos em folhas de crisântemo e pelargônio apresentou bom desempenho (R^2_{cv} de 0,83), também superior ao máximo R^2 encontrado nesse experimento para o mesmo atributo, que foi igual a 0,68 para o conjunto geral. Uma possível razão para o melhor desempenho é o número maior de amostras (578 amostras) em comparação ao presente trabalho. Além disso, os autores utilizaram o pré-tratamento SNV para esse modelo de melhor desempenho, pré-tratamento este que proporcionou modelo inferior aos modelos pré-tratados com OSC no presente trabalho.

Em relação ao teor de sacarose, Yendrek *et al.* (2017) conseguiram desenvolver um modelo PLSR para predição de sacarose em folha de milho com R^2_{cv} de 0,62 e RMSECV igual a 0,95, desempenho superior aos modelos desenvolvidos para “Keitt” para o mesmo atributo, com valores de R^2_{cv} iguais a 0,54, 0,42 e 0,55, para modelos PLSR sem e com seleção de variáveis e MLR. Desempenhos superiores foram encontrados em modelos PLSR desenvolvidos por Das *et al.* (2018) e Ely *et al.* (2019), e com R^2_{cv} de 0,78 em folhas de arroz e R^2_{cv} de 0,76 em folhas de diferentes vegetais (tomateiro, pepino, abobrinha, soja, manjeriçã, feijão e girassol).

Lohr *et al.* (2017) relataram dificuldade em desenvolver modelos preditivos para determinação de glicose, frutose e sacarose em folhas de crisântemo e pelargônio. Segundo Kemsley *et al.* (1992) a predição de açúcares pode ser dificultada devido à alta similaridade espectral entre diferentes açúcares. Soest (1994) sugere que a avaliação de carboidratos em tecido vegetal por meio da espectroscopia de infravermelho deveria ser baseada simplesmente na distinção entre carboidratos estruturais e não estruturais.

Modelos de regressão com seleção de variáveis

No presente estudo, nos mesmos dados em que foram aplicados o método PLSR, com e sem seleção de variáveis, também foram aplicados o método MLR (com seleção de variáveis), e com exceção dos atributos amido para a variedade Palmer, o desempenho do modelo aumentou. Diferente dos resultados encontrados no estudo desenvolvido por Das *et al.* (2018) em amostras de folhas de arroz, em modelos para predição de carboidratos, com R^2_{cv} iguais para ambos os modelos, igual a 0,72, e iguais 0,78 e 0,70 para predição de sacarose em modelos PLSR e MLR, respectivamente.

Os métodos de regressão podem apresentar diferentes desempenhos de acordo com cada tipo de amostra. Embora seja recomendado o método PLSR para medições multivariadas (JIA *et al.*, 2010), a utilização do método MLR para desenvolvimento de modelos preditivos é interessante devido à simplicidade de sua modelagem e fácil interpretação dos resultados, principalmente em tentativas preliminares (SAEYS *et al.*, 2019). Assim, ainda que os desempenhos dos modelos MLR e modelos como PLSR sejam semelhantes, por exemplo, o MLR se torna mais vantajoso, pois requer menor custo computacional de processamento, devido ao número reduzido de variáveis e propriedades do método.

Em relação às assinaturas identificadas neste estudo, aquelas que se encontram na região do visível, com comprimentos de onda entre 380 e 780 nm, estão associadas à presença dos pigmentos, em especial às clorofilas “a” e “b”, cujos picos de absorbância acontecem nessa região espectral (GOULD *et al.*, 2018). Já as assinaturas espectrais identificadas da faixa de 780 a 1400 nm estão relacionadas à própria estrutura das células que compõe o mesófilo. As assinaturas selecionadas entre 1400 a 2500 nm, normalmente estão relacionadas ao conteúdo de água na folha (NOVO, 2010; ABRANTES, 2019; JENSEN, 2009). Entretanto, segundo Dei Bei *et al.* (2017) sugerem que essas assinaturas podem estar associadas aos teores de amido e açúcares nas folhas.

Segundo Cen e He (2007), os carboidratos possuem quatro zonas principais de absorbância: em torno de 920 nm, devido a duas vibrações de alongamento assimétrico C-O-C, que corresponde a α e β ; em torno de 1100 nm para as vibrações C-O e C-C; em torno de 1470 e 1380 nm, devido a ligações C-H; e em cerca de 3220 nm, que ultrapassa a região do infravermelho próximo, e ocorre devido a uma ligação O-H da glicose. Bandas de absorção em 1450, 1790 e 2266 nm também podem ser associadas à sacarose, frutose e glicose em amostras de sucos de frutas (LANZA; LI, 2006).

Quanto aos modelos PLSR com seleção de variáveis, não houve melhora significativa no desempenho em relação aos modelos PLSR com a banda espectral completa, semelhante aos resultados encontrados por Frey *et al.* (2020), que obteve R^2_{cv} de 0,56 em modelos PLSR com todos os comprimentos de onda e R^2_{cv} de 0,18 com seleção de variáveis. Além disso, embora o PLSR não tenha em seu algoritmo básico a execução de seleção de fatores influentes, apresenta a propriedade de ponderação ascendente e decrescente de fatores, que pode tornar desnecessário o processo de seleção de fatores influentes por outros métodos (MARTENS; NAES, 1989; MEHMOOD *et al.*, 2012).

Contudo, apesar da seleção de variáveis poder, teoricamente, melhorar o desempenho do modelo, também pode ocorrer a eliminação de “redundâncias úteis” (informações importantes) para o modelo causando então a queda do seu desempenho. Além disso, diminuir o número de variáveis para predição também significa aumentar a influência de cada variável no modelo final (NØRGAARD *et al.*, 2000).

Outros métodos de regressão foram aplicados para os atributos amido, carboidrato e sacarose, utilizando apenas os melhores conjuntos de dados. Entretanto, nenhum dos novos modelos gerados (GP, MPL, M5P, RF e SVM) superou os modelos gerados com PLSR e MLR. Os modelos MPL, M5P e RF, para o atributo amido, apresentaram desempenho similar ao modelo MLR, todos com R^2_{cv} iguais a 0,92. Entre os modelos gerados para os outros atributos, os melhores modelos gerados foram GP e SVM, ambos com $R^2_{cv} = 0,63$, e M5P com $R^2_{cv} = 0,46$, para carboidrato e sacarose, respectivamente.

Embora tenham sido gerados modelos com desempenho similar aos modelos MLR para o atributo amido, o MLR requer menor custo computacional de processamento em relação aos modelos MPL, M5P e RF, tornando o uso do método MLR mais interessante nessa situação.

Em trabalhos que envolvem a determinação quantitativa de um composto em amostras por meio da espectroscopia, frequentemente se faz o uso do método PLSR devido à sua capacidade de trabalhar com um grande número de variáveis sem problemas de colinearidade (WOLD *et al.*, 2001; PAZ-KAGAN *et al.*, 2020; KAWAMURA *et al.*, 2017; LOHR *et al.*, 2017; GENISHEVA *et al.*, 2018), inclusive em amostras foliares (YENDREK *et al.*, 2017; ELY *et al.*, 2019; FREY *et al.*, 2020; SONG *et al.*, 2020). Também é comum a comparação entre modelos PLSR e outros modelos como: MLR, PCA, SVM (WU *et al.*, 2012; DAS *et al.*, 2018; SAMPAIO *et al.*, 2020; PENG *et al.*, 2019) ou combinado com outros algoritmos como *Boosting*, Kernel, algoritmo genético (MACEDO *et al.*, 2021; SHEN *et al.*, 2018). Contudo, não é comum a utilização dos métodos M5P e *Random Forest* na espectroscopia.

Existem vários fatores que influenciam na escolha do melhor método de regressão: objetivo da análise, tamanho da amostra, suposições inerentes à regressão (linearidade, normalidade) e avaliação dos parâmetros de desempenho gerados pelos modelos, como coeficiente de determinação, RMSEC e RPD (HAIR, 2005; ZHANG *et al.*, 2012; ELHAMDAOUI *et al.*, 2020). Isto é, além de considerar a aplicabilidade do método de regressão ao contexto do estudo, a escolha final do melhor método é feita em função dos parâmetros de desempenho dos diferentes modelos gerados para o estudo em questão. Ou

seja, quanto mais métodos de regressão forem aplicados, maiores são as chances de desenvolver um modelo preditivo com melhor desempenho.

Das *et al.* (2018), assim como o presente estudo, se propuseram a testar diferentes métodos de regressão, que apesar de possível, não são comumente aplicados para determinação de compostos em amostras, principalmente de carboidratos em folhas. Além do PLSR, MLR e SVM, foram desenvolvidos modelos *Random Forest*, ANN (*Artificial Neural Network*) e MARS (*Multivariate Adaptive Regression*) para predição de sacarose, açúcares redutores e totais. Entretanto, o modelo *Random Forest* apresentou desempenho inferior aos outros modelos desenvolvidos ($R^2_{cv} = 0,64$), diferente do que ocorreu com o presente trabalho, em que RF foi um dos métodos que proporcionou o melhor valor de R^2_{cv} (0,92).

Uma possível explicação é que o RF tenha sido favorecido pela utilização de um pré-tratamento previamente selecionado, capaz de remover informações que não seriam capazes de contribuir ou poderiam diminuir o desempenho do modelo, bem como pela seleção das variáveis com maior peso na construção do modelo. Semelhante aos resultados de Shen *et al.* (2018), que utilizaram o método PLSR para predição de cádmio em folhas de alface e os métodos KNN (*K-Nearest Neighbors*) e *Random Forest* para classificação das amostras quanto ao grau de contaminação. O modelo *Random Forest* utilizado para classificação obteve R^2_{cv} igual a 0,80 utilizando o espectro bruto (desempenho inferior ao KNN, com $R^2_{cv} = 0,84$). Após a aplicação de pré-tratamentos e PCA, seu desempenho superou o modelo de classificação KNN, com R^2_{cv} iguais a 0,96 e 0,92, respectivamente.

3.1.5 CONCLUSÕES

A espectroscopia se apresenta como uma ferramenta satisfatória para a determinação de atributos bioquímicos em folhas de mangueira (*Mangifera indica* L.). Para a determinação do teor de amido, utilizando conjunto formado por todas as variedades (Keitt, Shelly, Omer e Palmer), os modelos PLSR com banda espectral completa, e os modelos MLR, MPL e RF apresentaram R^2_{cv} superior a 0,90. Para carboidratos, apenas os modelos MLR apresentaram R^2_{cv} superior a 0,70. O modelo MLR também proporcionou o melhor modelo para o teor de sacarose, com $R^2_{cv} = 0,55$.

As assinaturas espectrais dos atributos amido e carboidratos, associadas aos modelos de MLR, são consideradas adequadas para construção de modelos de predição e proporcionam modelos com R^2_{cv} iguais a 0,92 e 0,71, respectivamente, superando os modelos PLSR.

Diversos métodos de regressão podem ser utilizados para predição de carboidratos solúveis, amido e sacarose em folhas de mangueira. Entretanto, de modo geral, modelos MLR com seleção de variáveis propiciaram modelos com melhor desempenho, além do baixo custo computacional.

3.1.6 REFERÊNCIAS

- ABRANTES, T. C. **Análise da deriva simulada de herbicidas auxínicos em soja através de índices de vegetação rgb obtidos por VANT**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2019.
- AITKENHEAD, M. J. *et al.* PHYLIS: A Low-Cost Portable Visible Range Spectrometer for Soil and Plants. **Sensors**, v. 17, n. 1, p. 99, jan. 2017.
- ALTHOFF, D. *et al.* Heuristic methods applied in reference evapotranspiration modeling. **Ciência e Agrotecnologia**, v. 42, p. 314–324, jun. 2018.
- ALVARES, C. A. *et al.* Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, p. 711–728, dez. 2013.
- AMARANTE, C. V. T. DO *et al.* Radiação, fotossíntese, rendimento e qualidade de frutos em macieiras “Royal Gala” cobertas com telas antigranizo. **Pesquisa Agropecuária Brasileira**, v. 42, n. 7, p. 925–931, jul. 2007.
- BANDARA, K. M. T. S. *et al.* A Low-Cost NDIR-Based N₂O Gas Detection Device for Agricultural Soils: Assembly, Calibration Model Validation, and Laboratory Testing. **Sensors**, v. 21, n. 4, p. 1189, jan. 2021.
- BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. **Appl Spectrosc**, v. 43, n. 5, p. 772–777, jul. 1989.
- BRASIL. **PORTARIA Nº 240, DE 12 DE MARÇO DE 2019 - Imprensa Nacional**. Disponível em: <<https://www.in.gov.br/materia>>. Acesso em: 11 mar. 2022.
- CANASSA, T. A.; LAMONATO, A. L.; RIBEIRO, A. V. Utilização da lei de Lambert-Beer para determinação da concentração de soluções. **Journal of Experimental Techniques and Instrumentation**, v. 1, n. 2, 6 jul. 2018.
- CARDOSO, J. R. F. Desenvolvimento de Estrutura Robótica para Aquisição e Classificação de Imagens (ERACI) de Lavoura de Cana-de-Açúcar. 29 jun. 2020.
- CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. **Trends in Food Science & Technology**, v. 18, n. 2, p. 72–83, fev. 2007.
- CEPEA. Anuário 2021-2022. **Hortifruti Brasil**, 2021.
- COLEZEA, M. *et al.* CLUeFARM: Integrated web-service platform for smart farms. **Computers and Electronics in Agriculture**, v. 154, p. 134–154, nov. 2018.
- COMINO, F. *et al.* Infrared spectroscopy as a tool for the assessment of soil biological quality in agricultural soils under contrasting management practices. **Ecological Indicators**, v. 87, p. 117–126, abr. 2018.

COSTA, D. **Desenvolvimento de protótipo optoeletrônico para determinação não destrutiva dos atributos de qualidade e estágio de amadurecimento em uvas viníferas.** Campinas: Universidade Estadual de Campinas, 2018.

CURRAN, P. J. Remote sensing of foliar chemistry. **Remote Sensing of Environment**, v. 30, n. 3, p. 271–278, 1 dez. 1989.

DALE, L. M. *et al.* Hyperspectral Imaging Applications in Agriculture and Agro-Food Product Quality and Safety Control: A Review. **Applied Spectroscopy Reviews**, v. 48, n. 2, p. 142–159, mar. 2013.

DAS, A. *et al.* Interrelations of growth regulators, carbohydrates and expression of flowering genes (FT, LFY, AP1) in leaf and shoot apex of regular and alternate bearing mango (*Mangifera indica* L.) cultivars during flowering. **Scientia Horticulturae**, v. 253, p. 263–269, 2019.

DAS, B. *et al.* Quantitative monitoring of sucrose, reducing sugar and total sugar dynamics for phenotyping of water-deficit stress tolerance in rice through spectroscopy and chemometrics. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 192, p. 41–51, mar. 2018.

DAVENPORT, T. L. Reproductive physiology of mango. **Brazilian Journal of Plant Physiology**, v. 19, p. 363–376, dez. 2007.

DE BEI, R. *et al.* Rapid measurement of total non-structural carbohydrate concentration in grapevine trunk and leaf tissues using near infrared spectroscopy. **Computers and Electronics in Agriculture**, v. 136, p. 176–183, abr. 2017.

DECHANT, B. *et al.* Estimation of photosynthesis traits from leaf reflectance spectra: Correlation to nitrogen content as the dominant mechanism. **Remote Sensing of Environment**, v. 196, p. 279–292, jul. 2017.

DUBOIS, M. *et al.* Colorimetric Method for Determination of Sugars and Related Substances. 1956.

ELHAMDAOUI, O. *et al.* Development of Fast Analytical Method for the Detection and Quantification of Honey Adulteration Using Vibrational Spectroscopy and Chemometrics Tools. **Journal of Analytical Methods in Chemistry**, v. 2020, p. e8816249, 23 dez. 2020.

ELY, K. S. *et al.* Spectroscopy can predict key leaf traits associated with source–sink balance and carbon–nitrogen status. **Journal of Experimental Botany**, v. 70, n. 6, p. 1789–1799, mar. 2019.

EMBRAPA. **Specsolo - Portal Embrapa.** Disponível em: <<https://www.embrapa.br/en/busca-de-solucoes-tecnologicas/-/produto-servico/3740/specsolo>>. Acesso em: 15 dez. 2021.

EMBRAPA. **Exportação de manga brasileira bate recorde em 2020, totalizando US\$ 246 milhões.** Disponível em: <<https://www.embrapa.br/busca-de-noticias/-/noticia/60585117/exportacao-de-manga-brasileira-bate-recorde-em-2020-totalizando-us-246-milhoes>>. Acesso em: 4 jun. 2021.

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

FIERRO, C. A.; ULLOA, M. A. Development reference stage for flower induction response to potassium nitrate in mango. **Acta Horticulturae**, n. Wageningen. n.291. p.71-78, 1991.

FILIPPUCCI, P. *et al.* Soil moisture as a potential variable for tracking and quantifying irrigation: A case study with proximal gamma-ray spectroscopy data. **Advances in Water Resources**, v. 136, p. 103502, fev. 2020.

FREY, L. A. *et al.* A Non-destructive Method to Quantify Leaf Starch Content in Red Clover. **Frontiers in Plant Science**, v. 11, 2020.

GANESAN, M.; ANDAVAR, S.; RAJ, R. S. P. Prediction of Land Suitability for Crop Cultivation Using Classification Techniques. **Brazilian Archives of Biology and Technology**, v. 64, 25 out. 2021.

GENISHEVA, Z. *et al.* New PLS analysis approach to wine volatile compounds characterization by near infrared spectroscopy (NIR). **Food Chemistry**, v. 246, p. 172–178, 25 abr. 2018.

GORRY, P. A. General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method. **Anal. Chem.**, v. 62, n. 6, p. 570–573, mar. 1990.

GOULD, K. *et al.* When are foliar anthocyanins useful to plants? Re-evaluation of the photoprotection hypothesis using Arabidopsis thaliana mutants that differ in anthocyanin accumulation. **Environmental and Experimental Botany**, v. 154, 1 mar. 2018.

HAIR, J. F. **Análise Multivariada de Dados - 6ª Edição**. 6. ed. [s.l.] 2005.

HAZIS, N. H. *et al.* Assessment of Carbohydrate Contents in Perlis Harumanis Mango Leaves during Vegetative and Productive Growth. **IOP Conference Series: Materials Science and Engineering**, v. 429, p. 012025, nov. 2018.

JENSEN, J. R. **Sensoriamento Remoto do Ambiente: Uma perspectiva em Recursos Terrestres**. [s.l.] Ed. Parêntese, 2009.

JIA, R.; MAO, Z.; CHANG, Y. **A nonlinear robust partial least squares method with application**. 2010 Chinese Control and Decision Conference. **Anais...** In: 2010 CHINESE CONTROL AND DECISION CONFERENCE. maio 2010.

KARLINASARI, L. *et al.* Near infrared (NIR) spectroscopy for estimating the chemical composition of (Acacia mangium Willd.) wood. **Journal of the Indian Academy of Wood Science**, 2014.

KAWAMURA, K. *et al.* Vis-NIR Spectroscopy and PLS Regression with Waveband Selection for Estimating the Total C and N of Paddy Soils in Madagascar. **Remote Sensing**, v. 9, p. 1081, 23 out. 2017.

KEMSLEY, E. K. *et al.* Quantitative analysis of sugar solutions using infrared spectroscopy. **Food Chemistry**, v. 44, n. 4, p. 299–304, jan. 1992.

- LAMMERS, K.; ARBUCKLE-KEIL, G.; DIGHTON, J. FT-IR study of the changes in carbohydrate chemistry of three New Jersey pine barrens leaf litters during simulated control burning. **Soil Biology and Biochemistry**, v. 41, n. 2, p. 340–347, fev. 2009.
- LANNA, A. C. *et al.* Mitigação dos estresses abióticos na agricultura mediada pela interação de microrganismos e plantas. **Documentos**, n. 319, 17 nov. 2021.
- LANZA, E.; LI, B. Application of Near Infrared Spectroscopy for Predicting the Sugar Content of Fruit Juices. **Journal of Food Science**, v. 49, p. 995–998, 25 ago. 2006.
- LI, B. *et al.* Prediction and monitoring of leaf water content in soybean plants using terahertz time-domain spectroscopy. **Computers and Electronics in Agriculture**, v. 170, p. 105239, mar. 2020.
- LI, X. *et al.* Determination of Hemicellulose, Cellulose and Lignin in Moso Bamboo by Near Infrared Spectroscopy. **Scientific Reports**, v. 5, p. 17210, 2015.
- LIMA, G. M. DE S. *et al.* Manejo da indução floral da mangueira “Palmer” com uso de uniconazole. **Ciência Rural**, v. 46, p. 1350–1356, 6 maio 2016.
- LOBO, J. T. *et al.* Biostimulants on Nutritional Status and Fruit Production of Mango ‘Kent’ in the Brazilian Semiarid Region. **HortScience**, v. 54, n. 9, p. 1501–1508, 1 set. 2019.
- LOHR, D. *et al.* Non-destructive determination of carbohydrate reserves in leaves of ornamental cuttings by near-infrared spectroscopy (NIRS) as a key indicator for quality assessments. **Biosystems Engineering**, v. 158, p. 51–63, jun. 2017.
- LUCAS, B. N.; SCHÚ, A. I.; DALLA NORA, F. M. Uso de smartphone como alternativa inovadora no controle de qualidade de alimentos: uma breve revisão. **Avanços em ciência e tecnologia de alimentos**, n. 21, p. 278–288, 1 abr. 2021.
- MARTENS, H.; MARTENS, M. Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). **Food Quality and Preference**, v. 11, n. 1, p. 5–16, 1 jan. 2000.
- MASSARUHÁ, S. M. F. S. *et al.* A transformação digital no campo rumo à agricultura sustentável e inteligente. In: **Embrapa Informática Agropecuária-Artigo em anais de congresso (ALICE)**. Brasília: EMBRAPA, 2020, 2020. p. 409.
- MEHMOOD, T. *et al.* A review of variable selection methods in Partial Least Squares Regression. **Chemometrics and Intelligent Laboratory Systems**, v. 118, p. 62–69, 15 ago. 2012.
- MENDES, W. DE S. *et al.* Geostatistics or machine learning for mapping soil attributes and agricultural practices. **Revista Ceres**, v. 67, p. 330–336, 28 ago. 2020.
- MENESES, P. R.; ALMEIDA, T.; BAPTISTA, G. M. **Reflectância dos materiais terrestres: análise e interpretação**. São Paulo: Oficina de Textos: [s.n.].
- MORA, C. R.; SCHIMLECK, L. R. On the selection of samples for multivariate regression analysis: application to near-infrared (NIR) calibration models for the prediction of pulp yield in *Eucalyptus nitens*. **Canadian Journal of Forest Research**, 2008.

- MOUAZEN, A. *et al.* Near Infrared Spectroscopy for Agricultural Materials: An Instrument Comparison. **Journal of Near Infrared Spectroscopy**, v. 13, p. 87–98, 1 abr. 2005.
- MUNAWAR, A. A. *et al.* Calibration models database of near infrared spectroscopy to predict agricultural soil fertility properties. **Data in Brief**, v. 30, p. 105469, jun. 2020.
- NAKAJIMA, S. *et al.* Quantification of starch content in germinating mung bean seedlings by terahertz spectroscopy. **Food Chemistry**, v. 294, p. 203–208, out. 2019.
- NEVES, L.; MORAES, D. M. Vigour And A-Amylase Analisis In Seeds Of Rice Cultivars Submitted To Several Treatments With Acetic Acid. **Revista de Ciências Agroveterinárias**, n. Lages, v. 4, n 1, p. 35–43, 2005.
- NØRGAARD, L. *et al.* Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. **Applied Spectroscopy**, v. 54, n. 3, p. 413–419, 1 mar. 2000.
- NOVO, E. M. L. M. **Sensoriamento remoto: princípios e aplicações**. 3ª edição ed. São Paulo: Blucher, 2010.
- OLIVEIRA, G. P. *et al.* Teores de carboidratos em mangueira “Ubá” submetida a diferentes doses de paclobutrazol. **Revista de Ciências Agrárias**, v. 41, n. 3, p. 749–756, 2018.
- OLIVEROS, N. *et al.* Predictive models of chlorophyll content in sugarcane seedlings using spectral images. **Engenharia Agrícola**, v. 41, p. 475–484, 15 set. 2021.
- ONU. **Population**. Disponível em: <<https://www.un.org/en/global-issues/population>>. Acesso em: 28 dez. 2020.
- PABÓN, R. E. C. **Espectroscopia de refletância e de imageamento aplicadas ao monitoramento de solos expostos em instalações petrolíferas**. Campinas: Universidade Estadual de Campinas, 2015.
- PAZ-KAGAN, T. *et al.* Assessing the nitrogen status of almond trees by visible-to-shortwave infrared reflectance spectroscopy of carbohydrates. **Computers and Electronics in Agriculture**, v. 178, p. 105755, nov. 2020.
- PENG, J. *et al.* Fast visualization of distribution of chromium in rice leaves by re-heating dual-pulse laser-induced breakdown spectroscopy and chemometric methods. **Environmental Pollution**, v. 252, p. 1125–1132, 1 set. 2019.
- PETISCO, C. *et al.* Near-infrared reflectance spectroscopy as a fast and non-destructive tool to predict foliar organic constituents of several woody species. **Anal Bioanal Chem**, v. 386, n. 6, p. 1823–1833, nov. 2006.
- PONZONI, F. J. Sensoriamento remoto no estudo da vegetação: diagnosticando a Mata Atlântica. In: **Instituto Nacional de Pesquisas Espaciais**. São José dos Campos: [s.n.].
- PRASAD, S. R. S. *et al.* Studies on Changes in Carbohydrate Metabolism in Regular Bearing and “Off” Season Bearing Cultivars of Mango (*Mangifera indica* L.) During Flowering. **International Journal of Fruit Science**, v. 14, n. 4, p. 437–459, out. 2014.

- REDA, R. *et al.* Comparing CalReg performance with other multivariate methods for estimating selected soil properties from Moroccan agricultural regions using NIR spectroscopy. **Chemometrics and Intelligent Laboratory Systems**, v. 211, p. 104277, abr. 2021.
- RODRÍGUEZ-PÉREZ, J. R. *et al.* Leaf water content estimation by functional linear regression of field spectroscopy data. **Biosystems Engineering**, Sensing and Control of Crop Water Status. v. 165, p. 36–46, jan. 2018.
- SAEYS, W. *et al.* Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. **Postharvest Biology and Technology**, v. 158, p. 110981, 1 dez. 2019.
- SAKANE, K. K.; MATSUURA, E. N.; SIMIONI, A. R. Espectroscopia no infravermelho como ferramenta para diferenciação de chá verde na agricultura orgânica e convencional. **Revista Tecnologia e Tendências**, v. 10, n. 2, p. 59–74, dez. 2019.
- SAMPAIO, P. S. *et al.* Identification of rice flour types with near-infrared spectroscopy associated with PLS-DA and SVM methods. **European Food Research and Technology**, v. 246, n. 3, p. 527–537, 1 mar. 2020.
- SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, jan. 1964.
- SCHIMLECK, L. R.; EVANS, R.; ILIC, J. Estimation of Eucalyptus delegatensis wood properties by near infrared spectroscopy. **Canadian Journal of Forest Research**, v. 31, n. 10, p. 1671–1675, 1 out. 2001.
- SCHOBER, P.; BOER, C.; SCHWARTE, L. Correlation coefficients: appropriate use and interpretation. **Anesthesia & Analgesia**, 2018.
- SERBIN, S. P. *et al.* Leaf optical properties reflect variation in photosynthetic metabolism and its sensitivity to temperature. **J Exp Bot**, v. 63, n. 1, p. 489–502, jan. 2012.
- SHAH, S. S. A. *et al.* Towards fruit maturity estimation using NIR spectroscopy. **Infrared Physics & Technology**, v. 111, p. 103479, dez. 2020.
- SHEN, T. *et al.* Rapid Determination of Cadmium Contamination in Lettuce Using Laser-Induced Breakdown Spectroscopy. **Molecules**, v. 23, n. 11, p. 2930, 2018.
- SHENK, J.; WESTERHAUS, M. O. Calibration the ISI way. **Near Infrared Spectroscopy: The future Waves. NIR Chichester**, p. 198–202, 1996.
- SILVA, M. M. L. *et al.* Effect of carbohydrate content on shoot maturation and yield of Palmer mango submitted to potassium fertilization and biostimulant | Research, Society and Development. **Agrarian and Biological Sciences**, v. 9, 2020.
- SKOLIK, P. *et al.* Attenuated total reflection Fourier-transform infrared spectroscopy coupled with chemometrics directly detects pre- and post-symptomatic changes in tomato plants infected with *Botrytis cinerea*. **Vibrational Spectroscopy**, v. 111, p. 103171, nov. 2020.
- SOEST, P. J. V. **Nutritional Ecology of the Ruminant**. [s.l.] Cornell University Press, 1994.

- SONG, X.-C. *et al.* Predicting the antioxidant capacity and total phenolic content of bearberry leaves by data fusion of UV-Vis spectroscopy and UHPLC/Q-TOF-MS. **Talanta**, v. 213, p. 120831, 1 jun. 2020.
- SOUZA, J. T. DE *et al.* Data mining and machine learning in the context of sustainable evaluation: a literature review. **IEEE Latin America Transactions**, v. 17, n. 03, p. 372–382, mar. 2019.
- TORRES, I. *et al.* Irrigation decision support based on leaf relative water content determination in olive grove using near infrared spectroscopy. **Biosystems Engineering**, v. 180, p. 50–58, abr. 2019.
- VAN HANDEL, E. Direct microdetermination of sucrose. **Analytical Biochemistry**, v. 22, n. 2, p. 280–283, fev. 1968.
- WALSH, K. B. *et al.* Visible-NIR ‘point’ spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. **Postharvest Biology and Technology**, v. 168, p. 111246, out. 2020.
- WILLIAMS, P. C. Implementation of near-infrared technology. In P. Williams & K. Norris (Eds.). In: **Near-infrared technology in the agricultural and food industries**. St. Paul, MN: American Association of Cereal Chemists, [s.n.].
- WILLIAMS, P. C.; SOBERING, D. C. How do we do it: a brief summary of the methods we use in developing near infrared calibration. **Near Infrared Spectroscopy: The Future Waves, NIR Publications**, n. Chichester, 185-188, 1996.
- WOLD, S.; SJOSTROM, M.; ERIKSSON, L. PLS-regression: A Basic Tool of Chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, p. 109–130, 1 out. 2001.
- WU, D. *et al.* Application of near infrared spectroscopy for the rapid determination of antioxidant activity of bamboo leaf extract. **Food Chemistry**, v. 135, n. 4, p. 2147–2156, 15 dez. 2012.
- YASSITEPE, J. E. DE C. T. *et al.* Genômica aplicada às mudanças climáticas: biotecnologia para a agricultura digital. - Portal Embrapa. In: **Geotecnologias na agricultura digital**. Brasília: EMBRAPA, 2020, 2020. p. 409.
- YENDREK, C. R. *et al.* High-Throughput Phenotyping of Maize Leaf Physiological and Biochemical Traits Using Hyperspectral Reflectance. **Plant Physiology**, v. 173, n. 1, p. 614–626, jan. 2017.
- ZHANG, Q.; LI, Q.; ZHANG, G. Rapid Determination of Leaf Water Content Using VIS/NIR Spectroscopy Analysis with Wavelength Selection. **Spectroscopy: An International Journal**, v. 27, n. 2, p. 93–105, 28 maio 2012.

3.2 DETERMINAÇÃO DE COMPOSTOS BIOQUÍMICOS EM FOLHAS DE MARACUJAZEIRO AMARELO COM USO DA ESPECTROSCOPIA VIS-NIR

RESUMO

O Brasil é o maior produtor e exportador mundial de maracujá, com quase 700 mil toneladas produzidas em 2020, dos quais cerca de 65% foi produzido no nordeste brasileiro. Compreender e observar os aspectos bioquímicos de uma planta é um marco essencial para o aumento da produtividade. Alguns compostos bioquímicos fornecem informações importantes sobre o estado fisiológico da planta, e monitorar esses parâmetros é imprescindível para alcançar os máximos valores de produtividade. Entretanto, métodos convencionais de análise são complexos e demorados. Deste modo, a espectroscopia pode otimizar a realização das análises de determinação desses atributos bioquímicos. Diante disso, o principal objetivo do trabalho foi determinar os teores de amido, carboidratos solúveis, sacarose, aminoácidos totais, proteínas, nitrato e nitrato redutase, em folhas de maracujazeiro amarelo (*Passiflora edulis*) por meio da espectroscopia Vis-NIR. O trabalho foi desenvolvido na região do Vale do São Francisco, seguindo as seguintes etapas: (1) amostragem das folhas de maracujazeiro; (2) aquisição dos espectros de refletância; (3) determinação dos compostos bioquímicos com uso dos métodos de referência; e (4) desenvolvimento dos modelos preditivos. Os melhores modelos de predição dos diferentes atributos bioquímicos foram selecionados com base no desempenho dos modelos, sendo os parâmetros estatísticos avaliados: coeficiente de determinação (R^2); erro quadrático médio de calibração (RMSEC); erro padrão de calibração (SEC); erro quadrático médio de validação cruzada (RMSECV); erro padrão de validação cruzada (SECV). Os métodos de regressão utilizados foram a *Partial Least Squares Regression* (PLSR) e *Multiple Linear Regression* (MLR). Os melhores modelos desenvolvidos foram a PLSR para amido e carboidratos, com R^2_{cv} iguais a 0,83 e 0,39, e a MLR para sacarose, aminoácidos totais, proteína, nitrato e nitrato redutase, com R^2_{cv} iguais a 0,98, 0,99, 0,86, 0,99 e 0,84, respectivamente. Deste modo, a espectroscopia é uma ferramenta com alto potencial para determinação de compostos bioquímicos em folhas de maracujazeiro.

PALAVRAS-CHAVE: Fruticultura. Métodos ópticos. Análises bioquímicas. Quimiometria.

ABSTRACT

Brazil is the world's largest producer and exporter of passion fruit, with almost 700 thousand tons produced in 2020, of which about 65% was produced in the Brazilian northeast. Understanding and observing the biochemical aspects of a plant is an essential milestone for increasing productivity. Some biochemical compounds provide important information about the physiological state of the plant, and monitoring these parameters is essential to achieve maximum productivity values. However, conventional methods of analysis are complex and time-consuming. In this way, spectroscopy can optimize the analysis of determination of these biochemical attributes. Therefore, the main objective of this work was to determine the levels of starch, soluble carbohydrates, sucrose, total amino acids, proteins, nitrate and nitrate reductase, in yellow passion fruit (*Passiflora edulis*) leaves by means of Vis-NIR spectroscopy. The work was carried out in the region of Vale do São Francisco, following the following steps: (1) sampling of passion fruit leaves; (2) acquisition of reflectance spectra; (3) determination of biochemical compounds using reference methods; and (4) development of predictive models. The best prediction models for the different biochemical attributes were selected based on the performance of the models, and the statistical parameters evaluated were: coefficient of determination (R^2); Mean Square Error of Calibration (RMSEC); calibration standard error (SEC); cross-validation mean square error (RMSECV); standard error of cross validation (SECV). The regression methods used were Partial Least Squares Regression (PLSR) and Multiple Linear Regression (MLR). The best models developed were the PLSR for starch and carbohydrates, with R^2_{cv} equal to 0.83 and 0.39, and the MLR for sucrose, total amino acid, protein, nitrate and nitrate reductase, with R^2_{cv} equal to 0.98, 0.99, 0.86, 0.99 and 0.84, respectively. Thus, spectroscopy is a tool with high potential for the determination of biochemical compounds in passion fruit leaves.

KEYWORDS: Fruticulture. Optical methods. Biochemical analysis. Chemometrics.

3.2.1 INTRODUÇÃO

Atualmente o Brasil é o maior produtor e exportador mundial da fruta, com quase 700 mil toneladas produzidas em 2020, dos quais cerca de 65% foi produzido no nordeste brasileiro (ABRAFRUTAS, 2021; IBGE, 2021). Embora seja o maior produtor, alguns fatores ainda são limitantes para o crescimento da produção do maracujá no Brasil, são esses: elevado custo da mão-de-obra, que é muito demandada pela cultura; grande ocorrência de pragas e doenças, que acarreta na diminuição da produtividade e aumento do custo de produção; falta de tecnologias adequadas para o sistema de produção, a utilização de variedades melhoradas, poucos estudos relacionados ao manejo do solo, água e planta, e tratos culturais (COELHO *et al.*, 2016; BENEDETTI; VIEIRA, 2021; PIRES *et al.*, 2011).

Para que novas tecnologias sejam desenvolvidas para uma cultura, é imprescindível que se conheça de forma minuciosa o seu ciclo, composição, bem como as alterações fisiológicas e bioquímicas no decorrer de seu desenvolvimento, e as respostas a determinadas condições ambientais diante de estresses bióticos e abióticos. Ou seja, compreender e observar os aspectos fisiológicos e bioquímicos de uma planta é um marco essencial para o aumento da produtividade (ELY *et al.*, 2019; SERBIN *et al.*, 2012; WHITE *et al.*, 2016).

Além disso, alguns compostos bioquímicos fornecem informações importantes sobre o estado fisiológico da planta. Por exemplo, a diminuição das reservas de amido nas folhas para manutenção das atividades metabólicas da planta pode indicar a redução da atividade fotossintética ocasionada por algum estresse (EARLES *et al.*, 2018). Para que não haja comprometimento de produtividade, é imprescindível o monitoramento constante da planta, mas através de avaliações rápidas e que contribuam de fato para a tomada de decisão no campo (KITIC' *et al.*, 2019).

Embora existam diversos métodos para determinação de compostos bioquímicos em amostras vegetais como, por exemplo, para quantificação de nutrientes e carboidratos em folhas, normalmente são métodos analíticos laboratoriais, que apesar de confiáveis, apresentam desvantagens, como alto custo, uso de reagentes (alguns de uso restrito), e preparo das amostras, processo que se torna ainda mais difícil quando se trabalha com um alto número de amostras (AITKENHEAD *et al.*, 2017; KITIC' *et al.*, 2019). Com isso, o desenvolvimento de métodos de análises mais rápidos, simples e não destrutivos é o ponto chave para otimizar os processos produtivos.

Nas últimas décadas, o sensoriamento proximal tem desempenhado um papel fundamental nos estudos do setor agrícola, em especial por meio da espectroscopia. Essa ferramenta otimiza os processos analíticos através da relação estabelecida entre os dados espectrais, obtidos pela radiação refletida ou absorvida, e as propriedades físico-químicas de interesse, por meio do desenvolvimento de modelos preditivos através de métodos de regressão (CANASSA *et al.*, 2018; FERREIRA, 2015). Além disso, esses modelos desenvolvidos podem ser fundamentais para desenvolvimento de protótipos e dispositivos óptico eletrônicos para obtenção de resultados bioquímicos diretamente no campo (KITIĆ *et al.*, 2019).

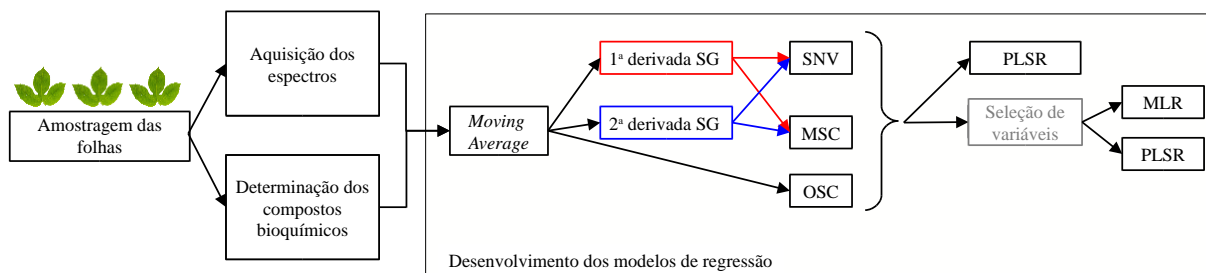
A espectroscopia Vis-NIR tem se destacado em diversas áreas do agronegócio na predição de diversos compostos bioquímicos vegetais, como carboidratos (DAS *et al.*, 2018; DE BEI *et al.*, 2017; FREY *et al.*, 2020), nitrogênio (LI *et al.*, 2016), proteínas (FERET *et al.*, 2021); teor de água (LI *et al.*, 2020); pigmentos fotossintéticos (NETO *et al.*, 2017; YENDREK *et al.*, 2017), celulose e lignina (PETISCO *et al.*, 2006), metabólitos secundários (SERBIN *et al.*, 2012).

Esses estudos podem contribuir para auxiliar diretamente no monitoramento da planta e no processo de tomada de decisão no campo. Diante disso, o presente trabalho teve como objetivo determinar os teores de amido, carboidratos solúveis, sacarose, aminoácidos totais, proteínas, nitrato e nitrato redutase, em folhas de maracujazeiro amarelo por meio da espectroscopia Vis-NIR.

3.2.2 MATERIAL E MÉTODOS

O presente estudo foi dividido em quatro etapas principais: amostragem, análises espectrais, determinação dos compostos bioquímicos, e desenvolvimento dos modelos de regressão, conforme a Figura 9. Após a amostragem, as folhas foram submetidas às análises espectrais para aquisição dos espectros de absorbância, e logo em seguida, às análises de referência, que são as análises convencionais de determinação dos compostos bioquímicos em análise (amido, carboidratos solúveis, sacarose, aminoácidos totais, proteínas, nitrato e nitrato redutase); os dados espectrais passaram por uma série de pré-tratamentos, cujo objetivo foi a remoção de informações irrelevantes, que não contribuem ou diminuem o desempenho dos modelos e após o pré-tratamento dos dados, métodos de regressão foram aplicados, para o desenvolvimento dos modelos de predição.

Figura 9 - Diagrama do processo de desenvolvimento dos modelos de predição para compostos bioquímicos em folhas de maracujazeiro.



Legenda: SNV - Standard Normal Variate; MSC - Multiplicative Scatter Correction; OSC - Orthogonal Signal Correction; PLSR - Partial Least Squares Regression; PCA - Principal Component Analysis; MLR - Multiple Linear Regression.

Amostragem das folhas

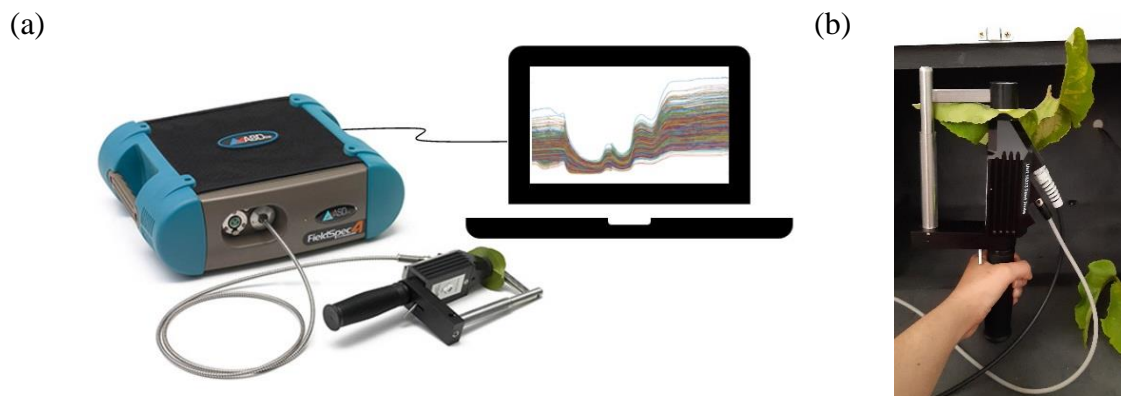
As folhas de maracujazeiro amarelo foram coletadas em uma área experimental da Universidade Federal do Vale do São Francisco, Campus Juazeiro, localizada na região do Submédio do Vale do São Francisco (9.4127° S, 40.5152° W). Essa região é caracterizada pela ocorrência de clima semiárido quente, classificado como BSh, segundo a classificação climática de Köppen (ALVARES *et al.*, 2013).

As amostras utilizadas neste experimento foram compostas por três folhas saudáveis, retiradas de plantas no estado vegetativo. As mudas com idade de 20 dias (25 cm) foram transplantadas em vasos, e as coletas de folhas realizadas 3 meses após o plantio. Imediatamente após a coleta, as amostras foram armazenadas em sacos plásticos devidamente identificados, imersos em caixas térmicas com gelo, transportados e congelados em freezers verticais (-20 °C) para manutenção de sua integridade até o momento da análise. Destaca-se que as amostras foram descongeladas até atingirem a temperatura ambiente de 25 °C, antes das análises espectrais e convencionais.

Aquisição dos espectros

O sistema de aquisição dos espectros de refletância foi constituído dos seguintes componentes: espectrômetro *FieldSpec3* (Analytical Spectral Devices, Boulder, Colorado, USA) que opera na faixa de comprimento de onda de 350 a 2500 nm, com resolução de 3 a 10 nm e tempo por varredura de 100 ms, equipado com *probe plant*; e computador com o software RS3 (Analytical Spectral Devices, Boulder, Colorado, USA), conforme a Figura 10. O sistema foi ligado para aquecimento por 30 minutos antes da aquisição dos espectros e as leituras espectrais foram feitas no limbo foliar do lado.

Figura 10 – Sistema óptico eletrônico constituído por um espectrômetro *FieldSpec3*, equipado com *probe plant*; e computador com o software RS3 em (a). Em (b), momento de análise da folha de mangueira utilizando a *probe plant*.



A *probe plant* possui uma referência branca, com a qual as operações de calibração são implementadas antes da aquisição dos dados. Cada espectro é composto por uma média de 30 varreduras executadas pelo equipamento. Os valores de refletância foram transformados em medições de absorbância ($\log(1/R)$), onde R significa refletância, utilizando o *software ViewSpec Pro* (*Analytical Spectral Devices*, Boulder, Colorado, USA).

Determinação dos compostos bioquímicos

Após as medições de refletância, as amostras de material foliar fresco (MF) foram submetidas às análises bioquímicas de referência para determinação dos teores de amido, carboidratos solúveis, sacarose, aminoácidos totais, proteínas, nitrato e nitrato redutase.

O teor de amido, expresso em mmol g^{-1} de MF, foi determinado segundo Dubois *et al.* (1956), em que foi utilizado material foliar fresco, etanol (80%) e ácido sulfúrico (50%), para a extração. As leituras foram realizadas em um espectrofotômetro a 620 nm.

O teor de carboidratos solúveis, expresso em mg g^{-1} de MF, também foi determinado segundo Dubois *et al.* (1956), em que foi utilizado material foliar fresco, solução tampão fosfato de potássio monobásico (0,1 M e pH 7,0), fenol (5%), e ácido sulfúrico. As leituras foram realizadas em um espectrofotômetro a 490 nm.

O teor de sacarose, expresso em mg g^{-1} de MF, foi determinado segundo Van Handel *et al.* (1968). Para a extração, em que foi utilizado material foliar fresco, solução de MCW (metanol, clorofórmio e água, na proporção 12:5:3), e posteriormente clorofórmio, água deionizada, hidróxido de potássio (30%) e solução de antrona (0,2%) em ácido sulfúrico. As leituras foram realizadas em um espectrofotômetro a 620 nm.

O teor de aminoácidos totais, expresso em $\mu\text{mol g}^{-1}$ de MF, foi determinado Yemm & Cocking (1995), em que foi utilizado material foliar fresco, solução tampão citrato (0,2 M e pH 5,0), cianeto de potássio (KCN), ninhidrina (5%) e etanol (60%). As leituras foram realizadas em um espectrofotômetro a 570 nm.

O teor de proteína, expresso em mg g^{-1} de MF, foi determinado segundo Bradford (1976), em que foi utilizado material foliar fresco, Polivinilpolipirrolidona (PVPP), solução tampão fosfato de potássio (pH 7,5), ácido etilenodiamino tetraacético (EDTA). As leituras foram realizadas em um espectrofotômetro a 280 nm.

O nitrato, expresso em mg g^{-1} de MF, foi determinado segundo Cawse (1967), em que foi utilizado material foliar fresco e ácido perclórico (5%). As leituras foram realizadas em um espectrofotômetro a 210 nm.

A enzima nitrato redutase, valor expresso em $\mu\text{mol g}^{-1}$ de MF, foi determinada segundo Jaworski (1971), em que foi utilizado material foliar fresco, tampão tris (0,1 M e pH 7,5) e cisteína. As leituras foram realizadas em um espectrofotômetro a 540 nm.

Desenvolvimento dos modelos de regressão

Os dados espectrais brutos de cada experimento foram submetidos a uma combinação de diferentes pré-tratamentos matemáticos: filtro de média móvel (*Moving Average*, MA) com tamanho de segmentos de 5, 11, 17, 23, 29 e 35 pontos; primeira e segunda derivada, obtidos pelo método de Savitzky-Golay (GORRY, 1990; SAVITZKY; GOLAY, 1964), com polinômio de segunda ordem e seis segmentos diferentes (5, 11, 17, 23, 29 e 35 pontos). Em seguida, os dados pré-tratados foram corrigidos pelos métodos de correção de dispersão multiplicativa (*Multiplicative Scatter Correction*, MSC) (ISAKSSON; NAES, 1988) e transformação variável normal padrão (*Standard Normal Variate Transformation*, SNV) (BARNES; DHANOA; LISTER, 1989). Por fim, foi aplicado uma correção de sinal ortogonal (*Orthogonal Signal Correction*, OSC) nos dados que foram unicamente filtrados pelo filtro de média móvel.

Assim, foram produzidas 61 matrizes de dados para desenvolvimentos de modelos para cada atributo foliar (amido, carboidratos solúveis, sacarose, aminoácidos totais, proteínas, nitrato e nitrato redutase): matriz bruta (sem pré-tratamentos), seis matrizes com filtros de média móvel, doze matrizes com primeira e segunda derivadas (em seis diferentes tamanhos de segmento), dezoito matrizes com SNV, dezoito com MSC e seis com OSC (apenas matrizes com dados filtrados).

Além de testar diferentes pré-tratamentos, a faixa espectral também foi dividida em três segmentos, formando assim quatro grupos espectrais de teste: FULL, incluindo todos

os comprimentos de onda (350-2500 nm); VNIR – *Visible and Near Infrared* (350-1000 nm); SWIR1 - *First Short Wavelength Infrared* (1001-1830 nm); e SWIR2 - *Second Short Wavelength Infrared* (1831-2500 nm).

A partir das matrizes pré-tratadas foram desenvolvidos os modelos de Regressão por Mínimos Quadrados Parciais (*Partial Least Squares Regression*, PLSR) para cada banda espectral e atributo foliar. Para isso, os dados brutos ou pré-tratados foram utilizados como variáveis de entrada independentes (X) e os atributos de interesse (amido, carboidrato e sacarose) como variáveis de entrada dependentes (Y). Foi aplicada a validação cruzada em todas as amostras dos conjuntos de dados e o número ótimo de fatores foi definido através do menor valor obtido do erro quadrático médio da raiz da validação cruzada (RMSECV). O algoritmo utilizado para a construção dos modelos foi Kernel, com 100 iterações.

Ao final do processamento, foram gerados 1708 modelos diferentes de PLSR. Após processamento, foram selecionados os melhores modelos para cada atributo bioquímico foliar, em suas respectivas faixas espectrais. Posteriormente, foi feita a seleção de variáveis através de uma PCA e exclusão das variáveis com menor influência e não significativas (p -valor $> 0,05$) para os modelos através do método Jack-Knife (MARTENS; MARTENS, 2000). Foram selecionados os comprimentos de ondas associados aos picos de absorbâncias em todas as componentes principais selecionadas, seguido da remoção das variáveis com menores coeficientes de regressão. O modelo foi recalculado a cada variável removida, de modo que, o objetivo foi encontrar modelos com os melhores parâmetros de desempenho e menor número de variáveis e que todas sejam significativas para a construção dos modelos. Os comprimentos de onda selecionados a partir dos picos de absorbâncias e aquelas submetidos à ANOVA e teste de significância a 5% são apresentados no Anexo B.

Em seguida foi aplicado o método MLR em todos os atributos em seus respectivos conjuntos de dados, seguido da reaplicação do método de regressão PLSR, utilizando apenas as variáveis selecionadas, nos melhores conjuntos de dados.

O desempenho dos modelos foi avaliado a partir da comparação dos parâmetros estatísticos, a saber: coeficiente de determinação (R^2); erro quadrático médio de calibração (RMSEC); erro padrão de calibração (SEC); erro quadrático médio de validação cruzada (RMSECV); erro padrão de validação cruzada (SECV). Estes parâmetros são definidos pelas equações de (9) a (13):

$$R^2(\text{Pearson}) = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}) * (y_i - \bar{y})^2}{(n - 1) * \sigma_r * \sigma_p} \quad (9)$$

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (10)$$

$$SEC = \sqrt{\frac{n * RMSEC^2}{n - 1}} \quad (11)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_{i^*})^2}{n}} \quad (12)$$

$$SECV = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i - bias)^2}{n - 1}} \quad (13)$$

Onde \hat{y}_i é o valor previsto pelo modelo de calibração; \hat{y}_{i^*} é o valor previsto pelo modelo na etapa de validação cruzada; y_i é o valor de referência; \bar{y} é a média dos valores de predição; n é o número de amostras na calibração ou validação; m é o número de amostras previstas; σ_r é o desvio-padrão dos valores de referência; σ_p é o desvio-padrão dos valores de predição.

A avaliação do parâmetro SECV foi realizada associando o desvio-padrão (SD) e os valores de amplitude do atributo bioquímico a ser analisado. Deste modo, o desvio preditivo residual (RPD) e a relação de erro de intervalo (RER) foram definidos pelas equações (14) e (15), conforme Williams (2001):

$$RPD = \frac{SD}{SECV} \quad (14)$$

$$RER = \frac{Alcance}{SECV} \quad (15)$$

Os melhores modelos são aqueles que apresentam os maiores valores de coeficiente de determinação da validação cruzada (R^2_{cv}), seguido dos menores valores de erro quadrático médio de validação cruzada (RMSECV) e maiores valores de desvio preditivo residual (RPD).

3.2.3 RESULTADOS

Análise química das folhas

A Tabela 6 apresenta a estatística descritiva para os compostos analisados (amido, carboidrato, sacarose, aminoácidos totais, proteína, nitrato e nitrato redutase) que caracterizam o conjunto de dados, isto é, o número de amostras, média, valores máximos e mínimos, amplitude, desvio-padrão e variância. É possível observar grande variabilidade para o atributo amido em relação aos outros atributos.

Tabela 6 - Estatística descritiva dos teores de amido (mmol g^{-1} de MF), carboidratos solúveis (mg g^{-1} de MF), sacarose (mg g^{-1} de MF), aminoácidos totais ($\mu\text{g g}^{-1}$ de MF), proteína (mg g^{-1} de MF), nitrato (mg g^{-1} de MF) e nitrato redutase ($\mu\text{g g}^{-1}$ de MF) em folhas de maracujazeiro.

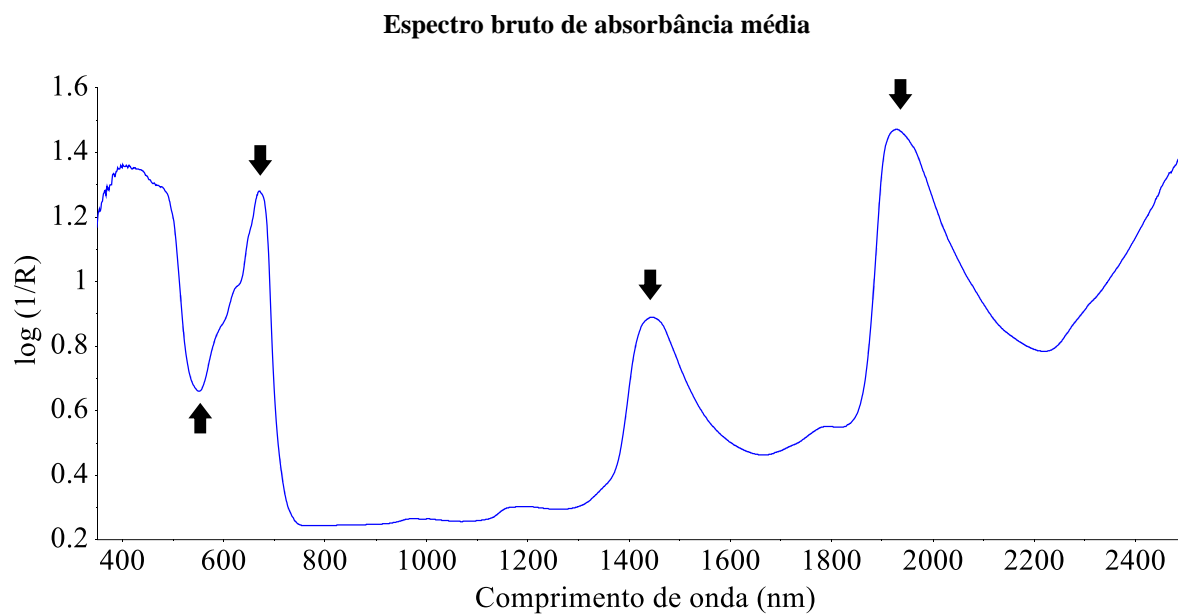
Parâmetro	Amostras	Média	Mín.	Máx.	Amplitude	SD	Variância
Amido		138,38	72,83	198,71	125,88	35,46	1257,67
CHOsol		10,32	5,73	11,88	6,15	1,47	2,16
Sacarose		263,12	249,82	267,12	17,30	2,66	7,10
AA totais	89	35,86	16,31	51,29	34,98	8,37	70,12
Proteína		29,25	3,10	46,42	43,32	10,69	114,24
Nitrato		17,58	12,91	23,03	10,12	2,49	6,22
Nitrato redutase		12,82	1,35	26,67	25,32	5,11	26,07

Legenda: MF – material foliar fresco; CHOsol – Carboidratos solúveis; AA totais – aminoácidos totais; Mín. – valor mínimo; Máx. – valor máximo; SD – desvio-padrão.

Análise exploratória do espectro

O espectro médio bruto de absorbância das folhas de maracujazeiro é mostrado na Figura 11. Esses espectros apresentam quatro principais pontos de absorção nas regiões, sendo eles uma faixa estável de absorção em torno de 460 nm, além dos picos próximos dos comprimentos de 670, 1450 e 1930 nm, como mostram os espectros médios dos dados brutos.

Figura 11 - Espectro bruto de absorvância média das amostras foliares de maracujazeiro amarelo.



Modelos de regressão PLSR

Foram desenvolvidos modelos PLSR para o conjunto de dados com folhas de maracujazeiro para os atributos amido, carboidrato, sacarose, aminoácidos totais, proteína, nitrato e nitrato redutase. A Tabela 7 apresenta os melhores resultados para cada atributo. Os melhores modelos de predição foram os desenvolvidos para os atributos amido, e proteína. Além disso, todos os modelos desenvolvidos que obtiveram melhor desempenho foram com os pré-tratamentos *Moving Average* e OSC.

Tabela 7 - Parâmetros estatísticos dos melhores modelos PLSR para predição de amido, carboidratos solúveis, sacarose, aminoácidos totais, proteína, nitrato e nitrato redutase em folhas de maracujazeiro.

Parâmetro	BoW	Pré-tratamento	F	calibração			validação			RPD
				R^2_{cal}	RMSEC	SEC	R^2_{cv}	RMSECV	SECV	
Amido	SWIR2	MA 17 pts, OSC	15	0,99	4,00	4,03	0,83	15,01	15,09	2,35
CHOsol	FULL	MA 5 pts, OSC	15	0,88	0,50	0,50	0,39	1,20	1,21	1,21
Sacarose	SWIR1	MA 11 pts, OSC	1	0,45	1,96	1,97	0,41	2,03	2,04	1,30
AA totais	SWIR1	MA 5 pts, OSC	8	0,97	1,38	1,38	0,70	4,63	4,66	1,80
Proteína	VNIR	MA 23 pts, OSC	15	0,98	1,63	1,64	0,82	4,57	4,60	2,32
Nitrato	SWIR1	MA 5 pts, OSC	8	0,98	0,37	0,37	0,62	1,54	1,55	1,61
N. redutase	VNIR	MA 23 pts, OSC	6	0,88	1,77	1,78	0,69	2,87	2,88	1,77

Legenda: PLSR – *Partial Least Squares Regression*; CHOsol – Carboidratos solúveis; AA totais – aminoácidos totais; BoW – banda de comprimento de onda; VNIR – banda espectral dos comprimentos de onda no intervalo de 350 a 1000 nm; FULL – espectro completo, que abrange os comprimentos de onda no intervalo de 350 a 2500 nm; MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*; F – fator; R^2_{cal} – coeficiente de determinação da calibração; RMSEC – erro quadrático médio de calibração; SEC – erro padrão de calibração; R^2_{cv} – coeficiente de determinação da validação cruzada; RMSECV – erro quadrático médio de validação cruzada; SECV – erro padrão de validação cruzada; RPD – desvio preditivo residual.

Para o atributo amido, a banda espectral que proporcionou os melhores modelos foi SWIR2 (1831 – 2500 nm), com o pré-tratamento *Moving Average* (17 pontos) e OSC. Esse obteve R^2_{cv} de 0,83 e RMSECV de 15,01, respectivamente. Os valores de RPD e RER foram iguais a 2,35 e 8,34, respectivamente.

Para o atributo carboidrato, a banda espectral que proporcionou os melhores modelos foi a banda FULL (espectro completo), com o pré-tratamento *Moving Average* (5 pontos) e OSC. Esse obteve R^2_{cv} de 0,39 e RMSECV de 1,20, respectivamente. Os valores de RPD e RER foram iguais a 1,21 e 5,08, respectivamente.

Para o atributo sacarose, a banda espectral que proporcionou os melhores modelos foi a banda SWIR1, com o pré-tratamento *Moving Average* (11 pontos) e OSC. Esse obteve R^2_{cv} de 0,41 e RMSECV de 2,03, respectivamente. Os valores de RPD e RER foram iguais a 1,30 e 8,47, respectivamente.

Para o atributo aminoácidos totais, a banda espectral que proporcionou os melhores modelos foi a banda SWIR1, com o pré-tratamento *Moving Average* (5 pontos) e OSC. Esse obteve R^2_{cv} de 0,70 e RMSECV de 4,63, respectivamente. Os valores de RPD e RER foram iguais a 1,80 e 7,51, respectivamente.

Para o atributo proteína, a banda espectral que proporcionou os melhores modelos foi a banda VNIR, com o pré-tratamento *Moving Average* (23 pontos) e OSC. Esse obteve

R^2_{cv} de 0,82 e RMSECV de 4,57, respectivamente. Os valores de RPD e RER foram iguais a 2,32 e 9,42, respectivamente.

Para o atributo nitrato, a banda espectral que proporcionou os melhores modelos foi a banda SWIR1, com o pré-tratamento *Moving Average* (5 pontos) e OSC. Esse obteve R^2_{cv} de 0,62 e RMSECV de 1,54, respectivamente. Os valores de RPD e RER foram iguais a 1,61 e 6,53, respectivamente.

Para o atributo nitrato redutase, a banda espectral que proporcionou os melhores modelos foi a banda VNIR, com o pré-tratamento *Moving Average* (23 pontos) e OSC. Esse obteve R^2_{cv} de 0,69 e RMSECV de 2,87, respectivamente. Os valores de RPD e RER foram iguais a 1,77 e 8,79, respectivamente.

Modelos de regressão com seleção de variáveis

O primeiro método de regressão aplicado foi o MLR, devido à simplicidade de sua modelagem e fácil interpretação dos resultados. Entretanto, em virtude de problemas de colinearidade, foi necessário fazer uma seleção de variáveis nos dados de entrada para aplicação do método MLR. A partir da seleção de variáveis pelo PCA e exclusão das variáveis com menor influência para o modelo através do método Jack-Knife feito anteriormente, foram gerados os modelos MLR para todos os atributos. A Tabela 8 apresenta os parâmetros de desempenho do modelo MLR para os diferentes atributos, bem como o número de variáveis selecionadas.

Tabela 8 - Parâmetros estatísticos dos melhores modelos MLR com variáveis selecionadas para predição de amido, carboidratos solúveis, sacarose, aminoácidos totais, proteína, nitrato e nitrato redutase em folhas de maracujazeiro.

Parâmetro	BoW	Pré-tratamento	Var	calibração			validação			RPD
				R ² _{cal}	RMSEC	SEC	R ² _{cv}	RMSECV	SECV	
Amido	SWIR2	MA 17 pts, OSC	7	0,68	20,85	20,01	0,62	22,03	22,08	1,61
CHOsol	FULL	MA 5 pts, OSC	8	0,36	1,24	1,18	0,22	1,32	1,32	1,12
Sacarose	SWIR1	MA 11 pts, OSC	66	1,00	0,16	0,08	0,98	0,37	0,37	7,29
AA totais	SWIR1	MA 5 pts, OSC	70	0,99	0,33	0,15	0,99	0,88	0,88	9,52
Proteína	VNIR	MA 23 pts, OSC	41	0,96	2,86	2,09	0,86	4,08	4,08	2,62
Nitrato	SWIR1	MA 5 pts, OSC	66	0,99	0,09	0,05	0,99	0,19	0,19	13,18
N. redutase	VNIR	MA 23 pts, OSC	40	0,95	1,48	1,09	0,84	2,06	2,06	2,48

Legenda: MLR – *Multiple Linear Regression*; CHOsol – Carboidratos solúveis; AA totais – aminoácidos totais; BoW – banda de comprimento de onda; VNIR – banda espectral dos comprimentos de onda no intervalo de 350 a 1000 nm; FULL – espectro completo, que abrange os comprimentos de onda no intervalo de 350 a 2500 nm; MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*; Var – número de variáveis selecionadas (comprimento de onda) com PCA F – fator; R²_{cal} – coeficiente de determinação da calibração; RMSEC – erro quadrático médio de calibração; SEC – erro padrão de calibração; R²_{cv} – coeficiente de determinação da validação cruzada; RMSECV – erro quadrático médio de validação cruzada; SECV – erro padrão de validação cruzada; RPD – desvio preditivo residual.

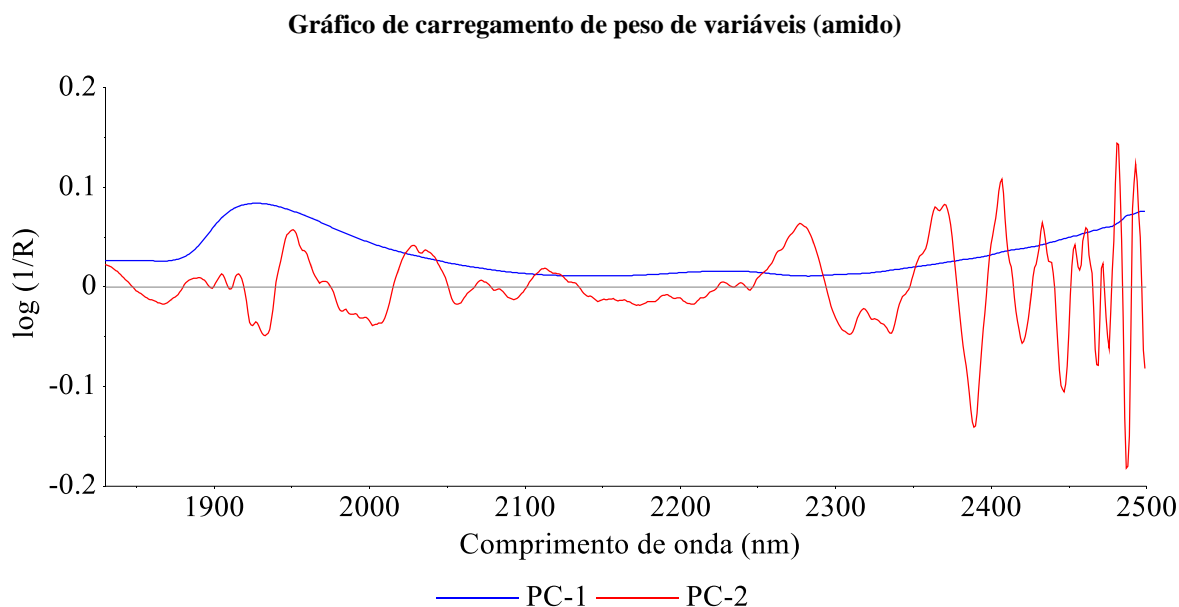
Com exceção dos modelos desenvolvidos para os atributos amido (R²_{cv} = 0,62) e carboidrato (R²_{cv} = 0,22), todos os modelos MLR gerados obtiveram desempenho superior aos modelos PLSR, devido à seleção dos componentes principais através da PCA e seleção de variáveis.

Os modelos gerados para sacarose, aminoácidos totais e nitrato obtiveram desempenho excelente com R²_{cv} iguais a 0,98, 0,99 e 0,99 e RPD iguais a 7,29, 9,52 e 13,18, respectivamente. Os modelos gerados para os atributos proteína e nitrato redutase também apresentaram bom desempenho, com R²_{cv} iguais a 0,86 e 0,84, e RPD iguais a 2,62 e 2,48, respectivamente. Os modelos com piores desempenhos foram gerados para os atributos amido e carboidrato, com valores de R²_{cv} iguais a 0,67 e 0,22, e RPD iguais a 1,68 e 1,12, respectivamente.

Os gráficos na Figura 12 representam o peso das variáveis independentes para cada atributo e suas respectivas componentes principais. A partir deles, é possível visualizar as evidências da possível associação entre os comprimentos de onda e os atributos de interesse.

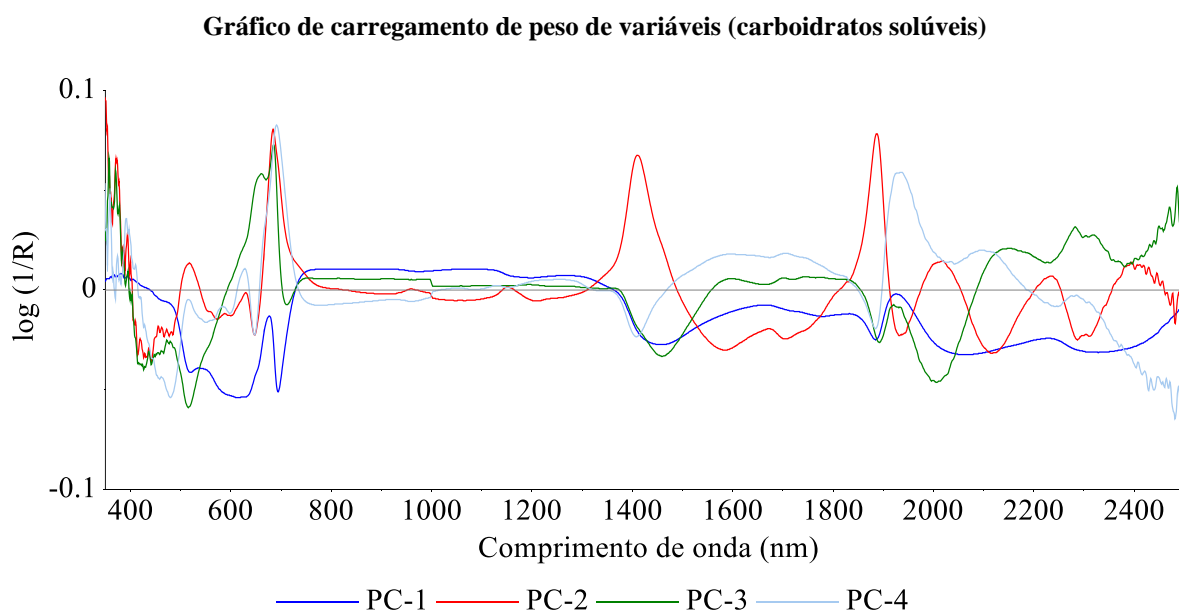
Figura 12 - Gráficos de carregamento de peso das variáveis independentes: (a) amido com pré-tratamento MA 17 pts, OSC; (b) carboidratos solúveis com pré-tratamento MA 5 pts, OSC; (c) sacarose com pré-tratamento MA 11 pts, OSC; (d) aminoácidos totais com pré-tratamento MA 5 pts, OSC; (e) proteína com pré-tratamento MA 23 pts, OSC; (f) nitrato com pré-tratamento MA 5 pts, OSC; (g) nitrato redutase com pré-tratamento MA 23 pts, OSC.

(a)



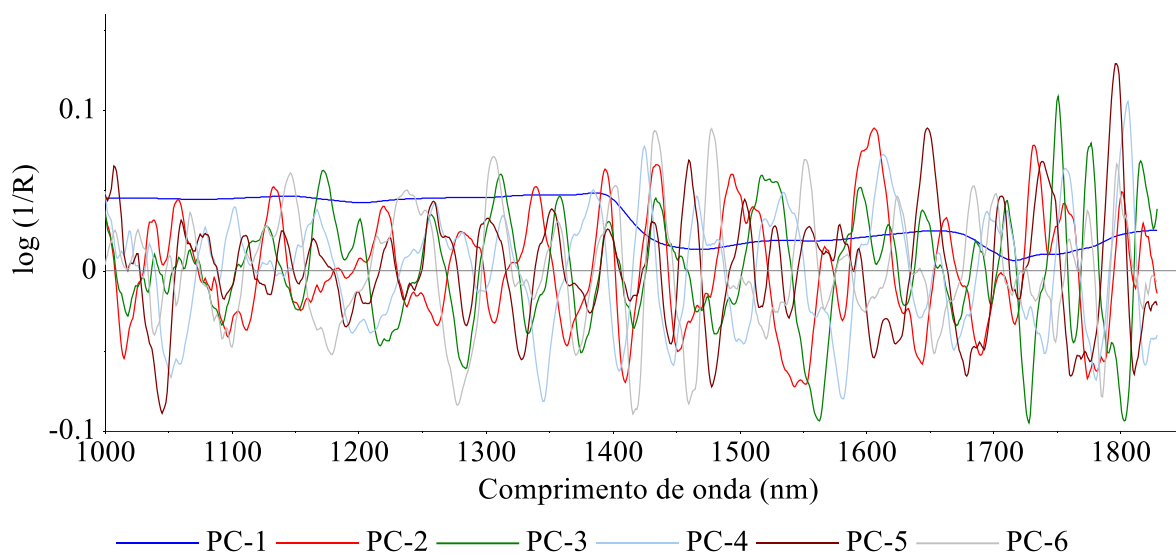
Legenda: PC – *Principal component* (componente principal); MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

(b)



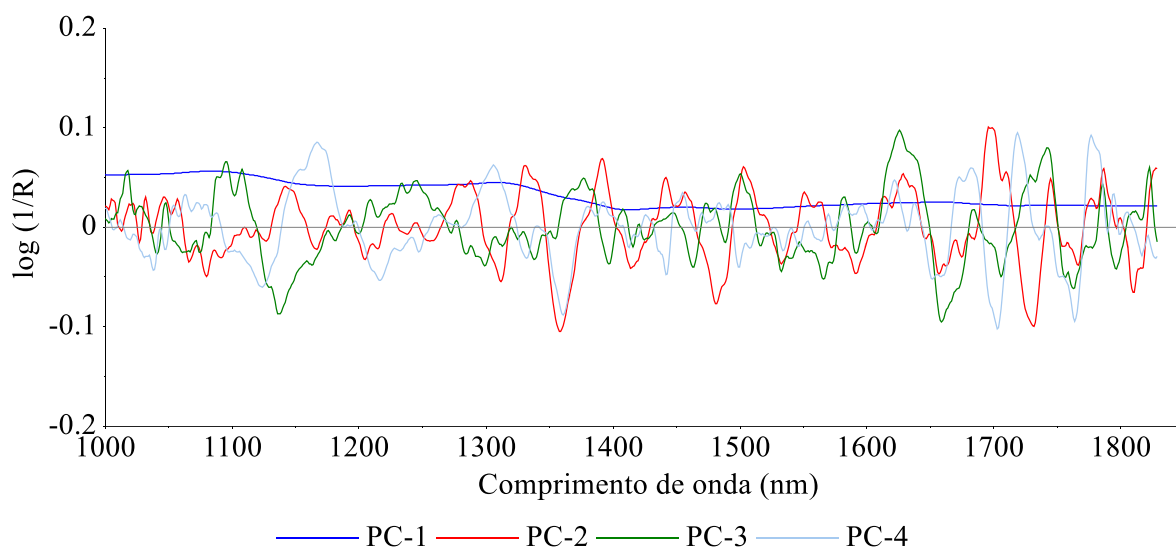
Legenda: PC – *Principal component* (componente principal); MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

(c)

Gráfico de carregamento de peso de variáveis (sacarose)

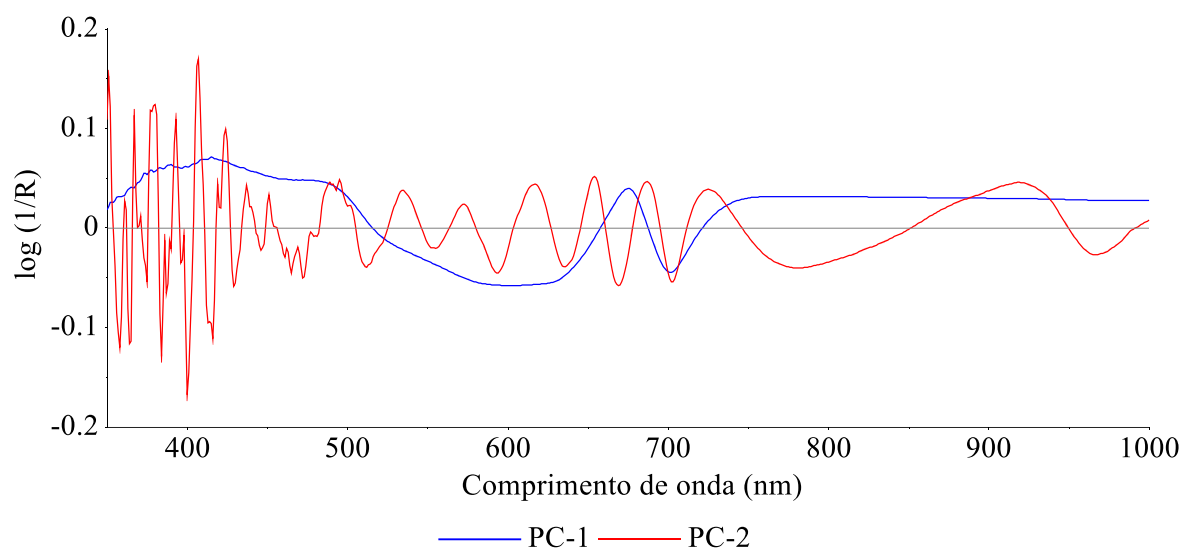
Legenda: PC – *Principal component* (componente principal); MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

(d)

Gráfico de carregamento de peso de variáveis (aminoácidos totais)

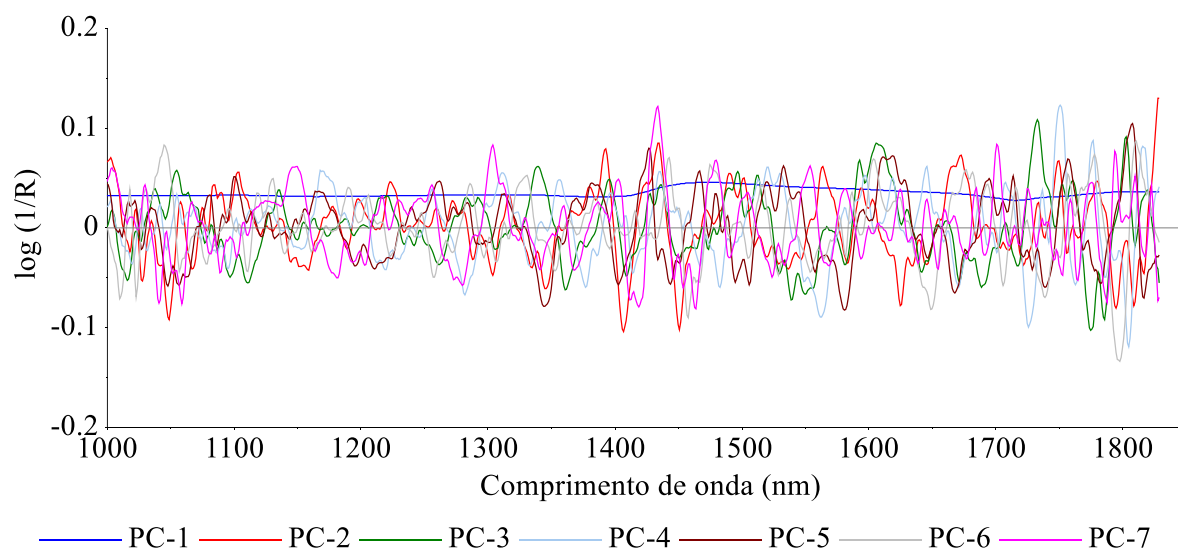
Legenda: PC – *Principal component* (componente principal); MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

(e)

Gráfico de carregamento de peso de variáveis (proteína)

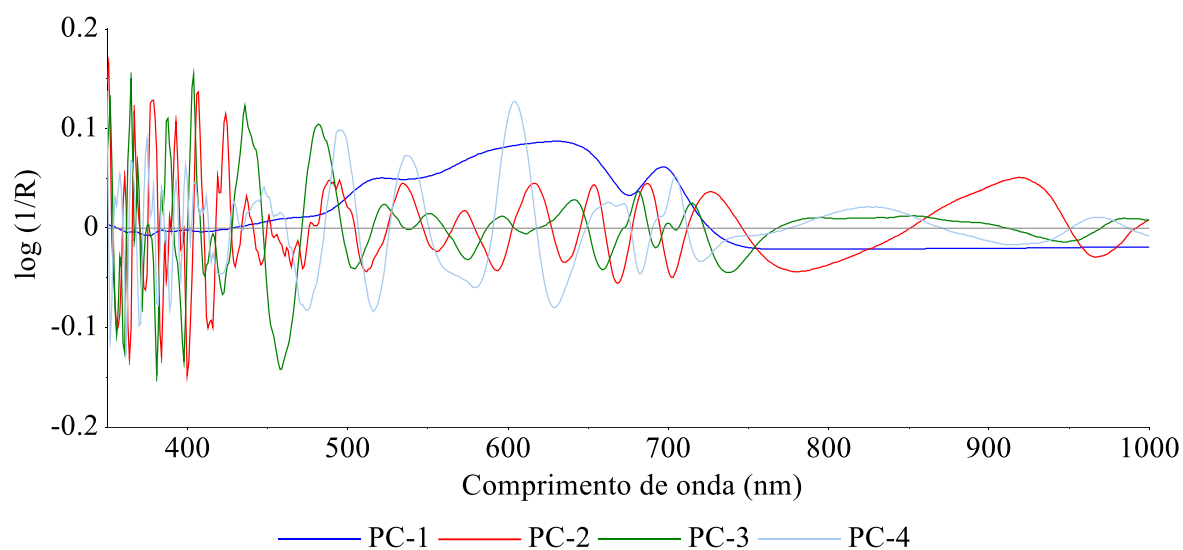
Legenda: PC – *Principal component* (componente principal); MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

(f)

Gráfico de carregamento de peso de variáveis (nitrato)

Legenda: PC – *Principal component* (componente principal); MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

(g)

Gráfico de carregamento de peso de variáveis (nitrato redutase)

Legenda: PC – *Principal component* (componente principal); MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

Os comprimentos de onda pré-selecionados, devido a maior correlação com os atributos analisados no estudo, bem como as assinaturas espectrais, ou seja, as variáveis independentes (comprimentos de onda) pré-selecionados, submetidos à ANOVA e teste de significância ($p\text{-valor} > 0,05$), são apresentados no Anexo B.

A partir da determinação das assinaturas espectrais, o método de regressão PLSR foi aplicado novamente, desta vez, utilizando apenas os comprimentos de ondas selecionados como variáveis independentes, cujos parâmetros de desempenho são apresentados na Tabela 9.

Tabela 9 - Parâmetros estatísticos dos melhores modelos PLSR com variáveis selecionadas para predição de amido, carboidrato, sacarose, aminoácidos totais, proteína, nitrato e nitrato redutase em folhas de maracujazeiro.

Parâmetro	BoW	Pré-tratamento	Var	F	calibração			validação			RPD
					R^2_{cal}	RMSEC	SEC	R^2_{cv}	RMSECV	SECV	
Amido	SWIR2	MA 17 pts, OSC	7	4	0,68	19,98	20,09	0,62	21,85	21,97	1,61
CHOsol	FULL	MA 5 pts, OSC	8	6	0,35	1,18	1,18	0,22	1,31	1,32	1,12
Sacarose	SWIR1	MA 11 pts, OSC	66	1	0,452	1,961	1,972	0,41	2,03	2,04	1,30
AA totais	SWIR1	MA 5 pts, OSC	70	1	0,63	5,04	5,07	0,62	5,15	5,18	1,62
Proteína	VNIR	MA 23 pts, OSC	41	1	0,62	6,54	6,57	0,60	6,69	6,72	1,59
Nitrato	SWIR1	MA 5 pts, OSC	66	1	0,36	1,99	2,00	0,33	2,03	2,04	1,22
N. redutase	VNIR	MA 23 pts, OSC	40	3	0,681	2,865	2,88	0,56	3,40	3,42	1,49

Legenda: PLSR – *Partial Least Squares Regression*; CHOsol – Carboidratos solúveis; AA totais – aminoácidos totais; BoW – banda de comprimento de onda; VNIR – banda espectral dos comprimentos de onda no intervalo de 350 a 1000 nm; FULL – espectro completo, que abrange os comprimentos de onda no intervalo de 350 a 2500 nm; MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*; Var – número de variáveis selecionadas (comprimentos de onda); F – fator; R^2_{cal} – coeficiente de determinação da calibração; R^2_{cv} – coeficiente de determinação da validação cruzada; RMSEC – erro quadrático médio de calibração; SEC – erro padrão de calibração; R^2_{cv} – coeficiente de determinação da validação cruzada; RMSECV – erro quadrático médio de validação cruzada; SECV – erro padrão de validação cruzada; RPD – desvio preditivo residual.

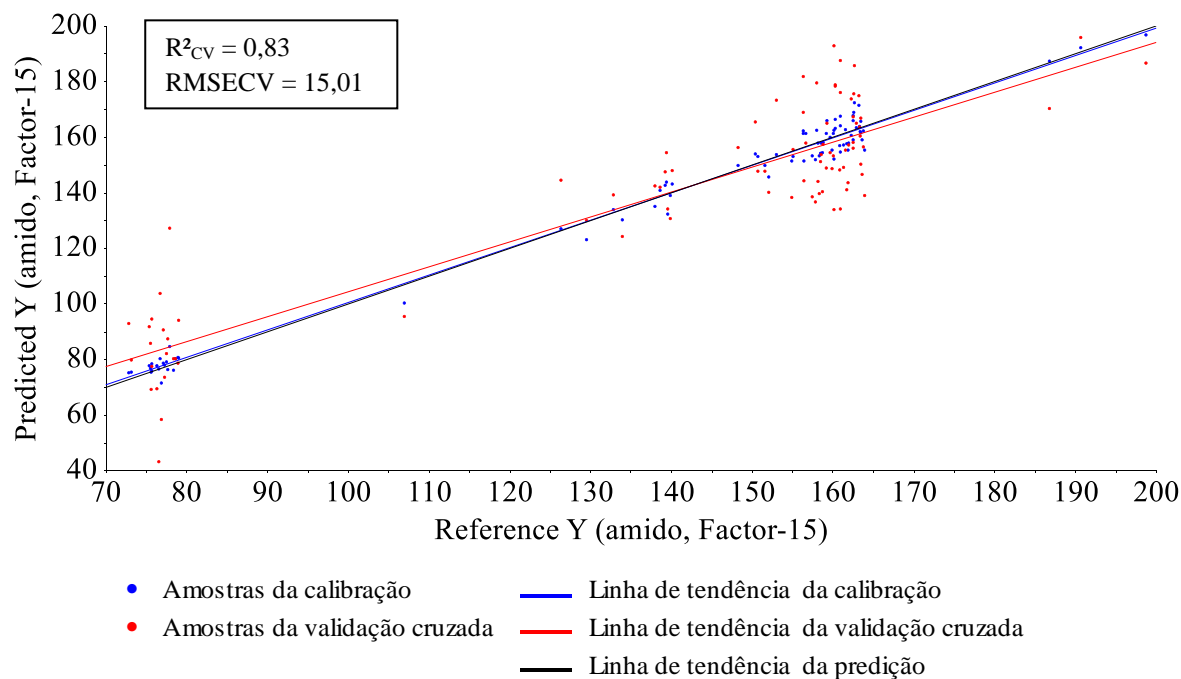
Para os atributos amido e carboidratos o desempenho foi semelhante entre os modelos desenvolvidos com seleção de variáveis, com R^2_{cv} iguais a 0,62 e 0,22, respectivamente. Entretanto, de modo geral, a seleção de variáveis em modelos PLSR não apresentou melhora no desempenho em relação aos modelos PLSR com espectro completo e MLR. Além disso, nenhum modelo obteve R^2_{cv} superior a 0,70.

Deste modo, os melhores modelos desenvolvidos foram PLSR com espectro completo para os atributos amido ($R^2_{cv} = 0,83$) e carboidratos ($R^2_{cv} = 0,30$), e modelos MLR para os atributos sacarose ($R^2_{cv} = 0,98$), aminoácidos totais ($R^2_{cv} = 0,99$), proteína ($R^2_{cv} = 0,86$), nitrato ($R^2_{cv} = 0,99$) e nitrato redutase ($R^2_{cv} = 0,84$). Os gráficos de dispersão de calibração e validação cruzada dos melhores modelos desenvolvidos são mostrados na Figura 13.

Figura 13 - Gráficos de dispersão de calibração e validação cruzada dos melhores modelos de predição: (a) amido com pré-tratamento MA 17 pts, OSC (MLR); (b) carboidratos solúveis com pré-tratamento MA 5 pts (MLR), OSC; (c) sacarose com pré-tratamento MA 11 pts, OSC (PLSR); (d) aminoácidos totais com pré-tratamento MA 5 pts, OSC (PLSR); (e) proteína com pré-tratamento MA 23 pts, OSC (PLSR); (f) nitrato com pré-tratamento MA 5 pts, OSC (PLSR); (g) nitrato redutase com pré-tratamento MA 23 pts, OSC (PLSR).

(a)

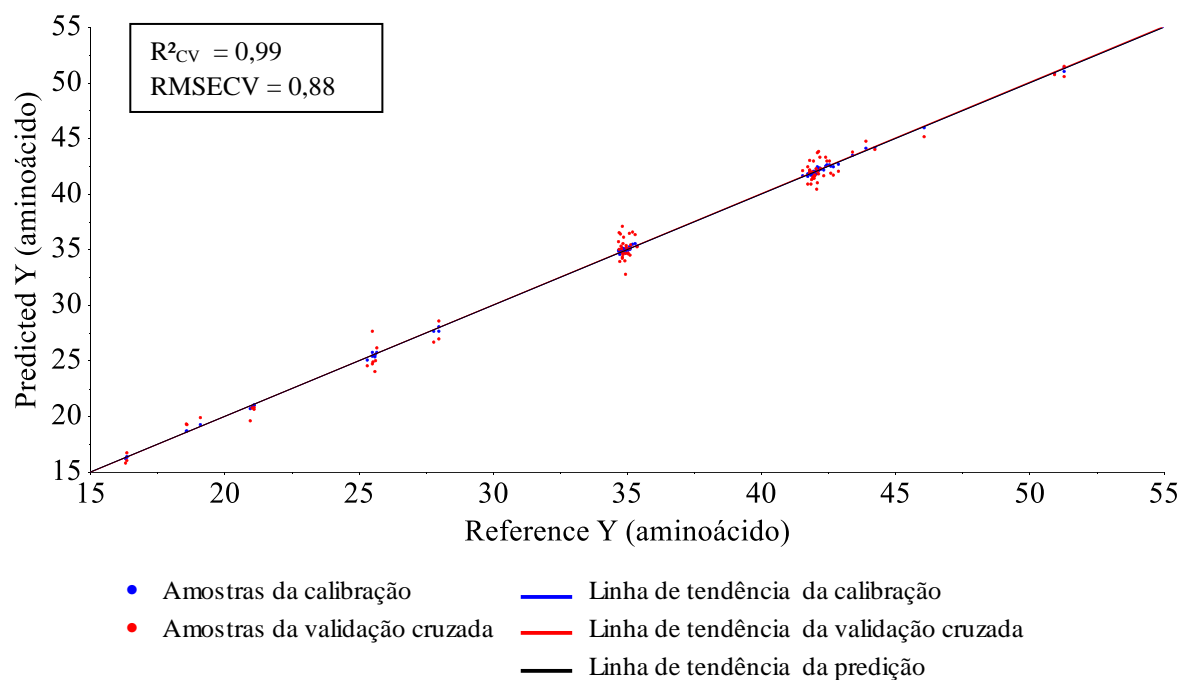
Gráficos de dispersão de calibração e validação cruzada (amido)



Legenda: R^2_{cv} – coeficiente de determinação da validação cruzada; RMSECV – erro quadrático médio de validação cruzada MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

(d)

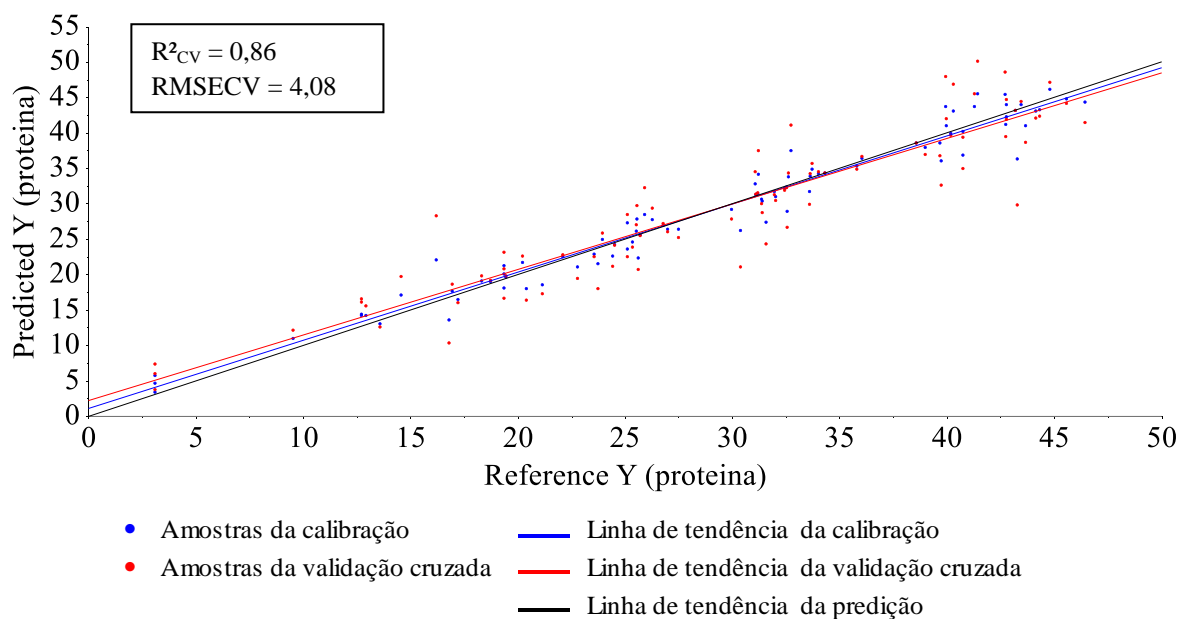
Gráficos de dispersão de calibração e validação cruzada (aminoácidos totais)



Legenda: R^2_{cv} – coeficiente de determinação da validação cruzada; RMSECV – erro quadrático médio de validação cruzada MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

(e)

Gráficos de dispersão de calibração e validação cruzada (proteína)



Legenda: R^2_{cv} – coeficiente de determinação da validação cruzada; RMSECV – erro quadrático médio de validação cruzada MA – *Moving Average* (filtro da média móvel) com diferentes tamanhos de segmento (5, 11, 17, 23, 29 ou 35 pontos); OSC – *Orthogonal Signal Correction*.

A partir das assinaturas espectrais identificadas nas bandas espectrais analisadas e dos melhores modelos preditivos desenvolvidos, é possível fazer a predição dos teores de amido, carboidrato e sacarose sem a necessidade de novas análises laboratoriais, considerando a precisão obtida para cada modelo.

3.2.4 DISCUSSÃO

Análise química das folhas

A maior variabilidade para amido pode ser destacada em relação aos outros atributos analisados. Segundo Souza *et al.* (2013), o acúmulo de açúcares nas folhas é diretamente influenciado pela fase fenológica das plantas, de modo que plantas em estado vegetativo tendem a apresentar maiores teores. Entretanto, essa variação pode ser atribuída a algum estresse sofrido por algumas plantas como, por exemplo, estresse hídrico ocasionado por irrigação desuniforme. Plantas com estresse hídrico tendem a diminuir suas reservas de amido nas folhas para manutenção das atividades metabólicas, em função da redução da atividade fotossintética ocasionada por algum estresse (EARLES *et al.*, 2018).

Análise exploratória do espectro

Segundo Ponzoni (2002), o comportamento espectral de uma folha é função de sua composição, morfologia e estrutura interna e apresenta diferenças desde que sejam de grupos geneticamente diferentes. Entretanto, esse comportamento no espectro Vis-NIR (350 - 2500 nm) é notavelmente semelhante entre todas as folhas de plantas, principalmente pela interação da luz com a clorofila e com as ligações moleculares O-H, C-H e N-H (CURRAN, 1989).

Os primeiros picos de absorvância, em torno de 460 e 670 nm podem estar associados à presença das clorofilas. Segundo Gould *et al.* (2018), a clorofila “a” apresenta absorção nos comprimentos de onda de 450 nm e 680 nm, a clorofila “b” de 475 nm e 675 nm, semelhante aos resultados do presente trabalho, exposto na Figura 11.

Novo (2010) propôs uma curva de refletância típica de uma folha verde sadia: na região do visível há uma alta absorvância de luz (em torno dos comprimentos 460 e 670 nm) devido às clorofilas, no infravermelho próximo a absorção diminui, resultado da interação da energia com o mesófilo, de modo que, quanto mais lacunosa for a estrutura interna foliar, menor a absorvância, e no infravermelho médio a absorvância volta a aumentar devido ao conteúdo de água na folha.

Considerando a curva de refletância proposta por Novo (2010), comprimentos de onda em torno de 1450 e 1930 nm estão correlacionados ao conteúdo de água. Entretanto, segundo De Bei *et al.* (2017) esses comprimentos podem estar associados a absorbância pelos teores de amido e açúcares nas folhas.

Modelos de regressão PLSR

Modelos PLSR com coeficiente de determinação superior a 0,9 ou mais próximo de 1 são considerados excelentes, pois indicam que a proporção da variabilidade dos dados de saída pode ser explicada pela variabilidade das variáveis de entrada (SCHOBER *et al.*, 2018; SHENK; WESTERHAUS, 1996). Entretanto, os melhores modelos PLSR gerados, para amido e proteína, apresentaram R^2_{cv} inferior a 0,9, com valores iguais a 0,83 e 0,82, com valores de RPD iguais a 2,35 e 2,32, respectivamente.

Segundo Williams e Sobering (1996), um bom modelo de predição deve apresentar valor de RPD próximo a três e RER igual ou superior a dez. Os melhores modelos, amido e proteína, apresentaram valores inferiores a três. De acordo com o RPD obtido, é possível analisar a estimativa de predição de um modelo: entre 3,1 e 4,9, o modelo é adequado para triagem; entre 5,0 e 6,4 para controle de qualidade; entre 6,5 e 8,0 para controle do processo; e por fim, acima de 8,0 é considerado adequado para qualquer aplicação (WILLIAMS, 2001). Deste modo, os modelos desenvolvidos para os compostos bioquímicos não apresentam os valores mínimos para as aplicações citadas por Williams (2001).

Entretanto, Li *et al.* (2015), consideram que, em aplicações agrícolas específicas, os valores de RPD classificam os modelos como: adequados para triagens preliminares e previsões iniciais, para valores superiores a 1,5 (SCHIMLECK *et al.*, 2001); satisfatório para a previsão, com valores entre 2,0 e 2,5 (KARLINASARI *et al.*, 2014); adequados para predição, com valores superiores a 3,0 (MOUAZEN *et al.*, 2005). Diante disso, os modelos para amido e proteína estariam classificados entre satisfatórios para predição, e os atributos aminoácidos totais, nitrato e nitrato redutase considerados adequados para triagens preliminares e previsões iniciais.

Em relação aos teores de amido, De Bei *et al.* (2017) e Ely *et al.* (2019) obtiveram modelos PLSR com desempenho semelhante aos modelos para folhas de maracujazeiro desse estudo ($R^2_{cv} = 0,83$): De Bei *et al.* (2017), em folhas de videira, com R^2_{cv} igual a 0,87, e Ely *et al.* (2019) em folhas de sete diferentes espécies vegetais (tomateiro, pepino, abobrinha, soja, manjeriço, feijão e girassol), com R^2 igual a 0,80. Burnett *et al.* (2021), em estudos com folha de abobrinha, obtiveram modelo com desempenho superior, com R^2_{cv} igual a 0,93.

Entretanto, embora tenham utilizado mais de 500 amostras, enquanto os modelos citados anteriormente utilizaram menos que 200 amostras, Lohr *et al.* (2017) e Frey *et al.* (2020) obtiveram modelo com desempenho inferior, com folhas de plantas ornamentais (crisântemo e pelargônio) e trevo vermelho, R^2_{cv} de 0,77 e 0,56, respectivamente.

Em relação aos carboidratos solúveis, o desempenho do modelo desenvolvido para este trabalho obteve baixo desempenho, com R^2_{cv} de 0,39. Diferente de outros estudos, com R^2_{cv} de 0,86 e 0,83, em folhas de videira e plantas ornamentais, nos trabalhos de De Bei *et al.* (2017) e de Lohr *et al.* (2017), respectivamente. Uma possível razão para o melhor desempenho é o número maior de amostra em comparação ao presente trabalho. Além disso, os autores utilizaram o pré-tratamento SNV para esse modelo de melhor desempenho, pré-tratamento este que proporcionou modelo inferior aos modelos pré-tratados com OSC no presente trabalho.

Quanto aos modelos PLSR para predição de sacarose desse estudo, assim como para carboidratos, o R^2_{cv} foi inferior a 0,5. Yendrek *et al.* (2017) conseguiram obter resultado superior, R^2_{cv} de 0,62 em folha de milho. Entretanto, desempenhos superiores foram encontrados em modelos PLSR desenvolvidos por Das *et al.* (2018) e Ely *et al.* (2019), e com R^2_{cv} de 0,78 em folhas de arroz e R^2_{cv} de 0,76 em folhas de diferentes vegetais (tomateiro, pepino, abobrinha, soja, manjericão, feijão e girassol).

Em relação aos aminoácidos totais, o desempenho do modelo desenvolvido para este trabalho obteve R^2_{cv} de 0,70, superior aos modelos desenvolvidos por Burnett *et al.* (2021) e Ely *et al.* (2019), com R^2_{cv} iguais a 0,53 e 0,58. O primeiro fator que pode ter contribuído para o desempenho inferior, foi a não utilização de pré-tratamentos dos dados. Além disso, as faixas espectrais utilizadas também foram diferentes, com intervalos de 350 - 2500 nm e 500 - 2500 nm, enquanto o presente trabalho obteve seu melhor desempenho utilizando apenas a faixa espectral SWIR1 (1001 - 1830 nm), que teoricamente é mais relevante para aminoácidos totais, além do uso de pré-tratamentos.

Situação semelhante ocorreu com os modelos para proteínas, Burnett *et al.* (2021) utilizando a mesma faixa espectral para aminoácidos totais e proteína, bem como a não utilização de pré-tratamento, obteve R^2_{cv} de 0,73, enquanto o presente trabalho obteve desempenho superior, com R^2_{cv} de 0,82, utilizando o espectro VNIR (350 - 1000 nm) com pré-tratamentos de dados.

Quanto aos modelos PLSR para predição de nitrato, nesse estudo foi obtido R^2_{cv} de 0,62. Itoh *et al.* (2011) obtiveram R^2_{cv} iguais a 0,83 e 0,82, utilizando os pré-tratamentos filtro de média e SNV, em folhas de espinafre e couve japonesa, respectivamente. Em relação

ao modelo PLSR para determinação de nitrato redutase em folhas vegetais, não há estudos semelhantes para comparação de desempenho.

Independente do atributo analisado, o desempenho dos modelos depende, além dos métodos de regressão e pré-tratamentos utilizados, também do tipo de amostra, da espécie vegetal em estudo (no caso de amostras foliares), objetivo da análise, tamanho do conjunto amostral, as suposições inerentes à regressão, como linearidade e normalidade. Ou seja, um modelo, seja PLSR ou outro, pode apresentar diferentes desempenhos, mesmo considerando a utilização das mesmas técnicas matemáticas, devido à influência dos diversos fatores (HAIR, 2005).

Modelos de regressão com seleção de variáveis

A utilização do método MLR para desenvolvimento de modelos preditivos é interessante devido à simplicidade de sua modelagem e fácil interpretação dos resultados, principalmente em tentativas preliminares. Esse método pode ser empregado quando o número de amostras é maior que o número de parâmetros a serem analisados do que o número de variáveis (SAEYS *et al.*, 2019).

Entretanto, normalmente não se utiliza método MLR para determinação quantitativa de um composto em amostras, devido à complexidade do processamento de dados e utilização de muitas variáveis (comprimentos de onda), sendo mais recomendado o método PLSR, inclusive por sua capacidade de resolver problemas de colinearidade e ruídos (Jia *et al.*, 2010).

Contudo, é possível corrigir esse problema com a utilização de ferramentas específicas, em destaque a seleção de variáveis. Esse processo é de extrema importância para melhorar a relação entre o desempenho do modelo e custo computacional, aumentar a previsão das variáveis selecionadas e melhorar a compreensão e interpretação dos modelos obtidos (YUN *et al.*, 2019). Diante disso, ainda que os desempenhos dos modelos MLR e modelos como PLSR, por exemplo, sejam semelhantes, o MLR se torna mais vantajoso, pois requer menor custo computacional de processamento.

No presente estudo, nos mesmos dados em que foram aplicados o método PLSR também foram aplicados o método MLR com seleção de variáveis. Com exceção dos modelos PLSR desenvolvidos para os atributos amido ($R^2_{cv} = 0,67$) e carboidrato ($R^2_{cv} = 0,22$) e todos os modelos PLSR com seleção de variáveis, todos os modelos MLR gerados obtiveram desempenho superior aos modelos PLSR, devido à seleção dos componentes principais através da PCA e seleção de variáveis.

Em relação às assinaturas identificadas, aquelas que se encontram na região do visível, com comprimentos de onda entre 380 e 780 nm, estão associadas à presença dos pigmentos, em especial às clorofilas “a” e “b”, cujos picos de absorvância acontecem nessa região espectral (GOULD *et al.*, 2018).

As assinaturas espectrais identificadas a partir do início da região do infravermelho próximo (780 – 2500 nm) até o comprimento em torno de 1400 nm estão relacionadas à própria estrutura das células que compõe o mesófilo. A partir de 1400 a 2500 nm, normalmente os comprimentos de onda estão relacionados ao conteúdo de água na folha (NOVO, 2010; ABRANTES, 2019, JENSEN, 2009). Entretanto, segundo Dei Bei *et al.* (2017) sugerem que essas assinaturas podem estar associadas aos teores de amido e açúcares nas folhas.

Quanto aos modelos PLSR com seleção de variáveis, não houve melhora no desempenho em relação aos modelos PLSR sem seleção de variáveis, semelhante aos resultados encontrados por Frey *et al.* (2020), que obteve R^2_{cv} de 0,56 em modelos PLSR com todos os comprimentos de onda e R^2_{cv} de 0,18 com seleção de variáveis. Além disso, embora o PLSR não tenha em seu algoritmo básico a execução de seleção de fatores influentes, apresenta a propriedade de ponderação ascendente e decrescente de fatores, que pode tornar desnecessário o processo de seleção de fatores influentes por outros métodos (MARTENS; NAES, 1989; MEHMOOD *et al.*, 2012).

Contudo, apesar da seleção de variáveis ter o potencial de melhorar o desempenho de um modelo, o desempenho pode ser comprometido devido à eliminação de informações consideradas “redundâncias úteis”. Ou seja, a seleção de variáveis pode remover informações que, a princípio, parecem redundantes ou desnecessárias, mas que contribuem para a construção do modelo. Além disso, diminuir o número de variáveis para predição também significa aumentar a influência de cada variável no modelo final (NØRGAARD *et al.*, 2000). Diante disso, é provável que tenha havido a remoção excessiva de informações importantes, ocasionando a queda do desempenho.

3.2.5 CONCLUSÕES

A espectroscopia se apresenta como uma ferramenta excelente e com alto potencial para predição de compostos bioquímicos em folhas de maracujazeiro por meio de modelos preditivos.

Os melhores modelos, com R^2_{cv} superior a 0,90, foram desenvolvidos para determinação dos teores de sacarose, aminoácidos totais e nitrato, com o método PLSR, e apresentaram valores de R^2_{cv} 0,98, 0,99 e 0,99, respectivamente.

Outros modelos são considerados satisfatórios para predição, com valores de R^2_{cv} iguais a 0,83, 0,86 e 0,85, com o método PLSR para amido, e com MLR para proteína e nitrato, respectivamente.

Embora modelos PLSR sejam mais comuns para essa finalidade, diversos métodos de regressão podem ser utilizados para predição de compostos bioquímicos vegetais. Entretanto, de modo geral, modelos MLR com seleção de variáveis propiciam os melhores modelos.

3.2.6 REFERÊNCIAS

- ABRAFRUTAS. **2021 é o ano internacional das frutas, legumes e verduras. O que isso quer dizer?**, 2021. Disponível em: <<https://abrafrutas.org/2021/04/2021-e-o-ano-internacional-das-frutas-legumes-e-verduras-o-que-isso-quer-dizer/>>. Acesso em: 4 jun. 2021
- ABRANTES, T. C. **Análise da deriva simulada de herbicidas auxínicos em soja através de índices de vegetação rgb obtidos por VANT**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2019.
- AITKENHEAD, M. J. *et al.* PHYLIS: A Low-Cost Portable Visible Range Spectrometer for Soil and Plants. **Sensors**, v. 17, n. 1, p. 99, jan. 2017.
- ALVARES, C. A. *et al.* Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, p. 711–728, dez. 2013.
- BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. **Appl Spectrosc**, v. 43, n. 5, p. 772–777, jul. 1989.
- BENEDETTI, A. R.; VIEIRA, M. L. C. Melhoramento de maracujá-doce: avaliação de progênies derivadas do cruzamento entre genótipos selecionados para a produção e qualidade de frutos. 2021.
- BRADFORD, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Analytical Biochemistry**, v. 72, n. 1, p. 248–254, 7 maio 1976.
- BURNETT, A. C.; SERBIN, S. P.; ROGERS, A. Source:sink imbalance detected with leaf- and canopy-level spectroscopy in a field-grown crop. **Plant, Cell & Environment**, v. 44, n. 8, p. 2466–2479, ago. 2021.
- CANASSA, T. A.; LAMONATO, A. L.; RIBEIRO, A. V. Utilização da lei de Lambert-Beer para determinação da concentração de soluções. **Journal of Experimental Techniques and Instrumentation**, v. 1, n. 2, 6 jul. 2018.
- CAWSE, P. A. The determination of nitrate in soil solutions by ultraviolet spectrophotometry. **Analyst**, v. 92, n. 1094, p. 311–315, 1 jan. 1967.
- COELHO, E.; AZEVÊDO, L.; UMZA-GUEZ, M. Fruto do maracujá: importância econômica e industrial, produção, subprodutos e prospecção tecnológica. **Cadernos de Prospecção**, v. 9, p. 347, 30 set. 2016.
- DAS, B. *et al.* Quantitative monitoring of sucrose, reducing sugar and total sugar dynamics for phenotyping of water-deficit stress tolerance in rice through spectroscopy and chemometrics. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 192, p. 41–51, mar. 2018.
- DE BEI, R. *et al.* Rapid measurement of total non-structural carbohydrate concentration in grapevine trunk and leaf tissues using near infrared spectroscopy. **Computers and Electronics in Agriculture**, v. 136, p. 176–183, abr. 2017.

- DUBOIS, M. *et al.* Colorimetric Method for Determination of Sugars and Related Substances. 1956.
- EARLES, J. M. *et al.* Extreme mid-winter drought weakens tree hydraulic-carbohydrate systems and slows growth. **New Phytologist**, v. 219, n. 1, p. 89–97, 2018.
- ELY, K. S. *et al.* Spectroscopy can predict key leaf traits associated with source–sink balance and carbon–nitrogen status. **Journal of Experimental Botany**, v. 70, n. 6, p. 1789–1799, mar. 2019.
- FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. [s.l.] Editora da Unicamp, 2015.
- FREY, L. A. *et al.* A Non-destructive Method to Quantify Leaf Starch Content in Red Clover. **Frontiers in Plant Science**, v. 11, 2020.
- GORRY, P. A. General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method. **Anal. Chem.**, v. 62, n. 6, p. 570–573, mar. 1990.
- GOULD, K. *et al.* When are foliar anthocyanins useful to plants? Re-evaluation of the photoprotection hypothesis using *Arabidopsis thaliana* mutants that differ in anthocyanin accumulation. **Environmental and Experimental Botany**, v. 154, 1 mar. 2018.
- IBGE. **Produção Agrícola Municipal 2020**. Rio de Janeiro, 2021: [s.n.].
- ISAKSSON, T.; NÆS, T. The Effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. **Applied Spectroscopy**, 1988.
- ITOH, H. *et al.* Development of Method for Non-destructive Measurement of Nitrate Concentration in Vegetable Leaves by Near-infrared Spectroscopy. **undefined**, 2011.
- JAWORSKI, E. G. Nitrate reductase assay in intact plant tissues. **Biochemical and Biophysical Research Communications**, v. 43, n. 6, p. 1274–1279, 18 jun. 1971.
- JENSEN, J. R. **Sensoriamento Remoto do Ambiente: Uma perspectiva em Recursos Terrestres**. [s.l.] Ed. Parêntese, 2009.
- JIA, R.; MAO, Z.; CHANG, Y. **A nonlinear robust partial least squares method with application**. 2010 Chinese Control and Decision Conference. **Anais...** In: 2010 CHINESE CONTROL AND DECISION CONFERENCE. maio 2010.
- KARLINASARI, L. *et al.* Near infrared (NIR) spectroscopy for estimating the chemical composition of (*Acacia mangium* Willd.) wood. **Journal of the Indian Academy of Wood Science**, 2014.
- KITIĆ, G. *et al.* A new low-cost portable multispectral optical device for precise plant status assessment. **Computers and Electronics in Agriculture**, v. 162, p. 300–308, 1 jul. 2019.
- LI, L. *et al.* Methods for estimating leaf nitrogen concentration of winter oilseed rape (*Brassica napus* L.) using in situ leaf spectroscopy. **Industrial Crops and Products**, v. 91, p. 194–204, 2016.

LI, X. *et al.* Review of NIR spectroscopy methods for nondestructive quality analysis of oilseeds and edible oils. **Trends in Food Science & Technology**, v. 101, p. 172–181, 2020.

LOHR, D. *et al.* Non-destructive determination of carbohydrate reserves in leaves of ornamental cuttings by near-infrared spectroscopy (NIRS) as a key indicator for quality assessments. **Biosystems Engineering**, v. 158, p. 51–63, jun. 2017.

MARTENS, H.; MARTENS, M. Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). **Food Quality and Preference**, v. 11, n. 1, p. 5–16, 1 jan. 2000.

MEHMOOD, T. *et al.* A review of variable selection methods in Partial Least Squares Regression. **Chemometrics and Intelligent Laboratory Systems**, v. 118, p. 62–69, 15 ago. 2012.

MOUAZEN, A. *et al.* Near Infrared Spectroscopy for Agricultural Materials: An Instrument Comparison. **Journal of Near Infrared Spectroscopy**, v. 13, p. 87–98, 1 abr. 2005.

NETO, A. J. *et al.* Vis/NIR spectroscopy and chemometrics for non-destructive estimation of water and chlorophyll status in sunflower leaves. **Biosystems Engineering**, v. 155, p. 124–133, 1 mar. 2017.

NØRGAARD, L. *et al.* Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. **Applied Spectroscopy**, v. 54, n. 3, p. 413–419, 1 mar. 2000.

NOVO, E. M. L. M. **Sensoriamento remoto: princípios e aplicações**. 3ª edição ed. São Paulo: Blucher, 2010.

PETISCO, C. *et al.* Near-infrared reflectance spectroscopy as a fast and non-destructive tool to predict foliar organic constituents of several woody species. **Anal Bioanal Chem**, v. 386, n. 6, p. 1823–1833, nov. 2006.

PIRES, M.; JOSÉ, A.; CONCEIÇÃO, A. **Maracujá: avanços tecnológicos e sustentabilidade**. [s.l: s.n.].

PONZONI, F. J. Sensoriamento remoto no estudo da vegetação: diagnosticando a Mata Atlântica. In: **Instituto Nacional de Pesquisas Espaciais**. São José dos Campos: [s.n.].

SAEYS, W. *et al.* Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. **Postharvest Biology and Technology**, v. 158, p. 110981, 1 dez. 2019.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, jan. 1964.

SCHIMLECK, L. R.; EVANS, R.; ILIC, J. Estimation of Eucalyptus delegatensis wood properties by near infrared spectroscopy. **Canadian Journal of Forest Research**, v. 31, n. 10, p. 1671–1675, 1 out. 2001.

SCHOBER, P.; BOER, C.; SCHWARTE, L. Correlation coefficients: appropriate use and interpretation. **Anesthesia & Analgesia**, 2018.

SERBIN, S. P. *et al.* Leaf optical properties reflect variation in photosynthetic metabolism and its sensitivity to temperature. **J Exp Bot**, v. 63, n. 1, p. 489–502, jan. 2012.

SHENK, J.; WESTERHAUS, M. O. Calibration the ISI way. **Near Infrared Spectroscopy: The future Waves. NIR Chichester**, p. 198–202, 1996.

SOUZA, E. R. *et al.* Variação de carboidratos em folhas da videira ‘Itália’ submetida a diferentes de níveis de desfolhas. **Revista Brasileira de Ciências Agrárias - Brazilian Journal of Agricultural Sciences**, v. 8, n. 4, p. 535–539, 28 dez. 2013.

VAN HANDEL, E. Direct microdetermination of sucrose. **Analytical Biochemistry**, v. 22, n. 2, p. 280–283, fev. 1968.

WHITE, A. C. *et al.* How can we make plants grow faster? A source–sink perspective on growth rate. **Journal of Experimental Botany**, v. 67, n. 1, p. 31–45, 2016.

WILLIAMS, P. C. Implementation of near-infrared technology. In P. Williams & K. Norris (Eds.). In: **Near-infrared technology in the agricultural and food industries**. St. Paul, MN: American Association of Cereal Chemists, [s.n.].

WILLIAMS, P. C.; SOBERING, D. C. How do we do it: a brief summary of the methods we use in developing near infrared calibration. **Near Infrared Spectroscopy: The Future Waves, NIR Publications**, n. Chichester, 185-188, 1996.

YEMM, E.; COCKING, E.; RICKETTS, R. The Determination of Amino-Acids with Ninhydrin. **Analyst**, v. 80, 1 jan. 1955.

YENDREK, C. R. *et al.* High-Throughput Phenotyping of Maize Leaf Physiological and Biochemical Traits Using Hyperspectral Reflectance. **Plant Physiology**, v. 173, n. 1, p. 614–626, jan. 2017.

YUN, Y.-H. *et al.* A hybrid variable selection strategy based on continuous shrinkage of variable space in multivariate calibration. **Analytica Chimica Acta**, v. 1058, p. 58–69, 13 jun. 2019.

4 DISCUSSÃO GERAL

No artigo 1, ficou evidenciado que a espectroscopia Vis-NIR (450 a 2500 nm) associada a modelos preditivos PLSR e MLR é uma excelente ferramenta para a quantificação não destrutiva de carboidratos solúveis, amido e sacarose em folhas de mangueira, das variedades Keitt, Omer, Shelly e Palmer.

Para o atributo amido, utilizando o conjunto geral (com todas as variedades de mangueira), os melhores modelos foram obtidos com o método MLR, cujo R^2_{cv} foi igual a 0,92 e RMSEC igual a 0,06. Para carboidratos, utilizando também conjunto geral, os melhores modelos foram obtidos com o método MLR, cujo R^2_{cv} foi igual a 0,71 e RMSEC igual a 33,81. Para sacarose, com o conjunto Keitt, o melhor modelo foi obtido com o método PLSR, cujo R^2_{cv} foi igual a 0,55 e RMSEC igual a 1,10.

O uso do MLR foi interessante por desenvolver modelos mais simples, mais fácil de ser analisado, além de apresentar menor custo computacional de processamento, ótimo para ser utilizado na espectroscopia Vis-NIR, desde que sejam realizados os procedimentos necessários no processamento (SAEYS *et al.*, 2019).

Além disso, o melhor desempenho para os atributos no conjunto geral (amido e carboidratos) pode ser atribuído ao fato de serem utilizados um número maior de amostras (302) em relação ao atributo sacarose (96), que foi mensurado apenas no conjunto Keitt. Segundo Mora e Shimleck (2008), maior variabilidade entre as amostras propicia o desenvolvimento de modelos matemáticos preditivos mais robustos, devido às suas maiores chances de apresentar todas as possíveis variações que possam ocorrer em aplicações reais (MORA; SHIMLECK, 2008).

No artigo 2, a espectroscopia VIS-NIR, associada a modelos preditivos PLSR e MLR, também se mostrou uma excelente ferramenta para predição de carboidratos solúveis, amido, sacarose, além de aminoácidos totais, proteína, nitrato e nitrato redutase.

Os melhores modelos obtidos foram desenvolvidos para os atributos sacarose, aminoácidos totais e nitrato, pois tiveram desempenho excelente, com R^2_{cv} superior a 0,90 e iguais 0,98, 0,99 e 0,99, respetivamente, pelo método MLR. Os modelos PLSR para amido e MLR para proteína e nitrato redutase também apresentaram bom desempenho, com R^2_{cv} , iguais 0,83, 0,86 e 0,85.

Embora ambos os artigos tenham o mesmo tipo de amostra (folhas vegetais), houve diferentes resultados para cada espécie analisada. Isso pode ocorrer porque, independente do atributo analisado, o desempenho dos modelos não depende apenas dos

métodos de regressão e pré-tratamentos utilizados, mas também do tipo e número de amostras, da espécie vegetal em estudo (para amostras de origem vegetal), entre outros fatores (HAIR, 2005). Por isso, cabe ao pesquisador analisar seu material de estudo e objetivo da pesquisa, para decidir os métodos de regressão e pré-tratamentos a serem utilizados em seus trabalhos, ou mesmo testar e comparar diferentes combinações de ferramentas matemáticas disponíveis para, então, avaliar os resultados obtidos para cada situação.

5 CONSIDERAÇÕES FINAIS

Os objetivos do trabalho foram todos atingidos e a hipótese se mostra verdadeira, para todas as condições estabelecidas. Portanto, a espectroscopia se apresenta como uma ferramenta de alto potencial para predição de compostos bioquímicos em folhas de mangueira e maracujazeiro por meio de modelos preditivos.

Foram desenvolvidos:

- I. Modelos preditivos para determinação de carboidratos solúveis, amido e sacarose em folhas de mangueira por meio da espectroscopia Vis-NIR.
 - Amido: $R^2_{cv} = 0,92$ e $RMSECV = 0,06$
 - Carboidratos solúveis: $R^2_{cv} = 0,71$ e $RMSECV = 33,80$
 - Sacarose: $R^2_{cv} = 0,55$ e $RMSECV = 1,10$
- II. Modelos preditivos para determinação de compostos bioquímicos em folhas de maracujazeiro por meio da espectroscopia Vis-NIR.
 - Amido: $R^2_{cv} = 0,83$ e $RMSECV = 15,01$
 - Carboidratos solúveis: $R^2_{cv} = 0,39$ e $RMSECV = 1,20$
 - Sacarose: $R^2_{cv} = 0,98$ e $RMSECV = 0,37$
 - Aminoácidos totais: $R^2_{cv} = 0,99$ e $RMSECV = 0,88$
 - Proteína: $R^2_{cv} = 0,86$ e $RMSECV = 4,08$
 - Nitrato: $R^2_{cv} = 0,99$ e $RMSECV = 0,19$
 - Nitrato redutase: $R^2_{cv} = 0,84$ e $RMSECV = 2,06$

6 REFERÊNCIAS

- ABRAFRUTAS. **2021 é o ano internacional das frutas, legumes e verduras. O que isso quer dizer?**, 2021a. Disponível em: <<https://abrafrutas.org/2021/04/2021-e-o-ano-internacional-das-frutas-legumes-e-verduras-o-que-isso-quer-dizer/>>. Acesso em: 4 jun. 2021
- ABRAFRUTAS. **De onde vem o que eu como: o Brasil é o maior produtor de maracujá e possui 150 variedades da fruta**, 2021b. Disponível em: <<https://abrafrutas.org/2021/07/de-onde-vem-o-que-eu-como-o-brasil-e-o-maior-produtor-de-maracuja-e-possui-150-variedades-da-fruta/>>. Acesso em: 20 nov. 2021
- ABRANTES, T. C. **Análise da deriva simulada de herbicidas auxínicos em soja através de índices de vegetação rgb obtidos por VANT**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2019.
- AITKENHEAD, M. J. *et al.* PHYLIS: A Low-Cost Portable Visible Range Spectrometer for Soil and Plants. **Sensors**, v. 17, n. 1, p. 99, jan. 2017.
- ALTHOFF, D. *et al.* Heuristic methods applied in reference evapotranspiration modeling. **Ciência e Agrotecnologia**, v. 42, p. 314–324, jun. 2018.
- ALVARES, C. A. *et al.* Köppen’s climate classification map for Brazil. **Meteorologische Zeitschrift**, p. 711–728, dez. 2013.
- AMARANTE, C. V. T. DO *et al.* Radiação, fotossíntese, rendimento e qualidade de frutos em macieiras “Royal Gala” cobertas com telas antigranizo. **Pesquisa Agropecuária Brasileira**, v. 42, n. 7, p. 925–931, jul. 2007.
- BANDARA, K. M. T. S. *et al.* A Low-Cost NDIR-Based N₂O Gas Detection Device for Agricultural Soils: Assembly, Calibration Model Validation, and Laboratory Testing. **Sensors**, v. 21, n. 4, p. 1189, jan. 2021.
- BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. **Appl Spectrosc**, v. 43, n. 5, p. 772–777, jul. 1989.
- BENEDETTI, Anderson R. Melhoramento de maracujá doce: avaliação de progênies derivadas do cruzamento entre genótipos selecionados para a produção e qualidade de frutos. 2021. 67 f. Dissertação (Mestrado) – Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba – SP, 2021.
- BORGOGNONE, D. *et al.* Effect of nitrogen form and nutrient solution pH on growth and mineral composition of self-grafted and grafted tomatoes. **Scientia Horticulturae**, Advances in Vegetable Grafting Research. v. 149, p. 61–69, 4 jan. 2013.
- BRADFORD, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Analytical Biochemistry**, v. 72, n. 1, p. 248–254, 7 maio 1976.
- BRASIL. **PORTARIA Nº 240, DE 12 DE MARÇO DE 2019 - Imprensa Nacional**. Disponível em: <<https://www.in.gov.br/materia>>. Acesso em: 11 mar. 2022.

BREDEMEIER, C.; MUNDSTOCK, C. Regulação da absorção e assimilação do nitrogênio nas plantas. **Ciência Rural - CIENC RURAL**, v. 30, 1 abr. 2000.

BRONDANI, G. *et al.* Determinação do teor de carboidratos em minicepas de *Eucalyptus benthamii*. **Journal of Biotechnology and Biodiversity**, v. 3, p. 51–60, 15 fev. 2012.

BURNETT, A. C.; SERBIN, S. P.; ROGERS, A. Source:sink imbalance detected with leaf- and canopy-level spectroscopy in a field-grown crop. **Plant, Cell & Environment**, v. 44, n. 8, p. 2466–2479, ago. 2021.

CANASSA, T. A.; LAMONATO, A. L.; RIBEIRO, A. V. Utilização da lei de Lambert-Beer para determinação da concentração de soluções. **Journal of Experimental Techniques and Instrumentation**, v. 1, n. 2, 6 jul. 2018.

CAPELA, J. M. V.; CAPELA, M. V.; MAGNANI, R. Generalização de um modelo linear de impedância eletroquímica. **Eclética Química**, v. 28, n. 1, p. 63–67, 2003.

CARDOSO, J. R. F. Desenvolvimento de Estrutura Robótica para Aquisição e Classificação de Imagens (ERACI) de Lavoura de Cana-de-Açúcar. 29 jun. 2020.

CARNEIRO, M. M. L. C. *et al.* Fotorrespiração e metabolismo antioxidante em plantas jovens de seringueira cultivadas sob diferentes fontes de nitrogênio (NO₃⁻ e NH₄⁺) - DOI:10.5039/agraria.v10i1a494. **Revista Brasileira de Ciências Agrárias (Agrária)**, v. 10, n. 1, p. 66–73, 30 jun. 2015.

CAWSE, P. A. The determination of nitrate in soil solutions by ultraviolet spectrophotometry. **Analyst**, v. 92, n. 1094, p. 311–315, 1 jan. 1967.

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

CEPEA. Anuário 2021-2022. **Hortifruti Brasil**, 2021.

CERVI, A. C. Espécies de *Passiflora* l. (Passifloraceae) publicadas e descritas nos últimos 55 anos (1950–2005) na América do Sul e principais publicações brasileiras. **Estudos de Biologia**, v. 27, n. 61, 24 nov. 2005.

COELHO, E.; AZEVÊDO, L.; UMZA-GUEZ, M. Fruto do maracujá: importância econômica e industrial, produção, subprodutos e prospecção tecnológica. **Cadernos de Prospecção**, v. 9, p. 347, 30 set. 2016.

COLEZEA, M. *et al.* CLUeFARM: Integrated web-service platform for smart farms. **Computers and Electronics in Agriculture**, v. 154, p. 134–154, nov. 2018.

COMINO, F. *et al.* Infrared spectroscopy as a tool for the assessment of soil biological quality in agricultural soils under contrasting management practices. **Ecological Indicators**, v. 87, p. 117–126, abr. 2018.

COSTA, D. **Desenvolvimento de protótipo optoeletrônico para determinação não destrutiva dos atributos de qualidade e estágio de amadurecimento em uvas viníferas**. Campinas: Universidade Estadual de Campinas, 2018.

- CRAWFORD, N. M. Nitrate: nutrient and signal for plant growth. **The Plant Cell**, v. 7, n. 7, p. 859–868, 1 jul. 1995.
- CURRAN, P. J. Remote sensing of foliar chemistry. **Remote Sensing of Environment**, v. 30, n. 3, p. 271–278, 1 dez. 1989.
- DALE, L. M. *et al.* Hyperspectral Imaging Applications in Agriculture and Agro-Food Product Quality and Safety Control: A Review. **Applied Spectroscopy Reviews**, v. 48, n. 2, p. 142–159, mar. 2013.
- DAS, A. *et al.* Interrelations of growth regulators, carbohydrates and expression of flowering genes (FT, LFY, AP1) in leaf and shoot apex of regular and alternate bearing mango (*Mangifera indica* L.) cultivars during flowering. **Scientia Horticulturae**, v. 253, p. 263–269, 2019.
- DAS, B. *et al.* Quantitative monitoring of sucrose, reducing sugar and total sugar dynamics for phenotyping of water-deficit stress tolerance in rice through spectroscopy and chemometrics. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 192, p. 41–51, mar. 2018.
- DAVENPORT, T. L. Reproductive physiology of mango. **Brazilian Journal of Plant Physiology**, v. 19, p. 363–376, dez. 2007.
- DA-WEN, S. **Infrared spectroscopy for food quality analysis and control**. 1. ed. [s.l.: s.n.].
- DE BEI, R. *et al.* Rapid measurement of total non-structural carbohydrate concentration in grapevine trunk and leaf tissues using near infrared spectroscopy. **Computers and Electronics in Agriculture**, v. 136, p. 176–183, abr. 2017.
- DECHANT, B. *et al.* Estimation of photosynthesis traits from leaf reflectance spectra: Correlation to nitrogen content as the dominant mechanism. **Remote Sensing of Environment**, v. 196, p. 279–292, jul. 2017.
- DUBOIS, M. *et al.* Colorimetric Method for Determination of Sugars and Related Substances. 1956.
- EARLES, J. M. *et al.* Extreme mid-winter drought weakens tree hydraulic-carbohydrate systems and slows growth. **New Phytologist**, v. 219, n. 1, p. 89–97, 2018.
- ELHAMDAOUI, O. *et al.* Development of Fast Analytical Method for the Detection and Quantification of Honey Adulteration Using Vibrational Spectroscopy and Chemometrics Tools. **Journal of Analytical Methods in Chemistry**, v. 2020, p. e8816249, 23 dez. 2020.
- ELY, K. S. *et al.* Spectroscopy can predict key leaf traits associated with source–sink balance and carbon–nitrogen status. **Journal of Experimental Botany**, v. 70, n. 6, p. 1789–1799, mar. 2019.
- EMBRAPA. **Specsolo - Portal Embrapa**. Disponível em: <<https://www.embrapa.br/en/busca-de-solucoes-tecnologicas/-/produto-servico/3740/specsolo>>. Acesso em: 15 dez. 2021.

EMBRAPA. **Exportação de manga brasileira bate recorde em 2020, totalizando US\$ 246 milhões**. Disponível em: <<https://www.embrapa.br/busca-de-noticias/-/noticia/60585117/exportacao-de-manga-brasileira-bate-recorde-em-2020-totalizando-us-246-milhoes>>. Acesso em: 4 jun. 2021.

FERREIRA, M. M. C. *et al.* Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, v. 22, p. 724–731, set. 1999.

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

FIERRO, C. A.; ULLOA, M. A. Development reference stage for flower induction response to potassium nitrate in mango. **Acta Horticulturae**, n. Wageningen. n.291. p.71-78, 1991.

FILIPPUCCI, P. *et al.* Soil moisture as a potential variable for tracking and quantifying irrigation: A case study with proximal gamma-ray spectroscopy data. **Advances in Water Resources**, v. 136, p. 103502, fev. 2020.

FLORENZANO, T. G. Imagens obtidas por sensoriamento remoto. In: **Imagens de satélite para estudos ambientais**. [s.l.] Oficina de textos, 2002.

FLORENZANO, T. G. In: **Iniciação em sensoriamento remoto**. [s.l.] Oficina de textos, 2007.

FREY, L. A. *et al.* A Non-destructive Method to Quantify Leaf Starch Content in Red Clover. **Frontiers in Plant Science**, v. 11, 2020.

GANESAN, M.; ANDAVAR, S.; RAJ, R. S. P. Prediction of Land Suitability for Crop Cultivation Using Classification Techniques. **Brazilian Archives of Biology and Technology**, v. 64, 25 out. 2021.

GENISHEVA, Z. *et al.* New PLS analysis approach to wine volatile compounds characterization by near infrared spectroscopy (NIR). **Food Chemistry**, v. 246, p. 172–178, 25 abr. 2018.

GORRY, P. A. General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method. **Anal. Chem.**, v. 62, n. 6, p. 570–573, mar. 1990.

GOULD, K. *et al.* When are foliar anthocyanins useful to plants? Re-evaluation of the photoprotection hypothesis using *Arabidopsis thaliana* mutants that differ in anthocyanin accumulation. **Environmental and Experimental Botany**, v. 154, 1 mar. 2018.

HAIR, J. F. **Análise Multivariada de Dados - 6ª Edição**. 6. ed. [s.l.] 2005.

HALLIDAY, D.; RESNICK, R.; WALKER, J. **Fundamentos de física - Gravitação, Ondas e Termodinâmica**. 10th edition ed. Rio de Janeiro: LTC, 2016a. v. 2

HALLIDAY, D.; RESNICK, R.; WALKER, J. **Fundamentos de física - Óptica e Física Moderna**. 10th edition ed. Rio de Janeiro: LTC, 2016b. v. 4

HAZIS, N. H. *et al.* Assessment of Carbohydrate Contents in Perlis Harumanis Mango Leaves during Vegetative and Productive Growth. **IOP Conference Series: Materials Science and Engineering**, v. 429, p. 012025, nov. 2018.

HURST, G. S. *et al.* **Spectroscopy** *Encyclopedia Britannica*, 2021. Disponível em: <<https://www.britannica.com/science/spectroscopy>>. Acesso em: 17 ago. 2021

IBGE. **Produção Agrícola Municipal 2020**. Rio de Janeiro, 2021: [s.n.].

ISAKSSON, T.; NÆS, T. The Effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. **Applied Spectroscopy**, 1988.

ITOH, H. *et al.* Development of Method for Non-destructive Measurement of Nitrate Concentration in Vegetable Leaves by Near-infrared Spectroscopy. **undefined**, 2011.

JAWORSKI, E. G. Nitrate reductase assay in intact plant tissues. **Biochemical and Biophysical Research Communications**, v. 43, n. 6, p. 1274–1279, 18 jun. 1971.

JENSEN, J. R. **Sensoriamento Remoto do Ambiente: Uma perspectiva em Recursos Terrestres**. [s.l.] Ed. Parêntese, 2009.

JIA, R.; MAO, Z.; CHANG, Y. **A nonlinear robust partial least squares method with application**. 2010 Chinese Control and Decision Conference. **Anais...** In: 2010 CHINESE CONTROL AND DECISION CONFERENCE. maio 2010.

KARLINASARI, L. *et al.* Near infrared (NIR) spectroscopy for estimating the chemical composition of (*Acacia mangium* Willd.) wood. **Journal of the Indian Academy of Wood Science**, 2014.

KAWAMURA, K. *et al.* Vis-NIR Spectroscopy and PLS Regression with Waveband Selection for Estimating the Total C and N of Paddy Soils in Madagascar. **Remote Sensing**, v. 9, p. 1081, 23 out. 2017.

KEMSLEY, E. K. *et al.* Quantitative analysis of sugar solutions using infrared spectroscopy. **Food Chemistry**, v. 44, n. 4, p. 299–304, jan. 1992.

KHOURI, Camila R. Atividade da redutase do nitrato, teores de nitrogênio e de carboidratos em cafeeiro influenciados pelo sombreamento e estágio fenológico. 2007. 63 f. Dissertação (Mestrado em Fitotecnia) – Universidade Federal de Viçosa, Campus Viçosa, Viçosa – MG, 2007.

KITIC', G. *et al.* A new low-cost portable multispectral optical device for precise plant status assessment. **Computers and Electronics in Agriculture**, v. 162, p. 300–308, 1 jul. 2019.

LAMMERS, K.; ARBUCKLE-KEIL, G.; DIGHTON, J. FT-IR study of the changes in carbohydrate chemistry of three New Jersey pine barrens leaf litters during simulated control burning. **Soil Biology and Biochemistry**, v. 41, n. 2, p. 340–347, fev. 2009.

LANDAU, E. C. *et al.* **Dinâmica da produção agropecuária e da paisagem natural no Brasil nas últimas décadas: produtos de origem vegetal**. [s.l.] Brasília, DF: Embrapa, 2020., 2020.

- LANNA, A. C. *et al.* Mitigação dos estresses abióticos na agricultura mediada pela interação de microrganismos e plantas. **Documentos**, n. 319, 17 nov. 2021.
- LANZA, E.; LI, B. Application of Near Infrared Spectroscopy for Predicting the Sugar Content of Fruit Juices. **Journal of Food Science**, v. 49, p. 995–998, 25 ago. 2006.
- LI, B. *et al.* Prediction and monitoring of leaf water content in soybean plants using terahertz time-domain spectroscopy. **Computers and Electronics in Agriculture**, v. 170, p. 105239, mar. 2020a.
- LI, L. *et al.* Methods for estimating leaf nitrogen concentration of winter oilseed rape (*Brassica napus* L.) using in situ leaf spectroscopy. **Industrial Crops and Products**, v. 91, p. 194–204, 2016.
- LI, X. *et al.* Determination of Hemicellulose, Cellulose and Lignin in Moso Bamboo by Near Infrared Spectroscopy. **Scientific Reports**, v. 5, p. 17210, 2015.
- LI, X. *et al.* Review of NIR spectroscopy methods for nondestructive quality analysis of oilseeds and edible oils. **Trends in Food Science & Technology**, v. 101, p. 172–181, 2020b.
- LIMA, G. M. DE S. *et al.* Manejo da indução floral da mangueira “Palmer” com uso de uniconazole. **Ciência Rural**, v. 46, p. 1350–1356, 6 maio 2016.
- LOBO, J. T. *et al.* Biostimulants on Nutritional Status and Fruit Production of Mango ‘Kent’ in the Brazilian Semiarid Region. **HortScience**, v. 54, n. 9, p. 1501–1508, 1 set. 2019.
- LOHR, D. *et al.* Non-destructive determination of carbohydrate reserves in leaves of ornamental cuttings by near-infrared spectroscopy (NIRS) as a key indicator for quality assessments. **Biosystems Engineering**, v. 158, p. 51–63, jun. 2017.
- LUCAS, B. N.; SCHÚ, A. I.; DALLA NORA, F. M. Uso de smartphone como alternativa inovadora no controle de qualidade de alimentos: uma breve revisão. **Avanços em ciência e tecnologia de alimentos**, n. 21, p. 278–288, 1 abr. 2021.
- LUZ, E. R. **Predição de propriedades de gasolinas usando espectroscopia FTIR e regressão por mínimos quadrados parciais**. Rio de Janeiro: Pontifícia Universidade Católica do Rio de Janeiro, 2003.
- MALINOWSKI, E. R. **Factor Analysis in Chemistry**. [s.l.] Wiley, 2002.
- MARSCHNER, H. **Marschner’s Mineral Nutrition of Higher Plants**. [s.l.] Academic Press, 2011.
- MARTENS, H.; MARTENS, M. Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). **Food Quality and Preference**, v. 11, n. 1, p. 5–16, 1 jan. 2000.
- MARTENS, H.; NAES, T. **Multivariate calibration**. Chichester, UK: Wiley, 1989.
- MARTUSCELLO, J. A. *et al.* Características produtivas e fisiológicas de capim-elefante submetido à adubação nitrogenada. **Archivos de Zootecnia**, v. 65, n. 252, p. 565–570, 15 out. 2016.

- MASSARUHÁ, S. M. F. S. *et al.* A transformação digital no campo rumo à agricultura sustentável e inteligente. In: **Embrapa Informática Agropecuária-Artigo em anais de congresso (ALICE)**. Brasília: EMBRAPA, 2020, 2020. p. 409.
- MEHMOOD, T. *et al.* A review of variable selection methods in Partial Least Squares Regression. **Chemometrics and Intelligent Laboratory Systems**, v. 118, p. 62–69, 15 ago. 2012.
- MENDES, W. DE S. *et al.* Geostatistics or machine learning for mapping soil attributes and agricultural practices. **Revista Ceres**, v. 67, p. 330–336, 28 ago. 2020.
- MENESES, P. R.; ALMEIDA, T.; BAPTISTA, G. M. **Reflectância dos materiais terrestres: análise e interpretação**. São Paulo: Oficina de Textos: [s.n.].
- MORA, C. R.; SCHIMLECK, L. R. On the selection of samples for multivariate regression analysis: application to near-infrared (NIR) calibration models for the prediction of pulp yield in *Eucalyptus nitens*. **Canadian Journal of Forest Research**, 2008.
- MOUAZEN, A. *et al.* Near Infrared Spectroscopy for Agricultural Materials: An Instrument Comparison. **Journal of Near Infrared Spectroscopy**, v.13, p. 87–98, 1 abr. 2005.
- MUNAWAR, A. A. *et al.* Calibration models database of near infrared spectroscopy to predict agricultural soil fertility properties. **Data in Brief**, v. 30, p. 105469, jun. 2020.
- NAKAJIMA, S. *et al.* Quantification of starch content in germinating mung bean seedlings by terahertz spectroscopy. **Food Chemistry**, v. 294, p. 203–208, out. 2019.
- NESPECA, M. G. Aplicação de espectroscopia no infravermelho próximo e análise multivariada para identificação e quantificação de hidrocarbonetos totais do petróleo em solo. 31 ago. 2018.
- NETO, A. J. *et al.* Vis/NIR spectroscopy and chemometrics for non-destructive estimation of water and chlorophyll status in sunflower leaves. **Biosystems Engineering**, v. 155, p. 124–133, 1 mar. 2017.
- NEVES, L.; MORAES, D. M. Vigour And A-Amylase Analysis In Seeds Of Rice Cultivars Submitted To Several Treatments With Acetic Acid. **Revista de Ciências Agroveterinárias**, n. Lages, v. 4, n 1, p. 35–43, 2005.
- NØRGAARD, L. *et al.* Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. **Applied Spectroscopy**, v. 54, n. 3, p. 413–419, 1 mar. 2000.
- NOVO, E. M. L. M. **Sensoriamento remoto: princípios e aplicações**. 3ª edição ed. São Paulo: Blucher, 2010.
- OLIVEIRA, G. P. *et al.* Teores de carboidratos em mangueira “Ubá” submetida a diferentes doses de paclobutrazol. **Revista de Ciências Agrárias**, v. 41, n. 3, p. 749–756, 2018.
- OLIVEIRA, M. *et al.* Structural and cytological aspects of mango floral induction using paclobutrazol. **Scientia Horticulturae**, v. 261, p. 1, 29 nov. 2019.

OLIVEROS, N. *et al.* Predictive models of chlorophyll content in sugarcane seedlings using spectral images. **Engenharia Agrícola**, v. 41, p. 475–484, 15 set. 2021.

ONU. **Population**. Disponível em: <<https://www.un.org/en/global-issues/population>>. Acesso em: 28 dez. 2020.

PABÓN, R. E. C. **Espectroscopia de refletância e de imageamento aplicadas ao monitoramento de solos expostos em instalações petrolíferas**. Campinas: Universidade Estadual de Campinas, 2015.

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

PAZ-KAGAN, T. *et al.* Assessing the nitrogen status of almond trees by visible-to-shortwave infrared reflectance spectroscopy of carbohydrates. **Computers and Electronics in Agriculture**, v. 178, p. 105755, nov. 2020.

PENG, J. *et al.* Fast visualization of distribution of chromium in rice leaves by re-heating dual-pulse laser-induced breakdown spectroscopy and chemometric methods. **Environmental Pollution**, v. 252, p. 1125–1132, 1 set. 2019.

PETISCO, C. *et al.* Near-infrared reflectance spectroscopy as a fast and non-destructive tool to predict foliar organic constituents of several woody species. **Anal Bioanal Chem**, v. 386, n. 6, p. 1823–1833, nov. 2006.

PIRES, M.; JOSÉ, A.; CONCEIÇÃO, A. **Maracujá: avanços tecnológicos e sustentabilidade**. [s.l: s.n.].

PONZONI, F. J. Sensoriamento remoto no estudo da vegetação: diagnosticando a Mata Atlântica. In: **Instituto Nacional de Pesquisas Espaciais**. São José dos Campos: [s.n.].

PRASAD, S. R. S. *et al.* Studies on Changes in Carbohydrate Metabolism in Regular Bearing and “Off” Season Bearing Cultivars of Mango (*Mangifera indica* L.) During Flowering. **International Journal of Fruit Science**, v. 14, n. 4, p. 437–459, out. 2014.

REDA, R. *et al.* Comparing CalReg performance with other multivariate methods for estimating selected soil properties from Moroccan agricultural regions using NIR spectroscopy. **Chemometrics and Intelligent Laboratory Systems**, v. 211, p. 104277, abr. 2021.

REFAEILZADEH, P.; TANG, L.; LIU, H. Cross-Validation. In: LIU, L.; ÖZSU, M. T. (Eds.). **Encyclopedia of Database Systems**. New York, NY: Springer, 2016. p. 1–7.

RINGNÉR, M. What is principal component analysis? **Nature Biotechnology**, v. 26, n. 3, p. 303–304, mar. 2008.

RODRÍGUEZ-PÉREZ, J. R. *et al.* Leaf water content estimation by functional linear regression of field spectroscopy data. **Biosystems Engineering, Sensing and Control of Crop Water Status**. v. 165, p. 36–46, jan. 2018.

SAEYS, W. *et al.* Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. **Postharvest Biology and Technology**, v. 158, p. 110981, 1 dez. 2019.

SAKANE, K. K.; MATSUURA, E. N.; SIMIONI, A. R. Espectroscopia no infravermelho como ferramenta para diferenciação de chá verde na agricultura orgânica e convencional. **Revista Tecnologia e Tendências**, v. 10, n. 2, p. 59–74, dez. 2019.

SAMPAIO, P. S. *et al.* Identification of rice flour types with near-infrared spectroscopy associated with PLS-DA and SVM methods. **European Food Research and Technology**, v. 246, n. 3, p. 527–537, 1 mar. 2020.

SANTOS, A. DE S. **Indução termoperiódica da nitrato redutase de membrana plasmática em abacaxizeiro (*Ananas comosus*)**. [s.l.] Universidade de São Paulo, 14 out. 2010.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, jan. 1964.

SCHIMLECK, L. R.; EVANS, R.; ILIC, J. Estimation of Eucalyptus delegatensis wood properties by near infrared spectroscopy. **Canadian Journal of Forest Research**, v. 31, n. 10, p. 1671–1675, 1 out. 2001.

SCHOBER, P.; BOER, C.; SCHWARTE, L. Correlation coefficients: appropriate use and interpretation. **Anesthesia & Analgesia**, 2018.

SERBIN, S. P. *et al.* Leaf optical properties reflect variation in photosynthetic metabolism and its sensitivity to temperature. **J Exp Bot**, v. 63, n. 1, p. 489–502, jan. 2012.

SHAH, S. S. A. *et al.* Towards fruit maturity estimation using NIR spectroscopy. **Infrared Physics & Technology**, v. 111, p. 103479, dez. 2020.

SHEN, T. *et al.* Rapid Determination of Cadmium Contamination in Lettuce Using Laser-Induced Breakdown Spectroscopy. **Molecules**, v. 23, n. 11, p. 2930, 2018.

SHENK, J.; WESTERHAUS, M. O. Calibration the ISI way. **Near Infrared Spectroscopy: The future Waves. NIR Chichester**, p. 198–202, 1996.

SILVA, A. O. DA *et al.* Da tradição a técnica: perspectivas e realidades da agricultura de derruba e queima na Amazônia. **Research, Society and Development**, v. 10, n. 1, p. e38310111799–e38310111799, 17 jan. 2021.

SILVA, M. M. L. *et al.* Effect of carbohydrate content on shoot maturation and yield of *Palmer mango* submitted to potassium fertilization and biostimulant | Research, Society and Development. **Agrarian and Biological Sciences**, v. 9, 2020.

SKOLIK, P. *et al.* Attenuated total reflection Fourier-transform infrared spectroscopy coupled with chemometrics directly detects pre- and post-symptomatic changes in tomato plants infected with *Botrytis cinerea*. **Vibrational Spectroscopy**, v. 111, p. 103171, nov. 2020.

SKOOG, D. *et al.* **Fundamentos de química analítica**. Tradução da 8ª edição norte-americana ed. São Paulo-SP: Editora Thomson, 2006.

SOEST, P. J. V. **Nutritional Ecology of the Ruminant**. [s.l.] Cornell University Press, 1994.

SOLOMONSON, L. P.; BARBER, M. J. Assimilatory Nitrate Reductase: Functional Properties and Regulation. **Annual Review of Plant Physiology and Plant Molecular Biology**, v. 41, n. 1, p. 225–253, 1990.

SONG, X.-C. *et al.* Predicting the antioxidant capacity and total phenolic content of bearberry leaves by data fusion of UV–Vis spectroscopy and UHPLC/Q-TOF-MS. **Talanta**, v. 213, p. 120831, 1 jun. 2020.

SOUZA, J. T. DE *et al.* Data mining and machine learning in the context of sustainable evaluation: a literature review. **IEEE Latin America Transactions**, v. 17, n. 03, p. 372–382, mar. 2019.

SOUZA, E. R. *et al.* Variação de carboidratos em folhas da videira ‘Itália’ submetida a diferentes de níveis de desfolhas. **Revista Brasileira de Ciências Agrárias - Brazilian Journal of Agricultural Sciences**, v. 8, n. 4, p. 535–539, 28 dez. 2013.

TAIZ, L. *et al.* **Fisiologia e desenvolvimento vegetal**. 6. ed. Porto Alegre: Artmed, 2017.

TININI, R. C. R. **Avaliação não destrutiva de sólidos solúveis por meio de imagens espectrais de uvas submetidas à desidratação parcial**. PhD Thesis—Campinas: Universidade Estadual de Campinas, 2015.

TORRES, I. *et al.* Irrigation decision support based on leaf relative water content determination in olive grove using near infrared spectroscopy. **Biosystems Engineering**, v. 180, p. 50–58, abr. 2019.

ULMER, T.; MACDOUGAL, J. M. **Passiflora: Passionflowers of the world**. Portland: Timber Press, 2004.

VAN HANDEL, E. Direct microdetermination of sucrose. **Analytical Biochemistry**, v. 22, n. 2, p. 280–283, fev. 1968.

VIANA, E. M. **Interação de nitrogênio e potássio na nutrição, no teor de clorofila e na atividade da redutase de nitrato em plantas de trigo**. Piracicaba: Universidade de São Paulo, 2007.

VIANA, E. M.; KIEHL, J. DE C. Doses de nitrogênio e potássio no crescimento do trigo. **Bragantia**, v. 69, p. 975–982, dez. 2010.

WALSH, K. B. *et al.* Visible-NIR ‘point’ spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. **Postharvest Biology and Technology**, v. 168, p. 111246, out. 2020.

WHITE, A. C. *et al.* How can we make plants grow faster? A source–sink perspective on growth rate. **Journal of Experimental Botany**, v. 67, n. 1, p. 31–45, 2016.

WILLIAMS, P. C. Implementation of near-infrared technology. In P. Williams & K. Norris (Eds.). In: **Near-infrared technology in the agricultural and food industries**. St. Paul, MN: American Association of Cereal Chemists, [s.n.].

WILLIAMS, P. C.; SOBERING, D. C. How do we do it: a brief summary of the methods we use in developing near infrared calibration. **Near Infrared Spectroscopy: The Future Waves, NIR Publications**, n. Chichester, 185-188, 1996.

WOLD, S.; SJOSTROM, M.; ERIKSSON, L. PLS-regression: A Basic Tool of Chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, p. 109–130, 1 out. 2001.

WU, D. *et al.* Application of near infrared spectroscopy for the rapid determination of antioxidant activity of bamboo leaf extract. **Food Chemistry**, v. 135, n. 4, p. 2147–2156, 15 dez. 2012.

YASSITEPE, J. E. DE C. T. *et al.* Genômica aplicada às mudanças climáticas: biotecnologia para a agricultura digital. - Portal Embrapa. In: **Geotecnologias na agricultura digital**. Brasília: EMBRAPA, 2020, 2020. p. 409.

YEMM, E.; COCKING, E.; RICKETTS, R. The Determination of Amino-Acids with Ninhydrin. **Analyst**, v. 80, 1 jan. 1955.

YENDREK, C. R. *et al.* High-Throughput Phenotyping of Maize Leaf Physiological and Biochemical Traits Using Hyperspectral Reflectance. **Plant Physiology**, v. 173, n. 1, p. 614–626, jan. 2017.

YUN, Y.-H. *et al.* A hybrid variable selection strategy based on continuous shrinkage of variable space in multivariate calibration. **Analytica Chimica Acta**, v. 1058, p. 58–69, 13 jun. 2019.

ZHANG, Q.; LI, Q.; ZHANG, G. Rapid Determination of Leaf Water Content Using VIS/NIR Spectroscopy Analysis with Wavelength Selection. **Spectroscopy: An International Journal**, v. 27, n. 2, p. 93–105, 28 maio 2012.

ANEXO A – VARIÁVEIS SELECIONADAS DO ARTIGO 1

Keitt	
Amido	
Pré-selecionadas	410, 415, 418, 424, 436, 455, 481, 504, 519, 528, 559, 573, 591, 609, 618, 635, 654, 664, 681, 695, 705, 727, 753, 785, 823, 881, 618, 635, 654 931, 965
Assinaturas espectrais	618, 635, 654
Carboidratos solúveis	
Pré-selecionadas	412, 450, 506, 514, 528, 539, 554, 568, 588, 608, 630, 651, 674, 690, 694, 711, 739
Assinaturas espectrais	539
Sacarose	
Pré-selecionadas	424, 434, 440, 453, 474, 497, 504, 514, 521, 527, 538, 557, 562, 587, 599, 625, 630, 644, 666, 673, 682, 694, 703, 720, 729, 740, 749, 782, 921, 971
Assinaturas espectrais	434, 440, 504, 521, 587, 625, 673, 682, 740, 782, 971
Omer, Shelly	
Amido	
Pré-selecionadas	381, 435, 481, 517, 535, 547, 563, 587, 609, 628, 661, 666, 675, 689, 697, 710, 735, 754, 817, 936
Assinaturas espectrais	563, 661, 735, 754, 936
Carboidratos solúveis	
Pré-selecionadas	424, 432, 452, 481, 484, 499, 508, 521, 525, 532, 538, 544, 554, 558, 566, 571, 589, 598, 608, 625, 634, 647, 649, 657, 665, 674, 677, 683, 689, 695, 701, 717, 728, 732, 733, 749, 779, 802, 870, 891, 934, 972
Assinaturas espectrais	499, 538, 544, 589, 598, 608, 634, 647, 665, 677, 934, 972
Palmer	
Amido	
Pré-selecionadas	484, 527, 528, 586, 627, 670, 694, 695, 697, 715, 753, 1004, 1087, 1138, 1302, 1361, 1401, 1471, 1715, 1828, 1837, 1891, 1925, 2010, 2112, 2173, 2242, 2302, 2492
Assinaturas espectrais	527, 586, 1004, 1087, 1302, 1401, 1715
Carboidratos solúveis	
Pré-selecionadas	410, 483, 527, 541, 585, 620, 630, 686, 697, 723, 757, 997, 1006, 1152, 1198, 1396, 1475, 1520, 1645, 1721, 1825, 1841, 1870, 1890, 1925, 2111, 2231, 2310
Assinaturas espectrais	757, 1645, 1870, 1925, 2231
Geral	
Amido	
Pré-selecionadas	408, 417, 464, 498, 508, 543, 588, 604, 623, 637, 658, 703, 757, 945, 992, 1006, 1078, 1127, 1156, 1275, 1360, 1461, 1628, 1654, 1702, 1721, 1788, 1827, 1864, 1884, 1910, 2015, 2025, 2089, 2140, 2204, 2241, 2276, 2313, 2331, 2381, 2500
Assinaturas espectrais	417, 508, 543, 604, 703, 945, 1006, 1156, 1360, 1628, 1654, 1788, 1864, 1910, 2500
Carboidratos solúveis	
Pré-selecionadas	399, 413, 427, 432, 456, 485, 493, 508, 519, 531, 550, 566, 573, 596, 598, 623, 644, 657, 670, 681, 687, 698, 706, 717, 723, 743, 765, 787, 887, 920, 937, 970
Assinaturas espectrais	427, 456, 485, 531, 566, 623, 644, 681, 698, 706, 723, 743, 765, 920

ANEXO B – VARIÁVEIS SELECIONADAS DO ARTIGO 2

Amido	
Pré-selecionadas	1867, 1889, 1901, 1906, 1911, 1915, 1925, 1934, 1950, 1969, 1970, 1982, 2006, 2029, 2034, 2038, 2055, 2071, 2095, 2112, 2149, 2177, 2200, 2211, 2247, 2280, 2310, 2319, 2339, 2364, 2372, 2391, 2408, 2423, 2434, 2448, 2455, 2459, 2462, 2470, 2473, 2477, 2483, 2489, 2494, 2500
Assinaturas espectrais	2029, 2034, 2055, 2071, 2112, 2211, 2247
Carboidratos solúveis	
Pré-selecionadas	458, 482, 521, 541, 617, 647, 679, 687, 693, 754, 975, 1109, 1155, 1183, 1315, 1372, 1416, 1451, 1587, 1664, 1713, 1845, 1888, 1930, 2010, 2048, 2119, 2240, 2288, 2334, 2383, 2395
Assinaturas espectrais	458, 482, 521, 687, 693, 975, 1183, 1845
Sacarose	
Pré-selecionadas	1010, 1026, 1040, 1047, 1053, 1059, 1069, 1076, 1080, 1087, 1098, 1105, 1112, 1134, 1146, 1149, 1156, 1163, 1168, 1171, 1181, 1185, 1192, 1201, 1308, 1313, 1323, 1330, 1339, 1348, 1364, 1378, 1388, 1396, 1404, 1410, 1417, 1426, 1437, 1447, 1453, 1461, 1470, 1481, 1494, 1506, 1519, 1531, 1544, 1553, 1568, 1582, 1597, 1608, 1614, 1631, 1644, 1652, 1659, 1673, 1679, 1686, 1692, 1704, 1712, 1718, 1729, 1736, 1742, 1750, 1754, 1763, 1781, 1787, 1799, 1802, 1813, 1819, 1828, 2500
Assinaturas espectrais	1010, 1026, 1040, 1053, 1059, 1069, 1076, 1080, 1087, 1105, 1134, 1146, 1156, 1163, 1168, 1171, 1181, 1185, 1308, 1323, 1330, 1339, 1348, 1364, 1378, 1388, 1396, 1404, 1410, 1417, 1426, 1437, 1461, 1470, 1481, 1494, 1519, 1531, 1544, 1553, 1568, 1582, 1597, 1608, 1614, 1631, 1644, 1652, 1659, 1673, 1686, 1692, 1704, 1712, 1718, 1729, 1736, 1750, 1754, 1763, 1781, 1787, 1799, 1802, 1813, 1828
Aminoácidos totais	
Pré-selecionadas	1016, 1024, 1039, 1050, 1058, 1068, 1072, 1082, 1087, 1096, 1104, 1137, 1145, 1159, 1168, 1177, 1194, 1204, 1216, 1226, 1238, 1247, 1259, 1280, 1288, 1301, 1309, 1313, 1322, 1334, 1342, 1360, 1378, 1394, 1399, 1412, 1417, 1444, 1455, 1464, 1475, 1483, 1489, 1498, 1504, 1518, 1528, 1533, 1553, 1566, 1575, 1584, 1593, 1608, 1616, 1621, 1631, 1649, 1661, 1670, 1676, 1681, 1686, 1697, 1705, 1711, 1719, 1734, 1745, 1747, 1760, 1765, 1780, 1788, 1801, 1813, 1819, 1824
Assinaturas espectrais	1016, 1024, 1039, 1050, 1058, 1068, 1072, 1082, 1087, 1096, 1104, 1137, 1159, 1168, 1177, 1194, 1204, 1216, 1226, 1247, 1259, 1280, 1288, 1301, 1309, 1322, 1334, 1342, 1360, 1394, 1399, 1412, 1417, 1444, 1455, 1464, 1475, 1489, 1498, 1504, 1518, 1533, 1553, 1566, 1584, 1593, 1608, 1616, 1621, 1631, 1649, 1661, 1670, 1676, 1681, 1686, 1697, 1711, 1719, 1734, 1745, 1747, 1760, 1765, 1780, 1788, 1801, 1813, 1819, 1824
Proteína	
Pré-selecionadas	358, 361, 365, 370, 379, 382, 387, 389, 393, 397, 400, 404, 407, 415, 419, 421, 423, 429, 436, 438, 447, 451, 457, 465, 469, 473, 477, 481, 484, 496, 505, 513, 520, 536, 552, 559, 568, 573, 582, 593, 605, 618, 632, 642, 653, 659, 670, 675, 681, 686, 703, 711, 717, 727, 736, 751, 779, 831, 856, 877, 946, 967, 1000
Assinaturas espectrais	358, 370, 379, 387, 393, 400, 407, 415, 419, 421, 423, 447, 451, 457, 465, 496, 505, 513, 536, 552, 559, 568, 573, 582, 605, 618, 632, 642, 653, 659, 681, 717, 727, 736, 751, 779, 831, 856, 877, 946, 967

Nitrato	
Pré-seleccionadas	1004, 1019, 1024, 1030, 1036, 1049, 1060, 1067, 1083, 1104, 1113, 1137, 1151, 1161, 1169, 1178, 1190, 1197, 1211, 1227, 1257, 1265, 1284, 1305, 1315, 1326, 1334, 1339, 1346, 1364, 1372, 1394, 1407, 1420, 1427, 1437, 1444, 1452, 1459, 1465, 1469, 1476, 1500, 1505, 1513, 1526, 1555, 1564, 1581, 1586, 1597, 1611, 1627, 1647, 1652, 1670, 1673, 1688, 1692, 1703, 1709, 1719, 1728, 1734, 1740, 1751, 1757, 1767, 1782, 1789, 1798, 1804, 1812, 1827
Assinaturas espectrais	1004, 1019, 1024, 1030, 1036, 1060, 1067, 1083, 1104, 1113, 1137, 1151, 1161, 1169, 1190, 1197, 1227, 1257, 1265, 1284, 1305, 1315, 1326, 1334, 1339, 1346, 1364, 1372, 1407, 1420, 1427, 1437, 1444, 1452, 1459, 1465, 1469, 1500, 1505, 1513, 1526, 1555, 1564, 1581, 1597, 1611, 1627, 1647, 1652, 1670, 1673, 1688, 1692, 1703, 1709, 1719, 1728, 1734, 1740, 1751, 1757, 1767, 1789, 1798, 1804, 1827
Nitrato redutase	
Pré-seleccionadas	356, 361, 364, 367, 370, 374, 379, 381, 384, 388, 393, 398, 400, 404, 406, 410, 414, 423, 429, 437, 446, 452, 459, 465, 469, 473, 481, 492, 497, 505, 512, 519, 534, 541, 555, 572, 580, 595, 604, 615, 629, 634, 644, 654, 661, 669, 676, 682, 688, 692, 698, 704, 715, 726, 734, 746, 776, 829, 920, 966
Assinaturas espectrais	356, 361, 364, 367, 370, 384, 388, 398, 404, 406, 410, 414, 423, 429, 452, 459, 473, 481, 497, 512, 519, 555, 572, 595, 604, 615, 629, 644, 654, 669, 676, 682, 688, 698, 715, 734, 776, 829, 920, 966