



MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E
TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA
E TECNOLOGIA GOIANO - CAMPUS URUTAÍ
CURSO DE LICENCIATURA EM QUÍMICA



DENISE DE FÁTIMA GONÇALVES

**APLICAÇÃO DE MÉTODOS DE SELEÇÃO DE VARIÁVEIS POR
INTERVALOS NA QUANTIFICAÇÃO DE BANHA DE PORCO EM
MANTEIGA DE CACAU**



Urutaí – GO

Março/2021

DENISE DE FÁTIMA GONÇALVES

Trabalho de Conclusão de Curso apresentado à Banca Examinadora do Curso de Licenciatura em Química do Instituto Federal de Educação, Ciência e Tecnologia Goiano – Campus Urutaí, como parte dos requisitos necessários para obtenção do título de Licenciado em Química.

Orientador: Prof. Dr. Lucas Caixeta Gontijo

Urutaí – GO

Março/2021

Sistema desenvolvido pelo ICMC/USP
Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas - Instituto Federal Goiano

G635a Gonçalves, Denise de Fátima
 Aplicação de Métodos de Seleção de Variáveis por
 Intervalos na Quantificação de Banha de Porco em
 Manteiga de Cacau / Denise de Fátima Gonçalves;
 orientador Lucas Caixeta Gontijo. -- Urutaí, 2021.
 67 p.

 Monografia (Graduação em Licenciatura em Química) --
 Instituto Federal Goiano, Campus Urutaí, 2021.

 1. Adulteração. 2. Banha de Porco. 3.
 Espectroscopia de Infravermelho Médio. 4. iPLS. 5.
 Manteiga de Cacau. I. Gontijo, Lucas Caixeta,
 orient. II. Título.

Responsável: Johnathan Pereira Alves Diniz - Bibliotecário-Documentalista CRB-1 nº2376

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR PRODUÇÕES
TÉCNICO-CIENTÍFICAS NO REPOSITÓRIO INSTITUCIONAL DO IF GOIANO**

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia Goiano, a disponibilizar gratuitamente o documento no Repositório Institucional do IF Goiano (RIIF Goiano), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IF Goiano.

Identificação da Produção Técnico-Científica

- | | |
|--|---|
| <input type="checkbox"/> Tese | <input type="checkbox"/> Artigo Científico |
| <input type="checkbox"/> Dissertação | <input type="checkbox"/> Capítulo de Livro |
| <input type="checkbox"/> Monografia – Especialização | <input type="checkbox"/> Livro |
| <input checked="" type="checkbox"/> TCC - Graduação | <input type="checkbox"/> Trabalho Apresentado em Evento |
| <input type="checkbox"/> Produto Técnico e Educacional - Tipo: | _____ |

Nome Completo do Autor: Denise de Fátima Gonçalves

Matrícula: 2017101221530056

Título do Trabalho: Aplicação de Métodos de Seleção de Variáveis por Intervalos na Quantificação de Banha de Porco em Manteiga de Cacau

Restrições de Acesso ao Documento

Documento confidencial: Não Sim, justifique: _____

Informe a data que poderá ser disponibilizado no RIIF Goiano: 24/03/2021

O documento está sujeito a registro de patente? Sim Não

O documento pode vir a ser publicado como livro? Sim Não

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

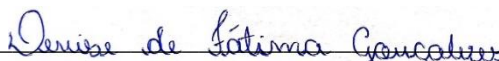
- o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Urutai _____

Local

23/03/2021 _____

Data



Assinatura do Autor e/ou Detentor dos Direitos Autorais

Ciente e de acordo:



Assinatura do(a) orientador(a)

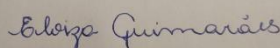
ATA DE DEFESA DO TRABALHO DE CONCLUSÃO DE
CURSO

Ata dos Trabalhos da Banca Examinadora do Trabalho de Conclusão de Curso (TC) da estudante **Denise de Fátima Gonçalves** para obtenção do título de Licenciada em Química pelo Instituto Federal de Educação, Ciência e Tecnologia Goiano – Campus Urutaí (IF Goiano/Urutaí). Integraram a Banca: **Profa. Dra. Eloiza Guimarães, Prof. Ms. Maxwell Severo da Costa e Prof. Dr. Lucas Caixeta Gontijo (orientador)**. Aos 19 (dezenove) dias do mês de março de 2021 às 14 horas realizou-se, via webconferência, a apresentação pública do TC pela estudante. O orientador abriu a sessão agradecendo a participação dos membros da Banca Examinadora. Em seguida convidou a estudante para que fizesse a exposição do trabalho intitulado: “**APLICAÇÃO DE MÉTODOS DE SELEÇÃO DE VARIÁVEIS POR INTERVALOS NA QUANTIFICAÇÃO DE BANHA DE PORCO EM MANTEIGA DE CACAU**”. Finalizada a apresentação, cada membro da Banca Examinadora realizou a arguição sobre o trabalho da estudante. Dando continuidade aos trabalhos, o orientador solicitou a todos que aguardassem enquanto a Banca Examinadora pudesse deliberar sobre o TC da candidata em outro ambiente virtual. Terminada a deliberação, o orientador leu a ata dos trabalhos declarando **aprovado** o TC da estudante. Em seguida, deu por encerrada a solenidade, da qual eu, **Lucas Caixeta Gontijo**, presidente da banca, lavrei a presente ata que vai assinada por mim e pelos demais membros da Banca Examinadora.



Prof. Dr. Lucas Caixeta Gontijo (IF Goiano/Urutaí)

Presidente da Banca



Profa. Dra. Eloiza Guimarães (SEE/MG)



Prof. Ms. Maxwell Severo da Costa (IF Goiano/Urutaí)

APLICAÇÃO DE MÉTODOS DE SELEÇÃO DE VARIÁVEIS POR
INTERVALOS NA QUANTIFICAÇÃO DE BANHA DE PORCO EM
MANTEIGA DE CACAU

Denise de Fátima Gonçalves

Trabalho de Conclusão de Curso apresentado à Banca Examinadora do Curso de Licenciatura em Química do Instituto Federal de Educação, Ciência e Tecnologia Goiano – Campus Urutaí, como parte dos requisitos necessários para obtenção do título de Licenciada em Química.


Defendido e aprovado em 19/03/2021.

Banca Examinadora



Prof. Dr. Lucas Caixeta Gontijo (IF Goiano/Urutaí)

Presidente da Banca



Profa. Dra. Eloiza Guimarães (SEE/MG)



Prof. Ms. Maxwell Severo da Costa (IF Goiano/Urutaí)

DEDICATÓRIA

Dedico este trabalho a Michel (*in memoriam*), por me inspirar com sua dedicação e por ter demonstrado, em sua breve passagem pela vida, o que significa “não desistir”.

AGRADECIMENTOS

Meus agradecimentos são voltados para todas as pessoas que passaram por minha vida e me fizeram ser que eu sou. Iniciando com meus pais, Juraneide e Vicente, que trabalharem duro durante anos com o único objetivo de fornecer o melhor (saúde, educação, inspiração e fé) a seus filhos. Agradeço ao meu irmão Daniel por, apesar das poucas palavras, sempre me ajudar da melhor forma, e também ao meu irmão Michel (*in memoriam*) por ter sido a primeira pessoa a confiar em meu potencial. Agradeço a minha vó Olívia por ter me criado até meus 6 anos de idade e por muitas vezes ter me levado e me buscado a pé, da fazenda até a escola. Jamais serei capaz de retribuir tudo o que fizeram por mim.

Agradeço especialmente ao meu namorado Renan, por ter sido paciente comigo, me escutado, incentivando, ajudando, aconselhando, estudando comigo se fosse necessário, buscando o melhor para mim, sendo meu melhor amigo e por ter sempre acreditado que eu conseguiria, me dizendo sempre “Tomara que consiga? Não. Você já conseguiu!” e “Não existe esse negócio de sorte”. Não esquecerei o que fez por mim, muito obrigada!

Fiz muitos amigos durante toda minha vida e gostaria de agradecer a todos, mas em especial, aqueles que estiveram comigo durante todo curso, principalmente a Renata, Sidimar e Ziom, pois compartilhamos muito mais que aulas e matérias, compartilhamos o almoço, lanche e janta, brincamos, estudamos, jogamos nosso precioso momento de “discórdia” e compartilhamos o mais importante: sonhos!

Estendo meus agradecimentos a todos os funcionários do curso de Licenciatura em Química do Instituto Federal Goiano-Campus Urutaí. Gostaria de destacar o corpo docente, pois estes me ensinaram o que é ser de fato “Professor”, sempre acessíveis, demonstrando paciência, compreensão, empatia e sensibilidade. A vocês o meu muito obrigada!

Agradeço ao meu orientador, Prof. Dr. Lucas Caixeta Gontijo, por sempre demonstrar uma imensa paciência, disponibilizar sempre um pouco de seu tempo para me ensinar e me ajudar, por ter aberto as portas da pesquisa científica para mim, ter me apresentado a “Quimiometria” e ter sido, muitas vezes, um “pai” para mim. Sempre será um membro muito importante da “Família Química” do IF Goiano-Campus Urutaí.

Finalmente, e mais importante, agradeço a Deus por ter me concedido o dom da vida, perseverança, fortaleza e fé. Me manteve de pé até em meus piores dias e

me proporcionou a oportunidade de conhecer todas as pessoas (principalmente as citadas acima) que me transformaram em uma pessoa melhor.

Obrigada Senhor!

EPÍGRAFE

*“É o tempo que gastaste com a tua rosa
que assim a torna tão importante.”*

O Pequeno Príncipe

Antoine de Saint-Exupéry

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS	xii
LISTA DE FIGURAS	xiv
LISTA DE TABELAS	xv
RESUMO.....	xvi
ABSTRACT	xvii
1. INTRODUÇÃO.....	18
2. REVISÃO BIBLIOGRÁFICA	21
2.1. Cacau.....	21
2.2. Adulteração	23
2.3. Espectroscopia de Infravermelho Médio (MIR)	27
2.4. Métodos de Análise Multivariada	29
2.5. Regressão por Quadrados Mínimos Parciais.....	32
2.6. Seleção de Variáveis por Intervalos (iPLS, biPLS e siPLS)	35
2.7. Validação do Modelo.....	39
3. MATERIAIS E MÉTODOS	42
3.1. Obtenção e Preparo das Amostras	42
3.2. Obtenção dos Espectros.....	42
3.3. Análise Quimiométrica	42
4. RESULTADOS E DISCUSSÕES.....	44
5. CONCLUSÃO	57
6. REVISÃO BIBLIOGRÁFICA	58

LISTA DE ABREVIATURAS E SIGLAS

ANVISA: Agência Nacional de Vigilância Sanitária

ASTM: Sociedade Americana de Testes e Materiais (do inglês, *American Society for Testing and Materials*)

biPLS: quadrados mínimos parciais por exclusão de intervalos (do inglês, *Backward Interval Partial Least Squares*)

CBA: Gordura alternativa a manteiga de cacau (do inglês, *cocoa butter alternative*)

CBE: Gordura equivalente a manteiga de cacau (do inglês, *cocoa butter equivalent*)

CBR: Repositor de manteiga de cacau (do inglês, *cocoa butter replacer*)

CBS: Substituto de manteiga de cacau (do inglês, *cocoa butter substitute*)

CG: Cromatografia Gasosa

CLS: Regressão por Quadrados Mínimos – Clássico (do inglês, *Classic Least Squares Regression*)

EJCR: Elipse de Confiança

EM: Espectrometria de Massas

FT: Transformada de Fourier (do inglês, *Fourier transform*)

FTMIR: Infravermelho Médio com Transformada de Fourier

GA: Algoritmo Genético

HATR: Acessório Horizontal de Reflectância Total Atenuada

iPLS: Quadrados Mínimos Parciais por Intervalos (do inglês, *Intervals Partial Least Squares*)

IR: Espectroscopia no Infravermelho (do inglês, *infra-red*)

iSPA-PLS: Algoritmo das Projeções Sucessivas para Seleção de Intervalos em PLS

IUPAC: União internacional da Química Pura e Aplicada (do inglês, *International Union of Pure and Applied Chemistry*)

MC: Manteiga de Cacau

MIR: Infravermelho Médio

MLR: Regressão Linear Múltipla (do inglês, *Multiple Linear Regression*)

MSC: Correção Multiplicativa de Sinal

NIPALS: Implementação Straight Foward de Modificação Estatisticamente Inspirada de PLS (do inglês, *Straight Foward Implementation of Statistically Inspired Modification of PLS*)

PCA: Análise de Componentes Principais

PCR: Regressão por Componentes Principais (do inglês, *Principal Components Regression*)

PLS: Regressão por Quadrados Mínimos Parciais (do inglês, *Partial Least Squares*)

R: Coeficiente de Correlação

RDC: Resolução da Diretoria Colegiada

RMN: Ressonância Magnética Nuclear

RMSEC: Erro Quadrático médio de Calibração

RMSECV: Erro Quadrático Médio de Validação Cruzada

RMSEP: Erro Quadrático Médio de Previsão

RMSEP: Erro Quadrático Médio de Previsão

SDV: Desvio padrão dos erros de validação

siPLS: quadrados mínimos parciais por intervalos sinérgicos (do inglês, *Synergy Interval Partial Least Squares*)

TAG: Triacilglicerol (do inglês, *triacylglycerol*)

Teste F: Distribuição de Fisher-Snedecor

Teste-t: Teste t-Student

VL: Varáveis Latente

LISTA DE FIGURAS

Figura 1. Fluxograma de processamento do cacau.	22
Figura 2. Composição dos ácidos graxos presentes na MC.	23
Figura 3. (a) Ácido Esteárico; (b) Ácido Palmítico; e (c) Ácido Oleico.	23
Figura 4. Diagrama esquemático de espectrofotômetro de transformada de Fourier.	28
Figura 5. Representação da construção da matriz X para modelagem multivariada	31
Figura 6. Fluxograma com esquema geral dos métodos multivariados mais populares dentro da Química e a indicação da sua utilização dependendo do conjunto de dados.	32
Figura 7. Gráfico do modelo iPLS do número de intervalos versus RMSECV.	36
Figura 8. a) Espectro original; b) Espectro após o Baseline.	44
Figura 9. a) RMSECV e RMSEC para cada VL; b) RMSEP e Erro Médio de Previsão para cada VL; c) RMSEC, RMSECV e RMSEP para cada VL.	45
Figura 10. Representação gráfica dos valores previstos versus os reais para o modelo PLS.	48
Figura 11. Espectro dividido em 25 intervalos equidistantes, cada um com o valor de VL ideal para o modelo.	49
Figura 12. Intervalo número 5 selecionado do modelo iPLS com variáveis de 389-485.	50
Figura 13. Região espectral que os intervalos foram selecionados.	51
Figura 14. Coeficiente de correlação do modelo biPLS.	51
Figura 15. Região espectral que os intervalos foram selecionados.	53
Figura 16. Coeficiente de correlação do modelo siPLS.	53
Figura 17. Valores reais versus previstos das amostras do conjunto de previsão. ...	55
Figura 18. EJCR para o modelo iPLS_25.	56

LISTA DE TABELAS

Tabela 1. Propriedades dos subgrupos de gorduras vegetais alternativas.	25
Tabela 2. Exemplo de Tabela de Exclusão obtida pelo biPLS em uma divisão espectral de 35 intervalos.	37
Tabela 3. Exemplo de resultado de saída fornecido pelo algoritmo siPLS para uma combinação entre dois intervalos.	39
Tabela 4. Capturância para cada VL.	46
Tabela 5. Valores de VL, R, RMSECV, RMSEC e RMSEP para o modelo PLS construído.	47
Tabela 6. O tipo de correlação associada aos valores de R^2	47
Tabela 7. Resultados para iPLS em cada divisão espectral.	48
Tabela 8. Resultados obtidos para os modelos biPLS.	50
Tabela 9. Resultados dos modelos siPLS para cada divisão espectral.	52
Tabela 10. Resultados dos modelos PLS, iPLS, biPLS e siPLS para quantificação de BP em MC.	54
Tabela 11. Cálculos de erros sistemáticos.	55

RESUMO

A manteiga de cacau é considerada um produto com alto valor agregado, sendo o ingrediente fundamental na produção de chocolates. Devido ao seu alto custo, matérias primas de menor valor podem ser adicionadas à manteiga de cacau, como por exemplo, margarina vegetal hidrogenada e banha de porco. Portanto, torna-se necessário o desenvolvimento de métodos analíticos para o controle de qualidade deste tipo de produto alimentício. Nesta perspectiva, este trabalho avaliou a utilização da espectroscopia no infravermelho médio (MIR) aliada à regressão por quadrados mínimos parciais (PLS) e métodos de seleção de variáveis por intervalos para quantificar banha de porco em manteiga de cacau. Neste estudo adulterou-se 68 amostras na faixa de 5,0 a 30,0% de banha de porco na manteiga de cacau, obteve-se espectros na região do Infravermelho Médio e todos os modelos foram construídos utilizando o software MATLAB 6.1. Os modelos PLS foram construídos centrando os dados na média e empregando o processo de validação cruzada por *venation blinds*. Os métodos de seleção de variáveis utilizados foram o de regressão por quadrados mínimos parciais por intervalos (iPLS- *Intervals Partial Least Squares*), regressão quadrados mínimos parciais por exclusão de intervalos (biPLS- *Backward Interval Partial Least Squares*) e quadrados mínimos parciais por sinergismo (siPLS- *Synergy Interval Partial Least Squares*), e buscou-se, entre eles, o modelo que apresenta-se o menor erro quadrático médio de validação cruzada (RMSECV) e o menor número de variáveis comparado ao modelo PLS global, ou seja, modelo contendo todas as variáveis espectrais (PLS). O modelo PLS foi construído com 6 variáveis latentes (VL), tendo 0.7756 de RMSECV, 0.5555 de erro quadrático médio de calibração (RMSEC) e 0.6917 de erro quadrático médio de previsão (RMSEP). O melhor modelo PLS obtido a partir dos métodos de seleção de variáveis foi o iPLS no qual os dados espectrais foram divididos em 25 intervalos. E, o melhor intervalo indicado foi o intervalo 5 contendo 97 variáveis. Este modelo foi construído com 5 VL e apresentou os seguintes parâmetros RMSECV de 0.4410, RMSEC de 0.3745 e RMSEP de 0.3805. Portanto, verificou-se que o modelo iPLS25 apresentou melhor capacidade preditiva que o modelo PLS global, sendo assim, este modelo pode ser empregado na quantificação de banha de porco em manteiga de cacau.

Palavras Chave: Adulteração; Banha de Porco; Espectroscopia de Infravermelho Médio; iPLS; Manteiga de Cacau; PLS; Quimiometria.

ABSTRACT

Cocoa butter is considered a product with high added value, being the fundamental ingredient in the production of chocolates. Due to its high cost, lower value raw materials can be added to cocoa butter, such as hydrogenated vegetable margarine and lard. Therefore, it is necessary to develop analytical methods for the quality control of this type of food product. In this perspective, this work evaluated the use of spectroscopy in the medium infrared (MIR) combined with the regression by partial least squares (PLS) and methods of selection of variables by intervals to quantify lard in cocoa butter. In this study, 68 samples in the range of 5.0 to 30.0% of lard were adulterated in cocoa butter, spectra were obtained in the Middle Infrared region and all models were built using the MATLAB 6.1 software. The PLS models were built by centering the data on the mean and employing the cross validation process by venation blinds. The methods of selection of variables used were regression by Intervals Partial Least Squares (iPLS), regression by Backward Interval Partial Least Squares (biPLS) and Synergy Interval Partial Least Squares (siPLS), and we sought, among them, the model that presents the lowest mean cross-validation square error (RMSECV) and the lowest number of variables compared to the global PLS model, that is, the model containing all spectral variables (PLS). The PLS model was built with 6 latent variables (VL), having 0.7756 of RMSECV, 0.5555 of mean square error of calibration (RMSEC) and 0.6917 of mean square error of forecast (RMSEP). The best PLS model obtained from the variable selection methods was the iPLS in which the spectral data were divided into 25 intervals. And, the best interval indicated was the X interval containing Y variables. This model was built with 5 VL and presented the following parameters RMSECV of 0.4410, RMSEC of 0.3745 and RMSEP of 0.3805. Therefore, it was found that the iPLS25 model had better predictive capacity than the global PLS model, therefore, this model can be used in the quantification of lard in cocoa butter.

Key words: Adulteration; Lard; Medium Infrared Spectroscopy; iPLS; Cocoa butter; PLS; Chemometrics.

1. INTRODUÇÃO

A manteiga de cacau (MC) é o ingrediente mais importante na formulação do chocolate, sendo de grande relevância a sua capacidade de fusão durante a degustação e responsável pela aromaticidade e sabor (SHUKLA, 2006). Trata-se de uma composição triacilglicérica, contendo teores de 10 a 24% de gordura (AFOAKWA, 2014; CABALLERO; FINGLAS; TOLDRÁ, 2015), composta principalmente de ácido oléico ($C_{18}H_{34}O_2$) e esteárico ($C_{18}H_{36}O_2$), que oferecem aspectos nutricionais adequados a saúde humana (LARSSON et al., 2006).

Segundo Dimick (1991), a MC é o ingrediente mais significativo na formulação do chocolate, e também o mais caro economicamente, chegando a um valor que corresponde a um terço do produto final, e também sendo um dos mais altos preços comparados a outras gorduras, devido a elevação de demanda no mercado contrastado a disponibilidade agrícola (JAHURUL et al, 2014). Mesmo este produto apresentando excelentes qualidades, tanto na saúde quanto nas propriedades sensoriais de um legítimo chocolate, seu alto valor de mercado faz com que este ingrediente seja substituído, parcialmente ou totalmente, por sucedâneos baratos. Estes são ricos em ácidos graxos saturados e trans, descaracterizando as propriedades físicas e/ou químicas do chocolate e comprometendo a pureza do produto. A substituição da manteiga ocorre, muitas das vezes, de forma negligente sem a devida informação nos rótulos e causando aversão do consumidor pelo produto final (BATISTA et al., 2016; TORBICA, 2016).

Nesta perspectiva, para detectar e quantificar esses sucedâneos utilizados no lugar da MC, geralmente usa-se cromatografia como técnica química (CSERHÁT et al., 2005). No entanto, esta técnica apresenta desvantagens, como por exemplo, o uso de altas temperaturas, necessidade de aquisição de padrões de cada componente a ser investigado, alta qualificação do operador do equipamento, várias etapas de preparo de amostras, demora no processo de análise e produção de resíduos (VISANI, 2017).

Por outro lado, a espectroscopia no infravermelho médio (MIR) vem mostrando eficácia quando introduzida como técnica de controle de qualidade, pois apresenta

excelentes vantagens, como por exemplo, rápida, sensível, simples e de baixo custo comparado a outras técnicas como RMN, sem necessidade de preparo de amostras e não destrutiva. Assim, devido a estas vantagens a espectroscopia MIR tem sido aplicada no controle de qualidade de vários produtos, como por exemplo, quantificação de adulterantes em combustíveis como o biodiesel e diesel (GUIMARÃES, 2018), no controle de qualidade de óleos utilizados como emolientes de cosméticos (SANTANA, 2015), na área farmacêutica analisando daptomicina em pó liofilizado para solução injetável (BELAVENUTO, 2016), e principalmente analisando adulterantes e/ou grupos funcionais em diversos alimentos, como na farinha de trigo (ROSSI, 2015), entre outros.

O principal método de calibração multivariada aplicada a dados de espectroscopia no infravermelho é a regressão por quadrados mínimos parciais (PLS). No entanto, existe a possibilidade de implementação de métodos de seleção de região espectral para melhorar significativamente o desempenho dos métodos de calibração de espectros totais. Esses métodos selecionam regiões específicas do espectro (um comprimento de onda ou um conjunto de comprimentos de onda) em que a colinearidade não é tão importante, enquanto gera modelos mais estáveis, robustos e mais simples de interpretar. Na prática, a filosofia está baseada na identificação de um subconjunto dos dados inteiros que produzirão erros de previsão mais baixos. Assim, em espectros vibracionais por exemplo, os comprimentos de onda que apenas induzem a ruídos, informações irrelevantes ou não-linearidades, podem ser eliminados (FERNANDES, et al. 2011).

Dentre os métodos de seleção de variáveis existentes pode-se destacar o método de mínimos quadrados parciais por intervalos (iPLS), proposto por Nørgaard et al. (2000). O método iPLS é uma extensão iterativa para o PLS, que desenvolve modelos locais PLS em subintervalos equidistantes de toda a região do espectro. Além do iPLS também foram propostos métodos de combinação de intervalos, como por exemplo, quadrados mínimos parciais por exclusão de intervalos (biPLS) (LEARDI; NØRGAARD, 2004; NØRGAARD, et al. 2000) e quadrados mínimos parciais por intervalos sinérgicos (siPLS) (MUNCK et al., 2001).

Assim, estes métodos têm sido aplicados em diversas amostras. Lopes (2015) obteve excelentes resultados para a caracterização de amostras brasileiras de óleo

de soja, em que foi determinado o índice de acidez desta amostra empregando a regressão por PLS e por iPLS. O modelo PLS foi construído utilizando 6 variáveis latentes (VLs), centrando os dados na média e tratados com correção multiplicativa de sinal (MSC). A faixa espectral empregada foi a região de 4000-800 cm^{-1} , totalizando 3200 variáveis, obtendo-se valor de erro quadrático médio de previsão (RMSEP) igual a 0,1459 mg KOH/g. Já para o modelo iPLS, a região selecionada foi de 1736-1687 cm^{-1} , totalizando apenas 49 variáveis. Neste modelo obteve-se valor de RMSEP igual a 0,1436 mg de KOH/g porém utilizando 3 VLs.

Dentro deste contexto, o presente trabalho teve como objetivo a utilização dos métodos de seleção de variáveis (iPLS, biPLS e siPLS) aliada à espectroscopia no infravermelho médio (MIR) na quantificação de banha de porco como adulterante em manteiga de cacau na faixa de concentração de 5,0 a 30,0% (m/m).

2. REVISÃO BIBLIOGRÁFICA

2.1. Cacau

O cacau é um fruto proveniente do cacaueiro, árvore nativa de florestas de clima tropical pertencente a espécie *Theobroma cacao*, que carrega o significado de “alimento dos deuses” já que eram cultivadas pelos povos maias e astecas. Suas sementes torradas e trituradas com água eram consumidas na forma de uma bebida amarga, chamada de “chocolatl”, que além de ser uma bebida sagrada, consideravam-na medicinal (ADEYEYE et al., 2010; AFOAKWA, 2016).

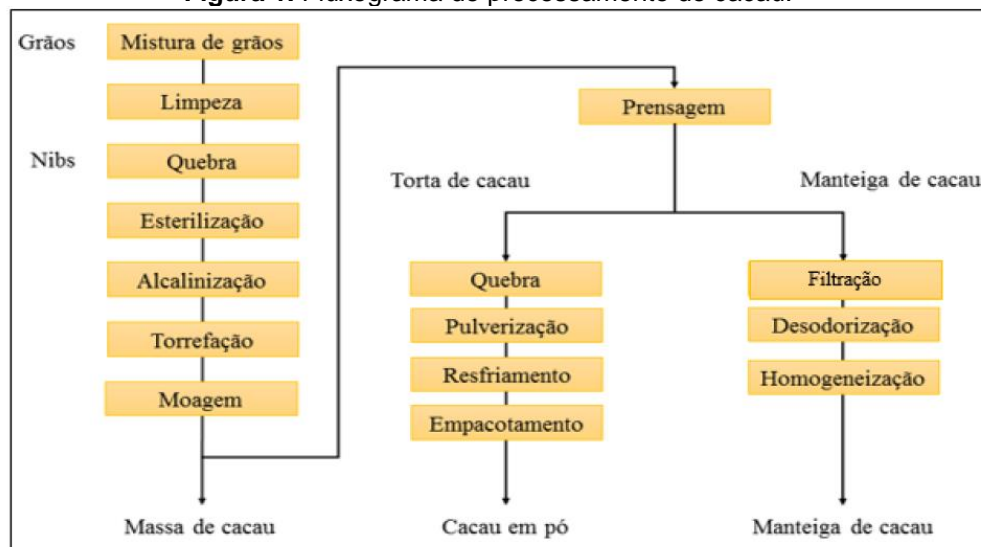
O Brasil já foi considerado um dos maiores produtores de cacau, chegando ao segundo lugar da produção mundial na década de 70, mas devido a incidências de pragas, como o fungo *Moniliophthora perniciosa*, houve queda de produção e o cacau deixou de ser uma das suas principais produções agrícolas. Ainda assim, esta cultura é predominante em pequenas regiões de propriedades familiares no norte e nordeste brasileiro, e ainda ocupa lugar de destaque no mercado internacional (LOPES et al., 2011 apud MARROCOS; MORAES; GOMES, 2018).

Seu grande valor e interesse em plantio está relacionado a sua venda em forma de *commodity*, onde utiliza-se principalmente de suas amêndoas para fabricação de produtos alimentícios, como o chocolate, manteiga de cacau e derivados (SANTOS; FONTES, 2020). Estes produtos, são provenientes de processos de fermentação, secagem e torrefação das amêndoas, passando por alguns procedimentos até chegarem ao ponto ideal do produto a ser industrializado. Por meio destes recursos obtêm-se a manteiga de cacau, produto com uso tanto alimentício quanto em cosméticos, com grande importância ao mercado, e principalmente, de grande relevância para a produção de um chocolate de qualidade (COHEN et al., 2004 apud FERREIRA et al., 2019).

O granulado de cacau é moído (nibs de cacau) após sua torrefação, gera uma massa de cacau chamada de liquor de cacau, que contém um grande teor de gordura (cerca de 54%), e quando se submete esse liquor a prensas hidráulicas a altas pressões é obtida a manteiga de cacau (MC) podendo conter teores de 10% a 24% de gordura total (AFOAKWA, 2014; CABALLERO; FINGLAS; TOLDRÁ, 2015). A

Figura 1 a seguir, mostra as etapas dos processos para obtenção da manteiga de cacau.

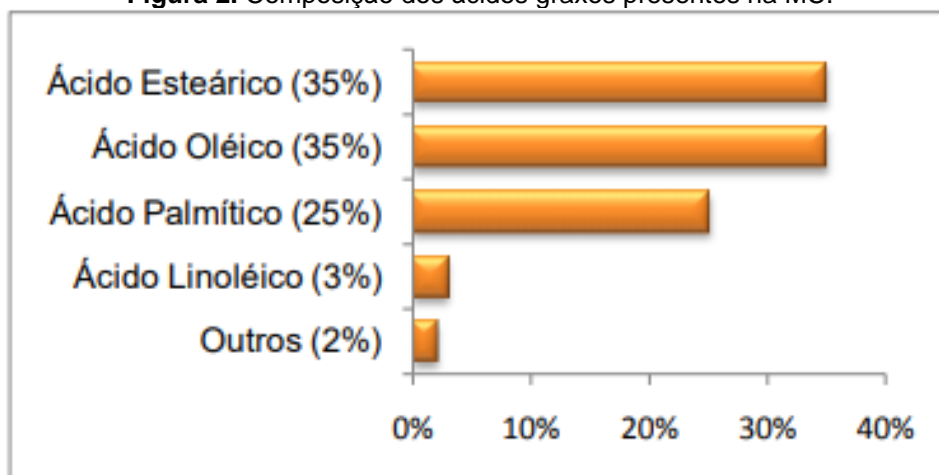
Figura 1. Fluxograma de processamento do cacau.



Fonte: MEURSING, 2009 apud SOUZA, 2017.

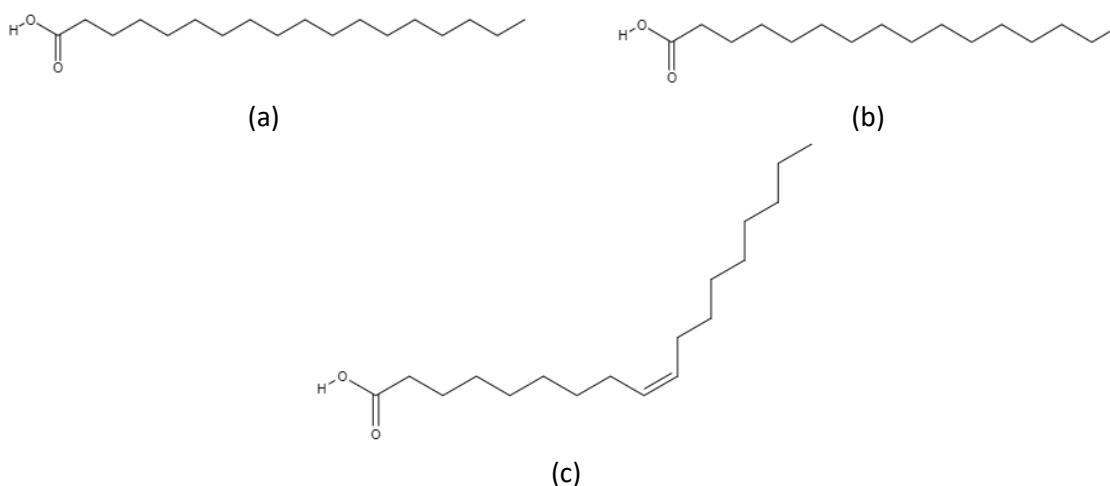
Após a prensagem, a MC apresenta uma coloração amarela com capacidade de derretimento a 35 °C, mesmo assim, é responsável pela dureza do chocolate a temperatura ambiente além da capacidade de derretimento quando em contato com a boca (capacidade polimórfica) e intensificador de aroma e sabor na degustação (QUAST, 2008). Ela é composta por diversos ácidos graxos com imensa importância nutricional para os consumidores, tendo como principais triglicerídeos o ácido palmítico, o esteárico e o ácido oleico (MARTIN, 1987). A porcentagem da composição dos ácidos graxos na MC está apresentada na Figura 2 e as fórmulas estruturais destes ácidos estão apresentados na Figura 3.

Figura 2. Composição dos ácidos graxos presentes na MC.



Fonte: SOUZA, 2010.

Figura 3. (a) Ácido Esteárico; (b) Ácido Palmítico; e (c) Ácido Oleico.



Fonte: A autora.

2.2. Adulteração

Devido a diversos fatores como as condições climáticas não favoráveis, a adulteração da MC por demais componentes denominados “gorduras alternativas” (CBA) ou sucedâneos, vem sendo algo decorrente na produção de chocolate e demais produtos derivados da MC. Estes componentes tratam-se de gorduras equivalentes e com propriedades físico-químicas semelhantes a MC, podendo substituí-la parcialmente desde que 25% dos sólidos totais tenham origem do cacau, de acordo

com a Resolução RDC 265 com vigência desde 22 de setembro de 2005 (ANVISA, 2020a; DE CLERCQ, 2016).

Os sucedâneos permitidos para a fabricação de chocolate são separados em três classificações de acordo com suas diferenças funcionais: CBE (equivalentes a manteiga de cacau – também sendo conhecida como alternativa para manteiga de cacau), CBR (repositor de manteiga de cacau) e CBS (substituto da manteiga de cacau) (LANNES et al, 1997). Esses sucedâneos devem melhorar o perfil nutricional do produto, diminuindo a quantidade de ácidos graxos saturados e insaturados além da queda do custo do produto (OLIVEIRA, 2016).

As CBEs apresentam compatibilidade e algumas vantagens sobre a MC, pois além de apresentarem polimorfismo e pontos de fusão semelhantes, a CBE contém maior resistência ao calor e padronização de dureza. São constituídas, muitas vezes, por gorduras não láuricas e não hidrogenadas, provenientes de uma combinação entre estearina de palma, estearina de shea (*Butyrospermum parkii*), estearina de ilipe (*Shorea stenoptera*), estearina da semente de manga (*Magnifera índica*), gordura de sal (*Butyrospermum parkii*) e a gordura de kokun (*Garcinia índica*), estes sendo sementes bastante encontradas em países e continentes de clima tropical, principalmente no oeste da África (OLIVEIRA, 2016; SMITH, 2001).

As CBRs são gorduras com constituintes por frações não láuricas mas que apresentam hidrogenação e fracionamento, além de alto conteúdo de ácidos graxos do tipo *trans*, entretanto, apresentam triacilgliceróis (TAGs) diferentes da MC, limitando a compatibilidade ao máximo de 30% na fase de gordura (PANTZARIS; BASIRON, 2002). As principais fontes de CBR são os óleos de palma, soja, canola, girassol, algodão e a oleína de shea. Já o grupo de CBS, não apresentam compatibilidade com a MC, onde são constituídas de frações láuricas ou por óleos líquidos parcialmente hidrogenados, com altas quantidades de ácidos graxos saturados como o ácido láurico e são provenientes do óleo de palmiste, óleo de coco e babaçu (OLIVEIRA, 2016; SMITH, 2001). As propriedades das gorduras alternativas estão descritas na Tabela 1.

Tabela 1. Propriedades dos subgrupos de gorduras vegetais alternativas.

Propriedades	CBE	CBR	CBS
Tipos de ácidos graxos	Gordura não láurica	Gordura não láurica	Gordura láurica
Propriedades físicas e químicas	Similar as propriedades químicas e físicas da MC, como perfil de fusão e polimorfismo	A distribuição de ácidos graxos é similar a MC, mas a estrutura de triglicerídeos é completamente diferente	Quimicamente diferente da MC, com algumas semelhanças físicas
Mistura com MC	Compatível em qualquer nível de mistura, sem alterar propriedades da MC	Apenas níveis baixos podem ser misturados com MC	Devido incompatibilidade, é indicada para substituição total da MC

Fonte: VERSTRINGE et al, 2012; NAIK; KUMARB, 2014 apud SOUZA, 2017.

Segundo a Codex Alimentarius (1981a, p. 6): “O uso de gorduras vegetais além da manteiga de cacau, de acordo com as disposições da Seção 2, deve ser indicado no rótulo em associação com o nome e/ou a representação do produto.” Apesar de no chocolate poder apresentar substituintes da gordura, a manteiga de cacau não pode, em si, haver substituintes, conforme a Codex Alimentarius (1981b, p. 1): “Nenhum aditivo é permitido neste produto.”

Embora todos estes tipos de substituições serem amparadas pela lei, de acordo com a Legislação Brasileira, aqueles que possuem outro tipo de gordura na formulação devem obter uma denominação de “chocolate composto” no rótulo, sendo que, qualquer rótulo que não apresentar as verdadeiras proporções da gordura na mistura da composição final do produto, é considerado adulterado e fraudulento (ANVISA, 2020b; GURDENIZ, 2008; BARIS et al, 2017).

A banha de porco (BP) trata-se, segundo RIISPOA (1952): “Entende-se por banha, genericamente, o produto obtido pela fusão de tecidos adiposos frescos de suínos ou de matérias-primas”, proveniente de gordura animal derivada de suínos. A

banha possui ácidos graxos, sendo boa parte saturados, compostos por ácido palmítico, esteárico e mirístico, e como insaturados o ácido oleico, linoleico e palmitoleico (SILVA; GIOIELLI, 2006). A utilização de BP como adulterante em alguns óleos vem sendo estudada, já que produz uma gordura barata que pode ser facilmente adicionada a gordura vegetal sem haver percepção do consumidor, e assim diminuindo o custo do produto final sem ao menos ter sua presença descrita no rótulo, o que a torna uma mistura ilegal (ROHMAN; CHE MAN, 2010).

Não foi comprovado nenhum malefício a saúde causado pela BP, porém esta não trata-se de gordura vegetal, mas sim animal, além de aumentar os níveis de colesterol do alimento (MARIKKAR et al, 2003). Portanto, a presença da BP se torna um problema quando adicionado irregularmente a MC, devido a sua composição de ácidos graxos e alta quantidade de gorduras saturadas e insaturadas, podendo causar problemas de saúde como alergias e aumento do colesterol.

Sua adição ilícita em alimentos também pode ferir a confiança e muitas vezes os costumes do consumidor, já que, há religiões que não permitem o consumo de qualquer derivado de suínos como, por exemplo, a halal e kosher (MANSOR et al, 2011), além da cultura vegana, que segundo Schinaider (2018, p. 26) “Os veganos (consumidores veganos) não consomem nenhum alimento de origem animal – carnes, laticínios, ovos, mel – e combatem essa exploração animal em vestuários, entretenimento, trabalho e experimentação animal.” e que vegano “não come alimentos de origem animal, carnes de todas as cores e tipos, ou que contenham qualquer resíduo: leites, queijos, salsichas, ovos, mel, banha, manteiga, etc.” (SOUZA, 2015; SCHINAIDER, 2018).

Devido a este tipo de adulteração irregular, vários países tem investido em técnicas e métodos que consigam identificar este tipo de fraude em produtos alimentícios, e o Brasil não fica de fora, pesquisa-se sobre adulterações em mel (AGUIAR, 2018), adulterações em suplementos alimentares (RIBEIRO, 2015), adulterações em bebidas alcólicas como vodcas (TEIXEIRA JÚNIOR, 2018).

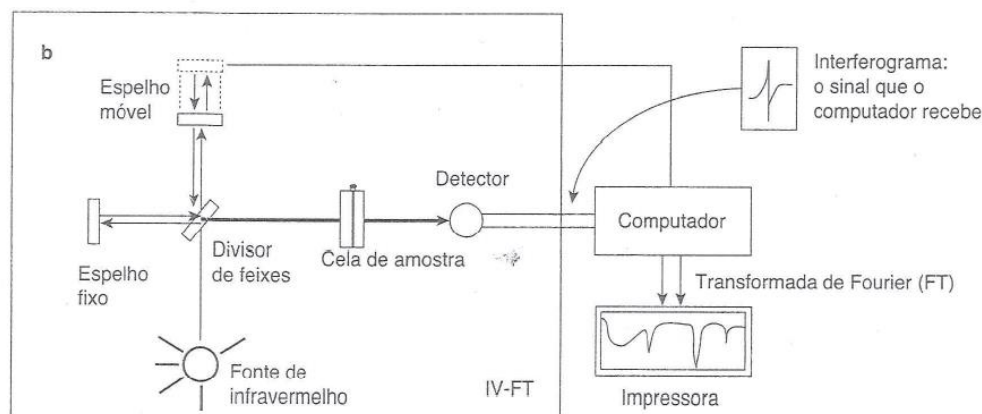
2.3. Espectroscopia de Infravermelho Médio (MIR)

A espectroscopia no infravermelho (IR) é uma técnica que consiste no estudo da interação da radiação eletromagnética, da região do infravermelho em um espectro eletromagnético, em forma de transições vibracionais ou rotacionais nas ligações entre os átomos de uma molécula (BARBOSA, 2007). A radiação do IR se concentra entre as regiões do visível e do micro-ondas no espectro, possui tamanho de aproximadamente 750 nm à 50000 nm. O IR é dividido em três partes: Infravermelho próximo (número de onda de 12800 a 4000 cm^{-1}); Infravermelho médio (4000 a 200 cm^{-1}) e Infravermelho distante (200 a 10 cm^{-1}), sendo que a região de maior uso se encontra na região do infravermelho médio (MIR) (PAVIA et al., 2010).

O espectrofotômetro de infravermelho é o nome dado ao equipamento em que se obtêm os espectros com absorção na região infravermelho, havendo dois tipos, sendo eles os instrumentos dispersivos e o de Transformada de Fourier (FT). Mesmo que os dois sejam bem próximos e apresentem espectros em uma região de 4000 a 400 cm^{-1} , o FT é mais rápido e apresenta vantagens, como sua maior característica uma operação matemática que permite a conversão de um sinal senoidal em frequência, simplificando as ondas coincidentes, amplificando os sinais de um composto no espectro e aproveitamento da energia radiante (PAVIA et al, 2010).

O equipamento de infravermelho FT, é composto por uma fonte de radiação, um separador de feixes, dois espelhos, um sendo fixo e o outro móvel, a célula em que se encontra a amostra, o detector, o conversor de sinal, um computador e o registrador. A radiação emitida pela fonte divide-se em dois feixes pelo separador de feixes, então o espelho fixo e o móvel refletem os feixes novamente para o separador, dessa forma obtêm-se uma sequência de interferências construtivas e destrutivas. Essas interferências variam devido o caminho óptico distinto ocasionado pelo movimento do espelho móvel, os feixes se combinam, atravessam a amostra e são detectados como um interferograma, contendo todas as informações do infravermelho na amostra, gerando assim, um espectro. As posições variadas do espelho, gera um espectro completo de infravermelho (SILVA, 2016; SILVERSTEIN et al, 2012). A Figura 4 apresenta o diagrama de um equipamento de infravermelho FT.

Figura 4. Diagrama esquemático de espectrofotômetro de transformada de Fourier.



Fonte: PÁVIA et al., 2010.

O espectro apresentado por um composto na técnica de infravermelho pode apresentar certa complexidade e geralmente haverá bandas características de determinados grupos funcionais ali presentes. Para o químico, estas bandas se tornam um grande recurso, permitindo uma análise detalhada, podendo então, determinar com certa precisão a identificação correta de amostras desconhecidas a partir de um composto padrão e conseguir efetivamente estipular as estruturas possíveis, já que, cada ligação haverá frequência de suas vibrações diferentes, e será dubitável que moléculas diferentes (que não sejam enantiômeras) apresentam um espectro igual (SILVERSTEIN et al, 2012). Para identificação, as regiões de 4000 a 200 cm^{-1} são tidas como as mais importantes, e dentre elas a faixa de 1200 a 600 cm^{-1} são consideradas bem úteis na análise, já que se apresenta como uma região de impressão digital, pois qualquer mudança na estrutura da molécula, será perceptível nesta região (HOLLER; SKOOG; CROUCH, 2009).

Devido a sua capacidade de identificação de compostos desconhecidos, o MIR vem apresentando muitas vantagens quando comparado a outras técnicas de identificação de compostos, como a Cromatografia Gasosa (CG), Espectrometria de Massas (EM) e Ressonância Magnética Nuclear (RMN) (GUIMARÃES, 2018). Suas vantagens são: rapidez e precisão na obtenção de um espectro; dispensabilidade de preparo de amostras, podendo utilizar tanto sólidos, líquidos, gases, materiais amorfos e polímeros; trata-se de uma técnica não destrutiva, conservando a amostra; tem grande sensibilidade em suas análises; equipamento relativamente pequeno e com preço acessível em relação a outras técnicas (MONTEIRO, 2007).

Devido a suas vantagens, o MIR vem sendo aplicado em diferentes trabalhos de identificação, como em áreas do ramo alimentício, apresentando trabalhos com grande relevância, alguns deles sendo: avaliação do perfil químico de vinhos e sucos de uvas tintas da região Sudeste do Paraná (BICAS et al, 2017), aplicação do MIR em detecção de adulterantes em produtos à base de frango (MARTINS, 2015); áreas medicinais: aplicação do MIR no controle de qualidade de medicamentos contendo diclofenaco de potássio (SOUZA; FERRÃO, 2006), aplicação do MIR em análise direto de tetraciclina em leite bovino (ARAUJO, 2007); áreas de cosméticos: determinação de quantificação de surfactantes em bactericidas (AMENDOLA, 2007); área de combustíveis: MIR e calibração multivariada para quantificação de biodiesel de soja em misturas de diesel (GONTIJO et al., 2014), aplicado na quantificação de biodiesel de *Jatropha* em misturas com diesel (SITOE et al, 2019).

Mesmo com todas suas vantagens e aplicabilidades, a interpretação do MIR, para muitos casos, ainda é complexa, e o desenvolvimento de métodos quimiométricos facilitou a compreensão e trouxe um tratamento estatístico e químico para os espectros, contribuindo com a retirada e tratamento de dados úteis para análise.

2.4. Métodos de Análise Multivariada

A quimiometria foi criada pelo químico Svante Wold em 1974 e seu surgimento se deu pela busca por ferramentas estatísticas, matemáticas e computacionais que convertessem conjunto de dados químicos em informações úteis. Segundo Malinowski (1991), a quimiometria pode ser definida como “uso de métodos matemáticos e estatísticos para o tratamento, a interpretação e previsão de dados químicos.” (FERREIRA et al., 1999; FERREIRA, 2015; WORKMAN et al., 2011).

Para uma análise eficiente dentro da quimiometria, deve-se buscar o desenvolvimento de modelos apropriados para quantificar uma propriedade que seja de interesse dentro de um sistema químico, satisfazendo os objetivos do analista (BARROS NETO et al, 2002). A maioria dos instrumentos de espectrofotometria fornecem como resposta a intensidade da luz medida pelo detector (absorção ou emissão de luz), algo muito subjetivo, que indiretamente pode fornecer uma variável química, que na maioria das vezes, se trata da concentração do composto na amostra.

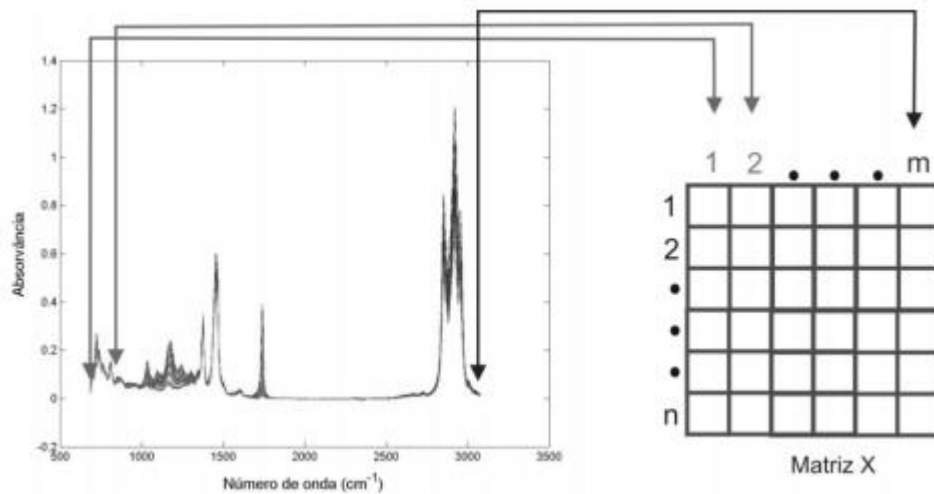
Para que o equipamento transforme a medida da luz em concentração, deve-se selecionar o conjunto de amostras em que as concentrações das espécies presentes sejam conhecidas, registrar no equipamento e realizar os cálculos concatenando as variáveis e criar um modelo matemático, denominado de regressão linear ou curva de calibração. Esse processo é conhecido como calibração (FERREIRA, 2015).

Calibração é dividida em dois grupos: calibração univariada e calibração multivariada. A calibração univariada é a mais utilizada em pesquisas e de fácil aplicação, onde há apenas uma medida instrumental para cada amostra de calibração. Mesmo sendo uma das mais aplicadas, ela é restrita, já que grandezas como absorvância de um composto, não deve haver qualquer interferente, pois este provoca desvios de linearidade no sistema (BRAGA, POPPI, 2004).

Na calibração multivariada duas ou mais respostas instrumentais são relacionadas com a propriedade de interesse, possibilitando que seja aplicado em análises com a presença de interferentes em seu meio, contudo, estes interferentes devem estar presentes nas amostras utilizadas no modelo de calibração, o que exige que haja atualizações no modelo multivariado. Mesmo com essa especificidade, este modelo se torna um bom substituto ao univariado, já que tem como vantagem sua aplicação em compostos com interferentes, rapidez e eficiência ao gerar resultados e realiza análises mesmo sem alta resolução (LIMA, 2015).

A resposta instrumental neste modelo de calibração é representada por uma matriz de dados (\mathbf{X}) e a propriedade de interesse, por um vetor (\mathbf{y}). Na Figura 5 está apresentado uma matriz \mathbf{X} de m colunas (variáveis, neste caso números de onda) e n linhas (número de amostras). Nesta matriz \mathbf{X} serão inseridos os valores de absorvância de cada número de onda por amostra. E o vetor \mathbf{y} conterà a resposta de interesse, sendo a maioria os valores da concentração do analito na amostra de interesse.

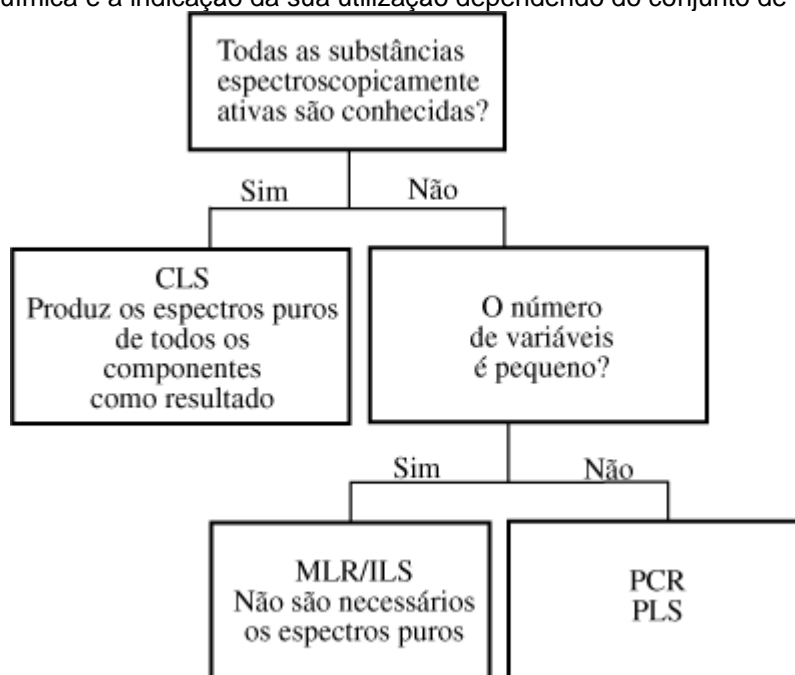
Figura 5. Representação da construção da matriz **X** para modelagem multivariada



Fonte: SOUZA, 2014.

A calibração multivariada é dividida em diferentes modelos e métodos, de acordo com os diferentes problemas e análises. Os mais populares são: Regressão por Quadrados Mínimos – Clássico (CLS); Regressão Linear Múltipla (MLR); Regressão por Componentes Principais (PCR); Regressão por Quadrados Mínimos Parciais (PLS). Dentre deles, o mais popular e presente em diversos trabalhos é o PLS (FERREIRA, 2015). A Figura 6, a seguir, mostra um fluxograma apresentando qual método multivariado é pertinente e indicado para o tipo de dado a ser analisado.

Figura 6. Fluxograma com esquema geral dos métodos multivariados mais populares dentro da Química e a indicação da sua utilização dependendo do conjunto de dados.



Fonte: FERREIRA et al, 1999.

2.5. Regressão por Quadrados Mínimos Parciais

A regressão por quadrados mínimos parciais é o método de regressão mais popular em toda a quimiometria e de grande aplicação nas espectroscopias de infravermelho. Esta foi desenvolvida por Herman Wold em 1970 voltada para a área de economia. A versão voltada para a quimiometria foi desenvolvida por seu filho Svante alguns anos depois (FERREIRA, 2015). Neste método, não há a necessidade de se conhecer todos os componentes presente na amostra, sendo capaz de realizar previsões mesmo na presença de interferentes, desde que, estes interferentes estejam presentes na construção do modelo (BEEBE; KOWALSKI, 1987).

O PLS trata-se de uma técnica de análise multivariada que relaciona inúmeras variáveis independentes (X) com uma ou mais variáveis respostas (Y), sempre baseando-se em fatores, em que X seria os valores de absorvância e Y a concentração ou outras propriedades da amostra (MACHADO, 2018). A principal função do PLS é construir modelos com capacidade preditiva baseando-se na existência de grande quantidade de variáveis independentes colineares (LUCENA, 2013).

O PLS é baseado na análise de componentes principais (PCA) em que se decompõe uma matriz de dados **X** e **Y** simultaneamente, em uma soma dos vetores **t** (escores) e **p** (pesos), conforme equações 1 e 2,

$$X = t_1 p_1^T + t_2 p_2^T + \dots + t_h p_h^T + E \quad \text{Equação 1}$$

$$Y = U Q^T + F \quad \text{Equação 2}$$

onde **E** é a parte não modelada na matriz **X**; **t** e **U** são matrizes de escores das matrizes **X** e **Y** respectivamente; **p** e **Q** são as matrizes de pesos de **X** e **Y**, respectivamente; **h** corresponde ao número de variáveis latentes (VL) e **E** e **F** correspondem às matrizes de resíduos (LENARDON,2016).

Para cada VL **h**, deve-se haver uma relação linear entre os escores de **X** e **Y** (Equação 3).

$$u_h = b_h t_h \quad \text{Equação 3}$$

onde **b_h** trata-se do vetor de coeficiente de regressão linear para cada VL, obtido através da equação 4.

$$b_h = \frac{u_h^T t_h}{t_h^T t_h} \quad \text{Equação 4}$$

ao obter a relação linear entre **t** e **u** juntamente com a minimização das matrizes de resíduos **E** e **F**, o melhor modelo terá sido encontrado.

O algoritmo NIPALS é o mais usado para construir o modelo de regressão por PLS, em que explicará a variância pelos valores das VLs, em que a primeira sempre será mais que a segunda VL, e assim por diante (GELADI; KOWALSKI, 1986). A variável latente (VL) é o fator que descreve a direção de variância máxima correlacionada com a concentração, sendo combinações lineares dos componentes principais (FERREIRA et al, 1999; LUCENA, 2013).

A VL é muito significativa para avaliação do melhor modelo PLS. É importante ter um número mediano de VLs na construção do modelo de calibração, pois uma

quantidade alta ajusta melhor um modelo, mas pode apresentar ruídos indesejados no espectro, já um número muito baixo de VL pode eliminar todos os ruídos, mas não modelará as informações desejadas e importantes (BORGES NETO, 2005).

Para que um modelo tenha validade, é necessário que o mesmo consiga prever, com bastante qualidade, a propriedade de interesse em novas amostras. Para isso, o modelo de calibração deve ser testado em um conjunto de amostra que tenha seus dados (**X** e **Y**) conhecidos, mas que não foram utilizados na construção da calibração, este é o que chamamos de validação externa (FERREIRA, 2015). A validação cruzada é o melhor método para se definir os padrões de validação da calibração. Nela são excluídos um ou mais números de amostras do conjunto de calibração do modelo PLS. A quantidade de amostras excluídas dependerá da quantidade de amostras de calibração utilizadas, normalmente, sendo 30% das amostras de calibração, se por acaso for mais de 20 amostras, se recomenda o método de “*venetian blinds*”, em que se retira do modelo duas ou mais amostras de uma só vez, e estes são previstas pelo modelo, sendo repetida até que todas as amostras sejam previstas. Após este processo calcula-se o erro quadrático médio de validação cruzada (RMSECV) através da equação 5,

$$RMSECV = \sqrt{\frac{\sum_{P=1}^P (y_P - \hat{y}_P)^2}{P}} \quad \text{Equação 5}$$

onde, y_P são os valores de referência para a propriedade de interesse, e o \hat{y}_P os valores previstos (FERREIRA, 2015; BRERETON, 2000).

Também se calcula o erro quadrático médio de previsão (RMSEP) e o erro quadrático médio de calibração (RMSEC), que são utilizados para a avaliar a capacidade do modelo de prever as amostras desconhecidas. O cálculo do RMSEP é semelhante ao do RMSECV, mas utiliza-se o número de amostras do conjunto teste (PAULO, 2013).

De nada adianta um modelo de calibração se sua capacidade de previsão da propriedade de interesse das amostras não for satisfatória. Devido a isso necessita-se fazer com que as novas amostras sejam obtidas, preparadas e processadas em equipamentos com um grau de semelhança aos da calibração de forma adequada.

Após isso, basta pré-processar os dados, prever a propriedade de interesse e aplicá-lo em novas amostras (FERREIRA, 2015).

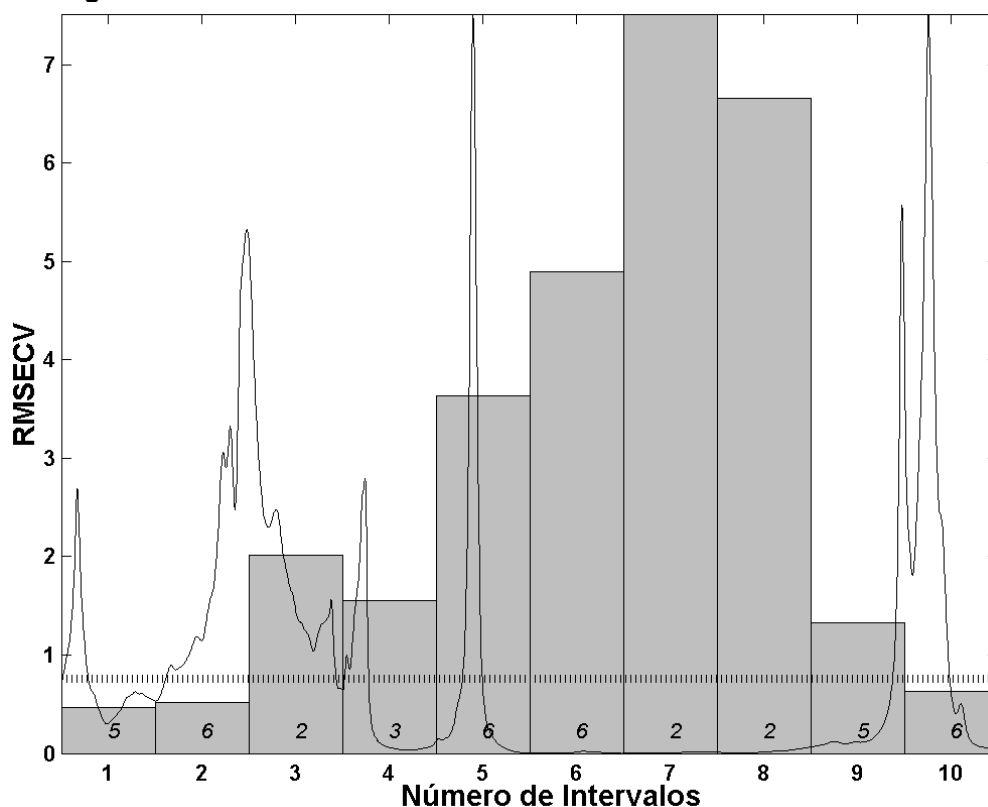
2.6. Seleção de Variáveis por Intervalos (iPLS, biPLS e siPLS)

Muitas vezes, selecionar algumas regiões espectrais dentro do modelo PLS pode ser satisfatório para melhorar o desempenho dos métodos de calibração, em que produzirá erros de previsão mais baixos, e poderá excluir do mesmo, ruídos não significantes. Seleciona-se um comprimento de onda específico, em cima dele produzirá modelos mais estáveis e de simples interpretação, onde a colinearidade não seja de total importância (FERNANDES et al., 2011; BALABIN; SMIRNOV, 2011).

Desta forma, diversas estratégias de seleção de variáveis surgiram dentro do PLS, dentre elas, quadrados mínimos parciais por intervalos (iPLS), quadrados mínimos parciais por exclusão de intervalos (biPLS) e quadrados mínimos parciais por intervalos sinérgicos (siPLS). Além destes, existem outros, como algoritmo genético (GA), algoritmo das projeções sucessivas para seleção de intervalos em PLS (iSPA-PLS), e outros mais, que não serão tratados neste trabalho, focando apenas nos três primeiros citados acima (FERNANDES, 2016).

O iPLS é uma extensão do PLS, em que se divide e seleciona regiões específicas no espectro com valores de validação menores que os do modelo global (PLS), construindo um modelo PLS particular para aquela região. Seu objetivo central é analisar todas as regiões espectrais do modelo global e conseguir apresentar as regiões em que as variáveis são mais relevantes e as que não estão dentro do interesse da pesquisa e/ou interferentes, estes dois últimos, podendo ser descartados e criando um modelo PLS apenas com as variáveis selecionadas (NØRGAARD et al, 2000). A Figura 7 a seguir, apresenta graficamente um estudo genérico de um modelo por iPLS.

Figura 7. Gráfico do modelo iPLS do número de intervalos versus RMSECV.



Fonte: A autora.

Na Figura 7 o espectro do modelo PLS global foi dividido em 10 intervalos equidistantes. As barras correspondem à estimativa dos valores de RMSECV para cada intervalo, os números dentro das barras são os valores das VLs, e a linha tracejada refere-se ao valor de RMSECV para o modelo global. A comparação dos modelos é feita em cima do parâmetro de validação RMSECV, ou seja, erros menores que a linha tracejada indicam melhores regiões para construir um novo modelo PLS (BORGES NETO, 2005; GONTIJO, 2016).

Além do modelo de iPLS, há também o biPLS, que trata-se de um método que divide o espectro em intervalos equidistantes como o iPLS, também com o objetivo de selecionar intervalos que tenha o melhor parâmetro, mas seu diferencial está em sua técnica, que após a região espectral ser dividida, aquele intervalo que apresentar o maior valor de RMSECV será descartado e construído um novo modelo com os restantes, será reavaliado e calculado um novo RMSECV para os demais restantes, e assim o processo se repetirá até sobrar apenas um intervalo (GONTIJO, 2016).

Tabela 2. Exemplo de Tabela de Exclusão obtida pelo biPLS em uma divisão espectral de 35 intervalos.

Número de Intervalos	Intervalo	RMSECV	Número de Variáveis
35	34	0.8963	2371
34	13	0.8437	2304
33	15	0.7965	2236
32	8	0.7471	2168
31	14	0.7019	2100
30	7	0.6578	2032
29	9	0.5997	1964
28	6	0.5718	1896
27	12	0.5531	1828
26	10	0.5476	1760
25	5	0.5441	1692
24	3	0.5413	1624
23	4	0.5353	1556
22	25	0.5323	1488
21	17	0.5313	1420
20	28	0.5304	1352
19	29	0.5296	1285
18	27	0.5285	1218
17	26	0.5281	1151
16	30	0.5274	1083
15	24	0.5268	1016
14	31	0.5266	948
13	23	0.5266	881
12	19	0.5268	813
11	18	0.5262	745
10	22	0.5270	677
9	20	0.5278	609
8	21	0.5296	541
7	11	0.5334	473
6	32	0.5356	405
5	16	0.5431	338
4	35	0.5659	270
3	33	0.6174	203
2	2	0.6337	136
1	1	0.7041	68

Fonte: A autora.

A Tabela 2 demonstra um exemplo de Tabela de Exclusão apresentado pelo modelo biPLS.. Neste contém o número de intervalos, o intervalo correspondente, o RMSECV do intervalo e por fim o número de variáveis. Selecionou o intervalo que

apresentou menor valor de RMSECV, e construiu-se o modelo com todos os intervalos abaixo do selecionado, no caso, intervalos [1:2:11:16:18:20:21:22:32:33:35], construindo a matriz **X**.

Deve-se analisar bem esta Tabela e interpretar os dados presentes, focando-se em alguns fatores, sendo eles: i) Os 35 intervalos que o modelo foi dividido (coluna 1); ii) primeira linha apresenta a estimativa do valor de RMSECV na presença de todos os intervalos no modelo (PLS global) e também indica o intervalo 34 como o que apresenta maior valor de RMSECV (coluna 2) e depois do 34, vem o intervalo 13, 15, e sucessivamente; iii) O valor de RMSECV presente na segunda linha da tabela, é um valor estimado para o modelo quando o intervalo 34 é excluído, na terceira linha, o valor é estimado para a exclusão do 34 e do 13, e assim sucessivamente; iv) o menor valor de estimativa de RMSECV corresponde a melhor combinação de intervalos, aqueles que juntos irão compor um novo modelo mais ajustado, neste caso, os intervalos 18, 22, 20, 21, 11, 32, 16, 35, 33, 2 e 1 serão a nova combinação de melhor RMSECV (0,5262); v) o número de variáveis (coluna 4) correspondente ao modelo inicia-se no menor número de intervalos, neste caso o 1, então irá de 1 até 68, o intervalo 2 de 69 até 136, o intervalo 33 de 2169 até e 2236, assim sucessivamente.

O siPLS, assim como o biPLS, trata-se de uma variação do iPLS, em que o analisador decide o número de divisões dos dados, e após isso, o algoritmo siPLS combina os intervalos em dois a dois, depois três a três, e por fim quatro a quatro, e o modelo PLS é calculado para cada uma das combinações, apresentando um valor de RMSECV estimado para cada um. Com o siPLS, é realizado todas as combinações possíveis, fornecendo uma tabela com os intervalos a serem combinados, valores de VLs e de RMSECV para cada combinação. A seguir, a Tabela 3 apresentando um modelo de resultado por siPLS.

Tabela 3. Exemplo de resultado de saída fornecido pelo algoritmo siPLS para uma combinação entre dois intervalos.

VL	Intervalos	RMSECV
4	[1:15]	0.6116
4	[1:5]	0.6224
4	[1:14]	0.6272
4	[1:3]	0.6450
4	[3:4]	0.6530
4	[1:4]	0.6598
4	[1:6]	0.6890
3	[1:9]	0.6978
3	[1:8]	0.6982
3	[1:10]	0.7072

Fonte: A autora.

A Tabela 3 exemplifica os resultados de saída de um modelo siPLS. Para esse conjunto de dados o estudo foi realizado dividindo os dados espectrais em 15 intervalos e realizando uma combinação de dois intervalos. Neste caso, a tabela mostra que a combinação dos intervalos 1 e 15 apresentou um menor valor de RMSECV e, portanto, sugere-se construir um novo modelo PLS com as variáveis destes intervalos.

Há bastante trabalhos que utilizam estes três modelos de seleção de variáveis, como sua aplicação em determinação do grau de viscosidade e teor de enxofre e nitrogênio em petróleo (SOARES, 2019), controle de qualidade de misturas de café arábica e robusta (ASSIS, 2018), caracterização de bebidas à base de soja (RECH, 2018), controle de qualidade de biodiesel de pinhão manso e de moringa em misturas com biodiesel (SITOE, 2019), na determinação de n,n-dietil-3-metilbenzamida em repelentes (SILVA, 2018), dentre outros mais.

2.7. Validação do Modelo

Uma das etapas importantes no desenvolvimento de metodologias é a validação. Em espectroscopia no infravermelho a norma ASTM E1655-17 (*American Society for Testing and Materials*) (2017) já apresenta algumas figuras de mérito necessárias para validar modelos quimiométricos por PLS.

A Exatidão demonstra o quanto os resultados encontrados estão próximos dos resultados aceitos como valores de referência ou reais (ANVISA, 2021). Para a exatidão a norma ASTM E1655 sugere o cálculo do RMSEP, conforme apresentado na equação 6,

$$RMSEP = \sqrt{\sum[(y' - y)^2/n]} \quad \text{Equação 6}$$

onde y' corresponde aos valores de referência, y aos valores previstos pelo modelo quimiométrico e n o número de amostras do conjunto de validação ou previsão.

Além do RMSEP, segundo a ASTM E1655-17, outros parâmetros podem avaliar a exatidão de um modelo, como testes t-Student para erros sistemáticos (*Bias*), já que este afeta diretamente a precisão e exatidão do modelo (MORSING; EKMAN, 1998). Segundo a IUPAC, o *bias* pode ser calculado pela diferença entre o valor real e o valor médio da população, sendo nenhum componente de erro aleatória. O teste t-Student deve ser aplicado a modelos que apresentam validação no nível de 95% para que avalie se possui um valor considerável de bias presente (VALDERRAMA, 2005). Calcula-se primeiramente o valor de bias médio para o conjunto de validação (Equação 7), depois o desvio padrão dos erros de validação (SDV) (Equação 8) e por fim, o valor do teste-t (Equação 9):

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

Em que n corresponde ao número de amostras do conjunto de validação.

$$SDV = \sqrt{\frac{\sum[(y_i - \hat{y}_i) - bias]^2}{n - 1}} \quad \text{Equação 8}$$

$$t_{bias} = \frac{|bias|\sqrt{n}}{SDV} \quad \text{Equação 9}$$

O valor de t_{bias} deve ser analisado, se apresentar um valor maior que o $t_{crítico}$ para n graus de liberdade com 95% de confiança, é indicativo que o valor dos erros sistemáticos deve ser considerado, pois sua presença no modelo será expressiva.

Entretanto, se o valor do t_{bias} for menor que o $t_{crítico}$, o erro sistemático no modelo pode ser desconsiderado (VALDERRAMA, 2005).

A linearidade e a comparação de modelos também são figuras de mérito importantíssimas para a validação de um modelo de calibração multivariada. A linearidade trata-se da capacidade dos resultados de um modelo serem proporcionais a concentração da amostra, onde a estimativa é dada a partir da correlação dos valores de referência e os valores estimados da propriedade de interesse, determinando-se uma reta (por quadrados mínimos) que melhor se ajusta ao modelo. Além da linearidade também pode-se avaliar o gráfico de resíduos, que verifica-se distribuição aleatória dos valores residuais das amostras de calibração e validação (PASSARI et al, 2011; VALDERRAMA et al, 2009).

No caso de comparação de modelos utiliza-se o teste F (equação 10) na comparação dos erros de previsão de modelos distintos, sendo $RMSEP_1 > RMSEP_2$.

$$F = \left(\frac{RMSEP_1}{RMSEP_2} \right)^2 \quad \text{Equação 10}$$

O $RMSEP_1$ corresponde ao valor de RMSEP do modelo global PLS, geralmente se tratando de um valor maior, o $RMSEP_2$ corresponde ao valor de RMSEP do modelo de seleção de variáveis (iPLS, biPLS ou siPLS), sendo este um valor menor. O F (também considerado como F_{calc}) é comparado com o valor estatístico de Fisher-Snedecor, onde compara-se os graus de liberdade com número de amostras utilizados na predição do modelo, tendo uma significância de 5%, em que o valor tabelado de F precisa ser superior ao de F_{calc} , caso não seja, não há evidências suficientes que este valor sigam a distribuição normal, sendo assim, o modelo com $RMSEP_2$ provém de melhor exatidão que o $RMSEP_1$ (FILGUEIRAS et al, 2014; GONTIJO, 2016).

3. MATERIAIS E MÉTODOS

3.1. Obtenção e Preparo das Amostras

A amostra de manteiga de cacau para fins culinários foi comprada no *site* do mercado livre da empresa Brotinho de Luz. A banha de porco (BP) foi adquirida no comércio local da cidade de Pires do Rio – GO.

Para preparação das amostras da mistura de MC/BP, primeiramente as amostras foram adicionadas em béqueres separados de 250 mL cada e fundidas em banho-maria a temperatura de $40\text{ }^{\circ}\text{C} \pm 2\text{ }^{\circ}\text{C}$. Após estarem totalmente líquidas, procedeu-se a mistura delas pesando-se uma massa correspondente de MC e BP afim de que as proporções entre elas pudessem ficar na faixa de 5,0 a 30,0% de banha de porco na manteiga de cacau. Na pesagem das amostras utilizou-se uma balança analítica de quatro casas decimais. O número de amostras preparadas nesta faixa totalizou 68 amostras, sendo 40 amostras separadas para o conjunto de calibração e 28 amostras para o conjunto de validação.

3.2. Obtenção dos Espectros

Os espectros MIR foram obtidos em triplicata, na região de 650 a 4000 cm^{-1} , resolução de 4 cm^{-1} e 16 varreduras, utilizando um espectrofotômetro FTMIR, marca PerkinElmer, modelo Spectrum Two, com acessório de reflectância total atenuada (HATR) (PikeTechnologies). As linhas de base espectral foram corrigidas empregando o método baseline na faixa de $1850 - 2570\text{ cm}^{-1}$ e $3200 - 4000\text{ cm}^{-1}$.

3.3. Análise Quimiométrica

A matriz de dados $m \times n$ foi construída, onde m é o número de amostra e n correspondem aos valores de absorbância em cada número de onda obtido, ou seja, de 650 a 4000 cm^{-1} . No entanto, devido aos sinais de ruído, foram eliminadas as variáveis correspondentes a 650 a 680 cm^{-1} e também a parte final do espectro, por não conter informações de interesse, ou seja, sem sinal analítico (3050 a 4000 cm^{-1}).

Para construção dos modelos de regressão por quadrados mínimos parciais (PLS) os dados foram centrados na média. No processo de validação externa foi

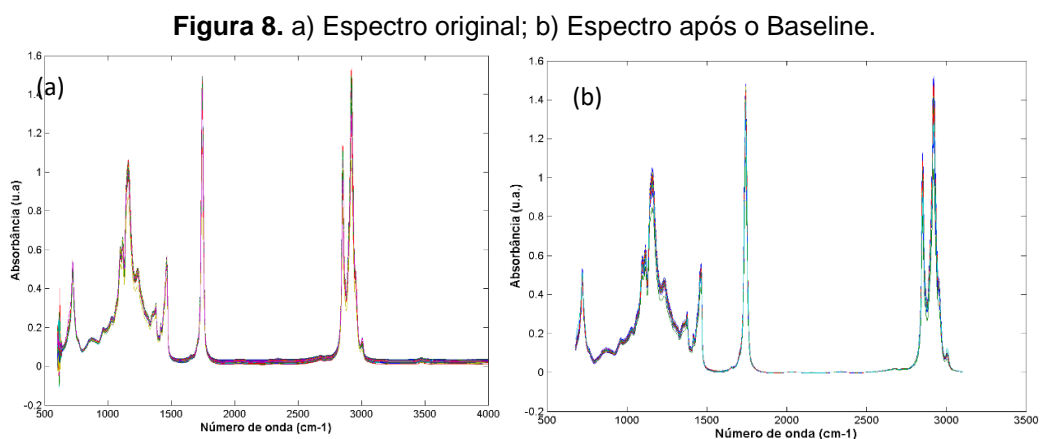
empregado o método por “*venetion blinds*”. O número de variáveis latentes foi escolhido conforme a porcentagem de capturança do bloco **X** (absorbâncias) e bloco **Y** (concentração) em comparação conjunta com o gráfico do erro quadrático médio de validação cruzada (RMSECV). A exatidão dos modelos foi determinada por meio do cálculo dos erros quadráticos médios de calibração (RMSEC), validação cruzada (RMSECV) e previsão (RMSEP). A existência de erros sistemáticos foi avaliada conforme diretrizes da norma ASTM E1655 (ASTM, 2017).

Construiu-se três modelos de seleção de variáveis por intervalos, sendo eles, a regressão por quadrados mínimos parciais por intervalos (iPLS), regressão por quadrados mínimos parciais por exclusão (biPLS) e regressão por quadrados mínimos parciais por sinergismo de intervalos (siPLS). Foram avaliados a divisão da região espectral em 5, 10, 15, 20, 25, 30 e 35 intervalos, e o modelo que apresentou capacidade preditiva melhor que o modelo PLS foi adotado como sendo o melhor modelo para quantificar banha de porco em manteiga de cacau.

O software MATLAB 6.1 (Mathworks Inc.) e os pacotes PLS Toolbox 3.5 e iToolbox 1.0 foram utilizados na concepção e execução dos procedimentos multivariados.

4. RESULTADOS E DISCUSSÕES

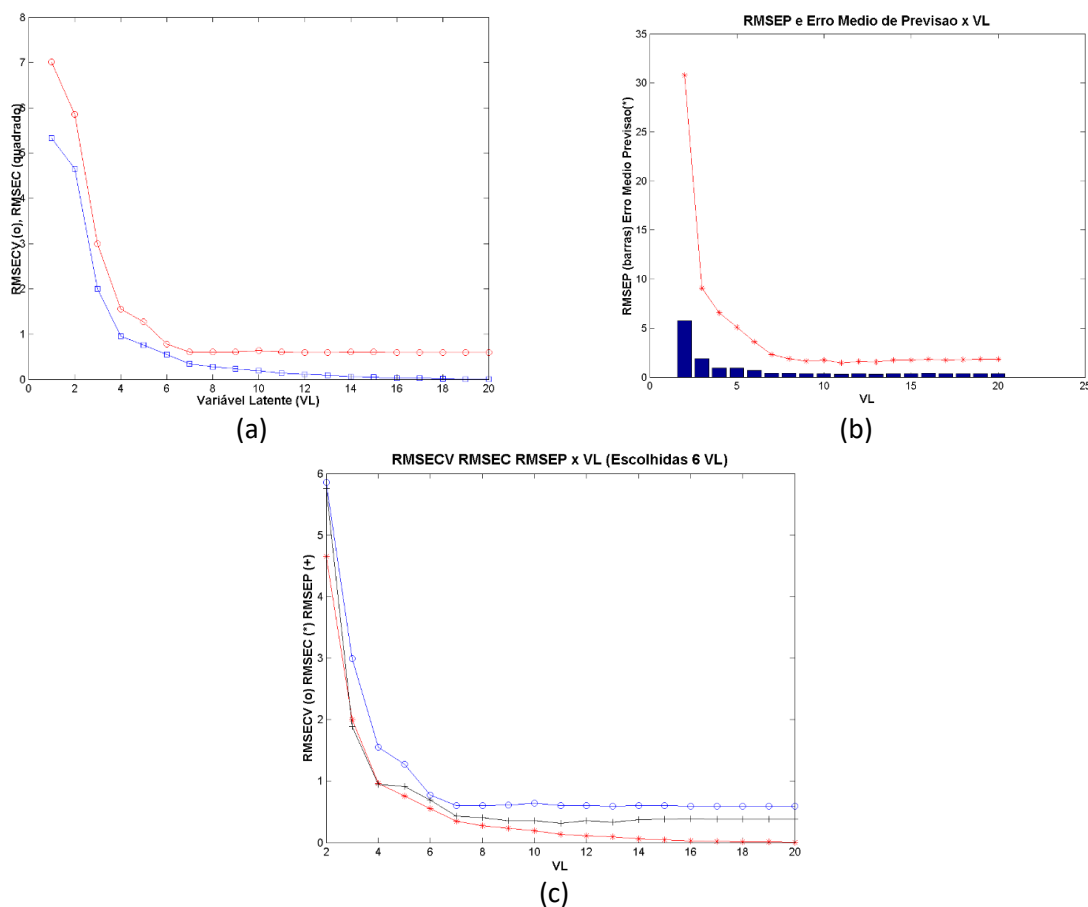
Na Figura 8 estão apresentados os espectros de infravermelho médio proveniente da MC adulterada com BP. A Figura 8a apresenta os espectros originais, sem nenhum ajuste. A Figura 8b apresenta os espectros após o processo denominado *baseline*. Neste processo, aplica-se um método matemático de derivada de uma constante, já que a derivada de uma constante é igual a zero, assim, retira-se partes do espectro em que se apresentam ruídos inerentes ao próprio equipamento ou nenhuma banda de absorção (MACHADO, 2018). Todo este processo tem a finalidade de obter apenas as respostas analíticas relevantes, minimizando e eliminando possíveis erros na construção do modelo. Além disso, as variáveis de 650 a 680 cm^{-1} e 3050 a 4000 cm^{-1} foram retiradas dos dados espectrais, conforme mostrado na Figura 8.



Fonte: A autora.

Após os tratamentos realizados nos espectros da MC/BP, o modelo PLS passou a ser construído. Na Figura 9 é apresentado o processo de escolha de variáveis latentes (VL) e a correlação entre os valores de erros obtidos.

Figura 9. a) RMSECV e RMSEC para cada VL; b) RMSEP e Erro Médio de Previsão para cada VL; c) RMSEC, RMSECV e RMSEP para cada VL.



Fonte: A autora

A Figura 9c mostra os valores de RMSECV, RMSEP e RMSEC para cada número de VL utilizada na construção do modelo, no qual observa-se que há uma queda acentuada dos erros até a quarta VL. A partir desta, deve-se verificar a tabela de capturância dos dados dos blocos X e Y, para não correr o risco de sobreajustar o modelo. O sobreajuste ocorre quando informações irrelevantes são adicionadas na construção do modelo. Isto é percebido por meio do ganho de informação ao modelo quando se adiciona uma nova variável latente. A tabela 4 apresenta a porcentagem de capturância na construção deste modelo.

Tabela 4. Capturância para cada VL.

VL	X-Block		Y-Block	
	Informação	Total	Informação	Total
1	29.56	29.56	40.78	40.78
2	65.85	95.41	13.98	54.76
3	2.51	97.92	36.90	91.66
4	0.40	98.32	6.42	98.08
5	0.90	99.22	0.72	98.80
6	0.65	99.87	0.56	99.36
7	0.03	99.90	0.39	99.75
8	0.02	99.92	0.10	99.84
9	0.02	99.94	0.04	99.88
10	0.01	99.95	0.04	99.92
11	0.01	99.95	0.04	99.96
12	0.01	99.96	0.01	99.97
13	0.01	99.97	0.01	99.98
14	0.00	99.97	0.01	99.99
15	0.01	99.98	0.00	100.00
16	0.00	99.98	0.00	100.00
17	0.00	99.98	0.00	100.00
18	0.00	99.99	0.00	100.00
19	0.00	99.99	0.00	100.00
20	0.00	99.99	0.00	100.00

Fonte: A autora

Observando a Tabela 4, percebe-se que a partir da VL 6 (a escolhida para o modelo), o ganho de informação para o modelo é baixo, e que no bloco **X** com 7 VL acrescentou apenas 0.03 de informação. É notável que com 18 VL, no bloco **X** apresentou um total de informações do modelo de 99.99 e no bloco **Y** de 100.00, mas este demonstrou valores de erro (RMSECV, RMSEC e RMSEP) distantes uns dos outros, e é importante que os valores destes sejam próximos para um modelo mais ajustado e com baixos erros.

Verificando a Tabela 4 percebe-se que até a sexta variável latente ainda é possível ganhar informações úteis ao modelo, sendo que a partir desta, percebe-se que não há mais ganho de informações relevantes. Portanto, optou-se em construir o modelo PLS utilizando as 6 VLs.

Os parâmetros de exatidão e linearidade do modelo PLS estão apresentados na Tabela 5.

Tabela 5. Valores de VL, R, RMSECV, RMSEC e RMSEP para o modelo PLS construído.

VL	R	RMSECV	RMSEC	RMSEP
6	0,99661	0,7756	0,5555	0,6917

Fonte: A autora.

Verifica-se na exatidão que os valores de RMSECV, RMSEC e RMSEP estão próximos, indicando que não há sobreajuste ou falta de ajuste ao modelo. Ou seja, à medida que se adiciona mais VLS ao modelo a tendência é o RMSEC ficar cada vez menor. Porém o RMSEP não acompanha essa diminuição indicando que o modelo está sobreajustado (neste caso).

Outro parâmetro de exatidão indicado é a elipse de confiança (EJCR), que diferente dos erros quadráticos, é livre de erros sistemáticos. Segundo Guimarães et al. (2015), o EJCR trata-se de um gráfico, em um formato elíptico, onde há comparação do intercepto e a inclinação de valores reais em função dos valores previstos entre os valores ideais, estabelecendo uma relação de magnitude dos erros experimentais e o grau de confiança escolhido.

Em relação a linearidade para que um modelo seja ideal, a linearidade destes pontos deveria ser perfeita, apresentando um coeficiente de correlação (R) próximo ou igual ao valor 1. A Tabela 6 a seguir mostra a interpretação de cada modelo quando analisado o R^2 do mesmo:

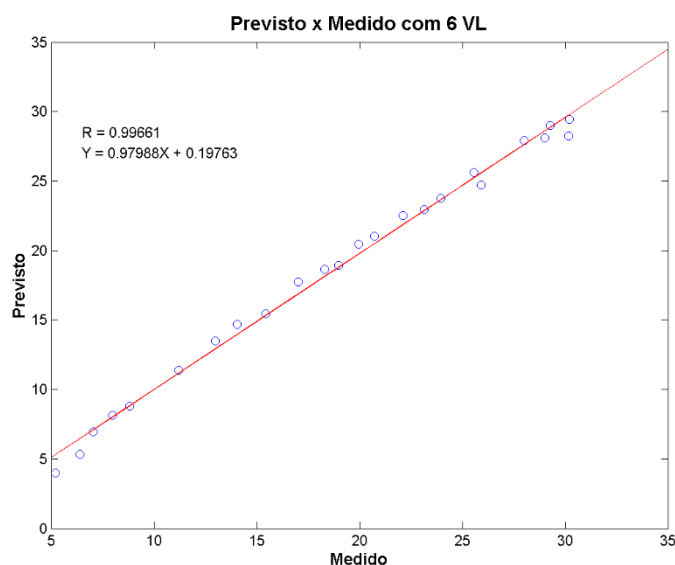
Tabela 6. O tipo de correlação associada aos valores de R^2

R^2	Correlação
=1	Perfeita
0,91<R<0,99	Fortíssima
0,61<R<0,91	Forte
0,31<R<0,60	Média
0,01<R<0,30	Fraca
=0	Nula

Fonte: MACHADO, 2018.

A Figura 10 apresenta o gráfico da linearidade do modelo PLS.

Figura 10. Representação gráfica dos valores previstos versus os reais para o modelo PLS.



Fonte: A autora.

Após a construção do modelo PLS estudou-se a aplicação de métodos de seleção de variáveis por intervalos (iPLS, biPLS e siPLS) com o objetivo de encontrar modelos com melhor capacidade preditiva.

Os resultados dos modelos iPLS obtidos das divisões espectrais em 5, 10, 15, 20, 25, 30 e 35 intervalos estão apresentados na tabela 7.

Tabela 7. Resultados para iPLS em cada divisão espectral.

Modelo iPLS	Intervalo selecionado	VL	Variáveis	Intervalo (cm ⁻¹)	RMSECV	RMSEC	RMSEP	R
iPLS 5	1	6	485	680-1164	0.4835	0.3474	0.3811	0.99889
iPLS 10	1	5	243	680-922	0.5204	0.3023	0.6186	0.99747
	10	4	241	2859-3100	0.9392	0.4882	0.4425	0.99871
iPLS 15	1	4	162	680-841	0.5439	0.3550	0.5719	0.99786
	3	6	161	1004-1165	0.4846	0.4065	0.4626	0.99849
	15	6	160	2940-3100	0.8459	0.4829	0.5025	0.99816
iPLS 20	1	4	122	680-801	0.5622	0.3636	0.5332	0.99810
	4	6	120	1044-1164	0.4276	0.3198	0.4592	0.99837
iPLS 25	1	5	97	680-776	0.5849	0.3558	0.5454	0.99800
	4	4	96	971-1067	0.5195	0.4513	0.6614	0.99734
	5	5	96	1068-1164	0.4410	0.3745	0.3805	0.99889

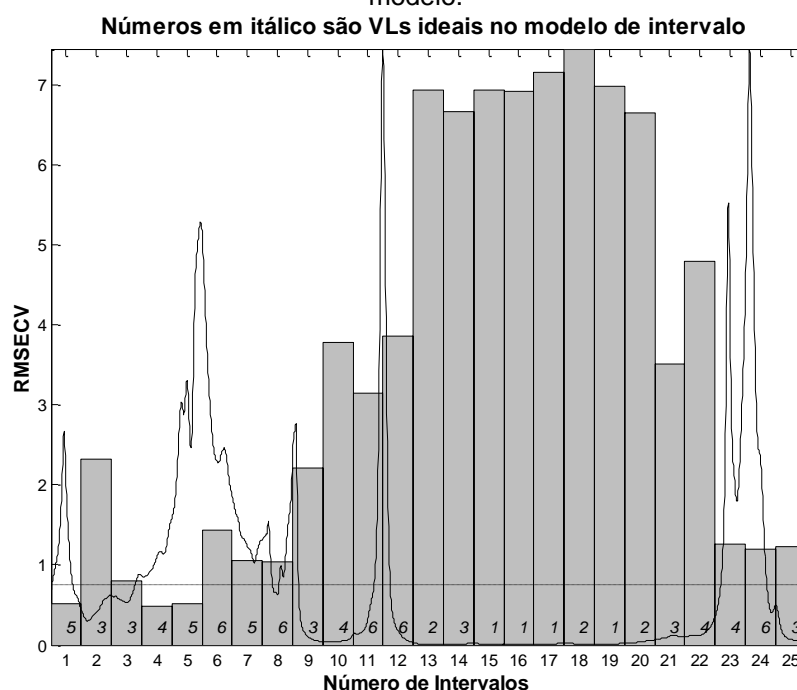
Tabela 7. Resultados para iPLS em cada divisão espectral.

iPLS 30	1	3	81	680-760	0.5869	0.4884	0.7162	0.99683
	5	4	80	1004-1084	0.6169	0.5688	0.6270	0.99693
	6	6	80	1085-1165	0.4133	0.3122	0.4794	0.99822
iPLS 35	1	4	70	680-749	0.5857	0.3993	0.5981	0.99788
	6	2	68	1003-1099	0.8506	0.5766	0.6768	0.99734
	7	5	68	1100-1168	0.5533	0.4708	0.7985	0.99539

Fonte: A autora

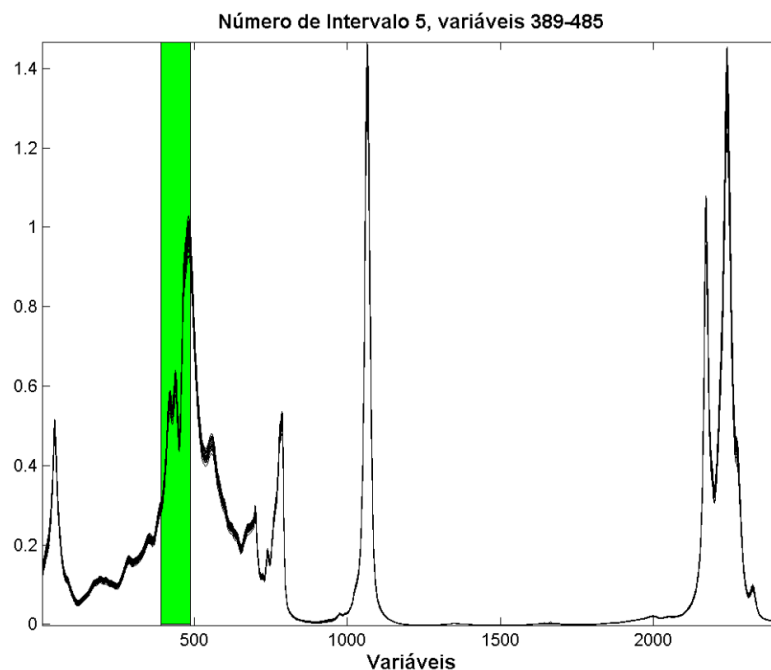
Analisando a Tabela 7 o modelo de regressão por quadrados mínimos parciais por intervalos (iPLS) que apresenta melhor capacidade preditiva é o modelo iPLS25. Neste caso, empregou-se a divisão espectral em 25 intervalos equidistantes e a região que apresentou menor erro de RMSECV se comparado ao modelo PLS global (todas as variáveis, ou seja, 2421) foi a região ou intervalo 5. Este intervalo refere-se às variáveis de 1068 – 1164 cm^{-1} (97 variáveis) A Figura 11 demonstra os resultados da aplicação do iPLS25 e a Figura 12 destaca o intervalo selecionado para construção do novo modelo PLS.

Figura 11. Espectro dividido em 25 intervalos equidistantes, cada um com o valor de VL ideal para o modelo.



Fonte: A autora.

Figura 12. Intervalo número 5 selecionado do modelo iPLS com variáveis de 389-485.



Fonte: A autora.

Os resultados dos modelos biPLS obtidos das divisões espectrais em 5, 10, 15, 20, 25, 30 e 35 intervalos estão apresentados na Tabela 8.

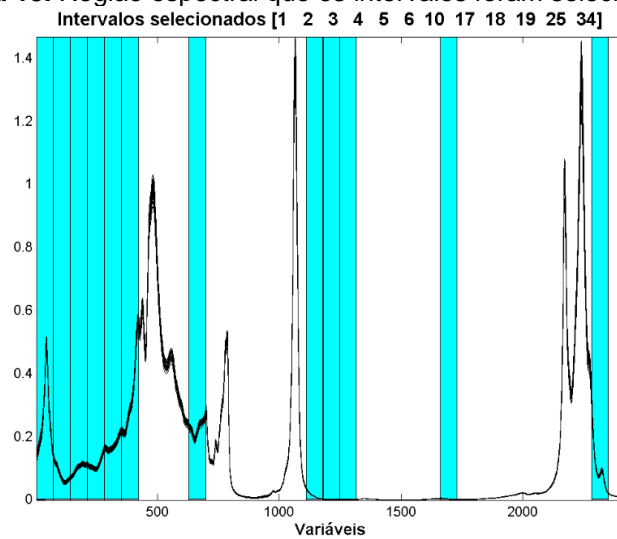
Tabela 8. Resultados obtidos para os modelos biPLS.

Modelos biPLS	Intervalos selecionados	Variáveis	RMSECV	RMSEC	RMSEP	R
biPLS5	[1]	485	0.4835	0.3474	0.3811	0.99889
biPLS10	[1:2]	727	0.4835	0.3474	0.3811	0.99889
biPLS15	[1:15]	485	0.4679	0.3818	0.5130	0.99822
biPLS20	[1:2:3:10:20]	722	0.4294	0.2637	0.4296	0.99865
biPLS25	[1:2:4:13:14:18:25]	775	0.4174	0.2744	0.3897	0.99893
biPLS30	[1:2:5:8:9:16:21:29]	809	0.4118	0.3070	0.4046	0.99877
biPLS35	[1:2:3:4:5:6:10:17:18:19:25:34]	903	0.4383	0.2739	0.3663	0.99898

Fonte: A autora.

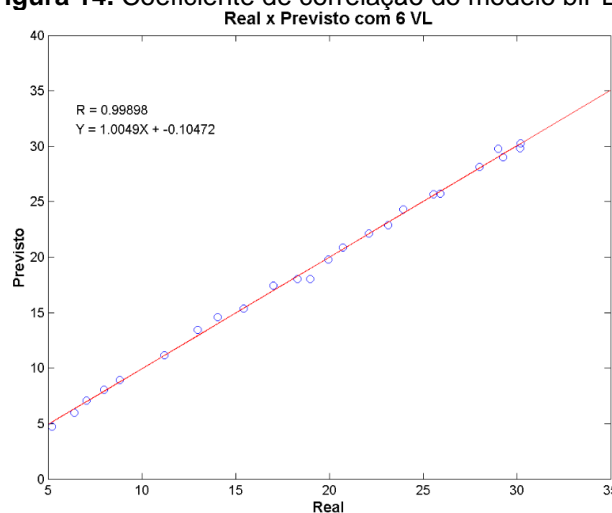
De acordo com a Tabela 8 verifica-se que o melhor modelo de regressão por quadrados mínimos parciais por exclusão de intervalar (biPLS), foram os intervalos [1:2:3:4:5:6:10:17:18:19:25:34], apresentando uma excelente capacidade preditiva. Neste, empregou-se a divisão espectral em 35 intervalos equidistantes, onde apresentou um menor valor de RMSEP e quantidade de variáveis (903 variáveis) comparado com o modelo PLS global, e os demais foram excluídos devido seu alto valor de RMSECV. A Figura 13 mostra os intervalos selecionados e a Figura 14 apresenta a correlação entre os valores previstos e medidos.

Figura 13. Região espectral que os intervalos foram selecionados.



Fonte: A autora.

Figura 14. Coeficiente de correlação do modelo biPLS.



Fonte: A autora.

Os resultados dos modelos siPLS obtidos das divisões espectrais em 5, 10, 15, 20, 25, 30 e 35 intervalos com as combinações 2, 3 e 4 estão apresentados na Tabela 9.

Tabela 9. Resultados dos modelos siPLS para cada divisão espectral.

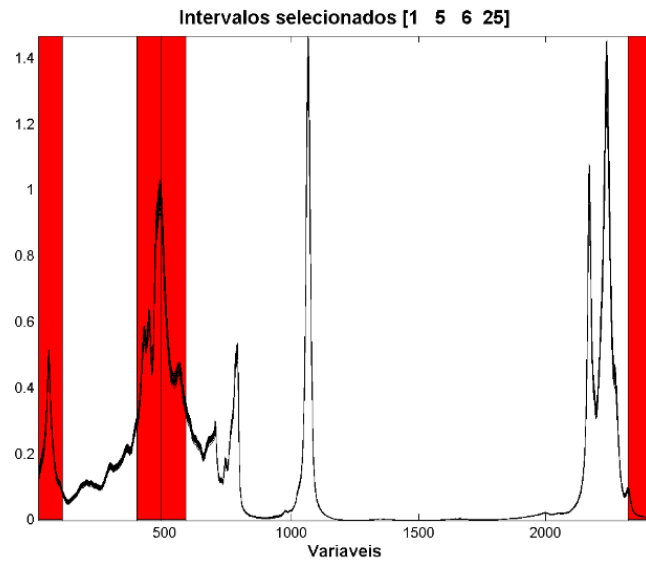
Modelos siPLS	Combinações de Intervalos	Intervalo	RMSECV	RMSEC	RMSEP	R
siPLS 5	Combinação 2	[1:3]	0.4771	0.3097	0.4082	0.99876
	Combinação 3	[1:3:4]	0.4779	0.3160	0.4057	0.99878
siPLS 10	Combinação 2	[1:2]	0.4835	0.3475	0.3811	0.99889
	Combinação 3	1:2:8	0.4747	0.3071	0.4560	0.99838
	Combinação 4	1:2:6:8	0.4885	0.3023	0.4551	0.99833
siPLS 15	Combinação 2	2:3	0.4161	0.3581	0.4380	0.99858
	Combinação 3	1:4:15	0.3893	0.3041	0.3866	0.99894
	Combinação 4	1:4:10:15	0.3897	0.3040	0.3866	0.99894
siPLS 20	Combinação 2	4:20	0.4791	0.3488	0.4071	0.99881
	Combinação 3	1:5:20	0.3969	0.3230	0.4009	0.99886
	Combinação 4	1:2:10:20	0.4416	0.2855	0.4451	0.99861
siPLS 25	Combinação 2	3:5	0.4445	0.3656	0.4367	0.99856
	Combinação 3	1:6:25	0.3821	0.3158	0.4036	0.99877
	Combinação 4	1:5:6:25	0.4317	0.2622	0.3326	0.99914
siPLS 30	Combinação 2	6:8	0.3707	0.2734	0.3726	0.99826
	Combinação 3	3:4:6	0.3604	0.2651	0.4721	0.99826
	Combinação 4	2:3:4:6	0.3924	0.2617	0.3801	0.99890
siPLS 35	Combinação 2	7:14	0.4343	0.3283	0.6505	0.99727
	Combinação 3	1:10:34	0.4236	0.2907	0.3654	0.99901
	Combinação 4	1:2:7:16	0.4785	0.3197	0.3767	0.99891

Fonte: A autora

Segundo a Tabela 9 o melhor modelo de regressão por quadrados mínimos parciais por sinergismo (siPLS) foi o que empregou-se a divisão espectral em 25 intervalos equidistantes, combinando-os em quatro em quatro, sendo eles os intervalos 1, 5, 6 e 25 apresentando um baixo valor de RMSEP e quantidade de variáveis (388 variáveis) comparado com o modelo PLS global. A Figura 15,

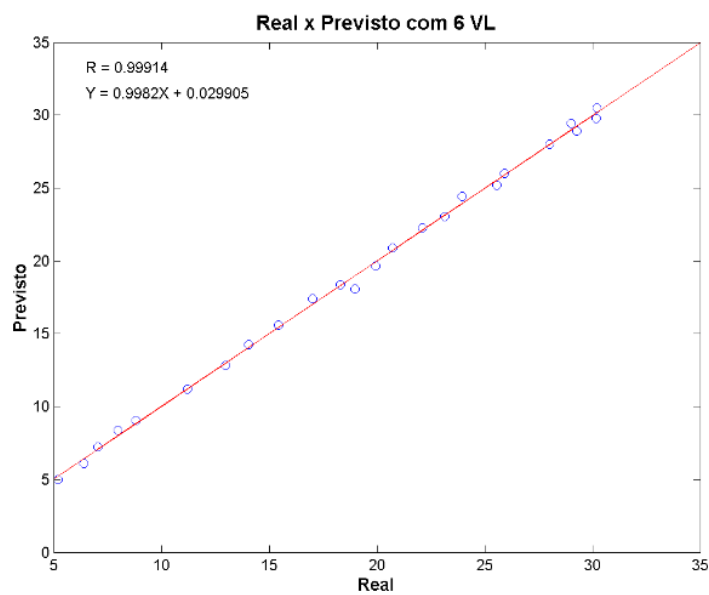
demonstra os intervalos selecionados pelo modelo siPLS e a Figura 16 a correlação entre os valores medidos e previstos.

Figura 15. Região espectral que os intervalos foram selecionados.



Fonte: A autora.

Figura 16. Coeficiente de correlação do modelo siPLS.



Fonte: A autora.

A Tabela 10 resume os melhores resultados para cada um dos métodos de seleção de variáveis por intervalos empregados e avaliados.

Tabela 10. Resultados dos modelos PLS, iPLS, biPLS e siPLS para quantificação de BP em MC.

Modelo	Variáveis	Intervalo	VL	RMSECV	RMSEC	RMSEP	Teste F
PLS global	2421		6	0.7756	0.5555	0.6917	
iPLS_25	96	1068:1164	5	0.4410	0.3745	0.3805	3,3046
biPLS_35	903	20	6	0.4383	0.2739	0.3663	3,5658
si4_PLS_25	388	1:5:6:25	6	0.4317	0.2622	0.3326	4,3250

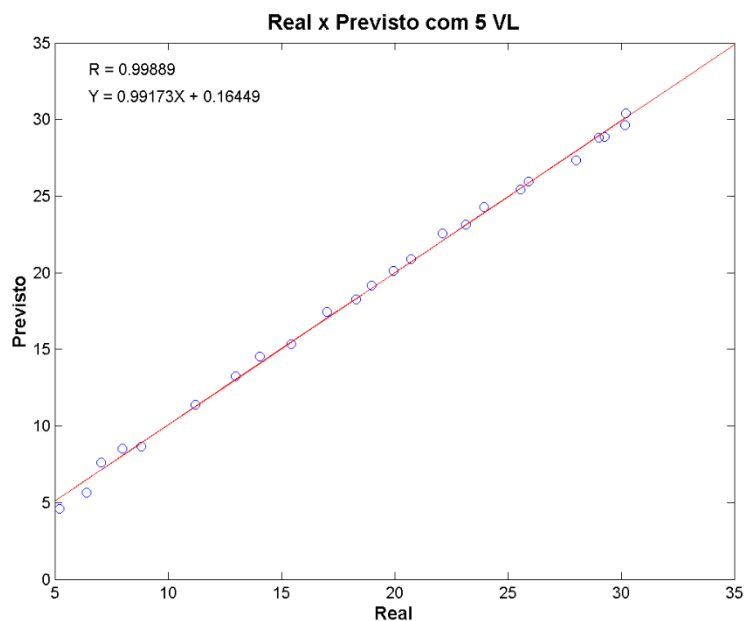
Fonte: A autora.

Na Tabela 10 foi analisado qual modelo apresentava o valor de RMSEP menor e mais próximo dos valores do RMSEC e RMSECV, aliado ao teste F que pode ser utilizado na comparação de modelos distintos multivariados. Neste cálculo o $RMSEP_1 > RMSEP_2$ e o valor calculado de F (F_{calc}) é comparado com o valor de distribuição de Fisher-Snedecor, chamado estatística F. O valor tabelado da estatística F deve ser inferior a F calc para que o método com $RMSEP_2$ apresente a melhor exatidão.

Devido ao valor tabelado ser 1,87 que é menor que 3,3046, 3,5658 e 4,3250 (F calculado), o método com $RMSEP_2$ apresenta melhor resultado, sendo assim os modelos de seleções de variáveis foram melhores preditivos que o modelo PLS global. Os três modelos apresentaram diferenças estatísticas, sendo que o si4_PLS_25 obteve o menor RMSEP, mas possui, juntamente com o biPLS_35, uma maior quantidade de variáveis, por este motivo, escolheu-se como melhor modelo aquele que apresentou menores números de variáveis, ou seja, iPLS_25.

A Figura 17 apresenta o coeficiente de correlação (R) entre os valores reais e os valores calculados pelo modelo iPLS_25 obtido no conjunto de previsão. A correlação apresentada a seguir é $R > 0,99$, valor próximo a 1 então indicando um modelo com ótima linearidade na faixa de concentração avaliada para quantificar BP em MC.

Figura 17. Valores reais versus previstos das amostras do conjunto de previsão.



Fonte: A autora.

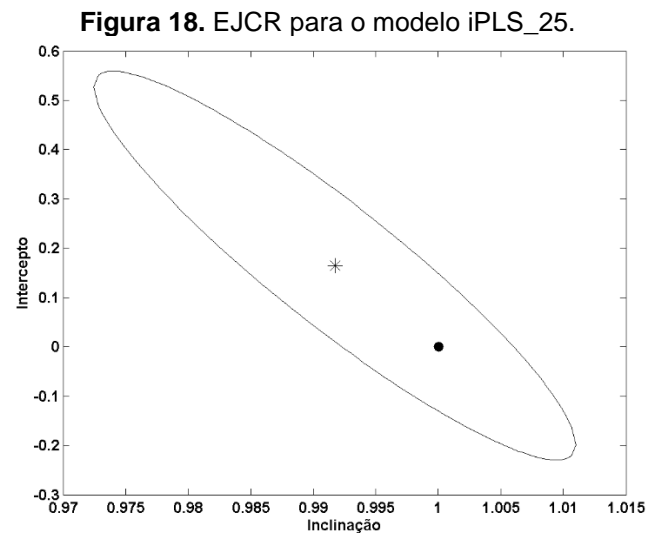
Verificou-se que o número mínimo de amostras está em conformidade com as normas e sem existência de erros sistemáticos no intervalo selecionado, pois de acordo com a norma ASTM E1655, é necessário que haja testes comprovando a exatidão do modelo, calculando-se o *bias*, o desvio padrão dos erros de validação (SDV), e por fim, a realização do test-*t* para avaliar se o “bias” é significativo, e neste caso, o $t_{\text{calculado}}$ foi menor que o $t_{\text{crítico}}$, com graus de liberdade igual a 21 com valores apresentados na Tabela 11. Isto significa que o modelo não apresenta erro sistemático.

Tabela 11. Cálculos de erros sistemáticos.

Modelo	iPLS_25
Bias	0.0089
SDV	0.3885
$t_{\text{crítico}}$	2.0796
$t_{\text{calculado}}$	0.1128

Fonte: A autora

Analisando-se a confiabilidade do modelo, foi realizada o teste de EJCR com 95% de confiança, onde o ponto ideal se encontrou dentro da elipse, demonstrando a ausência de diferenças significativas entre os valores reais e os previstos (HATANAKA et al., 2013), conforme apresentado na Figura 18.



Fonte: A autora.

5. CONCLUSÃO

A metodologia desenvolvida neste trabalho empregando a espectroscopia no infravermelho médio aliada ao método de regressão por quadrados mínimos parciais (PLS), demonstrou-se capaz de quantificar a banha de porco em manteiga de cacau na faixa de 5,00 a 30,00% (m/m). Além disso, verificou-se que os métodos de seleção de variáveis conseguiram encontrar regiões espectrais com melhores capacidades preditivas.

Tanto o modelo do PLS global quanto os métodos de seleção de variáveis iPLS, biPLS e siPLS, apresentam valores de R próximos a 1, demonstrando que os valores reais estão bem próximos daqueles previstos pela técnica, além de que os valores de RMSEP estão maiores que os valores de RMSEC, demonstrando que o modelo segue condições ideais para o desenvolvimento da metodologia de quantificação.

A aplicação de métodos de seleção de variáveis por intervalos mostrou-se eficiente na seleção de regiões espectrais que apresentassem melhor resultado. Este trabalho alvitra um modelo eficiente para determinar a adulteração de Banha de Porco em Manteiga de Cacau.

Assim, constatou-se que o melhor modelo obtido pelos métodos de seleção de variáveis (iPLS25) atendeu às exigências da norma ASTM E1655 indicando que a metodologia proposta pode ser empregada no controle de qualidade deste ramo alimentício.

6. REVISÃO BIBLIOGRÁFICA

ADEYEYE, E. I. et al. Effect of farm and industrial processing on the amino acid profile of cocoa beans. **Food Chemistry**, Ado Ekiti, v. 118, n. 2, p. 357-363, 2010.

AFOAKWA, E. O. **Chocolate Science and Technology**. 2 ed. Nova York: John Wiley & Sons, 2016.

AFOAKWA, E. O. **Cocoa Production and Processing Technology**. Boca Raton: CRC Press, 2014.

AGUIAR, A. C. S. **Panorama e perspectivas da cadeia produtiva do mel no Brasil**. 2018. 54 f. Trabalho de Conclusão de Curso (Graduação em Engenharia de Alimentos) – Faculdade de Engenharia Química, Universidade Federal de Uberlândia, Patos de Minas, 2018.

AMENDOLA, M. C. **Algoritmo genético e espectroscopia no infravermelho – algumas aplicações na indústria cosmética**. 2007. 110 f. Dissertação (Mestrado em Química Analítica) – Instituto de Química, Universidade de São Paulo, São Paulo, 2007.

ANVISA. Agência Nacional de Vigilância Sanitária, **Resolução RDC nº 265, de 22 de setembro de 2005**, Brasil. Disponível em: <<http://www.anvisa.gov.br>>. Acesso em: 19 de outubro de 2020 (a).

ANVISA. Agência Nacional de Vigilância Sanitária, **Resolução – CNNPA nº 12, de 1978, publicado no Diário Oficial da União de 24 de julho de 1978**. Disponível em: <https://www.agencia.cnptia.embrapa.br/Repositorio/legislacaoalimentos_000fgqfs0fv02wyiv8020uvkpsom46kk.htm>. Acesso em 08 de outubro de 2020 (b).

ANVISA. Agência Nacional de Vigilância Sanitária, **Resolução – RE nº 899, de 29 de maio de 2003**, Brasil. Disponível em: <http://bvsm.sau.gov.br/bvs/sau/legis/anvisa/2003/res0899_29_05_2003.html>. Acesso em 15 de fevereiro de 2021.

ARAUJO, T. P. **Emprego de espectroscopia no infravermelho e métodos quimiométricos para a análise direta de tetraciclina em leite bovino**. 2007. 96 f. Dissertação (Mestrado em Química) – Instituto de Química, Universidade Estadual de Campinas, Campinas, 2007.

ASSIS, C. **Aplicação de técnicas espectroscópicas, métodos quimiométricos, fusão de dados e seleção de variáveis no controle de qualidade de blends das espécies de café arábica e robusta**. 2018. 134 f. Tese (Doutorado em Ciências – Química) – Departamento de Química do Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, Belo Horizonte, 2018.

ASTM E1655-17, **Práticas Padrão para Análise Quantitativa Infrared Multivariate**, **ASTM International**, West Conshohocken, PA, 2017. Disponível em: <<http://www.astm.org>>. Acesso em: dia 06 de outubro de 2020.

BALADIN, R. M.; SMIRNOV, S. V. Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on diodisel data. **Analytica Chimica Acta**, v. 692, p. 63-72, 2011.

BARBOSA, L. C. A. **Espectroscopia no infravermelho na caracterização de compostos orgânicos**. Viçosa: Ed. UFV, 2007.

BARIS, K. N.; HUSSAIN, M. N.; BAKAR, J. SHARIF, Z.; KHIR, M. F. A.; ZOOLFAKAR, A. S. Classification and quantification of palm oil adulteration via portable NIR spectroscopy. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 173, p. 335-342, 2017.

BARROS NETO, B.; PIMENTEL, M. F.; ARAÚJO, M. C. U. Recomendações para calibração em química analítica – Parte I. Fundamentos e calibração com um componente (calibração univariada)*. **Química Nova**, v. 25, n. 5, p. 856-865, 2002.

BATISTA, N. N.; ANDRADE, D. P.; RAMOS, C. L.; DIAS, D. R.; SCHWAN, R. F. Antioxidant capacity of cocoa beans and chocolate assessed by FTIR, **Food Research International**, v. 90, p.313–319, 2016.

BEEBE, K. R.; KOWALSKI, B. R. Na Introduction to Multivariate Calibration and Analysis. **Analytical Chemistry**, v. 59, n. 17, p. 1007A – 1017^a, 1987.

BELAVENUTO, E. G. T. **Análise químico-farmacêutica e estudos microbiológico e de estabilidade de preparações injetáveis de daptomicina**. 2016. 302 f. Tese (Doutorado em ciências farmacêuticas) - Faculdade de Ciências Farmacêuticas- Universidade Estadual Paulista, Araraquara, 2016.

BICAS, T. C.; et al. Uso da Espectroscopia de Infravermelho Médio com Transformada de Fourier (IV-TF) e quimiometria pra classificação de vinhos e suco de uva. **Brazilian Journal of Food Research**, v. 8, n. 2, p. 89-97, 2017.

BORGES NETO, W. **Parâmetros de Qualidade de Lubrificantes e Óleo de Oliva Através de Espectroscopia Vibracional, Calibração Multivariada e Seleção de Variáveis**. Tese (Doutorado em Química) – Instituto de Química, Universidade Estadual de Campinas, Campinas, 2005.

BRAGA, J. W. B.; POPPI, R. J. Validação de Modelos de Calibração Multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo. **Química Nova**, v. 27, n. 6, p. 1004-1011, 2004.

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

CABALLERO, B.; FINGLAS, P.; TOLDRÁ, F. **Encyclopedia of Food and Health**. *Waltham*: Academic Press, 2015.

CODEX ALIMENTARIUS. **Standard for chocolate and chocolate products, 1981**. Disponível em: < http://www.fao.org/fao-who-codexalimentarius/sh-proxy/en/?lnk=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252FStandards%252FCXS%2B87-1981%252FCXS_087e.pdf>. Acesso em: 20 de outubro de 2020 (a).

CODEX ALIMENTARIUS. **Standard for cocoa butter, 1981**. Disponível em: < http://www.fao.org/fao-who-codexalimentarius/sh-proxy/en/?lnk=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252FStandards%252FCXS%2B86-1981%252FCXS_086e.pdf>. Acesso em: 20 de outubro de 2020 (b).

COHEN, K. O.; LUCCAS, V.; JACKIX, M. N. H. Revisão: Temperagem ou Pré-Cristalização do Chocolate. **Brazilian Journal of Food Technology**, v. 7, n. 1, p. 23-30, 2004.

CSERHÁTI, T., FORGÁCS, E., DEVL, Z., & MIKSIK, I. Chromatography in authenticity and traceability tests of vegetable oils and dairy products: a review. **Biomed. Chromatogr.**, 19, 183–190, 2005

DE CLERCQ, N. et al. Functionality of cocoa butter equivalents in chocolate products. **European Food Research and Technology**, Gent, p. 1-13, 2016.

DIMICK, P. S. Principles of cocoa butter crystallization. **The Manufacturing Confectioner**, 109-114, may, 1991.

FERNANDES, D. D. S. **Novas estratégias para seleção de variáveis por intervalos em problemas de classificação**. 2016. 137 f. Tese (Doutorado em Química) – Departamento de Química, Universidade Federal da Paraíba, 2016.

FERNANDES, D. D. S.; GOMES, A. A.; COSTA, G. B.; SILVA, G. W. B.; VÉRAS, G. Determination of biodiesel content in biodiesel/diesel blends using NIR and visible epectroscopy with variable selection. **Talanta**, v. 87, p. 30-34, 2011.

FERREIRA, D. N. S.; LOPES, V. C.; LOPES, V. C. S.; PEREIRA, I. L.; GÓES, A. O. S. Processo de Temperagem da Manteiga de Cacau: análises, conceitos e sinergias em uma Multinacional. **Brazilian Journals of Business**, Curitiba, v. 1, n. 3, p. 1207-1217, 2019.

FERREIRA, M. C.; ANTUNES, A.M.; MELGO, M. S.; VOLPE, P. L. O. Quimiometria I: Calibração Multivariada, um tutorial. **Química Nova**, v. 22, n. 5, p. 724-731, 1999.

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

FILGUEIRAS, P. R.; ALVES, J. C. L.; POPPI, R. J. Quantification of animal fat biodiesel in soybean biodiesel and B20diesel blends using near infrared spectroscopy and synergy interval support vector regression. **Talanta**, v. 119, p. 582-589, 2014.

GELADI, P.; KOWALSKI, B. R. Partial Least Squares Regression: A Tutorial. **Analytica Chimica Acta**, v. 185, p. 1-17, 1986.

GONTIJO, L. C.; GUIMARÃES, E.; MITSUTAKE, H.; SANTANA, F. B.; SANTOS, D. Q.; BORGES NETO, W. Quantification of soybean biodiesels em diesel blends according to ASMT E1655 using mid-infrared spectroscopy and multivariate calibration. **Fuel**, v. 117, p. 1111-1114, 2014.

GONTIJO, L. C. **Uso da espectrometria no infravermelho médio, calibração multivariada e seleção de variáveis por intervalos na quantificação de biodieseis em misturas com diesel**. 2016. 104 f. Tese (Doutorado em Química) – Instituto de Química, Universidade Federal de Uberlândia, Uberlândia, 2016.

GUIMARÃES, E. **Controle de qualidade de biodieseis de macaúba e algodão e suas misturas com diesel usando espectroscopia no infravermelho médio e cartas de controle multivariadas**. 2018. 95 f. Tese (Doutorado em Química) -Instituto de Química, Universidade Federal de Uberlândia, Uberlândia, 2018.

GUIMARÃES, E.; SANTANA, F. B.; GONTIJO, L. C.; MITSUTAKE, H.; SANTOS, D. Q.; BORGES NETO, W. Aplicação das Figuras de Mérito Multivariadas na Validação de Metodologias em Análises de Biocombustíveis empregando Espectroscopia no Infravermelho Médio e PLS. **Revista Virtual de Química**, v. 7, n. 6, p. 2242-2254, 2015.

GURDENIZ, G. **Chemometric Studies for Classification of Olive Oils and Detection of Adulteration**. 2008. 94 f. Tese (Mestre em Ciências em Engenharia de Alimentos)- The Graduate School of Engineering and Sciences, İzmir Institute of Technology, Turquia, 2008.

HATANAKA, R. R.; SEQUINEL, R.; GUALTIERI, C. E.; TERCINE, A. C. B.; FLUMIGNAN, D. L.; OLIVEIRA, J. E. Development and validation of an environmentally friendly attenuated total reflectance in the mid-infrared region method for the determination of ethanol content in used engine lubrication oil. **Talanta**, 2013, 109, 191.

HOOLER, F. J.; SKOOG, D. A.; CROUCH, S. R. Princípios de Análise Instrumental. Tradução da 6 ed. Porto Alegre: **Ed. Bookman**, 2009.

JAHURUL, M. H. A. Z Aidul, I. S. M.; NIK NORULAINI, N. A.; SAHENA, F.; KAMARUZZAMAN, B. Y.; KASHIF GHAFOR; OMAR, A. K. M. Cocoa butter replacers from blends of mango seed fat extracted by supercritical carbon dioxide and palm stearin. **Food Research International**, v. 65, p. 401-406, 2014.

LANNES, S. C. **Estudo das propriedades físico-químicas e de textura de chocolates**. 1997. Tese (Doutorado em Química) - Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 1997.

LARSSON, K.; QUINN, P.; SATO, K.; TIBERG, F. **Lipids: Structure, Physical Properties and Functionality**. England, Bridgwater: The Oily Press, v.19, 2006.

LEARDI, R.; NØRGAARD, L. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. **Journal Chemometrics**, v.18, p. 486-497, 2004.

LENARDON, L. **Ferramentas Quimiométricas Aplicadas no Desenvolvimento de Metodologia Analítica para Avaliação de Adulteração em Leite Bovino por Adição de Soro do Queijo por ATR-FTIR**. 2016. 109 f. Dissertação (Mestrado em Ciências Farmacêuticas) – Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.

LIMA, K. P. **Criação e atualização de modelos de calibração multivariada para determinação de carbono orgânico no solo por espectroscopia NIR**. 2015. 39 f. Trabalho de Conclusão de Curso (Licenciatura em Química) – Departamento Acadêmico de Química, Universidade Tecnológica Federal do Paraná, Campo Mourão, 2015.

LOPES, E. L. **Estratégias para a caracterização de amostras brasileiras de óleo de soja empregando ATR-FTIR e HPLC-ELSD aliadas a ferramentas quimiométricas**. 2015. 88 f. Dissertação (Mestrado em Química) - Instituto de Química, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015.

LOPES, U. V.; MONTEIRO, W. R.; PIRES, J. L.; CLEMENTE, D.; YAMADA, M.; GRAMACHO, K. P. Cacao breeding in Bahia, Brazil – strategies and results. **Crop Breeding and Applied Biotechnology**, Viçosa (MG), v. 11, n. (spe), p. 73-81, 2011.

LUCENA, D. V. **Algoritmos Evolutivo Multiobjetivo para Seleção de Variáveis em Problemas de Calibração Multivariada**. 2013. 56 f. Dissertação (Mestrado em Computação) – Instituto de Informática, Universidade Federal de Goiás, Goiânia, 2013.

MACHADO, C. C. S. **Metodologia de Detecção de Adulterações em Óleo de Macadâmia Usando Espectroscopia no Infravermelho Médio e Ferramentas Quimiométricas**. 2018. 33 f. Trabalho de Conclusão de Curso (Bacharelado em Química Industrial) – Instituto de Química, Universidade Federal de Uberlândia, Uberlândia, 2018.

MALINOWSKI, E. R. **Factor Analysis in Chemistry**. 2 ed., John Wiley & Sons: New York, 1991.

MANSOR, T. S. T.; CHE MAN, Y. B.; ROHMAN, A. Application of fast gas chromatography and Fourier transform infrared spectroscopy for analysis of lard

adulteration in virgin coconut oil. **Food Anal. Methods**, v. 4, p. 365–372, 2011. <http://dx.doi.org/10.1007/s12161-010-9176-y>.

MARIKKAR, J. M. N.; GHAZALI, H. M.; LONG, K.; LAI, O. M. Lard uptake and its detection in selected food products deep-fried in lard. **Food Research International**, v.36, p. 1047–1060, 2003. <http://dx.doi.org/10.1016/j.foodres.2003.08.003>.

MARROCOS, T. R.; MORAES, M. E. B.; GOMES, R. L. Diagnóstico dos padrões de certificação socioambiental do cacau na Bahia, Brasil. **Revista Brasileira de Gestão e Desenvolvimento Regional**, Taubaté, São Paulo, v. 14, n. 3, p. 76-100, 2018.

MARTIN JUNIOR, R. A. **Chocolate**. *Adv. Food Res.*, New York, v.31, p.211-342, 1987.

MARTINS, O. J. S. **Aplicação de espectroscopia no infravermelho médio para detecção de adulteração em produto a base de frango**. 2015. 151 f. Dissertação (Mestrado em Ciências dos Alimentos) – Faculdade de Farmácia, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

MEURSING, E. H. DeZaan-Cocoa and Chocolate Manual. **ADM Cocoa International**, Switzerland, Suíça: ADM, 2009.

MONTEIRO, G. M. M. **Análise de Revestimentos Orgânicos por Espectroscopia de Infravermelho**. 2007. 99 f. Dissertação (Mestrado em Engenharia de Materiais) – Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, 2007.

MORSING, T.; EKMAN, C. Comments on construction of confidence intervals in connection with partial least squares, **Journal of Chemometrics**, v.12, p. 295-299, 1998.

MUNCK, L.; PRAM NIELSEN, J.; MØLLER, B.; JACOBSEN, S.; SØNDERGAARD, I.; ENGELSEN, S.B.; NØRGAARD, L.; BRO, R. Exploring the phenotypic expression of a regulatory proteome-altering gene by spectroscopy and chemometrics. **Analytica Chimica Acta**, v. 446, p. 171-186, 2001.

NAIK, B.; KUMARB, V. Cocoa Butter and Its Alternatives: A Reveiw. **Journal of Bioresource Engineering and Technology**, v.1, p. 07-17, 2014.

NØRGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L; ENGELSEN, S. B. Interval partial least-square regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. **Appl. Spectrosc.**, v. 54, n. 3, p. 413-419, 2000.

OLIVEIRA, G. M. **Obtenção e caracterização de bases lipídicas destinadas à estabilização do poliformismo de lipídios com ênfase na manteiga de cacau**. 2016. 163 f. Tese (Doutorado em Engenharia Química) – Faculdade de Engenharia Química, Universidade Estadual de Campinas, Campinas, 2016.

PANTZARIS, T. P.; BASIRON, Y. The lauric (coconut and palm kernel) oils. In: GUNSTONE, F. D. *Vegetable Oils in Food Technology Composition, Properties and Uses*. Boca Raton, Florida, U.S.A: **CRC Press LLC**, p. 157-201, 2002.

PASSARI, L. M. Z. G.; SOARES, P. K.; BRUNS, R. E.; SCARMINIO, I. S. Estatística aplicada à química: dez dúvidas comuns. **Química Nova**, v. 34, n. 5, p. 888-892, 2011.

PAULO, J. M. **Determinação de parâmetros de qualidade de gasolinas automotivas utilizando espectroscopia de emissão por chama e métodos quimiométricos**. 2013. 150 f. Tese (Doutorado em Ciências – Química) – Departamento de Química, Universidade de Federal de Minas Gerais, Belo Horizonte, 2013.

PAVIA, D. L.; et al. *Introdução à Espectroscopia*, 1ª ed.; **Cengage Learning**: São Paulo, SP, 2010.

QUAST, L. B. **Estudo do efeito da adição de gorduras alternativas na cristalização da manteiga de cacau**. 2008. 127p. Tese (Doutorado em Engenharia Química), Universidade Estadual de Campinas, Faculdade de Engenharia Química, Campinas, SP.

RECH, A. M. **Caracterização de Bebidas à Base de Soja Empregando Espectroscopia no Infravermelho Médio com Transformada de Fourier por Reflexão Total Atenuada e Quimiometria**. 2018. 84 f. Dissertação (Mestrado em Química) – Instituto de Química, Universidade do Rio Grande do Sul, Porto Alegre, 2018.

RIBEIRO, A. F. F. **Adulteração de suplementos alimentares por adição de compostos com ação farmacológica**. 2015. 28 f. Monografia (Estágio Curricular do Mestrado Integrado em Ciências Farmacêuticas) – Faculdade de Farmácia, Universidade de Coimbra, Coimbra, 2015.

RIISPOA. Regulamento da Inspeção Industrial e Sanitária de Produtos de Origem Animal, 1952. Disponível em: <<https://www.sertaobras.org.br/wp-content/uploads/2010/11/RIISPOA.pdf>>. Acesso em: 19 de outubro de 2020.

ROHMAN, A.; CHE MAN, Y. B. FTIR spectroscopy combined with chemometrics for analysis of lard in the mixtures with body fats of lamb, cow and chicken. **International Food Research Journal**, v. 17, p. 519–526, 2010.

ROSSI, A. C. **Análise espectroscópica de amostras de trigo por infravermelho médio FTIR e NIR e construção de modelos multivariados de calibração por PLS**. 2015. 65 f. Trabalho de Conclusão de Curso-Departamento de Química, Universidade Tecnológica Federal do Paraná, Pato Branco, 2015.

SANTANA, F. B. **Uso da espectroscopia no infravermelho médio e análise discriminante por quadrados mínimos parciais na determinação de adulterações em óleos de andiroba, prímula e rosa mosqueta**. 2015, 77 f. Dissertação (Mestrado

em Química)-Instituto de Química, Universidade Federal de Uberlândia, Uberlândia, 2015.

SANTOS, F. C. G.; FONTES, M. J. V. Do Cacau ao Chocolate: estudo de caso em pequena indústria familiar da cadeia de valor do cacau no território litoral sul da Bahia – TLS. **Revista Livre de Sustentabilidade e Empreendedorismo**, v. 5, n. 6, p. 196-216, 2020.

SCHINAIDER, A. D. **Consumo Vegano: o estado da arte e o comportamento do consumidor baseado na teoria do comportamento planejado (TCP)**. 2018. 84 f. Dissertação (Mestrado em Agronegócio) – Centro de Estudos e Pesquisa em Agronegócio (CEPAN), Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

SHUKLA, V. K. S. Cocoa butter, cocoa butter equivalents, and cocoa butter substitutes. **Handbook of Functional Lipids**. Taylor & Francis Group LLC, p. 279-307, 2006.

SILVA, D. I. O. **Uma metodologia baseada em espectrometria UV-Vis e quimiometria para a determinação de n,n-dietil-3-metilbanzamida em repelentes**. 2018, 72 f. Dissertação (Mestrado em Química) – Departamento de Química, Universidade Federal da Paraíba, João Pessoa, 2018.

SILVA, R. C.; GIOIELLI, L. A. Propriedades físicas de lipídios estruturados obtidos a partir de banha e óleo de soja. **Revista brasileira de ciências farmacêuticas**, v. 42, n. 2, 2006.

SILVA, V. V. **Identificação de óleo vegetal em amostras de óleo lubrificante usado ou contaminado (OLUC): um estudo para aplicação de metodologia de espectroscopia de infravermelho médio com transformada de Fourier (FTIR)**. 2016. 35 f. Monografia (Licenciado em Química) - Departamento de Química, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Bauru, 2016.

SILVERSTEIN, R. M.; et al. Identificação Espectrométrica de Compostos Orgânicos. 7ª ed.; **LTC**: Rio de Janeiro, RJ, 2012.

SITOE, B. V. **Aplicação de métodos de seleção de variáveis para o controle de qualidade de biodiesel de pinhão manso e de moringa em misturas com diesel usando espectrometria no infravermelho médio**. 2019. 116 f. Tese (Doutorado em Química) – Laboratório de Quimiometria do Triângulo, Universidade Federal de Uberlândia, Uberlândia, 2019.

SITOE, B. V.; MÁQUINA, A. D. V.; GONTIJO, L. C.; OLIVEIRA, L. R.; SANTOS, D. Q.; BORGES NETO, W. Quantification of *Jatropha* methyl biodiesel em mixtures with diesel using mid-infrared spectrometry and interval variable selection methods. **Analytical Letters**, v. 53, n. 4, p. 589-605, 2019.

SMITH, K. W. Cocoa butter and cocoa butter equivalents. In: GUNSTONE, F. D. **Structured and Modified Lipids**. New York: Marcel Dekker, p. 401-422, 2001.

SOARES, A. L. **Determinação do grau API e teor de enxofre e nitrogênio em petróleo usando ATR-FTIR e métodos quimiométricos**. 2019. 53 f. Trabalho de Conclusão de Curso (Bacharelado em Química) – Departamento de Química, Universidade Federal de Santa Maria, Santa Maria, 2019.

SOUZA, A. S. **Avaliação de estabilidade térmica e oxidativa de chocolates amargos**. 2010. Dissertação de Mestrado em Ciência e Tecnologia de Alimentos. Universidade Federal da Paraíba – UFP, João Pessoa.

SOUZA, C. S. **Efeito da Substituição Parcial de Manteiga de Cacau por Gordura CBE no Perfil de Compostos Voláteis de Chocolate Meio Amargo**. 2017. 95 p. Dissertação (Mestrado em Ciência dos Alimentos) – Departamento de Ciência e Tecnologia de Alimentos, Universidade Federal de Santa Catarina, Florianópolis, 2017.

SOUZA, J. S.; FERRÃO, M. F. Aplicação e espectroscopia no infravermelho no controle de qualidade de medicamentos contendo diclofenaco de potássio. Parte I: Dosagem por regressão multivariada. **Revista Brasileira de Ciências Farmacêuticas**, v. 42, n. 3, p. 439-445, 2006.

SOUZA, L. M. **Uso de espectroscopia no infravermelho médio, calibração multivariada e seleção de variáveis na quantificação de adulterantes em biodiesel e suas misturas com diesel**. 2014. 104 f. Dissertação (Mestrado em Química) – Instituto de Química, Universidade Federal de Uberlândia, Uberlândia, 2014.

SOUZA, W. A. Benefícios da Dieta Vegana para a Saúde. **Revista Vida & Saúde Vegana**. n. 6, p. 10 – 13, 2015.

TEIXEIRA JÚNIOR, O. J. **Desenvolvimento de Método Analítico combinado com ferramentas quimiométricas para identificação de Adulterações em Vodcas**. 2018. 61 f. Dissertação (Mestrado em Química) – Instituto de Química, Universidade Estadual Paulista, Araraquara, 2018.

TORBICA, A. et al. Solid fat content, pre-crystallization conditions, and sensory quality of chocolate with addition of cocoa butter analogues. **International Journal of Food Properties**, Novi Sad, v. 19, n. 5, p. 1029-1043, 2016.

VALDERRAMA, P. **Avaliação de figuras de mérito em calibração multivariada na determinação de parâmetros de controle de qualidade em indústria alcooleira por espectroscopia no infravermelho próximo**. 2005. 135 f. Dissertação (Mestrado em Química) - Instituto de Química, Universidade Estadual de Campinas, Campinas, 2005.

VERSTRINGE, S. et al. Enzymatic and other modification techniques to produce cocoa butter alternatives. **Cocoa Butter and Related Compounds**, Gent, p. 444, 2012.

VISANI, V. **Teor de óleo em cápsulas gelatinosas por espectroscopia no infravermelho médio e técnicas quimiométricas**. 2017. 142 f. Tese (Doutorado em Ciências, concentração em Química Analítica) – Departamento de Química, Universidade Federal da Paraíba, João Pessoa, 2017.

WORKMAN, J.; LAVINE, B.; CHRISMAN, R.; KOCH, M. Process Analytical Chemistry. **Analytical Chemistry**, v.83, p. 4557-4578, 2011.