



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE ALIMENTOS

FABIELLE NEGRÃO FERREIRA

**DISCRIMINAÇÃO DA ORIGEM GEOGRÁFICA E QUANTIFICAÇÃO DE
PARÂMETROS DE QUALIDADE EM SEMENTES DE CACAU (*Theobroma cacao*)
CULTIVADAS NA AMAZÔNIA UTILIZANDO ESPECTROSCOPIA NO
INFRAVERMELHO PRÓXIMO E FERRAMENTAS QUIMIOMÉTRICAS**

BELÉM-PA

2022

FABIELLE NEGRÃO FERREIRA

**DISCRIMINAÇÃO DA ORIGEM GEOGRÁFICA E QUANTIFICAÇÃO DE
PARÂMETROS DE QUALIDADE EM SEMENTES DE CACAU (*Theobroma cacao*)
CULTIVADAS NA AMAZÔNIA UTILIZANDO ESPECTROSCOPIA NO
INFRAVERMELHO PRÓXIMO E FERRAMENTAS QUIMIOMÉTRICAS**

Tese de Doutorado apresentada ao Programa de Pós-graduação em Ciência e Tecnologia de Alimentos, Instituto de Tecnologia de Alimentos, Universidade Federal do Pará, como requisito para obtenção do título de Doutora em Ciência e Tecnologia de Alimentos.

Orientadora: Prof^ª. Dr^ª. Alessandra Santos Lopes
Coorientador: Prof. Dr. Marcos Enê Chaves Oliveira

BELÉM-PA

2022

**Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)**

N385d Negrão Ferreira, Fabielle.
DISCRIMINAÇÃO DA ORIGEM GEOGRÁFICA E
QUANTIFICAÇÃO DE PARÂMETROS DE QUALIDADE EM
SEMENTES DE CACAU (*Theobroma cacao*) CULTIVADAS NA
AMAZÔNIA UTILIZANDO ESPECTROSCOPIA NO
INFRAVERMELHO PRÓXIMO E FERRAMENTAS
QUIMIOMÉTRICAS / Fabielle Negrão Ferreira. — 2022.
96 f. : il. color.

Orientador(a): Prof^ª. Dra. Alessandra Santos Lopes
Coorientador(a): Prof. Dr. Marcos Enê Chaves Oliveira
Tese (Doutorado) - Universidade Federal do Pará, Instituto de
Tecnologia, Programa de Pós-Graduação em Ciência e Tecnologia
de Alimentos, Belém, 2022.

1. Cacau. 2. NIR. 3. Infravermelho. 4. Chocolate. 5.
Métodos Analíticos. I. Título.

CDD 338.17374

FABIELLE NEGRÃO FERREIRA

**DISCRIMINAÇÃO DA ORIGEM GEOGRÁFICA E QUANTIFICAÇÃO DE
PARÂMETROS DE QUALIDADE EM SEMENTES DE CACAU (*Theobroma cacao*)
CULTIVADAS NA AMAZÔNIA UTILIZANDO ESPECTROSCOPIA NO
INFRAVERMELHO PRÓXIMO E FERRAMENTAS QUIMIOMÉTRICAS**

Tese de Doutorado apresentada ao Programa de Pós-graduação em Ciência e Tecnologia de Alimentos, Instituto de Tecnologia de Alimentos, Universidade Federal do Pará, como requisito para obtenção do título de Doutora em Ciência e Tecnologia de Alimentos.

Orientadora: Prof^a. Dr^a. Alessandra Santos Lopes
Coorientador: Prof. Dr. Marcos Enê Chaves Oliveira

Avaliada em: 03/08/2022 Conceito: EXC

BANCA EXAMINADORA

Prof^a. Dr^a. Alessandra Santos Lopes
Orientadora
PPGCTA/ITEC/UFPA

Prof. Dr. Marcos Enê Chaves Oliveira
Coorientador
Embrapa – Amazônia Oriental

Prof. Dr. Raul Nunes de C. Júnior
Membro Interno
PPGCTA/ITEC/UFPA

Prof^a. Dr^a. Deusa Do S. T. Costa Leitão
Membro Externo
ETPP-JK

Prof. Dr. Renan Campos Chisté
Membro Interno
PPGCTA/ITEC/UFPA

Prof. Dr. Heronides Adonias D. Filho
Membro Externo
PPGQ/ICEN/UFPA

BELÉM-PA

2022

Dedico esta conquista os meus pais, que me trouxeram ao mundo, me amaram e deram o máximo de si para que eu pudesse chegar até aqui.

AGRADECIMENTOS

Ao Deus criador de tudo.

Ao Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos (PPGCTA) da Universidade Federal do Pará (UFPA), pela oportunidade de ensino gratuito e de qualidade e pela conclusão deste doutorado.

À CAPES que me concedeu a bolsa de estudos, que tornou possível minha permanência no Programa.

Ao Projeto Cacau (3472) pelo suporte financeiro para a coleta de amostras e compra de reagente para as análises.

À minha orientadora Alessandra Santos Lopes pela oportunidade, pela confiança, pelo compartilhamento dos seus conhecimentos, e pela disponibilidade em me orientar.

Ao meu coorientador Marcos Enê pela sua paciência e dedicação.

Aos membros da banca examinadora, Renan Campos Chisté, Raul Nunes de Carvalho Júnio, Heronides Adonias Dantas Filho, Deusa Do Socorro Teixeira Costa Leitão.

Aos amigos do LABIOTEC, que foram pessoas importantíssimas para o andamento desta pesquisa, Evellyn, Silvana, Thais Sena, Alan, Felipe.

As minhas queridas amigas que amo infinitamente, Deusa e Francilha.

Ao meu amigo Gilson Chagas, que me apoiou desde o início do projeto e ficou ao meu lado até o último instante.

A todas as pessoas que dão suporte de forma indireta e que tornam o ambiente de pesquisa mais agradável, que são os auxiliares de limpeza, os técnicos, os guardas, a equipe do RU, etc.

À Hadriane Pombo, secretária super gentil, competente e dedicada, que executa brilhantemente seu trabalho junto à Coordenação do programa.

À minha melhor amiga, Marley Antonia Silva da Silva.

Aos meus segundos pais Wanderley Lima e Maria do Livramento de Oliveira. Obrigada pelo carinho, amizade, motivação, por sempre acreditarem no meu potencial e por me acolherem como uma filha.

Ao meu pai José Luiz Gonçalves Ferreira, que não está mais entre nós, e que sempre teve orgulho da pessoa que me tornei, e minha mãe Maria do Socorro Negrão Ferreira pelo amor e apoio incondicional.

Que Deus nos abençoe.

Meus sinceros agradecimentos.

Fabielle Negrão Ferreira

RESUMO GERAL

Este trabalho discorre sobre a aplicação da técnica de Espectroscopia Vibracional no Infravermelho Próximo (NIR) em amostras de cacau de diferentes regiões do estado do Pará (Medicilândia, Tucumã e Tomé-açu) com o intuito de prever componentes majoritários e discriminar estas diferentes regiões geográficas. A técnica NIR foi associada a análise exploratória dos dados espectrais, testando-se diferentes modelos de regressão e de discriminação. Esta pesquisa está dividida em três capítulos. O primeiro capítulo faz uma revisão sobre os conceitos referente à técnica NIR, e explora os mais diversos artigos sobre aplicação da técnica NIR em cacau para diferentes avaliações, tais como, para análise de componentes majoritários, identificação de fraudes e discriminação de cacau em diferentes regiões do planeta. No segundo capítulo desenvolveram-se modelos de predição para nitrogênio, umidade e lipídios totais em amostras não fermentadas, uma vez que, determinados componentes do cacau não são alterados durante a fermentação, com isso estes componentes majoritários das sementes puderam ser rapidamente preditos pela espectroscopia FT-NIR utilizando regressão PLS. No terceiro artigo apresentou-se aplicação de NIR associada a análise exploratória dos dados espectrais por componentes principais, para discriminação das sementes de cacau sob cinco diferentes condições de tratamento (cru, secos fermentados, secos não fermentados e fermentado desengordurado), das três diferentes regiões geográficas supracitadas, e obteve-se os melhores resultados ao avaliar as amostras de cacau fermentadas das três regiões.

Palavras-chave: Predição; Discriminação Geográfica; Espectroscopia no Infravermelho Próximo.

ABSTRACT

This work discusses the application of the Near Infrared Vibrational Spectroscopy (NIR) technique in cocoa samples from different regions of the state of Pará (Medicilândia, Tucumã and Tomé-açu) with the aim of predicting major components and discriminating these different geographic regions. The NIR technique was associated with exploratory analysis of spectral data, testing different regression and discrimination models. This research is divided into three chapters. The first chapter reviews the concepts related to the NIR technique, and explores the most diverse articles on the application of the NIR technique in cocoa for different evaluations, such as, for analysis of major components, identification of fraud and discrimination of cocoa in different regions. from the planet. In the second chapter, prediction models were developed for nitrogen, moisture and total lipids in non-fermented samples, since certain cocoa components are not altered during fermentation, so these major components of the seeds could be quickly predicted by FT spectroscopy -NIR using PLS regression. The third article presents the application of NIR associated with exploratory analysis of spectral data by principal components, for discrimination of cocoa beans under five different treatment conditions (raw, dry fermented, dry unfermented and fermented defatted), from the three different regions mentioned geographic areas, and the best results were obtained when evaluating the fermented cocoa samples from the three regions.

Keywords: Prediction; Geographic Discrimination; Near Infrared Spectroscopy.

LISTA DE ILUSTRAÇÕES

CAPÍTULO I – Aplicação de técnicas espectrais de infravermelho próximo para avaliar a qualidade e atributos composicionais do cacau: uma visão geral

Figura 1 – Esquema simplificado da influência da água durante análise por espectroscopia no infravermelho próximo 22

CAPÍTULO II - Predição rápida e simultânea de umidade, nitrogênio total e lipídeos em cacau da Amazônia (*Theobroma cacao*) não fermentado utilizando a técnica FT-NIR

Figura 1 – Região espectral utilizada para o modelo de predição de umidade 53

Figura 2 – Região espectral utilizada para o modelo de predição de nitrogênio total . 54

Figura 3 – Região espectral utilizada para o modelo de predição de lipídios 55

Figura 4 – Valores reais *versus* valores preditos (NIR) para os conjuntos de validação interna e externa (conjunto teste) de: a) umidade, b) nitrogênio total, e c) umidade, expressos em g/100 g da amostra 58

CAPÍTULO III – Geographical Discrimination of Ground Amazon Cocoa by Near-Infrared Spectroscopy: Influence of Sample Preparation

Figure 1 – Absorption spectrum of a) RC, b) DU, c) DF, d) DFD, in the range of 3600 to 12500 cm^{-1} 70

Figure 2 – Absorption spectrum of the separation regions of a) RC, c) DU, e) DF and their respective scores in the PCs graphs (b, d and f) 72

Figure 3 – Scores of DFD cocoa samples 76

Figure 4 – Absorption spectrum of the DFD separation regions (a and c) and their respective scores (b and d), for the two selected optimal separation regions 77

Figure 5 – Absorption spectrum highlighting the two separation regions of DFD samples (1st and 2nd region) 78

LISTA DE TABELAS

CAPÍTULO I – Aplicação de técnicas espectrais de infravermelho próximo para avaliar a qualidade e atributos composicionais do cacau: uma visão geral

Tabela 1 – Componentes preditos em cacau e subprodutos utilizando a técnica por espectroscopia vibracional no infravermelho próximo no período de 1989 a 2018 29

Tabela 2 – Compostos fenólicos preditos em cacau utilizando a técnica de análise por Espectroscopia Vibracional no Infravermelho Próximo 34

Tabela 3 – Discriminação geográfica de cacau sob diferentes características de amostragem apresentada por diferentes autores 36

CAPÍTULO II – Predição rápida e simultânea de umidade, nitrogênio total e lipídeos em cacau da Amazônia (*Theobroma cacao*) não fermentado utilizando a técnica FT-NIR

Tabela 1 – Análises físicas e químicas avaliadas em cacau úmido e seco 52

Tabela 2 – Análises físicas e químicas avaliadas em cacau *para* cada região 52

Tabela 3 – Parâmetros estatísticos avaliados para validação externa de umidade, nitrogênio e lipídeos utilizando PLSR 55

Tabela 4 – Comparação entre os valores reais de concentração de lipídeos, nitrogênio e umidade, e os preditos utilizando modelos NIR 57

CAPÍTULO III – Geographical Discrimination of Ground Amazon Cocoa by Near-Infrared Spectroscopy: Influence of Sample Preparation

Table 1 – Spectra collected by region and category: raw cocoa (RC), dried unfermented cocoa (DU), dried fermented cocoa (DF) and dried fermented degreased (DFD) 68

SUMÁRIO

INTRODUÇÃO GERAL	13
OBJETIVOS	16
GERAL	16
ESPECÍFICOS	16
CAPÍTULO I – APLICAÇÃO DE TÉCNICAS ESPECTRAIS DE INFRAVERMELHO PRÓXIMO PARA AVALIAR A QUALIDADE E ATRIBUTOS COMPOSICIONAIS DO CACAU: UMA VISÃO GERAL	17
RESUMO	17
1 INTRODUÇÃO	19
2 FUNDAMENTOS BÁSICOS DA ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO	20
3 TRATAMENTO QUIMIOMÉTRICO DOS DADOS ESPECTRAIS	22
3.1 PRÉ-TRATAMENTO DOS DADOS E QUÍMIOMETRIA PARA CALIBRAÇÃO MULTIVARIADA	22
3.2 SELEÇÃO DAS AMOSTRAS PARA CALIBRAÇÃO	23
3.3 VALIDAÇÃO E MODELAGEM DO MODELO NIR DESENVOLVIDO	24
3.4 NÚMERO DE FATORES E DETECÇÃO DE OUTLIERS	26
4 APLICAÇÃO DA TÉCNICA NIR EM CACAU	27
4.1 PREVISÃO DA COMPOSIÇÃO QUÍMICAS, MUDANÇAS BIOQUÍMICAS E MICROBIOLÓGICAS NO CACAU NÃO FERMENTADO, DURANTE E APÓS A FERMENTAÇÃO UTILIZANDO A TÉCNICA NIR	28
4.2 IDENTIFICAÇÃO DA VARIEDADE DE CACAU	31
4.3 IDENTIFICAÇÃO DE FRAUDES	32
4.4 PREVISÃO DE PROPRIEDADES SENSORIAIS DO CACAU	32
4.5 DETERMINAÇÃO DO CONTEÚDO FENÓLICO DO CACAU	33
4.6 DIFERENCIAÇÃO RÁPIDA E IDENTIFICAÇÃO DA ORIGEM GEOGRÁFICA DE CACAU	34
5 CONCLUSÃO	37

SUMÁRIO

AGRADECIMENTOS	37
REFERÊNCIAS	38
CAPÍTULO II – PREDIÇÃO RÁPIDA E SIMULTÂNEA DE UMIDADE, NITROGÊNIO TOTAL E LIPÍDEOS EM CACAU DA AMAZÔNIA (<i>Theobroma cacao</i>) NÃO FERMENTADO UTILIZANDO A TÉCNICA FT-NIR	45
RESUMO	45
1 INTRODUÇÃO	46
2 MATERIAL E MÉTODOS	48
2.1 OBTENÇÃO E PREPARAÇÃO DAS AMOSTRAS	48
2.2 ANÁLISES FÍSICO-QUÍMICAS	48
2.3 AQUISIÇÕES ESPECTRAIS	49
2.4 DESENVOLVIMENTO DOS MODELOS DE PREDIÇÃO	50
3 RESULTADOS E DISCUSSÃO	51
3.1 ANÁLISES FÍSICO-QUÍMICAS	51
3.2 INFORMAÇÕES ESPECTRAIS	52
3.3 MODELOS DE PREDIÇÃO	55
4 CONCLUSÃO	59
REFERÊNCIAS	60
CAPÍTULO III – GEOGRAPHICAL DISCRIMINATION OF GROUND AMAZON COCOA BY NEAR-INFRARED SPECTROSCOPY: INFLUENCE OF SAMPLE PREPARATION	63
1 INTRODUCTION	64
2 MATERIAL AND METHODS	66
2.1 SELECTING COLLECTION REGIONS	66
2.2 SAMPLE COLLECTION AND PREPARATION	66
2.3 DEFATTING OF COCOA SAMPLES	67
2.4 OBTAINING NIR SPECTRA	68

SUMÁRIO

2.5 DISCRIMINATION MODEL DEVELOPMENT BY GEOGRAPHICAL ORIGIN	68
3 RESULTS AND DISCUSSION	69
3.1 SPECTRAL ANALYSIS	69
3.2 DISCRIMINATION MODELS FOR RAW COCOA	70
3.3 DISCRIMINATION MODELS FOR DRIED UNFERMENTED COCOA .	72
3.4 DISCRIMINATION MODELS FOR FERMENTED COCOA	73
3.5 DISCRIMINATION MODELS FOR DEGREASED COCOA	75
4 CONCLUSIONS	79
ACKNOWLEDGEMENTS	79
FUNDING SOURCES	80
ETHICAL GUIDELINES	80
STATEMENTS	80
REFERENCES	80
CONCLUSÃO GERAL	86
ANEXO	87
ANEXO A – ARTIGO PUBLICADO	87

INTRODUÇÃO GERAL

A técnica de espectroscopia vibracional na região do Infravermelho Próximo (NIR) apresenta-se como uma excelente opção, para processos que requerem rapidez, exatidão e baixo custo (CHEN et al., 2009). Métodos espectroscópicos em análise quantitativa estão de forma geral, associados ao uso de ferramentas quimiométricas, as quais auxiliam na identificação e quantificação de componentes de interesse. Tais ferramentas matemáticas e estatísticas envolvendo calibração multivariada associadas a técnica NIR são ideais para o monitoramento e controle de qualidade de produtos agroindustriais (VARMUZA et al., 2009).

A técnica NIR já vem sendo testada para controle de qualidade de sementes de cacau e seus subprodutos para predição de lipídeos, proteínas, umidade (VASELÁ et al., 2007; TEYE & HUANG, 2014), açúcar (COSTA FILHO, 2009), pH e índice de fermentação (TEYE et al., 2015), identificação de fraudes por adição de banha de porco em chocolates (CHE MAN et al., 2005; TEYE et al., 2014). Os modelos preditivos utilizados para tais avaliações geralmente são os de regressão por mínimos quadrados parciais (PLS) e por componentes principais (PCR) (MEVIK; WEHRENS, 2007).

Além de análises preditivas, os dados espectrais também podem ser utilizados para avaliações mais simples, que inicialmente não requerem correlações com outros métodos analíticos, como por exemplo, para discriminação de espécies e origem geográfica do cacau, utilizando análise de componentes principais (Principal Components Analysis -PCA) (TEYER et al., 2013)

O método PCA é a base da maioria dos métodos modernos para tratamento de dados multivariados. Esta técnica permite extrair de um determinado conjunto de dados informações relevantes para que sejam mais bem entendidas (FERREIRA et al., 1999; CAMO, 2011). Proporciona uma visão geral da principal informação interpretável contida em uma tabela multidimensional (CAMO, 2011).

A técnica NIR apresenta inúmeras vantagens no que se refere à rapidez e precisão, tem muitas aplicações na indústria alimentar e é comumente utilizada para discriminação de espécies e determinação da composição química em diversas matrizes alimentares.

REFERÊNCIAS

CAMO ASA. **The Unscrambler** Software 10.1, trial edn, 2011.

CHE MAN, Y.B.; SYAHARIZA, Z.A.; MIRGHANI, M.E.S.; JINAP, S.; BAKAR, J. Analysis of potential lard adulteration in chocolate and chocolate products using Fourier transform infrared spectroscopy. **Food Chemistry** 90, 815–819. 2005.

CHEN, Q.; ZHAO, J.; LIN, H. Study on discrimination of Roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition, **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy** 72, 845–850. 2009.

COSTA FILHO, P. A. DA. Rapid determination of sucrose in chocolate mass using near infrared spectroscopy, **Analytica Chimica Acta** 631 (2) 206–211. 2009.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O. Quimiometria I. calibração multivariada, um tutorial. **Química Nova**, São Paulo, vol. 22, n. 5, 1999.

MEVIK, B-H., WEHRENS, R. The pls Package: Principal Component and Partial Least Squares Regression in R. **Journal of Statistical Software**. Volume 18. 2007.

TEYE, E.; HUANG, X.; DAI, H.; CHEN, Q. Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy** 114, 183–189. 2013.

TEYE, E.; HUANG, X.; LEI, W.; DAÍ, H. Feasibility study on the use of Fourier transform near-infrared spectroscopy together with chemometrics to discriminate and quantify adulteration in cocoa beans, **Food Research International** 55, 288–293. 2014.

TEYE, E.; HUANG, X.; SAM-AMOH, L. K.; TAKRAMA, J.; BOISON, D.; BOTCHWAY, F.; KUMI, F. Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis, **Food chemistry** 176, 403–410. 2015.

VARMUZA, K.; FILZMOSER, P. **Introduction to multivariate statistical analysis in chemometrics**, New York . CRC Press/Taylor & Francis Group, 2009.

VESELÁ, A.; BARROS, A. S.; SYNYTSYA, A.; DELGADILLO, I.; ČOPÍCOVÁ, J.; COIMBRA, M. A. Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder, **Analytica Chimica Acta** 601, 77–86. 2007.

OBJETIVOS

GERAL

✓ Avaliar a utilização da técnica NIR para identificação da origem geográfica de cacau e predição de seus componentes majoritários.

ESPECÍFICOS

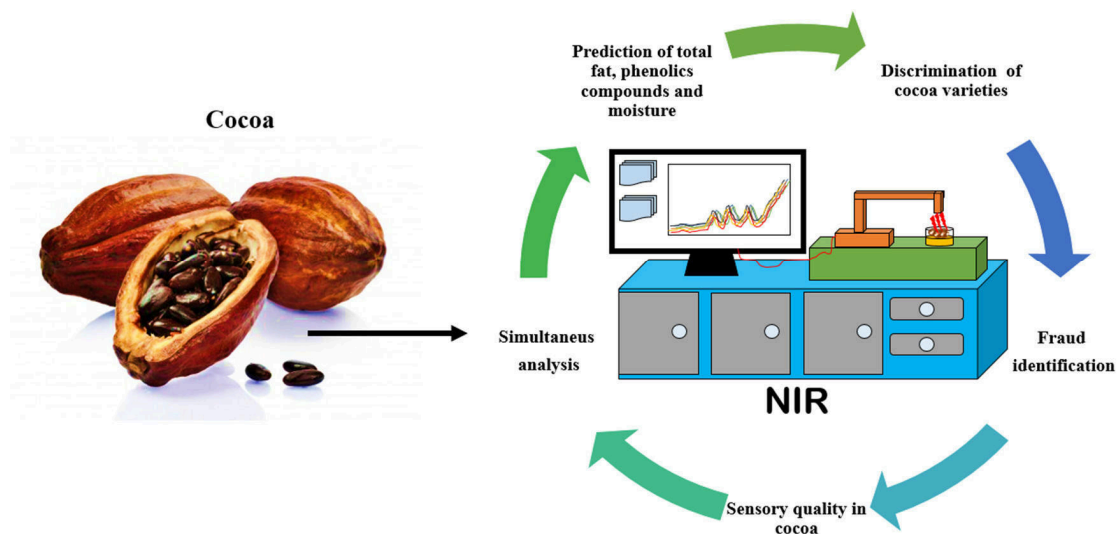
- ✓ Realizar um levantamento bibliográfico dos artigos que utilizam a técnica nir em cacau.
- ✓ Quantificar pelo método tradicional de análise, o teor de humidade, lipídios e proteínas nas amostras de cacau
- ✓ Correlacionar os dados das análises tradicionais com os dados espectrais NIR coletados.
- ✓ Analisar amostras cruas, secas, fermentadas e fermentadas desengorduradas.
- ✓ Aplicar Análise de Componentes principais nos dados NIR coletados para verificar a formação de grupos em função da origem geográfica das amostras das diferentes regiões.

CAPITULO I

APLICAÇÃO DE TÉCNICAS ESPECTRAIS DE INFRAVERMELHO PRÓXIMO PARA AVALIAR A QUALIDADE E ATRIBUTOS COMPOSICIONAIS DO CACAU: UMA VISÃO GERAL

Manuscrito do artigo a ser submetido no Journal of Near Infrared Spectroscopy (Qualis B1, Capes).

RESUMO



Este artigo de revisão apresenta, de maneira sucinta, conceitos referentes à Técnica analítica que utiliza a região do Infravermelho Próximo (NIR). Tem como eixo principal a aplicação de NIR para predição de componentes químicos, discriminação de amostras e identificação de fraudes em cacau e seus subprodutos. Dentre os componentes químicos mais caracterizados nos artigos destacaram-se os compostos fenólicos (totais, taninos, flavanóis) e lipídios, além de propriedades físico-químicas como umidade. Estes componentes são os de maior interesse na indústria de beneficiamento do cacau. Através da revisão bibliográfica, pode-se observar o potencial de NIR para a indústria cacauceira, uma vez que a maior parte das análises convencionais são demoradas e onerosas, em contra partida o NIR oferece diversas vantagens

importantes para o aperfeiçoamento do processo industrial, reduzindo os impactos e aumentando a eficiência.

Palavras-chave: Revisão bibliográfica; NIR; Cacau.

DESTAQUES

- NIR é uma técnica de predição com análises simultâneas
- Predição simultânea de gordura total, compostos fenólicos e umidade.
- NIR possibilita a discriminação de variedades de cacau.
- NIR possibilita a identificação de fraudes em cacau.
- NIR prediz qualidade sensorial em cacau.

1 INTRODUÇÃO

A região do Infravermelho Próximo (NIR) apresenta-se como uma excelente opção, para processos industriais que requerem rapidez, exatidão e baixo custo (CHEN et al., 2009). Esta técnica tem sido utilizada em análises qualitativas e quantitativas em sementes de cacau e seus subprodutos, obtendo sucesso em suas aplicações (DAVIES et al., 1991; CHE MAN et al., 2005; TEYE et al., 2014).

NIR já vem sendo empregado no controle de qualidade de sementes de cacau e seus subprodutos para predição de lipídeos, nitrogênio, umidade (VESELÁ et al., 2007; TEYE & HUANG, 2015), açúcar (COSTA FILHO, 2009), pH e índice de fermentação em amêndoas (TEYE et al., 2015) e identificação de fraudes (CHE MAN et al., 2005; TEYE et al., 2014). Além de análises preditivas, o infravermelho próximo é utilizado para outros fins em análise de amêndoas de cacau, como por exemplo, identificação de sua origem geográfica (TEYE et al., 2013).

Métodos espectroscópicos em análise quantitativa estão de forma geral, associados ao uso de ferramentas quimiométricas, as quais auxiliam na identificação e quantificação de componentes de interesse. Tais ferramentas matemáticas e estatísticas envolvendo calibração multivariada associadas a técnica NIR são ideais para o monitoramento e controle de qualidade de produtos agroindustriais (VARMUZA & FILZMOSER, 2009).

Dentre os vários elementos que compõem as amêndoas de cacau, seus lipídeos são os constituintes de maior valor e que mais interfere na qualidade das amêndoas. Diferenças na composição química da manteiga de cacau causam diferenças importantes na cinética de cristalização. Enquanto que proporção entre os teores de ácidos graxos saturados e insaturados é de fundamental importância na consistência do chocolate (FOUBERT et al., 2006; COHEN & JACKIX, 2005).

Outro elemento importante na composição destas sementes são os compostos fenólicos, que têm sido largamente estudados em razão dos efeitos benéficos à saúde, como atividade antioxidante, importante na prevenção de reações oxidativas e na formação de radicais livres, bem como na proteção contra danos ao DNA (WOLLGAST & ANKLAM, 2000).

A técnica NIR apresenta inúmeras vantagens no que se refere à rapidez e precisão, tem muitas aplicações na indústria alimentar e é comumente utilizada para discriminação de

espécies e determinação da composição química em diversas matrizes alimentares. Desta forma, o objetivo geral desta revisão bibliográfica é fazer um levantamento de dados sobre a espectroscopia no infravermelho próximo e sua aplicação na predição de diversos componentes em cacau.

2 FUNDAMENTOS BÁSICOS DA ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO

A espectroscopia no infravermelho próximo, mais conhecido como NIR, é um método analítico quantitativo e qualitativo sofisticado, baseado em fundamentos básicos da física como radiação eletromagnética e da química. A química é reconhecida como a aplicação de métodos estatísticos e matemáticos para calibração multivariada e análise de dados químicos e físicos (PASQUINI, 2003; POREP, KAMMERER & CARLE, 2015).

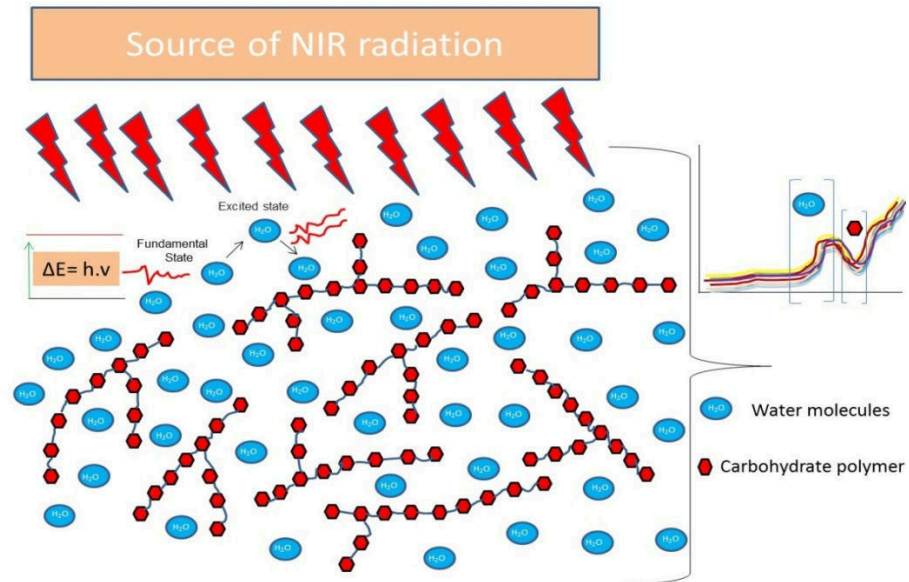
De modo geral, a espectroscopia baseia-se no fato de que as moléculas absorvem radiação em quantidades discretas de energia, com absorção apenas em determinadas frequências, referentes às características do corpo em estudo. Se a molécula receber radiação eletromagnética com “exatamente” a mesma energia de uma dessas vibrações, então a luz será absorvida. Desta forma, grupos funcionais e estruturas de macromoléculas podem ser identificados (SKOOG, HOLLER & NIEMAN, 2005).

A região do infravermelho corresponde aproximadamente à parte do espectro eletromagnético situada entre as regiões do visível e das micro-ondas. As aplicações da espectroscopia de infravermelho ocorrem em três principais categorias do espectro eletromagnético: nas regiões espectrais do infravermelho próximo (NIR), infravermelho médio (MIR) e infravermelho distante (MID) (BELLON, VIGNEAU & SÉVILA, 1994; CEN & HE, 2007). Não existe um limite específico definindo onde termina uma região e começa outra, é comum encontrar na literatura diferentes valores para cada faixa, entretanto, a IUPAC (International Union of Pure and Applied Chemistry) define a região NIR entre 780 a 2500 nm. Para que a espectroscopia na região do infravermelho próximo possa ser utilizada em análises qualitativas e quantitativas é importante a interpretação e o conhecimento sobre as bandas espectrais nessa região do espectro (CEN & HE, 2007).

NIR oferece muitas vantagens em comparação às técnicas tradicionais de análise, como medições sem preparação previa da amostra, aplicações em linhas industriais, aplicações on-line, análises rápidas, diretas e econômicas (PASQUINI, 2003; POREP, KAMMERER & CARLE, 2015; GREDILLA et al., 2016). De acordo com o modo de coleta das informações espectrais, as interações entre a radiação eletromagnética e a matérias podem ser obtidas através de diversas formas. As mais comumente empregadas são a transmitância/absorbância, transflectância, refletância e interância, dependendo do preparo da amostra e do instrumento usado para análise (SALGUERO-CHAPARRO et al., 2013). O modo mais utilizado para substâncias sólidas é a refletância e para líquidos a transmitância/absorbância.

Moléculas como a água absorvem de forma precisa e constante a radiação NIR. Quando o material analisado é ou está úmido, os resultados da análise podem ser afetados, pois as bandas de absorção da água se sobrepõem as bandas dos demais componentes, comprometendo a eficiência da análise NIR (NICOLAÏ et al, 2007). Na **Figura 1**, pode ser observada uma representação esquemática da influência das moléculas de água na avaliação de outros componentes da amostra utilizando a técnica NIR.

Figura 1 – Esquema simplificado da influência da água durante análise por espectroscopia no infravermelho próximo.



Fonte: Autor

Grupos químicos importantes como fenóis, anéis aromáticos, além de gorduras, açúcares e complexo químicos de glicoproteínas, tornam o meio rico em espectros sobrepostos e difíceis de serem compreendidos (TEYE et al., 2013). No entanto as aplicações de técnicas estatísticas, principalmente as análises multivariadas, auxiliam na organização espectral e na elaboração de informações mais precisas.

3 TRATAMENTO QUIMIOMÉTRICO DOS DADOS ESPECTRAIS

3.1 PRÉ-TRATAMENTO DOS DADOS E QUÍMIOMETRIA PARA CALIBRAÇÃO MULTIVARIADA

Os espectros do infravermelho próximo são caracterizados por serem sobrepostos e por consideráveis desvios da linha de base. Estes desvios são um fenômeno como diversas causas

intrínsecas e extrínsecas como a dispersão da luz nas amostras, variações na temperatura externa, densidade da amostra, viscosidade da amostra, umidade, tamanho e porosidade das partículas, além do ruído espectral provenientes do espectrometro. Desta forma, o pré-tratamento dos dados se faz necessário para melhorar a correlação linear de absorbância com a concentração do analito (LONG, ENGEL & SIEMENS, 2008).

No contexto da necessidade de métodos sofisticados para avaliar e melhorar a resolução de espectros provenientes das análises NIR, a quimiometria surgiu como uma proposta inovadora e moderna para enfrentar esses desafios. A quimiometria é aplicação de métodos estatísticos e matemáticos para calibração multivariada e análise de dados químicos e físicos. O desenvolvimento de um modelo de classificação e calibração multivariada são metodologias importantes e amplamente relatadas na literatura para diversos fins, como na caracterização da fermentação de sementes de cacau (ACULEY et al., 2010), na previsão do teor de cálcio nos vinhos espumantes (VÉSTIA et al., 2019), autenticidade e rastreabilidade em bebidas (KAMILOGLU, 2019), entre outras aplicações.

A calibração multivariada é uma técnica estatística de correlação quantitativa, que correlaciona os dados de espectros obtidos das amostras com os dados químicos ou físicos obtidos por técnicas convencionais de análise. O modelo de calibração multivariada pode ser usado para prever resultados quantitativos de interesse, com ampla faixa de aceitabilidade, correlacionando os resultados com diversos parâmetros estatísticos confiáveis que corroboram os resultados com dados precisos, rápidos e com erro mínimo. Após a construção do modelo de calibração multivariada, as técnicas convencionais de análise podem ser substituídas por uma técnica preditiva de alta confiabilidade e com mínimo impacto ao meio ambiente e ao operador (FU & YING, 2016).

3.2 SELEÇÃO DAS AMOSTRAS PARA CALIBRAÇÃO

As coleções de amostra para a calibração e construção do modelo de predição multivariada, devem estar bem distribuídas e variadas. A qualidade da equação de calibração é influenciada por vários parâmetros relacionados com a seleção das amostras, portanto, esta é uma etapa crucial no procedimento de calibração. A uniformidade e a qualidade dos dados de

calibração são essenciais, assim, as amostras devem ser representativas e devem fazer parte da mesma população em estudo, bem como devem ser previstas as possíveis variações, tudo dentro da ampla faixa da população de interesse (RINNANA et al., 2014).

3.3 VALIDAÇÃO E MODELAGEM DO MODELO NIR DESENVOLVIDO

Após a seleção das variáveis e modelagem dos espectros em equações adequadas e representativas da amostra, os modelos de regressão e as equações de calibração podem ser testados, avaliados e validados quanto ao seu desempenho na predição do método desenvolvido. Dentre os parâmetros utilizados para avaliar a robustez de um modelo NIR de predição podemos destacar os valores dos coeficientes de determinação de calibração (R^2), erro padrão médio de validação cruzada (RMSECV) e de predição (RMSEP), a relação de desempenho do desvio (RPD) e bias. Abaixo podemos observar as equações utilizada para a aquisição de cada parâmetro.

$$\mathbf{RMSECV} = \sqrt{\frac{1}{M} \cdot \sum_{i=1}^M (y_i^{meas} - y_i^{pred})^2} \quad (1)$$

$$\mathbf{RMSEP} = \sqrt{\frac{1}{M} \cdot \sum_i^M (y_i^{ref} - y_i^{est})^2} \quad (2)$$

$$\mathbf{R^2} = \left(1 - \frac{1}{RPD^2}\right) \times 100 \quad (3)$$

$$\mathbf{RPD} = \frac{SD}{SEP_{bias}} = \frac{\sqrt{\frac{1}{M-1} \sum_{i=1}^M (y_i^{meas} - y_m)^2}}{\sqrt{\frac{1}{M-1} \sum_{i=1}^M (y_i^{meas} - y_i^{pred} - bias)^2}} \quad (4)$$

$$\mathbf{Bias} = \frac{\sum_{i=1}^M y_i^{meas} - y_i^{pred}}{M} \quad (5)$$

Onde:

M = número de amostras;

y_i^{meas} = Valor de concentração medido da amostra i ;

y_i^{pred} = Valor de concentração previsto da amostra i ;

y_m = Valor médio da concentração.

y_i^{ref} = Valor de referência para amostra i

y_i^{est} = resultado estimado do modelo para amostra i

O erro quadrático médio de calibração (RMSEC) é uma importante estimativa preditiva do erro nos valores do conjunto de calibração, sendo considerada uma avaliação interna que prediz com coerência o erro do modelo e não o erro da predição (QU et al., 2015). O erro médio quadrático de validação cruzada (RMSECV) consiste na seleção aleatória de dados do conjunto de calibração, com posterior eliminação dos mesmos, os dados restantes são ajustados a uma equação de calibração. Este procedimento é repetido continuamente, até que todas as amostras de dados de calibração tenham sido excluídas pelo menos uma vez do processo (DALE et al., 2013).

O erro médio quadrático de predição (RMSEP) pode ser determinado como sendo a raiz quadrada do quadrado médio dos resíduos para um número de amostras presentes no conjunto de validação, onde o resíduo é resultado da diferença entre os valores previstos menos os valores de referência para as amostras de fora do conjunto de calibração. Portanto, o erro médio quadrático de predição (RMSEP), é a previsão das variações no conjunto de valores de referência, previstos a partir de um conjunto de validação. Para que o conjunto de calibração represente os dados de forma confiável, o conjunto teste deve ser independente. No caso do cacau faz-se necessário que as amostras sejam provenientes de um campo de produção ou localidade diferente, uma vez que as análises devem cobrir toda a faixa espectral relevante de amostras em teste, visto que as equações de calibração exibirão propriedades variáveis para diferentes conjuntos de teste (DALE et al., 2013; LI et al., 2016).

O erro sistemático entre os valores preditos e os valores de referência é calculado através do parâmetro, viés. Diversos erros padrões encontrados em modelos NIR, como o erro padrão de previsão (SEP), o erro padrão de calibração (SEC) e o erro padrão de validação cruzada (SECV), respectivamente, podem ser corrigidos por Viés (QU et al., 2015). O coeficiente de determinação (R^2), é um dos parâmetros estatísticos mais relevantes e significativos para uma análise crítica dos dados e de modelos de predição. Pois, fornece uma boa estimativa da qualidade e adequação dos ajustes feitos no modelo de predição. O coeficiente de determinação (R^2), fornece valores entre 0–1, onde 0 indica que não há nenhuma correlação e 1 revela que existe uma forte correlação linear entre os valores reais e os valores preditos com o modelo NIR. Outro parâmetro estatístico de desvio padrão útil e importante é a relação de previsão para desvio (RPD), valores mais altos deste parâmetro indicam uma melhor capacidade de predição do modelo (VANDEGINSTE et al., 1997). Por fim, o parâmetro bias é a média do desvio entre

o conjunto de dados originais e os valores preditos (TSUCHIKAWA & KOBORI, 2015; MILLER & MILLER, 1993).

3.4 NÚMERO DE FATORES E DETECÇÃO DE OUTLIERS

Uma etapa importante e crucial para a construção e calibração de um bom modelo NIR, passa pela seleção ideal do número de fatores, componentes ou variáveis latentes que serão usados. Essas variáveis devem incluir o máximo possível de informações sobre a amostra. Assim, se forem utilizados poucos fatores, pode ser que estes não sejam suficientes para abranger toda uma complexidade de informações espectrais relevantes, ideais para estabelecer um modelo confiável. Por outro lado, o excesso de número de fatores pode gerar redundâncias e ruídos provenientes da variabilidade espectral. Mas, a estimativa de parâmetros e as previsões são mais precisas, quanto maior o número de amostras dispostas no conjunto de calibração (XU et al., 2013; PIERNA et al., 2016).

Algumas ferramentas podem ser utilizadas para aprimorar a escolha do número de fatores necessário à construção de um modelo NIR. Por exemplo, a validação cruzada pode ser utilizada para prever o número de fatores. O teste de previsão é desejável, e pode ser feita pela plotagem do RMSEP versus o número de fatores. A validação cruzada repetida, validação cruzada alinhada e a validação cruzada dupla, representam estratégias de otimização dos complexos modelos de regressão, que por vez são mais eficientes na estimativa de erros de previsão (MORA & SCHIMLECK, 2008).

Como todo modelo que usa estimativas estatísticas, alguns dados acabam por não se encaixarem no modelo, essas amostras são chamadas de outliers. Essas amostras são caracterizadas por terem um alto valor de resíduo entre o valor previsto e o valor real, além de espectros anômalos. A retirada dos outliers pode melhorar a qualidade preditiva do modelo. No entanto, este procedimento deve ser feito com muita atenção, pois, às vezes, esses outliers possuem alguma variação que pode ser útil a qualidade do modelo, e desta forma sua retirada acaba por enfraquecer a uniformidade e a eficiência do modelo desenvolvido (HE et al., 2015; LIN et al., 2017).

Assim como na escolha do número de fatores, onde podem ser utilizadas ferramentas estatísticas para melhor definir a escolha, ferramentas estatísticas também são utilizadas para avaliar a calibração para a presença de outliers. Por exemplo, Ballabio e Consonni (BALLABIO & CONSONNI, 2013), descrevem que a estatística de alavancagem e as parcelas de T^2 e Q de Hotelling (soma dos quadrados de resíduos), podem ajudar na uniformidade de calibração. A estatística T^2 se fundamenta na soma dos escores quadrados normalizados, o que determina a variação em cada amostra de calibração do modelo. Enquanto que a estatística Q , determina o quanto cada amostra está de conformidade com o modelo desenvolvido. Desta forma grandes valores de Q , são indicativos de que o modelo não foi suficiente para cobrir todas as amostras. Por outro lado, grandes valores de T^2 e as estatística de alavancagem mostram que a amostra possui grande influência na calibração do modelo NIR desenvolvido.

4 APLICAÇÃO DA TÉCNICA NIR EM CACAU

Avaliação de sementes de cacau utilizando NIR, não é um método de análise tão recente. Uma das primeiras pesquisas foi publicada em 1989, para a avaliação do potencial de NIR para o controle de qualidade de produtos de cacau em pó, predizendo umidade, gordura e sacarose (PERMANYER & PEREZ, 1989). Em 1991, foi utilizado para avaliação sensorial, utilizando um grande número de amostras de chocolate de uma ampla variedade de grãos de cacau de diferentes áreas geográficas (DAVIES et al., 1991). Logo em 1999, a técnica foi avaliada para a predição da qualidade sensorial e medição da cristalização da gordura do cacau na linha de produção de manteiga e chocolate (BOLLIGER; ZENG & WINDHAB, 1999). A partir disto, pode-se observar o avanço nas pesquisas que utilizaram a técnica NIR (Tabela 1).

NIR elevou o patamar das análises em cacau, possibilitando acompanhar e realizar avaliações simultâneas em todo o processo de produção, como, avaliação da composição de cacau não fermentado (ÁLVAREZ, et al., 2012; TROGNITZ et al., 2013), monitoramento do processo de fermentação (ACULEY et al., 2010; HUE et al., 2014; SUNOJ, IGATHINATHANE & VISVANATHAN, 2016), monitoramento das mudanças microbiológicas que ocorrem durante a fermentação do cacau (NIELSEN, NIELSEN & VAN DEN BERG, 2008) monitoramento da secagem e do efeito do método de secagem (ACULEY

et al., 2010; SUNOJ, IGATHINATHANE & VISVANATHAN, 2016; HUE et al., 2014), discriminação geográfica (TEYE et al., 2013) e detecção de fraudes (QUELAL-VASCONEZ et al., 2018).

4.1 PREVISÃO DA COMPOSIÇÃO QUÍMICAS, MUDANÇAS BIOQUÍMICAS E MICROBIOLÓGICAS NO CACAU NÃO FERMENTADO, DURANTE E APÓS A FERMENTAÇÃO UTILIZANDO A TÉCNICA NIR.

Uma série de componentes puderam ser preditos no cacau utilizando a técnica NIR (Tabela1), dentre eles, os componentes com maior destaque são respectivamente, o teor de gordura total, seguido de compostos fenólicos e umidade.

Tabela 1 – Componentes preditos em cacau e subprodutos utilizando a técnica por espectroscopia vibracional no infravermelho próximo no período de 1989 a 2018.

Amostra analisada	Componentes preditos	Referência*
Cacau seco não fermentado	Procyanidins B2, procyanidins B5, procyanidins C1, umidade,	(1)
	Cafeína, (-) – epicatequina, gordura, teobromina	(2, 1)
Cacau em diferentes níveis de fermentação	Nitrogênio amoniacal (NH ₃)	(3)
	Índice de fermentação	(4, 5)
	pH, polifenóis totais	(5)
Cacau seco fermentado	Atividade antioxidante, Índice de fermentação	(6)
	Acidez, conteúdo de casca, proteína	(7)
	Ácido láctico, ácidos orgânicos	(8)
	pH	(8,7)
	Cafeína, teobromina, umidade	(1, 7)
	Polifenóis totais	(8,6)
	Gordura	(9, 1, 7)
	Procyanidins B2, procyanidins B5, procyanidins C1. (-) – epicatequina	(1) (1, 8)
Liquor de cacau	(-) – epicatequina, (+) – catequina, oligômeros de procianidina (monômero a decâmero)	(10)
Cacau durante o armazenamento	Índice de fermentação, polifenóis totais, pH.	(5)
Cacau em pó	Gordura, umidade	(11, 12)
	Nitrogênio total	(11)
	Sacarose	(12)

*1-TROGNITZ et al., 2013; 2- ÁLVAREZ, et al., 2012; 3- HUE et al., 2014; 4- ACULEY et al., 2010; 5- SUNOJ, IGATHINATHANE & VISVANATHAN, 2016; 6- CAPORASO et al., 2018; 7- HASHIMOTO et al., 2018; 8- KRÄHMER et al., 2015; 9- TEYE & HUANG, 2015, 10- WHITACRE et al., 2003; 11- VESELÁ et al., 2007; 12- PERMANYER & PEREZ, 1989.

Antes mesmo da fermentação do cacau, a qualidade do fruto do cacau pode ser predita (ÁLVAREZ, et al., 2012; TROGNITZ et al., 2013), através da quantificação de compostos fenólicos. Apesar do potencial antioxidante de compostos fenólicos, sua intensa concentração no cacau atribui para o amargor e adstringência, proporcionando produtos com características sensoriais pouco apreciadas ao paladar do consumidor mais exigente (TROGNITZ et al., 2013).

A quantificação da manteiga do cacau por NIR também merece destaque, pois altas concentrações de gordura podem agregar valor ao cacau, além disso, pode ser predita antes mesmo da fermentação, uma vez que não sofre variação durante a fermentação. Álvarez et al. (2012), testou a previsão de componentes não voláteis em grãos de cacau não fermentados (seco ao sol), de uma população genética diversificada, para quantificar gordura, cafeína, teobromina e (-) epicatequina. O modelo de predição construído para estas amostras provou que a espectroscopia NIR pode ser usada como um método rápido para determinar estes componentes antes mesmo da fermentação. Logo, Trognitz et al. (2013), também avaliaram sementes de cacau não fermentadas de diferentes genótipos em diferentes períodos de colheita. Além dos compostos preditos anteriormente, também conseguiram prever umidade, Procyanidins B2, Procyanidins B5 e Procyanidins C1, e ainda, puderam diferenciar características próprias para os diferentes momentos de colheitas.

Após a colheita do cacau, as sementes passam pelo processo de fermentação que pode perdurar por até sete dias. O acompanhamento desta etapa é fundamental para a obtenção de cacau de alta qualidade através da formação de precursores de aroma (TROGNITZ et al., 2013). Assim trabalhos recentes (ACULEY et al., 2010; HUE et al., 2014; SUNOJ, IGATHINATHANE & VISVANATHAN, 2016), conseguiram, durante todo o processo de fermentação, medir os índices de fermentação do cacau com eficiência. Diferentes técnicas convencionais foram utilizadas pelos autores para a posterior correlação com os dados NIR, dentre elas, técnicas por espectrofluorometria (BioView), espectrofotometria UV-vis e técnica de Conway e Byrne (CONWAY & BYRNE, 1933), através da quantificação de nitrogênio amoniacal (NH_3).

Aculey et al. (2010), também obtiveram através da aplicação da análise de componentes principais (PCA), a formação de grupos com base nos tempos de fermentação. Sunoj et al. (SUNOJ, IGATHINATHANE & VISVANATHAN, 2016), também avaliaram a previsão de outros parâmetros de qualidade como, pH e polifenóis totais nos grãos durante a fermentação. Tais parâmetros sofrem modificação durante esta etapa e puderam ser preditos com sucesso utilizando a técnica NIR. Estas mudanças bioquímicas e físicas que acontecem no interior das sementes durante a fermentação são provocadas por uma série de fatores, incluindo a atividade microbiana que ocorre fora da semente, particularmente, na polpa do cacau. O processo é complexo e envolve atividades de vários grupos diferentes de microrganismos (NIELSEN, NIELSEN & VAN DEN BERG, 2008).

Nielsen et al. (2008), relataram o primeiro estudo que utilizou NIR para monitorar as mudanças microbiológicas que ocorrem durante a fermentação do cacau. Eles avaliaram as micro-populações bacterianas e eucarióticas durante a fermentação pela técnica que utiliza eletroforese em gel com gradiente desnaturante (DGGE) e correlacionaram com as mudanças dentro dos grãos determinadas pelo NIR usando análise de dados multivariada, e obtiveram excelentes resultados.

Para o cacau fermentado pode-se observar um maior número de pesquisas, uma vez que geralmente, o cacau é adquirido seco fermentado, em grandes volumes, as técnicas rápidas que utilizam o infravermelho próximo apresentam-se como excelente alternativa para o controle de qualidade e rastreabilidade da origem do cacau fermentado. Além disso, também pode ser aplicado para analisar e/ou controlar o cacau durante o armazenamento (SUNOJ, IGATHINATHANE & VISVANATHAN, 2016) e depois do processamento (VESELÁ et al., 2007).

4.2 IDENTIFICAÇÃO DA VARIEDADE DE CACAU

Geralmente, a variedade do cacau é avaliada em decorrência das diferentes características que podem proporcionar ao cacau. O estudo realizado por Teye et al. (TEYE et al., 2015A), mostraram sua novidade na possibilidade de combinar os dados de língua eletrônica (ET) e NIR para uma classificação precisa das variedades de grãos de cacau. Os sensores únicos (NIRS e ET) tiveram uma taxa de classificação entre 83 e 93%, enquanto que a fusão de dados (ET-NIR) teve uma taxa de classificação de 100%. Neste caso, a técnica NIR quando associada a técnica de língua eletrônica mostrou-se claramente mais eficiente do que, quando se utilizou as técnicas isoladas.

Barbin et al. (2018), predisseram cor e composição química e cacau fermentado para caracterizar diferentes variedades do fruto. Nesta pesquisa os espectros NIR das amostras exibiram perfis característicos para cada variedade e a análise de componentes principais identificou grupos diferentes para cada variedades. Mais uma vez, a eficácia da técnica NIR foi comprovada para identificação de variedades de cacau.

4.3 IDENTIFICAÇÃO DE FRAUDES

Modelos de predição de fraudes já foram testados para identificação de fraudes em cacau fermentado (TEYE et al., 2014), cacau em pó (QUELAL-VASCONEZ et al., 2018), chocolate (CHE MAN et al., 2005). Tais modelos apresentaram-se eficientes e precisos para as predições de fraudes em cacau e seus subprodutos.

O cacau em pó também é um produto altamente valioso que pode ser adulterado com matérias-primas de baixo custo. Quelal-Vasconez, et al. (2018), testaram a adulteração de pó de cacau com farinha de alfarroba, já que pequenas quantidades dessa farinha não alteram as características de cor, aroma e sabor do produto final. Como resultado foi alcançado precisão de classificação de 100%, com isso, puderam-se distinguir, com precisão, pós de cacau puro de amostras adulteradas.

Che Man et al. (2005), empregaram a técnica para detecção de banha de porco em formulações de chocolate, obtiveram excelentes resultados e puderam prever diferentes concentrações de banha de porco nas formulações. Teye et al. (2014), avaliaram performance da técnica na identificação de fraudes na produção de chocolate através da adição de sementes não fermentadas sobre as amêndoas fermentadas, as quais influenciavam na qualidade do chocolate. A técnica demonstrou alta precisão nos resultados das análises quando correlacionado aos métodos tradicionais.

4.4 PREVISÃO DE PROPRIEDADES SENSORIAIS DO CACAU

Após a fermentação do cacau, as amêndoas sofrem o processo de secagem, esse processo é fundamental para a formação do aroma e sabor característico do chocolate. Aroma e sabor são os principais fatores que determinam a qualidade potencial do cacau, eles frequentemente dependem das condições ambientais específicas (colheitas, fermentações) e não apenas do genótipo (TROGNITZ et al., 2013).

As predições da qualidade sensorial do cacau por NIR já foi anteriormente testadas por Davies et al. (1991), com base na origem geográfica, tipo de amostra e grau de fermentação. O modelo proposto por Davies et al. (1991), identificou apenas 64% das amostras de baixa qualidade, enquanto rejeitava 20% das amostras de qualidade aceitáveis. Em vista da incerteza inerente a todos os dados sensoriais, o autor sugeriu pesquisas mais profundas. No entanto, em 2013, a qualidade sensorial do cacau foi novamente avaliada por provadores experientes em amostras de nove genótipos de cacau da cidade de Waslala (Nicarágua), os principais compostos químicos de sabor também foram quantificados e correlacionados com os resultados de NIRS, e desta vez, NIR correspondeu às expectativas para aplicação como método alternativo de análise (TROGNITZ et al., 2013).

4.5 DETERMINAÇÃO DO CONTEÚDO FENÓLICO DO CACAU

De acordo com relatórios recentes (FIGUEROA-HERNÁNDEZ et al., 2019; VALADEZ-CARMONA et al., 2018; MAZZUTTI et al., 2018), os compostos fenólicos são importantes metabolitos secundários com inúmeras atividades biológicas como atividade antioxidante, antitumoral e anti-inflamatória. Estes compostos são uma das classes de metabolitos mais estudados, pois estão associados a diversas rotas de biossíntese, mecanismos de supressão e regulação, além de serem moléculas associativas, formando inúmeros compostos complexos, com redes de deslocamento de elétrons, o que confere propriedades de cor, aroma e sabor (QUIROZ-REYES & FOGLIANO, 2018).

O número de estudos sobre predição de compostos fenólicos presentes no cacau utilizando a técnica NIR tem aumentado consideravelmente nos últimos anos (EFRAIM, ALVES, JARDIM, 2011). Os diversos modelos de predição avaliados obtiveram resultados satisfatórios (Tabela 2), provando que a técnica pode ser utilizada para predição de inúmeros componentes em cacau.

Tabela 2 - Compostos fenólicos preditos em cacau utilizando a técnica de análise por Espectroscopia Vibracional no Infravermelho Próximo

Compostos fenólicos	Característica da amostra	Referências*	
Totais	Cacau durante a fermentação	(1)	
	durante o armazenamento		
	Cacau fermentado	(2,3)	
Taninos	Procyanidins (B2, B5, C1).	Cacau fermentado e não fermentado	(4)
	Oligômeros de procianidina	Liquor de cacau	(5)
Flavanóis	(-) –epicatequina	Cacau não fermentado	(5)
		Cacau fermentado	(4, 2)
	(+) – catequina	Liquor de cacau	(6)
		Liquor de cacau	(6)

* 1- SUNOJ, IGATHINATHANE & VISVANATHAN, 2016; 2- KRÄHMER et al., 2015; 3- CAPORASO et al., 2018; 4- TROGNITZ et al., 2013; 5- ÁLVAREZ, et al., 2012; 6- WHITACRE et al., 2003

Dentre os principais compostos fenólicos preditos para sementes de cacau, taninos e flavanóis são evidentes. Taninos, destacam-se, principalmente os procyanidins B2, procyanidins B5, procyanidins C1 e oligômeros de procianidina (monômero a decâmero) (TROGNITZ et al., 2013; WHITACRE et al., 2003). Dentre os flavanóis, (-) – epicatequina e (+) – catequina (ÁLVAREZ, et al., 2012; TROGNITZ et al., 2013; WHITACRE et al., 2003, KRÄHMER et al., 2015). Contudo, a maior parte das pesquisas optou por quantificar e prever a concentração total de compostos fenólicos (SUNOJ, IGATHINATHANE & VISVANATHAN, 2016; KRÄHMER et al., 2015, HUANG et al., 2014, CAPORASO et al., 2018).

4.6 DIFERENCIAÇÃO RÁPIDA E IDENTIFICAÇÃO DA ORIGEM GEOGRÁFICA DE CACAU

O cacau comercial de qualidade é comercializado como cacau a granel, onde cada lote representa o cacau de vários agricultores. O conhecimento da origem geográfica é frequentemente reconhecido e apreciado pela indústria de alimentos e consumidores, é um fator

importante que pode agregar valor às amêndoas de cacau (TEYE et al., 2013). A identificação da origem geográfica agrega valor ao produto, facilita a identificação de campos de produção de qualidade, traz investimentos em pesquisa, agrega reconhecimento ao produtor e desenvolvimento social e econômico ao local.

Geralmente nos estudos em questão (Tabela 3), são primeiramente avaliados componentes ou características de interesse que podem se diferenciar entre as regiões, como por exemplo, cor, compostos voláteis (ACULEY et al., 2010), fenólicos, antioxidantes, índice de fermentação (HUE et al., 2014; WHITACRE et al., 2003; CAPORASO et al., 2018), gordura (TEYE & HUANG, 2015) e características sensoriais (DAVIES et al., 1991), para que possam correlacionar com os dados espectrais NIR do cacau de cada região. Dentre os tratamentos matemáticos de avaliação da origem geográfica, a análise dos componentes principais (PCA) é utilizada com maior frequência para extrair informações relevantes dos dados espectrais e fornecer tendências de cluster visíveis.

Teye et al. (2013), foi o primeiro a realizar um estudo mais específico sobre métodos de classificação multivariada, para a classificação e discriminação de grãos de cacau de acordo com as diferentes regiões de cultivo em Gana. Diferentes métodos de classificação multivariada foram testados, ressaltando a qualidade das informações derivadas dos dados espectrais após a varredura. Modelos matemáticos ideais conseguiram extrair o máximo de informações relevantes, o que foi de fundamental importância para classificação e discriminação de grãos de cacau de acordo com as diferentes regiões de cultivo em Gana.

Tabela 3 – Discriminação geográfica de cacau sob diferentes características de amostragem apresentada por diferentes autores.

Regiões estudadas	Objeto de estudo	Característica da amostra	Referências*
África Ocidental (Costa do Marfim; Camarões), América do Sul, Malásia, W. Samoa, Sri Lanka, Nova Guiné	Odor e Flavor	Sementes cruas e torradas, massa de chocolate e chocolate acabado.	DAVIES et al., 1991
Gana (Western, Brong-Ahafo, Ashianti, Central, Volta e Leste)	Teste de corte, cor, compostos voláteis	Sementes durante a fermentação e amêndoas completamente fermentadas.	ACULEY et al., 2010
Estados Unidos, Reino Unido, Oeste africano, Sudeste Asiático, América Central	Compostos Fenólicos	Liquor de cacau	WHITACRE et al., 2003.
Gana (Ashanti, Brong-Ahafo, Central, Oriental, Volta, norte ocidental e sul ocidental)	Diferença entre quatro métodos de classificação multivariada: LDA, KNN, BPANN, SVM.	Cacau fermentado e seco	TEYE et al., 2013
Camarões, Equador, Gana, Indonésia, Madagascar, República Dominicana, Trinidad e Tobago	Índice de fermentação por nitrogênio amoniacal (NH ₃)	Cacau em diferentes níveis de fermentação	HUE et al., 2014
Diferentes regiões de cultivo de cacau de Gana	Gordura total	Cacau fermentado e seco pronto para exportação	TEYE & HUANG, 2015
Brasil, Camarões, Equador, Gana, Indonésia, Costa do Marfim, México, Nigéria, Venezuela	Índice de fermentação (FI), polifenóis totais (TP) e atividade antioxidante (AA)	Cacau fermentado e seco	CAPORASO et al., 2018

5 CONCLUSÃO

A aplicação da técnica NIR é bastante simples, contudo o trajeto de aquisição de informações espectrais e de dados analíticos tradicionais é demorado e oneroso. No entanto, uma vez que o banco de informações espectrais e analíticas é obtido e o modelo matemático é construído, as análises tornam-se simples, rápidas e econômicas. A partir destas etapas, pode-se realizar o mapeamento da qualidade e rastreamento da origem deste fruto, como mostrado anteriormente. Além das análises correlacionadas acima, inúmeras outras informações podem ser obtidas, isso dependerá da necessidade de avaliação para um controle de qualidade desejado. Uma vez que foi comprovada a eficiência da técnica NIR em diversas aplicações com cacau, pode-se presumir que este método de análise poderá suprir praticamente todas as necessidades analíticas durante o controle de qualidade do fruto e em outras etapas do processamento da indústria. A qualidade do cacau pode ser eficientemente monitorada durante toda a etapa de produção do cacau fermentado, até mesmo nas etapas posteriores a fermentação, para o controle de qualidade dos seus subprodutos. Assim a técnica NIR é uma ferramenta muito eficiente, rápida e ecologicamente correta, o que contribui para o aumento da eficiência dos processos industriais sem causar impactos negativos no meio ambiente.

AGRADECIMENTOS

Os autores agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e ao Instituto Tecnológico da Vale (Cacau P2) (Brasil) pelo suporte financeiro, e ao Programa de Pós Graduação em Ciência e Tecnologia de Alimentos da Universidade Federal do Pará (PPGCTA/UFPA).

REFERÊNCIAS

ACULEY, P. C.; SNITKJAER, P.; OWUSU, M.; BASSOMPIERE, M.; TAKRAMA, J.; NØRGAARD, L.; NIELSEN, D. S. Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics. *Journal of Food Science*, 75, S300–S307, 2010.

ÁLVAREZ, C.; PÉREZ, E.; CROS, E.; LARES, M., ASSEMAT, S.; BOULANGER, R.; DAVRIEUX, F. The use of near infrared spectroscopy to determine the fat, caffeine, theobromine and (–)-epicatechin contents in unfermented and sun-dried beans of Criollo cocoa, *Journal of Near Infrared Spectroscopy* 20, 307–315, 2012.

B.G.M. VANDEGINSTE, D.L. MASSART, L.M.C. BUYDENS, S. DE JONG, P.J. LEWI, J. SMEYERS-VERBEKE, *Handbook of chemometrics and qualimetrics: Part A*, in: B.G.M. Vandeginste, S.C. Rutan (Eds.), *Data Handling in Science and Technology*, Elsevier Science B.V., Amsterdam, Netherlands 20 (1997).

BALLABIO, D.; CONSONNI, V. Classification tools in chemistry. Part 1: linear models. PLS DA. *Analytical Methods*, 5, 3790-3798, 2013.

BARBIN, D. F.; MACIEL, L. F.; BAZONI, C. H. V.; RIBEIRO, M. S.; CARVALHO, R. D. S.; BISPO, E. S.; MIRANDA, M. P. S.; HIROOKA, E. Y. Classification and compositional characterization of diferente varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses, *Journal Food Sci. Technol.* 55 (7), 2457–2466, 2018.

BELLON, V.; VIGNEAU, J. L.; SÉVILA, F. Infrared and near-infrared technology for the food industry and agricultural uses: on-line applications. *Food Control*, 5, 21-27, 1994.

BOLLIGER, S.; ZENG, Y.; WINDHAB, E. J. In-line Measurement of Tempered Cocoa Butter and Chocolate by Means of Near-Infrared Spectroscopy, *JAOCS* 76 (6), 659-667, 1999.

CAPORASO, N.; WHITWORTH, M. B.; FOWLER, M. S.; FISK, I. D. Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans, *Food Chemistry* 258, 343–351, 2018.

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

CHE MAN, Y.B.; SYAHARIZA, Z.A.; MIRGHANI, M.E.S.; JINAP, S.; BAKAR, J. Analysis of potential lard adulteration in chocolate and chocolate products using Fourier transform infrared spectroscopy, *Food Chemistry* 90, 815–819. 2005.

CHEN, Q.; ZHAO, J.; LIN, H. Study on discrimination of Roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 72, 845–850. 2009.

COHEN, K. O.; JACKIX, M. N. H. Estudo do liquor de cupuaçu, *Ciência e Tecnologia dos Alimentos* 25 (1) (2005) 182-190, 2005.

CONWAY, E. J.; BYRNE, A. An absorption apparatus for the microdetermination of certain volatile substances, the micro-determination of ammonia. *Biochemistry Journal*, 27 (2), 419–429, 1933.

COSTA FILHO, P. A. da. Rapid determination of sucrose in chocolate mass using near infrared spectroscopy, *Analytica Chimica Acta* 631 (2) 206–211, 2009.

DALE, L. M.; THEWIS, A.; BOUDRY, C.; ROTAR, I.; DARDENNE, P.; BAETEN, V.; PIERNA, J. A. F.; Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: a review, *Appl. Spectrosc. Rev.* 48, 142-159, 2013.

DAVIES, A.M.C.; FRANKLIN, J.G.; GRANT, A.; GRIFFITHS, N.M; SHEPHERD, R.; FENWICK, G.R. Prediction of chocolate quality from near-infrared spectroscopic measurements of the raw cocoa beans, *Vibrational Spectroscopy* 2, 161-172. 1991.

EFRAIM, P.; ALVES, A.; JARDIM, D. C. P. Revisão: Polifenóis em cacau e derivados: teores, fatores de variação e efeitos na saúde, *Braz. J. Food Technol*, Campinas 14 (3), 181-201, 2011.

FIGUEROA-HERNÁNDEZ, C.; MOTA-GUTIERREZ, J.; FERROCINO, I.; HERNÁNDEZ-ESTRADA, Z. J.; GONZÁLEZ-RÍOS, O.; COCOLIN, L.; SUÁREZ-QUIROZ, M. L. The challenges and perspectives of the selection of starter cultures for fermented cocoa beans, *International Journal of Food Microbiology* 301, 41–50, 2019.

FOUBERT, I.; VANROLLEGHEM, P. A.; THAS, O.; DEWETTINK, K. Influence of chemical composition on the isothermal cocoa butter crystallization, *Journal of Food Science* 69 (9), 2006.

FU, X.P.; YING, Y.B. Food safety evaluation based on near infrared spectroscopy and imaging: a review, *Crit. Rev. Food Sci. Nutr.* 56, 1913-1924, 2016.

GREDILLA, A.; VALLEJUELO, S. F. O.; ELEJOSTE, N.; DIEGO, A.; MADARIAGA, J. M. Non destructive Spectroscopy combined with chemometrics as a tool for Green Chemical Analysis of environmental samples: a review, *Trac. Trends Anal.Chem.* 76, 30-39, 2016.

HASHIMOTO, J.C.; LIMA, J. C.; CELEGHINI, R. M. S.; NOGUEIRA, A. B.; EFRAIM, P.; POPPI, R. J.; PALLONE, J. A. L. Quality Control of Commercial Cocoa Beans (*Theobroma cacao* L.) by Near-infrared Spectroscopy, *Food Analytical Methods* 11 (5), 1510–1517, 2018.

HE, Z. H.; LI, M. C.; MA, Z. H. Design of a reference value-based sample-selection method and evaluation of its prediction capability, *Chemometr. Intell. Lab. Syst.* 148, 72-76, 2015.

HUANG, X.; TEYE, E.; SAM-AMOA, L. K.; HAN, F.; YAO, L.; TCHABO, W. Rapid measurement of total polyphenols content in cocoa beans by data fusion of NIR spectroscopy and electronic tongue, *Anal. Methods*, 2014.

HUE, C.; BRAT, P.; GUNATA, Z.; SAMANIEGO, I.; SERVENT, A.; MOREL, G.; KAPITAN, A.; BOULANGER, R.; DAVRIEUX, F. Near Infra-red Characterization of Changes in Flavan-3-ol Derivatives in Cocoa (*Theobroma cacao* L.) as a Function of Fermentation Temperature, *Journal of Agricultural and Food Chemistry* 62, 10136–10142, 2014.

HUE, C.; GUNATA, Z.; BERGOUNHOU, A.; ASSEMAT, S.; BOULANGER, R.; SAUVAGE, F. X.; DAVRIEUX, F. Near infrared spectroscopy as a new tool to determine cocoa fermentation levels through ammonia nitrogen quantification, *Food Chemistry* 148, 240–245, 2014.

J.C. MILLER, J.N. MILLER, *Statistics for Analytical Chemistry*, Ellis Horwood Ltd., Chichester, England 3 (1993) 233 p.

KAMILOGLU, S. Authenticity and traceability in beverages. *Food Chemistry*. 277 (2019) 12-24.

KRÄHMER;. ENGEL, A.; KADOW, D.; ALI, N.; UMAHARAN, P.; KROH, L. W.; SCHULZ, H. Fast and neat – Determination of biochemical quality parameters in cocoa using near infrared spectroscopy. *Food Chemistry* 181, 152–159, 2015.

LI, J. L.; SUN, D. W.; CHENG, J. H. Recent advances in nondestructive analytical techniques for determining the total soluble solids in fruits: a review, *Compr. Rev. Food Sci. Food Saf.* 15, 897-911, 2016.

LIN, Z. D.; WANG, Y. B.; WANG, R. J.; WANG, L. S.; LU, C. P.; ZHANG, Z. Y.; SONG, L. T.; LIU, Y. Improvements of the Vis-NIR model in prediction of soil organic matter content using spectral pretreatments, sample selection, and wavelength optimization, *J. Appl. Spectrosc.* 84, 529-534, 2017.

LONG, D. S. ENGEL, R. E.; SIEMENS, M. C. Measuring grain protein concentration with in-line near infrared reflectance spectroscopy. *Agronomy Journal*, 100, 247-252, 2008.

MAZZUTTI, S.; RODRIGUESA, L. G. G.; MEZZOMO, N.; VENTURI, V.; FERREIRA, S. R. S. Integrated green-based processes using supercritical CO₂ and pressurized ethanol applied to recover antioxidant compounds from cocoa (*Theobroma cacao*) bean hulls, *The Journal of Supercritical Fluids* 135, 52–59, 2018.

MORA, C. R.; SCHIMLECK, L. R. On the selection of samples for multivariate regression analysis: application to near-infrared (NIR) calibration models for the prediction of pulp yield in *Eucalyptus nitens*, *Can. J. For. Res.* 38, 2626-2634, 2008.

NICOLAÏ, B. M.; BEULLENS, K.; BOBELYN, E.; PEIRS, A.; SAEYS, W.; TORRICELLI; VANDENDRIESSCHE, A. T.; VERBOVEN, P. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biology and Technology*, 46, 99-118, 2007.

NIELSEN, D. S.; NIELSEN, P. S.; VAN DEN BERG, F. Investigating the fermentation of cocoa by correlating denaturing gradient gel electrophoresis profiles and near infrared spectra, *Int. J. Food Microbiol.* 125, 133–40, 2008.

PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications, *J. Braz. Chem. Soc.* 14, 198-219, 2003.

PERMANYER, J. J.; PEREZ, M. L. Compositional Analysis of Powdered Cocoa Products by Near Infrared Reflectance Spectroscopy, *Journal of Food Science* 54 (3), 768-769, 1989

PIERNA, J. A. F.; VINCKE, D.; BAETEN, V.; GRELET, C.; DEHARENG, F.; DARDENNE, P. Use of a multivariate moving window PCA for the untargeted detection of contaminants in agro-food products, as exemplified by the detection of melamine levels in milk using vibrational spectroscopy, *Chemometr. Intell. Lab. Syst.* 152, 157-162, 2016.

POREP, J. U.; KAMMERER, D. R.; CARLE, R. On-line application of near infrared (NIR) spectroscopy in food production, *Trends Food Sci. Technol.* 46, 211-230, 2015.

QU, J. H.; LIU, D.; CHENG, J. H.; SUN, D. W.; MA, J.; PU, H. B.; ZENG, X. A. Applications of near-infrared spectroscopy in food safety evaluation and control: a review of recent research advances, *Crit. Rev. Food Sci. Nutr.* 55, 2015.

QUELAL-VASCONEZ, M. A.; PEREZ-ESTEVE, E.; ARNAU-BONACHERA, A.; BARAT, J. M.; TALENS, P. Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy, *Food Control* 92, 183-189, 2018.

QUIROZ-REYES, C. N.; FOGLIANO, V. Design cocoa processing towards healthy cocoa products: The role of phenolics and melanoidins, *Journal of Functional Foods* 45, 480-490, 2018.

RINNANA, Å.; ANDERSSON, M.; RIDDERC, C.; ENHELSENA, S. B. Recursive weighted partial least squares (rPLS): an efficient variable selection method using PLS, *J. Chemometr.* 28, 439-447, 2014.

SALGUERO-CHAPARRO, L.; BAETEN, V.; FERNANDEZ-PIERNA, J. A.; PENARODRÍGUEZ, F. Near infrared spectroscopy (NIRS) for on-line determination of quality parameters in intact olives. *Food Chemistry*, 139, 1121-1126, 2013.

SKOOG, D. A.; HOLLER, F. J.; NIEMAN, T. A. "Principles of Instrumental Analysis", 5th Edition, Saunders College Publishing. Harcourt Brace College Publishers 725-750, 840-848, 2005.

SUNOJ, S.; IGATHINATHANE, C.; VISVANATHAN, R. Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy, *Computers and Electronics in Agriculture* 124, 234–242, 2016.

TEYE, E.; HUANG, X.; DAI, H.; CHEN, Q. Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 114, 183–189, 2013.

TEYE, E.; HUANG, X. Novel Prediction of Total Fat Content in Cocoa Beans by FT-NIR Spectroscopy Based on Effective Spectral Selection Multivariate Regression, *Food Analytical Methods* 8 (4), 945-953, 2015.

TEYE, E.; HUANG, X.; LEI, W.; DAI, H. Feasibility study on the use of Fourier transform near-infrared spectroscopy together with chemometrics to discriminate and quantify adulteration in cocoa beans, *Food Research International* 55, 2014.

TEYE, E.; HUANG, X.; SAM-AMOAHAH, L. K.; TAKRAMA, J. BOISON, D.; BOTCHWAY, F.; KUMI, F. Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis, *Food chemistry* 176, 403 -410, 2015.

TEYE, E.; HUANG, X.; TAKRAMA, J.; HAIYANG, G. Integrating nir spectroscopy and electronic tongue together with chemometric analysis for accurate classification of cocoa bean varieties, *Journal of Food Process Engineering* 37, 560–566, 2015A

TROGNITZ, B.; CROS, E.; ASSEMAT, S.; DAVRIEUX, F.; FORESTIER-CHIRON, N; AYESTAS, E.; KUANT, A.; SCHELDEMAN, X.; HERMANN, M. Diversity of Cacao Trees in Waslala, Nicaragua: Associations between Genotype Spectra, Product Quality and Yield Potential, *PLoS ONE* 8 (1), 2013.

TSUCHIKAWA, S.; KOBORI, H. A review of recent application of near infrared spectroscopy to wood science and technology, *J. Wood Sci.* 61, 213-220, 2015.

VALADEZ-CARMONA, L.; ORTIZ-MORENO, A.; CEBALLOS-REYES, G.; MENDIOLA, J. A.; IBÁÑEZ, E. Valorization of cacao pod husk through supercritical fluid extraction of phenolic compounds, *The Journal of Supercritical Fluids* 131, 99–105, 2018.

VARMUZA, K., FILZMOSE, P. Introduction to multivariate statistical analysis in chemometrics, New York . CRC Press/Taylor & Francis Group (2009).

VESELÁ, A.; BARROS, A. S.; SYNITSYA, A.; DELGADILLO, I.; ČOPÍCOVÁ, J.; COIMBRA, M. A. Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder, *Analytica Chimica Acta* 601, 77–86, 2007.

VÉSTIA, J.; BARROSO, J. M.; FERREIRA, H.; GASPAR, L.; RATO, A. L. Predicting calcium in grape must and base wine by FT-NIR spectroscopy. *Food Chemistry*. 276 (2019) 71-76.

WHITACRE, E.; OLIVER, J.; VAN DEN BROEK, R.; VAN ENGELLEN, P.; KREMERS, B.; VANDER HORST, B.; STEWART, M.; JANSENBEUVINK, A. Predictive Analysis of Cocoa Procyanidins Using Near-Infrared Spectroscopy Techniques, *Journal of Food Science* 68 (9), 2618-2622, 2003.

WOLLGAST, J.; ANKLAM, E. Review on polyphenols in *Theobroma cacao*: changes in composition during the manufacture of chocolate and methodology for identification and quantification, *Food Research International*, Essex 33 (6), 423-447, 2000.

XU, L.; YAN, S. M.; CAI, C. B.; YU, X. P. Untargeted detection and quantitative analysis of poplar balata (PB) in Chinese propolis by FT-NIR spectroscopy and chemometrics, *Food Chem.* 141, 4132-4137, 2013.

CAPITULO II

PREDIÇÃO RÁPIDA E SIMULTÂNEA DE UMIDADE, NITROGÊNIO TOTAL E LIPÍDEOS EM CACAU DA AMAZÔNIA (*Theobroma cacao*) NÃO FERMENTADO UTILIZANDO A TÉCNICA FT-NIR

Manuscrito do artigo a ser submetido no Journal of Near Infrared Spectroscopy
(Qualis B1, Capes).

RESUMO

No Brasil, os estados do Pará e da Bahia são os maiores produtores de cacau. O estado do Pará tradicionalmente se destaca por uma produção de cacau baseada, quase em sua totalidade, na agricultura familiar e com produto elevada qualidade. Esta qualidade está relacionada a uma série de fatores que englobam desde condições climáticas até as técnicas de fermentação. Os parâmetros de qualidade que não são influenciados pela fermentação podem ser avaliados logo após a colheita, como o teor de lipídeos, que representa um dos componentes de maior valor no cacau, o teor de umidade de sementes *in natura*, para avaliação do rendimento do cacau após a secagem, e o teor de nitrogênio, um indicador de compostos nitrogenados que influenciam no aroma e sabor do cacau após a fermentação. Para essa avaliação, técnicas rápidas de análise utilizando a espectroscopia no infravermelho próximo podem ser utilizadas, poupando tempo, minimizando a utilização de solventes poluentes e reduzindo o custo de análises. Nesse contexto, este trabalho desenvolveu modelos de predição para nitrogênio, umidade e lipídios totais em sementes de cacau não fermentadas. Os resultados foram avaliados pelo erro quadrático-médio, coeficiente de correlação, relação e desempenho de desvio. Os resultados obtidos demonstram que tais componentes das sementes de cacau não fermentadas podem ser rapidamente preditos pela espectroscopia FT-NIR utilizando regressão PLS.

Palavras chave: *Theobroma cacao*; NIR; nitrogênio total; umidade; lipídeos.

1 INTRODUÇÃO

O cacau é um importante produto agrícola do estado do Pará por ser uma espécie nativa da região, ser produzido principalmente pela agricultura familiar e, dessa forma, ser uma importante fonte de geração de renda sustentável para os agricultores locais. Em 2016, o estado despontou como o maior produtor de cacau do Brasil, atingindo a produção de aproximadamente 118 mil toneladas de cacau em amêndoas e, assim, ultrapassando a produção do estado da Bahia, até então principal produtor (MENDES, 2016).

Neste contexto, torna-se interessante a implantação de técnicas rápidas de análises que permitem avaliar as sementes ainda na pós-colheita e, dessa forma, prever determinados componentes que são essenciais para a qualidade do cacau fermentado (SUNOJ, IGATHINATHANE, VISVANATHAN, 2016; TEYE, et al., 2015).

A gordura, ou manteiga do cacau, é um dos componentes mais requeridos nas sementes do cacau e agrega valor ao produto. A composição da manteiga do cacau influencia diretamente características reológicas e na textura do chocolate, e é o constituinte mais importante para a maioria dos fabricantes e processadores de chocolate. A gordura do cacau é composta por cerca de 60% de ácidos graxos saturados (esteárico, palmítico) e 35-43% de ácidos graxos insaturados (ácidos oléico, linoleico, linolênico) (AFOAKWA, 2007). O teor total de gordura nas sementes do cacau está relacionado ao grau de maturação do fruto (ÁLVAREZ et al., 2012) e independe do tempo de fermentação, com isso pode ser avaliada ainda no campo.

Outro parâmetro de qualidade que pode ser levado em consideração é o teor de proteínas, que está entre um dos itens relevantes para o processo de obtenção de sabor e cor do chocolate. Isto ocorre porque ao final da fermentação, com a morte das amêndoas, há o rompimento. As frações proteicas do cacau têm sido implicadas influenciando tanto o potencial bioativo quanto as propriedades sensoriais do cacau e dos produtos de cacau (RAWEL et al., 2019). A prática da fermentação no processo de amêndoas de cacau tem um papel muito importante também por produzir peptídeos bioativos. A proteólise que ocorre durante a fermentação do cacau pode produzir fragmentos peptídicos que trazem benefícios à saúde, como anti-hipertensivos, antiobesogênico e antidiabético (HALIZA et al., 2019; CORONADO-CÁCERES et al., 2020). Assim, a fermentação do grão de cacau pode não apenas ser estudada para a produção de alimentos e bebidas palatáveis, mas também para a produção de produtos funcionais saudáveis (DOMÍNGUEZ-PÉREZ, et al., 2020).

Tradicionalmente, a determinação do teor de gorduras e proteínas é efetuada por métodos de química úmida que demandam elevado tempo de análise, são caros e levam à geração de resíduos poluentes. Por exemplo, o método soxleth (AOAC Official Method 963.15) para o teor de gorduras utiliza solventes como hexano, éter de petróleo, etanol ou isopropanol (ElKhorri et al.). Por outro lado, para determinação de proteínas, o método Kjeldahl (AOAC Official Method 970.22.) faz uso de ácidos fortes como o ácido sulfúrico e o ácido clorídrico, além de outros reagentes.

Atualmente, técnicas espectroscópicas como NIRs (Near Infrared Spectroscopy) contornam os problemas dos métodos tradicionais citados anteriormente, acrescentando capacidade de realizar múltiplas análises efetuadas simultaneamente. Os métodos analíticos baseados na tecnologia NIRs vêm ganhando espaço e sendo aplicados na indústria alimentar, particularmente no controle da qualidade das matérias primas agroindustriais.

A tecnologia NIRs oferece a possibilidade de medição de propriedades físicas e químicas de amostras e uma vez que o modelo de avaliação é calibrado, a operação do espectrômetro NIR se torna simples (VESELÁ et al., 2007).

A calibração de modelos NIRs requer um conjunto amostral com as propriedades analíticas desejadas previamente determinadas. O tamanho deste conjunto amostral varia com a complexidade da amostra e sua consequente interferência no sinal do analito de interesse, não havendo valores fixos para o mesmo (METROHM, 2014). Além disso, para predição de componentes em um sistema de múltiplos componentes, são utilizados métodos de calibração multivariada, tais como Regressão Linear Múltipla (MLR), Regressão de componentes principais (PCR), ou regressão de Mínimos Quadrados Parciais (PLSR). Dentre eles o mais comumente utilizado é o método de regressão PLS (CONZEN, 2006).

A quantificação de componentes em cacau utilizando a técnica NIRs já vem sendo avaliada principalmente no cacau fermentado (TEYE et al., 2013). Dentre os parâmetros de qualidade avaliados e seus subprodutos, podem-se destacar a predição de lipídeos, nitrogênio, umidade (VESELÁ et al., 2007; TEYE & HUANG, 2014), açúcar (COSTA FILHO, 2009), pH e índice de fermentação (TEYE et al., 2015).

Considerando que a tecnologia NIRs tem potencial de ser utilizado para determinação de propriedades em cacau não fermentado, de forma semelhante ao cacau fermentado, o

presente trabalho teve como objetivos desenvolver modelos NIRs para predição de umidade, nitrogênio e lipídios em cacau úmido e seco não fermentado de diferentes regiões do estado do Pará, Brasil.

2 MATERIAL E MÉTODOS

2.1 OBTENÇÃO E PREPARAÇÃO DAS AMOSTRAS

Os frutos de cacau foram colhidos nos municípios de Tucumã (sul do estado do Pará), Medicilândia (oeste do estado), e Tomé-Açu (na região central do estado). Foram coletados 50 frutos em uma propriedade familiar para cada município, totalizando 150 amostras. Os frutos maduros foram colhidos aleatoriamente nas bordas e no centro do plantio, utilizando a técnica de colheita tradicional de cada produtor. Os frutos colhidos foram quebrados, despoldados, as testas foram removidas e apenas os cotilédones foram utilizados. As amostras foram armazenadas a $-22\text{ °C} \pm 1\text{ °C}$ até as análises físico-químicas e espectroscópicas.

Cada conjunto de cotilédones obtidos por fruto foram moídos, obtendo-se de uma amostra por fruto, depois cada amostra foi peneirada utilizando uma malha de 600 μm . As amostras foram avaliadas a partir de duas categorias: amostras úmidas não fermentadas ($n=150$) e amostras secas não fermentadas ($n=150$).

2.2 ANÁLISES FÍSICO-QUÍMICAS

A determinação da umidade foi realizada por gravimetria secando-se 5 g da amostra até peso constante em estufa com circulação de ar a $105\text{ °C} \pm 1\text{ °C}$. As amostras secas nestas condições foram utilizadas para a determinação de nitrogênio total (NT) que foi quantificado

utilizando 0,1 g de amostra de cotilédone seco pelo método de Dumas em aparelho Leco 528 LC (ETHERIDGE, PESTI & FOSTER, 1998).

A determinação de lipídeos foi efetuada pelo método 963.15 da AOAC (2006) utilizando-se 3 gramas de amostra seca processada em extrator tipo soxhlet no qual a extração ocorreu durante 12 horas, utilizando éter de petróleo como solvente a uma temperatura de 55 ± 1 °C. Todas as análises foram realizadas em triplicata e os resultados foram expressos em g/100g de amostra.

Os resultados de umidade, lipídeos e nitrogênio total foram avaliados estatisticamente utilizando o software STATISTICA, buscando avaliar diferenças significativas entre os frutos dos diferentes municípios.

2.3 AQUISIÇÕES ESPECTRAIS

Os espectros NIR foram obtidos utilizando espectrômetro MPA FT-NIR da BRUKER. Os dados espectrais foram adquiridos no modo de absorbância, na faixa espectral de 3500 cm^{-1} a 12500 cm^{-1} , com 16 cm^{-1} de resolução e média de 32 varreduras por espectro. Utilizou-se o acessório de vidro de 3mL para leitura de amostras sólidas em que cerca de um grama de cotilédone moído e peneirado foi utilizado para realização da leitura espectral. As amostras foram analisadas em temperatura de 25 ± 1 °C.

Para o modelo NIR de predição de umidade foram utilizadas as amostras moídas de cotilédonesúmidos. Para o modelo de predição de nitrogênio total foram utilizadas somente amostras de cotilédones secos, enquanto que para o modelo de predição do teor de lipídeos foram utilizadas tanto amostras de cotilédones secos quanto de úmidos.

2.4. DESENVOLVIMENTO DOS MODELOS DE PREDIÇÃO

O modelo de predição para umidade em frutos úmidos não fermentados foi desenvolvido utilizando espectros de 100 amostras de cotilédones das quais 90 foram utilizadas como conjunto de calibração final do modelo e 10 amostras foram utilizadas como conjunto teste. Estes espectros tiveram pré-tratamento com Normalização vetorial SNV (Standard Normal Variate).

O modelo para a predição do teor de lipídeos foi construído com 79 amostras de cotilédones secos e 44 amostras de cotilédones úmidos. Destas amostras, 110 (70 secas e 40 úmidas) ficaram no conjunto de calibração final do modelo e 13 (9 secas e 4 úmidas) foram utilizadas como conjunto teste. Novamente no pré-processamento foi utilizada a normalização vetorial SNV.

Para a predição do teor de nitrogênio total foram utilizados espectros de 41 amostras de cotilédones dos quais 34 ficaram no conjunto final de calibração do modelo e 7 ficaram como conjunto teste. Neste caso, foi feito pré-processamento de forma a utilizar a primeira derivada para o desenvolvimento do modelo.

Os modelos de PLSR para predição de umidade, nitrogênio e lipídios foram avaliados através dos valores dos coeficientes de determinação de calibração (R^2), erro padrão médio de validação cruzada (RMSECV) e de predição (RMSEP), a relação de desempenho do desvio (RPD) e bias. Os parâmetros foram calculados de acordo com as equações:

$$\mathbf{RMSECV} = \sqrt{\frac{1}{M} \cdot \sum_{i=1}^M (y_i^{meas} - y_i^{pred})^2} \quad (1)$$

$$\mathbf{RMSEP} = \sqrt{\frac{1}{M} \cdot \sum_{i=1}^M (y_i^{ref} - y_i^{est})^2} \quad (2)$$

$$\mathbf{R}^2 = \left(1 - \frac{1}{RPD^2}\right) \times 100 \quad (3)$$

$$\mathbf{RPD} = \frac{SD}{SEP_{bias}} = \frac{\sqrt{\frac{1}{M-1} \cdot \sum_{i=1}^M (y_i^{meas} - y_m)^2}}{\sqrt{\frac{1}{M-1} \cdot \sum_{i=1}^M (y_i^{meas} - y_i^{pred} - bias)^2}} \quad (4)$$

$$\mathbf{Bias} = \frac{\sum_{i=1}^M y_i^{meas} - y_i^{pred}}{M} \quad (5)$$

Onde:

M = número de amostras;

y_i^{meas} = Valor de concentração medido da amostra i ;

y_i^{pred} = Valor de concentração previsto da amostra i ;

y_m = Valor médio da concentração.

y_i^{ref} = Valor de referência para amostra i

y_i^{est} = resultado estimado do modelo para amostra i

O processamento de dados e desenvolvimento dos modelos NIRs descritos foram efetuados utilizando-se os recursos do software OPUS 6.5 da BRUKER (BrukerOptics, Ettlingen, Alemanha).

3 RESULTADOS E DISCUSSÃO

3.1 ANÁLISES FÍSICO-QUÍMICAS

Na Tabela 1 pode-se observar os valores médios, mínimos e máximos, bem como os desvios padrão das análises físico-químicas de cacau não fermentado utilizadas para o desenvolvimento dos modelos NIR deste trabalho. Foram feitas 100 análises de umidade, cujo percentual variou entre 28% e 40%, 41 análises de nitrogênio total, cujo percentual variou entre 2.1% e 2.9%, e 123 análises de lipídios (79 amostras secas e 44 amostras parcialmente secas), cujo percentual variou entre 27% e 54%.

Tabela 1 – Análises físicas e químicas avaliadas em cacau úmido e seco.

(%)	N	Média	Faixa	SD
Umidade¹	100	32.82	28.00 – 40.70	3.94
Nitrogênio²	41	2.58	2.14 – 2.90	0.16
Lipídeos²	79	48.79	43.50 – 54.40	4.11
Lipídeos³	44	38.12	27.40 – 53.50	7.32

¹ Amostras úmidas; ² Amostras de secas; ³ Amostras parcialmente secas

Na tabela 2 pode-se visualizar os resultados das análises físico-químicas discriminados por município. Nesta tabela observa-se que não há diferença estatisticamente significativa entre as regiões avaliadas para os três componentes.

Tabela 2 – Análises físicas e químicas avaliadas em cacau para cada região.

MUNICÍPIO	UMIDADE (%)	NITROGÊNIO (%)	LIPÍDIOS (%)
TOMÉ-AÇU	33,17 ± 3,52 ^a	2,57 ± 0,18 ^a	47, 81 ± 2,77 ^a
MEDICILÂNDIA	33,14 ± 3,32 ^a	2,54 ± 0,13 ^a	46,84 ± 2,40 ^{a,b}
TUCUMÃ	30,67 ± 2,79 ^a	2,63 ± 0,17 ^a	42,84 ± 7,29 ^b

Médias seguidas da mesma letra em cada coluna não diferem significativamente entre si pelo teste de Tukey a 5% de significância.

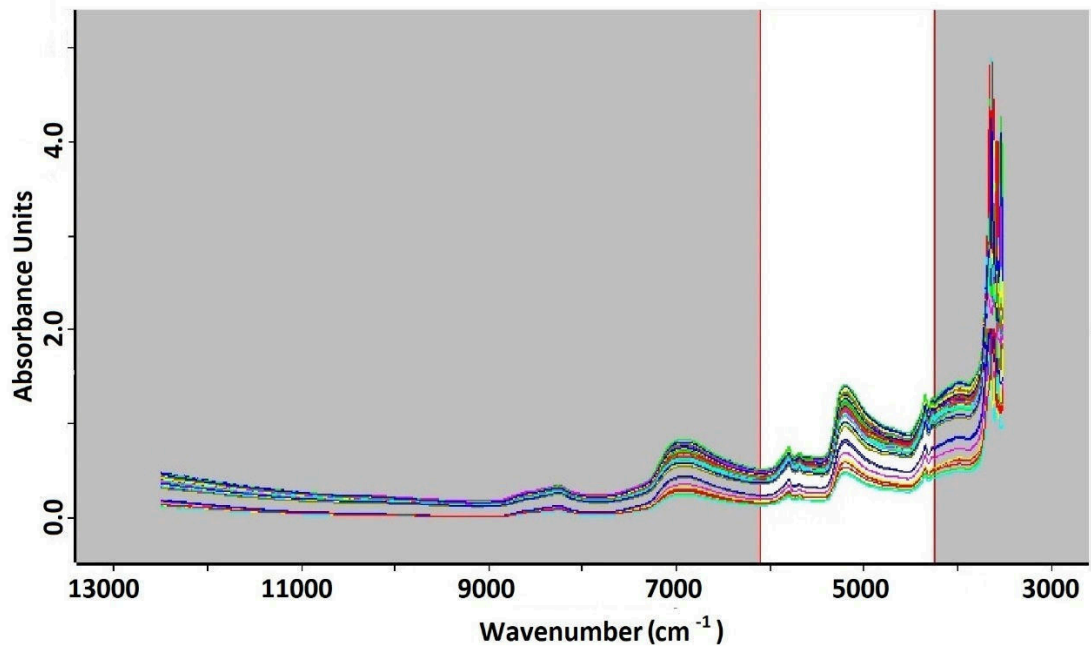
Do ponto de vista comercial, os teores mais elevados de lipídeos dos frutos de Tomé-Açu e Medicilândia valorizam os mesmos. Provavelmente, os altos teores de gordura nos frutos de Medicilândia somam-se a outras características para tornar o cacau dessa região um dos mais valorizados no Brasil.

3.2 INFORMAÇÕES ESPECTRAIS

Nas figuras 1 a 3 são apresentadas as regiões espectrais utilizadas para o desenvolvimento dos modelos preditos NIRs para umidade, nitrogênio total e lipídeos nas amostras de cacau não fermentado utilizadas neste trabalho.

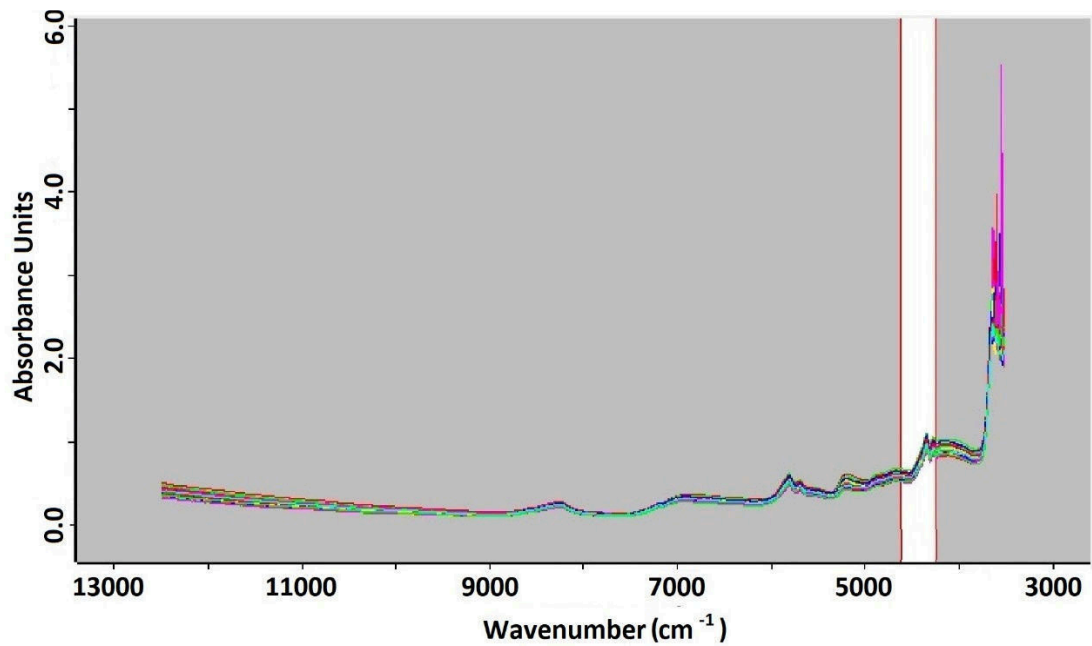
Observa-se na figura 1 que a faixa espectral para predição do teor de umidade localizou-se entre 6102 cm^{-1} a 4243 cm^{-1} , correspondendo a bandas de absorção das ligações O-H, N-H e C-H e coerentes com a complexidade das amostras de cacau que apresentam água, nitrogênio e outros compostos orgânicos

Figura 1 – Região espectral utilizada para o modelo de predição de umidade.



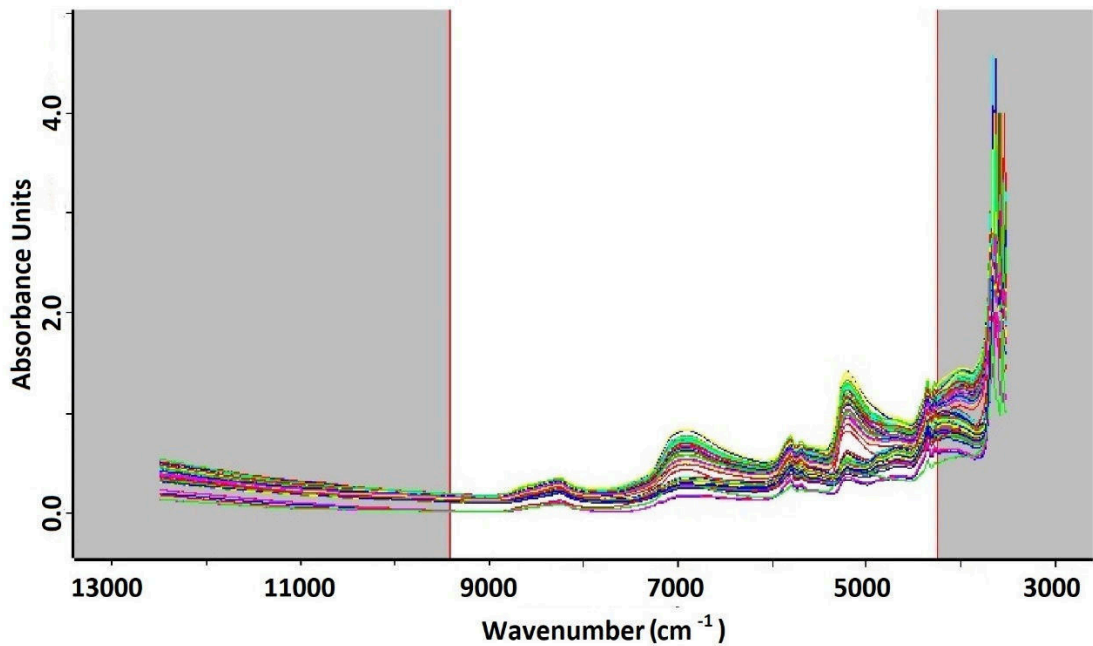
Na figura 2 é apresentada a faixa espectral utilizada para o desenvolvimento do modelo preditivo de nitrogênio total, que ficou compreendida entre 4605 cm^{-1} a 4250 cm^{-1} . Esta faixa espectral está coerente com a variável predita pelo fato de que nesta região há absorção das ligações N-H típicas de proteínas.

Figura 2 – Região espectral utilizada para o modelo de predição de nitrogênio total.



O modelo para a predição de lipídeos utilizou uma ampla faixa espectral compreendida entre 9404 cm^{-1} a 4243 cm^{-1} , conforme apresentado na figura 3. Este resultado é comparável ao de Teye & Huang (2014) que desenvolveu modelo para predição de lipídeos em amostras de cacau fermentado utilizando valores espectrais variando de 5000 cm^{-1} a 8000 cm^{-1} .

Figura 3 – Região espectral utilizada para o modelo de predição de lipídios.



3.3. MODELOS DE PREDIÇÃO

Na tabela 3 são apresentados os parâmetros de validação dos modelos desenvolvidos para predição de umidade, nitrogênio total e teor de lipídeos nas amostras deste trabalho. As eficiências dos modelos de predição podem ser avaliadas através da comparação entre os parâmetros estatísticos de validação interna e externa, como R^2 , RMSECV, RMSEP, RPD.

Tabela 3 – Parâmetros estatísticos avaliados para validação externa de umidade, nitrogênio e lipídeos utilizando PLSR.

	N	Fatores	R^2	RMSECV	RMSEP	RPD	Bias
Umidade	90	9	88.29	1.16	1.39	2.93	0.07
Nitrogênio	34	9	92.25	0.04	0.04	3.59	-0.01
Lipídios	110	10	94.04	1.83	2.00	4.09	-0.01

*N, number of samples; RMSECV, Root Mean Square Error of Cross Validation; R^2 , coefficient of correlation; RMSEP, Root Mean Square Error of Prediction; RPD, ratio of performance to deviation.

O R^2 é um indicativo de linearidade, indica em percentagem, o quanto o modelo consegue explicar os valores observados. Quanto mais tende a 100%, mais próximos os dados

observados estarão do modelo desenvolvido (VANDEGINSTE et al., 1997). Contudo, para se obter um bom modelo de predição, os parâmetros estatísticos como RPD e RMSEP devem ser avaliados em conjunto.

Os resultados apresentados na tabela 3 revelam que há uma forte correlação entre os métodos analíticos utilizados para a determinação de umidade, nitrogênio e lipídeos e os espectros NIR das amostras de cotilédone de cacau. Isto é verificado pelo valor de RPD (ratio performance deviation), definido como a razão entre o desvio padrão do método analítico e o erro de predição padrão do modelo NIRs (RMSEP), que para todos os modelos ficou superior a 2.9, sendo que alguns autores consideram que valores superiores a 2.5 já indicam uma boa correlação entre método analítico e espectro NIR (CONZEN, 2006; HUE et al., 2014; SUNOJ, IGATHINATHANE, VISVANATHAN, 2016) enquanto outros mais rigorosos sugerem valores em torno de três (WILLIAMS, 2001; FEARN, 2002).

Os valores médios de RMSECV e de RMSEP apresentados para a umidade, nitrogênio e lipídeos, sugerem que há uma expectativa de que os modelos sejam capazes de prever os parâmetros modelados com erros inferiores a 5%, valor de erro aceitável na grande maioria dos casos de determinações físico-químicas relacionadas a avaliações de qualidade de produto. Além disso, em cada modelo desenvolvido obteve-se valores de RMSECV e RMSEP muito próximos, indicando um bom equilíbrio entre o conjunto de calibração e o conjunto teste, requisito importante para um bom modelo preditivo NIR robusto (VESELÁ et al., 2007).

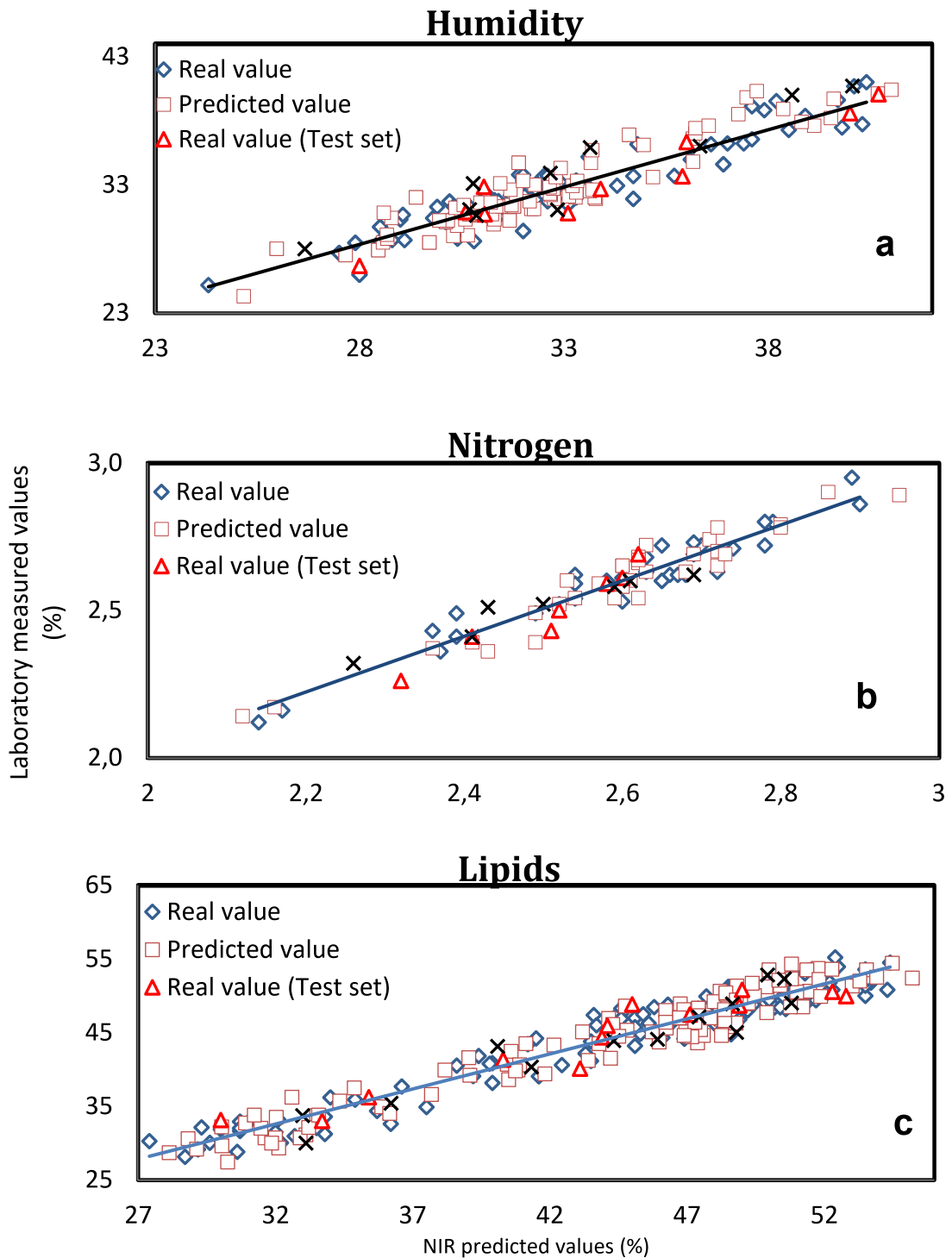
Na tabela 4 apresenta-se uma comparação entre média, valores mínimos e máximos preditos pelos métodos NIR desenvolvidos neste trabalho e os mesmos valores quantificados nas amostras com o método analítico. Os resultados desta tabela confirmam novamente a adequação dos modelos desenvolvidos para quantificação de umidade, nitrogênio e lipídeos em amostras de cacau.

Tabela 4 – Comparação entre os valores reais de concentração de lipídeos, nitrogênio e umidade, e os preditos utilizando modelos NIR.

	Valores	UMIDADE				NITROGÊNIO				LIPÍDEOS			
		Méd	Min	Max	DP	Méd	Min	Max	DP	Méd	Min	Max	DP
Validação	Reais	32,50	24,30	40,40	3,39	2,60	2,14	2,90	0,18	44,08	27,40	54,40	7,49
	Preditos	32,38	25,17	41,01	3,25	2,60	2,12	2,95	0,17	44,08	28,12	55,21	7,34
Predição	Reais	33,31	26,66	40,06	4,04	2,51	2,32	2,62	0,11	43,51	30,00	52,80	7,03
	Preditos	33,31	26,66	40,06	4,04	2,50	2,26	2,69	0,15	43,86	32,99	50,81	6,50

As figuras 4a a 4c apresentam as correlações entre valores reais e previstos para a umidade, nitrogênio e lipídeos baseados nos modelos NIRs desenvolvidos. Estas figuras reforçam os resultados obtidos anteriormente da adequação dos modelos desenvolvidos para predição dos parâmetros umidade, nitrogênio total e lipídeos em amostras de cacau não fermentado.

Figura 4 – Valores reais *versus* valores preditos (NIR) para os conjuntos de validação interna e externa (conjunto teste) de: a) umidade, b) nitrogênio total, e c) umidade, expressos em g/100g da amostra.



4 CONCLUSÃO

Neste trabalho foram desenvolvidos modelos NIRs para predição dos teores de umidade, nitrogênio total e lipídeos em amostras de cacau não fermentado de três regiões do estado do Pará, Brazil. As faixas espectrais utilizadas para o desenvolvimento destes modelos apresentaram semelhanças com faixas espectrais utilizadas para predição dos mesmos parâmetros em amostras de cacau não fermentado, sendo somente mais estreitas para o teor de nitrogênio e umidade. Os valores de RMSEP, RMSEC e RPD obtidos indicam a adequação e robustez dos modelos desenvolvidos e mostram que estes modelos se apresentam como ferramentas analíticas úteis para avaliação de parâmetros de qualidade em amostras de cacau não fermentado.

REFERÊNCIAS

AFOAKWA, E. O., PATERSON, A., & FOWLER, M. Factors influencing rheological and textural qualities in chocolate - a review. *Trends in Food Science & Technology*, 18(6), 290–298, 2007.

ÁLVAREZ, C., PÉREZ, E., CROS, E., LARES, M., ASSEMAT, S., BOULANGER, R., & DAVRIEUX, F. The use of near infrared spectroscopy to determine the fat, caffeine, theobromine and (–)- epicatechin contents in unfermented and sun-dried bean sof Criollo cocoa. *Journal of Near Infrared Spectroscopy*, 20, 307-315, 2012.

AOAC. Association of Official Analytical Chemistry – Official Method 963.15. 16 ed. Washington. Retrieved from [http:// www.aoac.org](http://www.aoac.org), 2006.

AOAC. Association of Official Analytical Chemistry. Official Method 970.22. 16 ed. Washington. Retrieved from [http:// www.aoac.org](http://www.aoac.org), 2006.

CALIGIANI, A. ET AL. Influence of fermentation level and geographical origin on cocoa bean oligopeptide pattern. *Food Chemistry*, v. 211, p. 431–439, 2016.

CONZEN, J. P. Multivariate calibration, a practical guide for developing methods in the quantitative analytical chemistry. Ettlingen, Germany: Bruker Optik Gm bH, 2 ed, p.118, 2006.

CORONADO-CÁCERES, LUIS JORGE; RABADÁN-CHÁVEZ, ; MOJICA, ; HERNÁNDEZ-LEDESMA, BLANCA; QUEVEDO-CORONA, LUCÍA; LUGO CERVANTES, EUGENIA. Cocoa (*Theobroma cacao* L.) Seed Proteins' Anti-Obesity Potential through Lipase Inhibition Using In Silico, In Vitro and In Vivo Models. *Foods*, 9(10), 1359, 2020.

COSTA FILHO, P. A. da. Rapid determination of sucrose in chocolate mass using near infrared spectroscopy. *Analytica Chimica Acta*, 631, 206–211, 2009.

CRUZ, J. F. M. et al. Bioactive compounds in different cocoa (*Theobroma cacao*, L) cultivars during fermentation. *Food Science and Technology*. Campinas, v. 35, n. 2, p. 279-284, 2015.

DOMÍNGUEZ-PÉREZ, LEYDY A.; BELTRÁN-BARRIENTOS, LILIA M.; GONZÁLEZ-CÓRDOVA, AARÓN F.; HERNÁNDEZ-MENDOZA, ADRIÁN; VALLEJO-CORDOBA,

BELINDA. Artisanal cocoa bean fermentation: From cocoa bean proteins to bioactive peptides with potential health benefits. *Journal of Functional Foods* 73, 2020.

ELKHORI, S., PARÉ, J. R. J., BÉLANGER, J. M. R. & PÉREZ, E. The micro waves sisted process (MAPTM1): extraction and determination of fat from cocoa powder and cocoa nibs. *J Food Eng* 79:1110–1114, 2007.

ETHERIDGE, R. D., PESTI, G. M., & FOSTER, E. H. A comparison of nitrogen values obtained utilizing the Kjeldahl nitrogen and Dumas combustion methodologies (Leco CNS 2000) on samples typical of an animal nutrition analytical laboratory. *Animal Feed Science and Technology*, 73, 21-28, 1998.

FEARN, T. Assessing calibrations: SEP, RPD, RER and R². *NIR News*, 13(6), 12–14, 2002.

HALIZA, W.; PURWANI, E Y.; FARDIAZ, D.; THENAWIDJAJA, M. Potency of Protein Cocoa Beans as Food Bioactive Precursor to Prevent Hypertension. *IOP Conf. Series: Earth and Environmental Science* 309, (2019) 012049, 2019.

HUE, C., BRAT, P., GUNATA, Z., SAMANIEGO, I., SERVENT, A., MOREL, G., KAPITAN, A., BOULANGER, R., & DAVRIEUX, F. Near Infra-red Characterization of Changes in Flavan-3-ol Derivatives in Cocoa (*Theobroma cacao* L.) as a Function of Fermentation Temperature. *Journal of. Agricultural and Food Chemistry*, 62, 10136–10142, 2014.

MENDES, F. A. T. Relatório anual da previsão de safra no Estado do Pará. MAPA/CEPLAC, 2016.

METROHM NIR SYSTEMS. NIR Spectroscopy - A guide to near-infrared spectroscopic analysis of industrial manufacturing processes. Metrohm AG. P.43, 2014.

RAWEL, HARSHADRAI M.; HUSCHEK, GERD; SAGU, SOREL TCHEWONPI; HOMANN, THOMAS. Cocoa Bean Proteins – Characterization, Changes and Modifications due to Ripening and Post-Harvest Processing. *Nutrients*, 11, 428; doi:10.3390/nu11020428, 2019.

SUNOJ, S., IGATHINATHANE, C., & VISVANATHAN, R. Non destructive determination of cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in Agriculture*, 124, 234–242, 2016, 2016.

TEYE, E. & HUANG, X. Novel Prediction of Total Fat Content in Cocoa Beans by FT-NIR Spectroscopy Based on Effective Spectral Selection Multivariate Regression. *Food Analytical Methods*. 2014.

TEYE, E., HUANG, X., DAI, H. & CHEN, Q. Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 114, 183–189, 2013.

TEYE, E., HUANG, X., SAM-AMOAH, L. K. TAKRAMA, J., BOISON, D., BOTCHWAY, F. & KUMIF. Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. *Food Chemistry*, 176, 403–410, 2015.

VANDEGINSTE, B. G. M., MASSART, D. L., BUYDENS, L. M. C., JONG, S. DE, LEWI, P.J., & SMEYERS-VERBEKE, J. Handbook of chemometrics and qualimetrics: Part A, in: Vandeginste, B. G. M., & Rutan, S. C. (Eds.), *Data Handling in Science and Technology*, vol. 20A, first ed., Elsevier Science B.V., Amsterdam, Netherlands, 1997.

VESELÁ, A., BARROS, A. S., SYNYTSYA, A., DELGADILLO, I., ČOPÍCOVÁ, J., & COIMBRA, M. A. Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Analytica Chimica Acta*, 601, 77–86, 2007.

WILLIAMS, P. C. Implementation of near-infrared technology. In Williams, P. C. & Norris, K. H. (Eds.), *Near Infrared Technology in the Agricultural and Food Industries* (145–169). St. Paul, Minnesota, USA: American Association of Cereal Chemist, 2001.

CAPITULO III

GEOGRAPHICAL DISCRIMINATION OF GROUND AMAZON COCOA BY NEAR INFRARED SPECTROSCOPY: INFLUENCE OF SAMPLE PREPARATION

Manuscrito do artigo publicado no Journal of Food Quality (Qualis A4, Capes).

DOI: <https://doi.org/10.1155/2022/8126810>

Reprodução neste documento isenta de permissão de acordo com instruções de Creative Commons Attribution License, Hindawi Editors (Anexo A).

ABSTRACT

This work presents the application of the NIR technique associated with exploratory analysis of spectral data by main principal components for the discrimination of Amazon cocoa ground seeds. Cocoa samples from different geographic regions of the state of Pará, Brazil (Medicilândia, Tucumã and Tomé-Açu) were evaluated. The samples collected from each region were divided into four groups distinguished by the treatment applied to the samples, which were: fermented (1- with fat and 2- fat-free), and unfermented (3- with moisture and 4- dried). Each set of samples was analyzed separately in order to identify the influence of moisture, fermentation and fat on the geographical differentiation of the three regions. From the results obtained, it can be observed that it was not possible to differentiate the samples of seeds not fermented by geographic origin. However, fermentation was crucial for efficient discrimination, providing more defined clusters for each geographic region. The presence of fat in the seeds was determinant to obtain the best model of geographic discrimination.

Keywords: fermentation; geographical origin; quality; *Theobroma cacao*.

1 INTRODUCTION

Cocoa (*Theobroma cacao* L.) is one of the most important commercial crops in the world, contributing to the maintenance of inputs to various industries, such as food, pharmaceuticals and cosmetics, contributing to the capital turnover of millions of dollars per year in addition to significantly influencing the world economy and income generation in several developing countries [1].

It is the primary raw material for the production of chocolate, which contains functional and bioactive groups, such as polyphenols, responsible for anticancer bioactivity [2], vasodilator [3], in addition to antioxidant bioactivity [4,5] and anti-inflammatory [6].

The quality of cocoa seeds and various food products are correlated with several aspects, including genotype, crop conditions and climate. Knowledge of geographical origin is often recognized and appreciated by the food industry and consumers and is an important factor that can add value to cocoa beans [7].

Unfermented cocoa seeds have an estimated chemical composition of 1% organic acids, 1% caffeine, 1-2% theobromine, 2-3% sucrose, 2-3% cellulose, 4-6% pentosans, 4-6% starch, 5 to 6% polyphenols, 10 to 15% proteins, 30 to 32% lipids and 32 to 39% moisture [1, 8]. These compounds profiles differ even before fermentation, according to the region where they were cultivated.

The fermentation and drying stages are crucial in cocoa processing, especially for the formation of colour, aroma and flavour of chocolate. These steps lead to profound changes in the chemical structure of cocoa seeds, changing the profile of phenolic, sugars, peptides, triglycerides, among others [9-12]. Fat is the main constituent of fermented cocoa seeds composition and its fatty acid profile is changed during fermentation and directly affects the texture, viscosity, melting behavior, aroma and flavour of the food produced from it [13].

According to the report from the Executive Committee of the Cocoa Plantation Plan (Ceplac), the state of Pará has reinforced its position as the largest cocoa producer in Brazil in 2017 [14], and this production is dispersed in different localities, which produce fermented cocoa seeds with different sensory characteristics.

Geographical identification can help in the quality control and traceability management of these seeds. This identification and geographical authenticity can be performed through specific laboratory analyzes, which are mostly time-consuming, tedious and require chemical products that are sometimes harmful to the environment [15-17].

The Infrared (IR) Spectroscopy technique, particularly in the Near Infrared Region (NIR), is a simple, fast, non-chemical waste method and requires minimal sample handling. This method has been used effectively to determine the origin of various food and non-food products [18].

Numerous food matrices have already been discriminated by their geographical origins using the NIR technique, among them, products such as honey [19], alcoholic beverages [20], mushrooms [21], green tea [22], butter [23], wines [24] and others.

The efficiency of the NIR technique has already been tested on cocoa from different regions of the world, and several components of the cocoa have been analyzed and correlated with the NIR method, such as: color, volatile, phenolic compounds, antioxidants, fermentation index, fat [7,17,25-29].

In addition, the multivariate classification methods (LDA, KNN, BPANN, SVM) have also been tested by Teye et al. [7] for the discrimination of cacao originated by different regions. In this study, Teye et al. [7] used 194 samples and evaluated the grain discrimination from seven different regions (Ashanti, Brong-Ahafo, Central, Eastern, Volta, north west and south west), according to the different growing regions in Ghana. PCA was used to test the models, as it provides relevant information about the trend of the samples and formation of clusters [17,25-29].

The purpose of this article was to use the NIR technique to evaluate the influence of sample preparation on the geographical discrimination of cocoa fruits in the cities of Medicilândia, Tucumã and Tomé-Açu, that are the main important producing regions in the state of Pará, Brazil.

2 MATERIAL AND METHODS

2.1 SELECTING COLLECTION REGIONS

The region's selection for the execution of this present work was based on the production volume and the geographical distinction between them. Cities from the three main producing regions of the state of Pará (Medicilândia, Tucumã and Tomé-Açu); and geographically distant from each other were chosen. In the *Transamazon region*, southwest of Pará, the city of Medicilândia (geographic coordinates 03°26'45" S and 52°53'20" W) was chosen, while in the southeastern and northeastern Pará, the cities of Tucumã (geographic coordinates 06°44'52" S and 51°09'39" W and Tomé-Açu (geographic coordinates 02°28'41.3" S and 48°16'50.7" W) were chosen, respectively. The production quality was the second criterion, where the region of Medicilândia stands out in the international market for having superior quality.

2.2 SAMPLE COLLECTION AND PREPARATION

A total of 189 samples were used, 117 samples of raw cocoa fruits (RC) and 72 dried and ground fermented cocoa samples (DF) were used. Samples were collected in three rural properties in the cities of Tomé-açu (RC = 40, DF = 20), Tucumã (RC = 40, DF = 21) and Medicilândia (RC = 37, DF = 31), corresponding to three different geographic regions of the state of Pará, Brazil. 15 samples were used to external prediction of the classification model performance.

Raw cocoa samples were collected following the traditional harvesting practice of each region, choosing random fruits at the edges and at the center of the plantation.

After the harvest, the fruits were broken and stripped; the cotyledons were separated and milled individually in a multifunctional mill (model A11, IKA, Staufen, Germany). The milled material was then sieved in a 600 μm mesh and stored at $-22\text{ }^{\circ}\text{C}$ until the analysis.

Fermented cocoa samples were obtained after the fermentation and drying processes, according to the traditional methods of each producer in the three regions mentioned above. These samples were dried to constant weight in an oven (model 80 series, Lucadema, São José do Rio Preto, SP, Brazil) at $80\text{ }^{\circ}\text{C}$ for moisture standardization. After the drying process, the dried fermented cocoa cotyledons were milled individually and then sieved in a 600 μm mesh and stored at $-22\text{ }^{\circ}\text{C}$ until the analysis.

After the NIR analysis, the 117 samples of raw cocoa (RC) were dried to constant weight in an oven under air circulation at $80\text{ }^{\circ}\text{C}$. After drying these samples, the spectra were again obtained.

2.3 DEFATTING OF COCOA SAMPLES

Dried fermented cocoa samples (56 samples) were exhaustively degreased with petroleum ether (Synth, Diadema, SP, Brazil) at temperature of $55 \pm 1\text{ }^{\circ}\text{C}$ (method 963.15, AOAC), for 24 hours, according to the Association of Official Analytical Chemists – AOAC [30]. After removing the fat, DF samples were kept for 24 hours at $80\text{ }^{\circ}\text{C}$ in an air circulation oven (model 80 series, Lucadema, Brazil) to remove the solvent. After solvent removal, dried fermented degreased cocoa samples (DFD) were stored at $-22\text{ }^{\circ}\text{C}$ until analysis.

Through the procedures described above, it was possible to evaluate cocoa samples in 4 categories: raw cocoa (117), dried unfermented cocoa (117), dried fermented cocoa (72) and dried fermented degreased cocoa (56) (Table 1).

Table 1 – Spectra collected by region and category: raw cocoa (RC), dried unfermented cocoa (DU), dried fermented cocoa (DF) and dried fermented degeared (DFD).

	RC	DF	DU	DFD
Tomé-Açú	40	20	40	19
Tucumã	40	21	40	19
Medicilândia	37	31	37	18
Total	117	72	117	56

2.4 OBTAINING NIR SPECTRA

NIR spectra were obtained using MPA FT-NIR spectrometer (Bruker Optics, Ettlingen, Germany). The spectral data were acquired using absorbance mode in the spectral range from 3500 to 12500 cm^{-1} , with 16 cm^{-1} resolution and average of 32 scans per spectrum. For the samples spectral reading, vials with a volume of 3 ml were used with capacity for about 1 g of milled and sieved cotyledon. The spectral reading was performed at ± 25 °C.

2.5 DISCRIMINATION MODEL DEVELOPMENT BY GEOGRAPHICAL ORIGIN

The spectroscopy software used to construct of discrimination models was OPUS 6.5 (Bruker Optics, Ettlingen, Germany). The spectral data were previously processed in OPUS 6.5, according to the type of sample. For the samples of cocoa RC, DF and DFD, the vector normalization pre-treatment (SNV- standard normal variate) was performed, while for the DU samples, SNV treatment was performed along with the first derivative.

After the application of pre-treatments, the data were processed to develop the discrimination model by geographical origin of each set of samples. The exploratory analysis method, PCA, was applied to multivariate data to make visual inspection of results more evident.

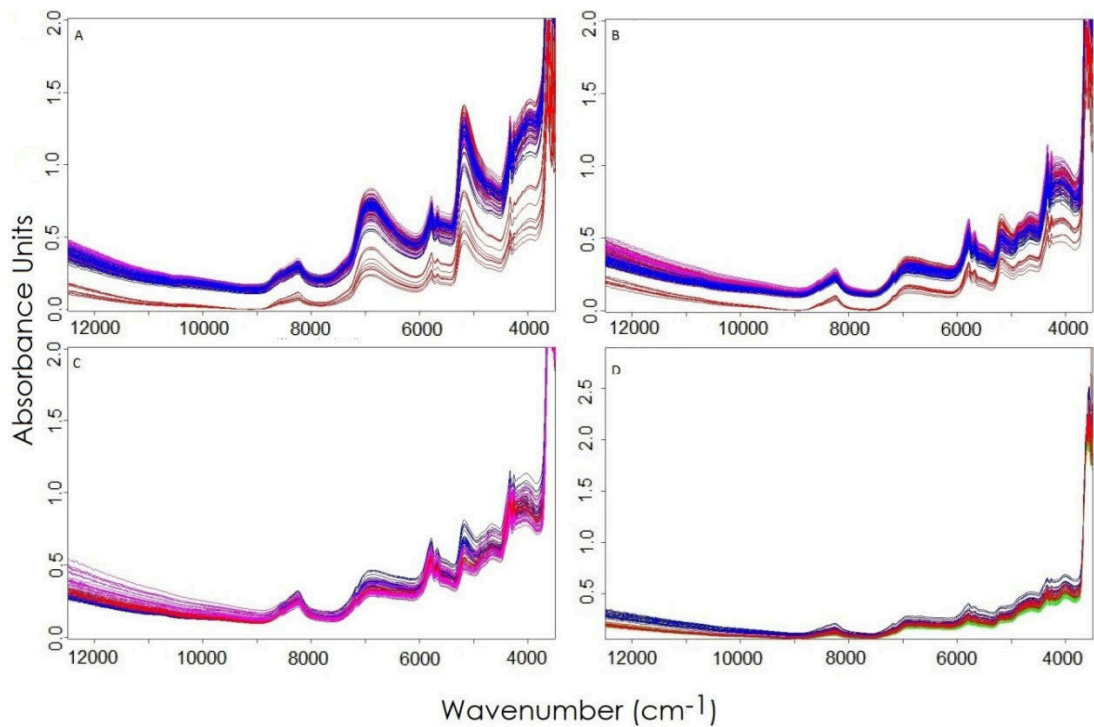
The spectral region that has best differentiated the geographic regions was chosen by the software operator, based on information related to the sample chemical composition, such as water content, fat and protein.

3 RESULTS AND DISCUSSION

3.1 SPECTRAL ANALYSIS

The complete NIR spectra obtained from RC, DU, DF and DFD samples are shown in Figure 1. The prominent peaks can be observed around 9000-8000, 8500-6000, 6000-5000 and 5000-4500 cm^{-1} within the NIR range. In the RC spectrum (Figure 1a), two intense absorption peaks are observed, in 5312 and 7202 cm^{-1} , which are related to the water content in the samples, according to [7], corresponding to the region of the first overtone of the O-H stretch and O-H deformation. The water absorption bands should be removed to reduce interference with the chemical structures corresponding to the groups CH and CO and combinations of amine groups [31]. With that, we could see in figure 1b and 1c, that after removing the water, the absorption peaks of other components in the samples became more evident.

Figure 1 – Absorption spectrum of a) RC, b) DU, c) DF, d) DFD, in the range of 3600 to 12500 cm^{-1} .



The regions corresponding to 9000-7500, 6000-6030, 6030-4000, 4950-4500 and 4500-3850 cm^{-1} were used for all the calibration model development.

None of the spectra presented in Figure 1 was possible to verify the differences between fruits from different analysed regions, even in the fermented samples. These differences were only noticed after mathematical treatments through the application of statistical analysis techniques for model development.

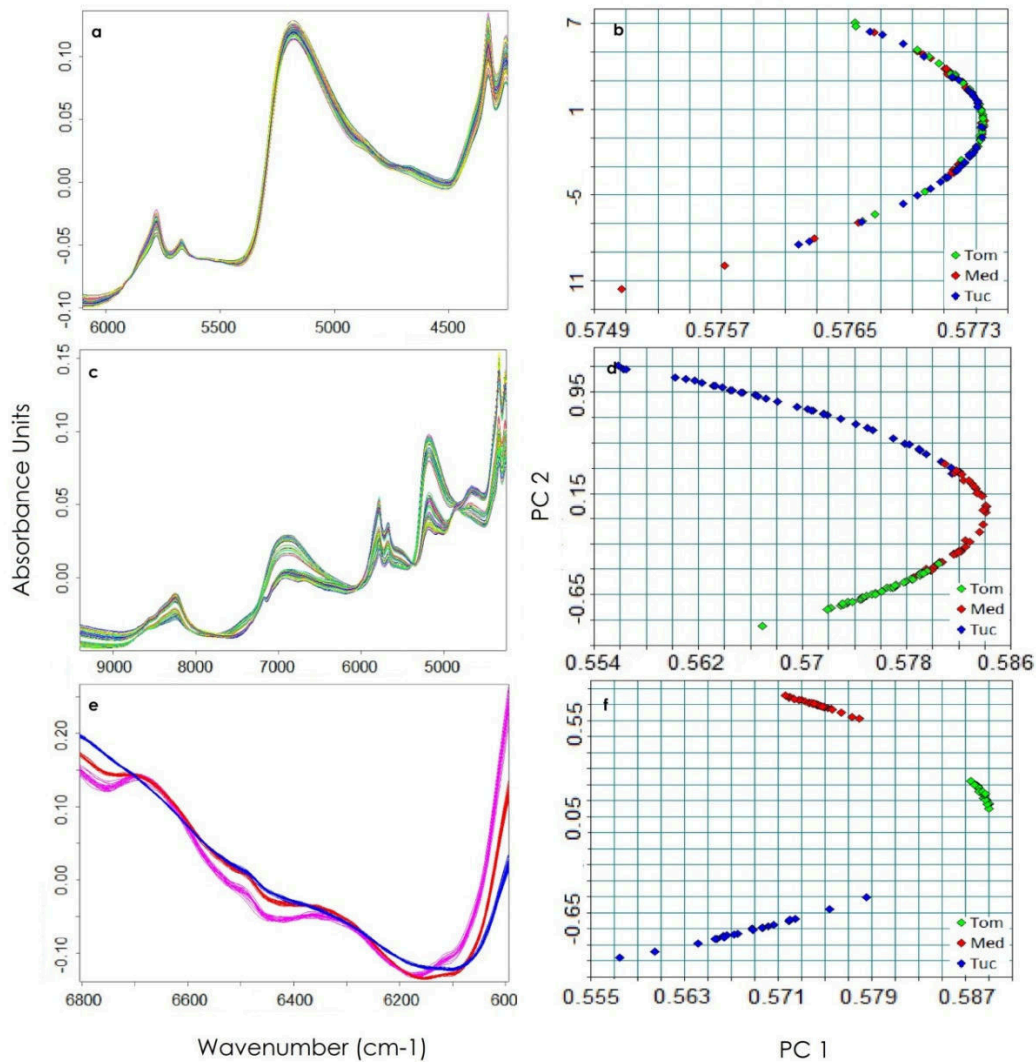
3.2. DISCRIMINATION MODELS FOR RAW COCOA

For the RC cocoa samples, it was impossible that the spectral band (6200-4000 cm^{-1}) could discriminate groups of the fruits from different regions, as observed in Figure 2a.

The graph of the scores presented in Figure 2b, referring to the spectral range between 4000 and 6200 cm^{-1} , shows the lack of differentiation between the three regions' scores. Bands referring to the phenol group, which can be observed in the regions corresponding to 3562-3322 cm^{-1} , aromatic ring related bands, referring to the region of 2925-2854 cm^{-1} (attributed to the CH stretch of the aromatic ring) and 1645 -1544 cm^{-1} (attributed to C in the aromatic ring) [7], were overlapped. This effect is directly related to the amount of water and fats in the RC cocoa samples.

Therefore, high absorption of radiation NIR by water present in samples is a factor that contributes negatively in analyses, since the water spectra overlap with the other spectra of interest, making it challenging to construct an analytical and statistical model of geographic discrimination of the in nature Amazonian cocoa.

Figure 2 – Absorption spectrum of the separation regions of a) RC, c) DU, e) DF and their respective scores in the PCs graphs (b, d and f).



3.3 DISCRIMINATION MODELS FOR DRIED UNFERMENTED COCOA

After the removal of water by drying, the DU samples have shown a trend towards the formation of groups related to discrimination by geographic region. The best spectral range for the DU discrimination was between 4300 and 9300 cm⁻¹ observed in Figure 2c.

These spectra, located between 4300-4597, 4902-5199, 5805-6102 and 6406-9300 cm^{-1} in the whole spectrum for DU are associated with the carbonyl group spectral regions; corresponding to stretching combination (CH_2 and CH), first harmonic of CH present in the aromatic ring, combination of C-C and C-N and second harmonic of N-H [15]. These molecular vibrations are caused by functional groups corresponding to polyphenols, alkaloids, vicillin class globulins, proteins, amines, acids, polysaccharides and other aromatic compounds [7]. These diverse functional groups are the digital identity of each sample from different Amazon regions.

The differentiation between the samples from Tomé-Açu and Tucumã was evident, as observed in Figure 2d. However, the scores related to the set of Medicilândia samples get mixed with the scores of the samples from other two regions. A number of factors can explain the formation of characteristic groups of cocoa, such as chemical composition [32, 33], degree of fermentation [34], genotype [35], among other factors, which may be associated with the region of origin of the seeds [36].

According to [37], cocoa, like other food products, has its characteristics influenced by locality. This influence is directly related to the main components of alkaloid polyphenols, proteins, amines, polysaccharides, acids and diverse aromatic compounds, as previously mentioned [7].

3.4 DISCRIMINATION MODELS FOR FERMENTED COCOA

The fermentation of cocoa beans is essential, due to the diverse microbial processes developed as consequence of changes in temperature, pH and oxygen availability, promoting significant biochemical changes in the type and concentration of flavour precursors in cocoa beans [37]. Is a crucial step in the formation of quality sensory attributes. Chemical and biochemical complex changes and interactions occur in the cocoa beans during drying, storage and fermentation, contributing to the complexity and identity of the Amazon cocoa [38].

The fermentation produces flavour and aroma precursors, such as free amino acids and peptides from enzymatic degradation of proteins and sugars that reduce the enzymatic

degradation of sucrose, as well as a significant increase in volatile compounds such as organic acids, esters, alcohols and aldehydes, after and during the fermentation of cocoa beans. The stoichiometry of their progenitors influences the concentration of these precursors; therefore, this is a response to the nutritional quality of the cultivated soil, from which the nutrients necessary for the synthesis of their parents were collected, revealing the good indicators of the production region [39].

Fermentation was a crucial factor for forming of the characteristics of each region, because unlike the dried unfermented samples (DU), the dried fermented samples (DF), whose spectra are shown in Figure 2e, allowed good discrimination between groups from different regions. The best spectral discrimination range observed was 6000 to 6800 cm^{-1} .

These results allow one to verify that there is a strong influence of the fermentation on the biochemical characteristics of the samples, in order to allow a clear distinction of them, as shown in Figure 2f.

Caligiani *et al.* [10], reported that the progressive fermentation of cocoa seeds causes the hydrolysis of peptides in amino acids, reducing the violet colour of the cotyledons at the end of the fermentation. The phenolic compounds are reduced upon drying; this reduction is mainly attributed to the enzymatic action, followed by non-enzymatic reactions due to quinone polymerization, with pH increase and high capitation of O_2 during sun drying [9].

The study of Sirbu *et al.* [12], estimated the increase of lipid content during fermentation from 3–6% when compared to the value before fermentation. In addition, the authors observed a change in the triacylglycerol profile caused by fermentation. Some polar triacylglycerols, such as derivatives of hydroxyl allyl fatty acid (now *cacaic acid*) presented in unfermented cocoa seeds, were not found in dried fermented cocoa seeds, thus allowing chemical differentiation between unfermented and fermented cocoa seeds.

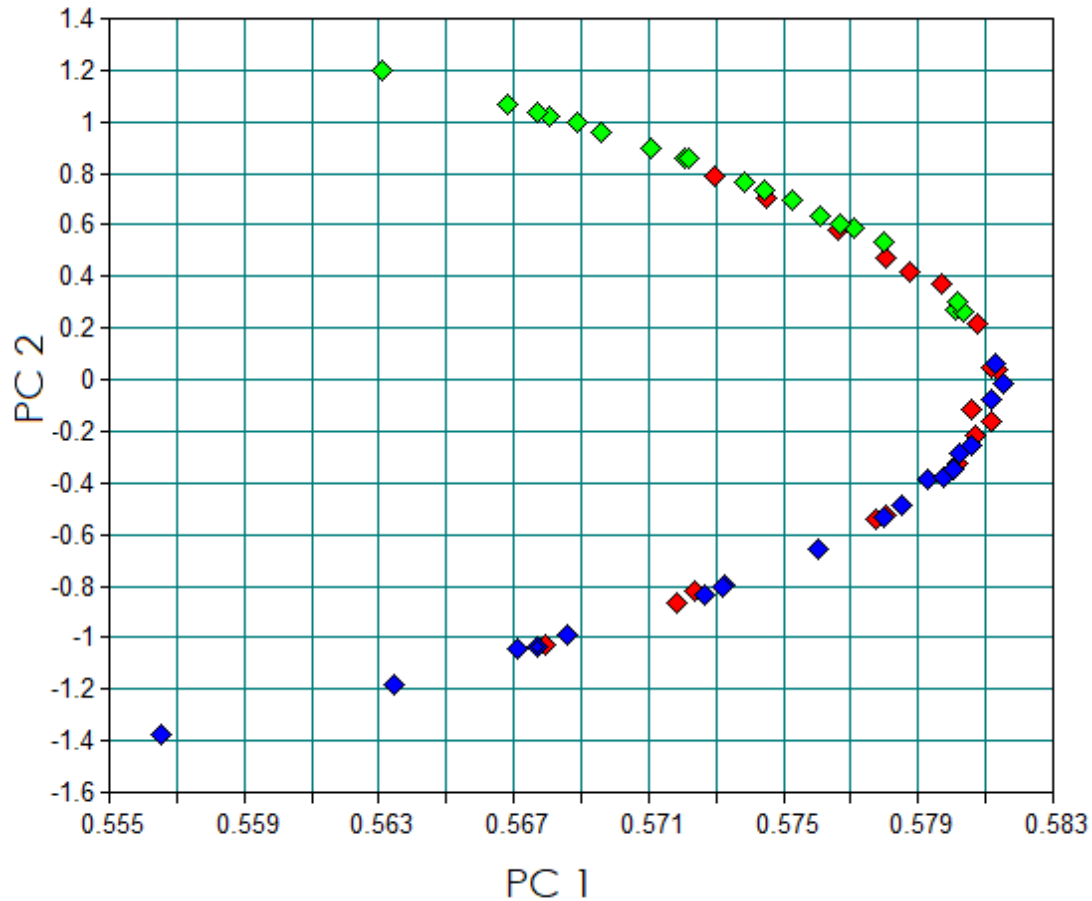
According to the literature, we can see that fermentation causes changes in the profile and concentration of various components of cocoa. Such components or concentrations of components can be characteristic for each region, and because of this it is possible to carry out geographic identification. The NIR technique associated with chemometrics can previously identify the similarity and difference in various types of samples and classify them, making it possible to visually identify groups that have similar or distinct characteristics.

In this work, we used 15 external samples (not used in the construction of the model) to be able to test the efficiency of the constructed discrimination model (5 samples from Tomé-Açú, 5 from Tucumã and 5 from Medicilândia). 100% of the samples were correctly classified according to their regions of origin.

3.5 DISCRIMINATION MODELS FOR DEGREASED COCOA

In order to evaluate the importance of the lipid content of the samples in geographic differentiation, it was made the degreasing of the DF cocoa samples and the spectral evaluation was carried out again, using the same spectral treatment applied to discriminate the non-degreased fermented samples. Scores of the degreased samples are shown in Figure 3.

Figure 3 – Scores of DFD cocoa samples.



As observed in Figure. 3, even though the mathematical model treatment was used for DF samples (dried fermented), it was not observed the separation between the three groups, however a partial separation between two regions (Tomé-Açu and Tucumã) was observed, which were the same visualized in the DU samples. This shows that the fat removal made it difficult to classify the samples in three different geographic regions using the DFD sample discrimination model, validating partially the results of [40], who have shown a strong correlation of fatty acid composition in the geographic differentiation. In our research, we observed that it was possible to discriminate fermented samples more easily with their original butter constitution.

The lipids and proteins storage cells in cocoa seeds have complex cytology, composed of a compacted cytoplasm with multiple vacuoles, where proteins and lipids can be found, as well as other components such as starch granules, which are important for the definition of specific flavour and aroma characteristics. These biochemical and cytological characteristics of the cells change in percentage and morphological characteristics from region to region, affecting by several physical, chemical and climatic factors [39].

Since there was an indication of partial separation of DFD samples, it was developed a model from two separation steps. In the first separation step, the spectral region of 5950 to 6835 cm^{-1} was used (Figure 4a), in which the regions of Tomé-Açu and Tucumã were separated (Figure 4b). In the second step of the model, the sample data for the two groups that did not show clear separation in the first step (Medicilândia and Tucumã), the spectral region of 7100 and 9300 cm^{-1} was used (Figure 4d). Two spectral regions were used to build the complete separation model of the DFD samples, as shown in Figure 5.

Figure 4 – Absorption spectrum of the DFD separation regions (a and c) and their respective scores (b and d), for the two selected optimal separation regions.

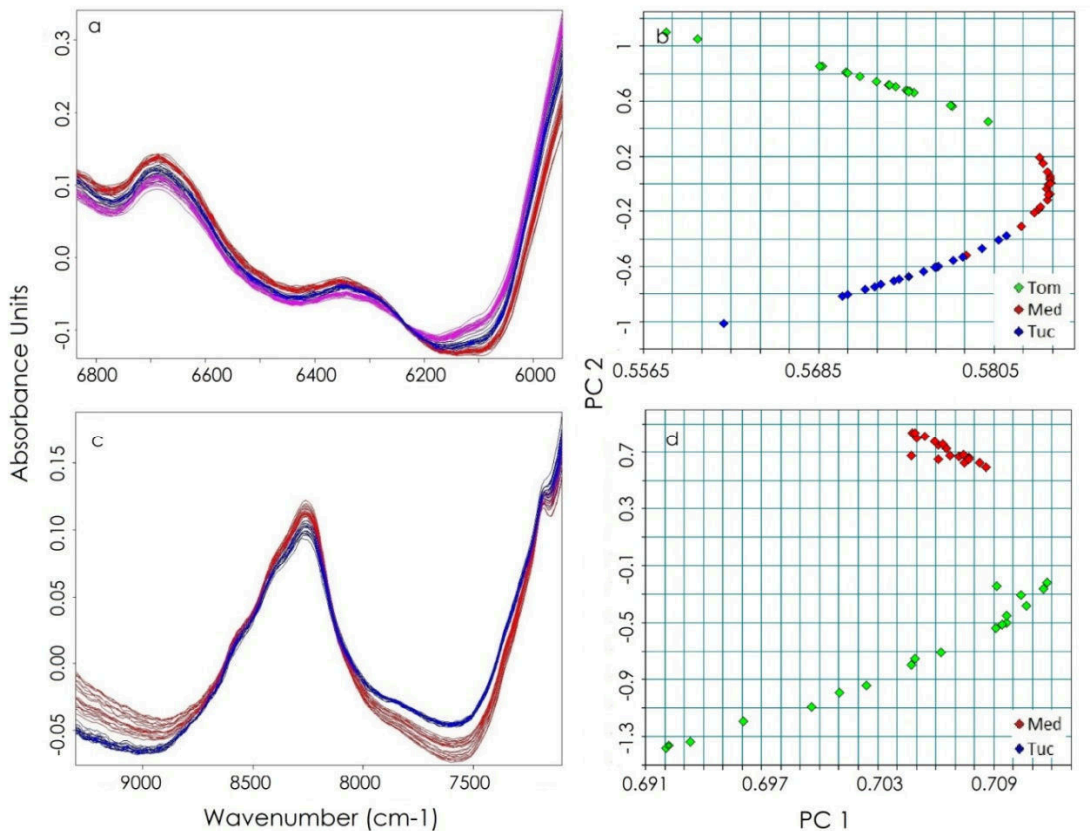
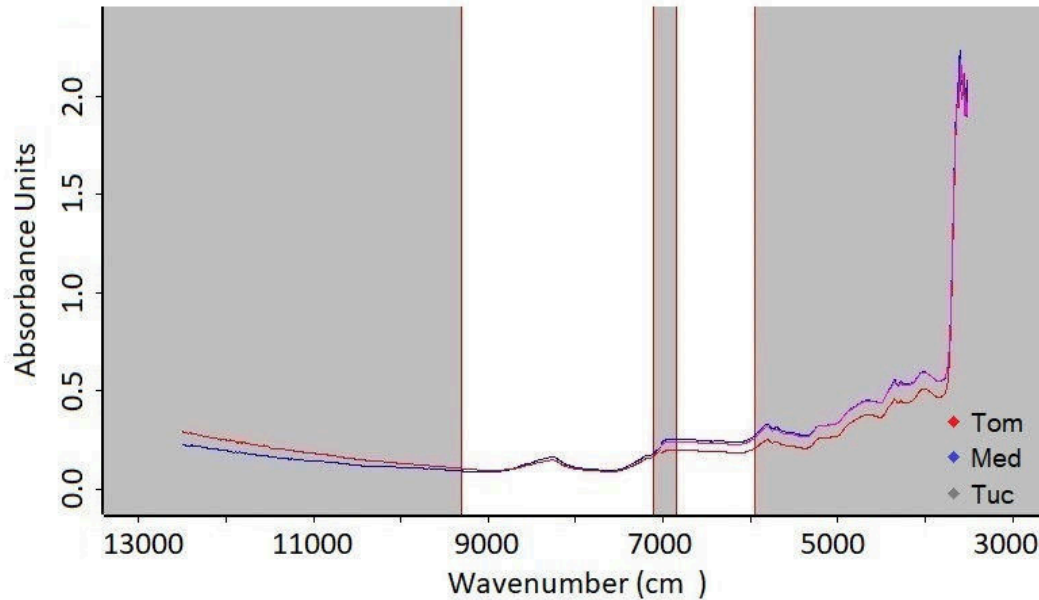


Figure 5 – Absorption spectrum highlighting the two separation regions of DFD samples (1st and 2nd region).



The DFD cocoa seed sample separation model was performed in two stages. For the first stage it was used the spectral data of the cocoa from the three regions (Medicilândia, Tucumã and Tomé-Açu) (Fig. 4 b), for the second stage it was used the spectral data of the cocoa from Medicilândia and Tucumã, whose separation it was less evident in the first stage (Fig. 4c). Thus, comparing to the previous separation model (of DF seeds), we could observe that the lipid fraction had considerable influence, however it was not essential for geographic identification.

According to Sirbu et al. [12], after fermentation, cocoa presents an different profile of triacylglycerol and starch granules, which influence the spectroscopic behavior. And after the fat removal, the other components that changed during fermentation remain there, making geographic differentiation possible [12].

The application of the NIR technique is quite simple, however the path to acquire spectral information and analytical data is time-consuming and costly. Once the database of spectral and analytical information is obtained, and the mathematical model can be constructed through the use of computational algorithms that plan and / or optimize such experimental procedures. From this, this elaborated model can be used to analyse external samples only with spectral information. The spectral data will be correlated with the data previously obtained and

the result is estimated. In this way, the analysis becomes simple, fast and economical. It is enough that, after the construction of the prediction model, the spectra of the material in question are obtained and the model is applied.

4 CONCLUSIONS

The results showed that it was possible to evaluate the influence of sample preparation on the geographic discrimination of cocoa from different geographic regions, in the Pará state, using the NIR technique.

The importance of the fermentation process for geographic identification was observed, and it was noticed provided specific characteristics for each region evaluated. The fermented cocoa lipids were crucial in the formation of groups in the fermented samples, also indicating that fermentation is a crucial step to develop specific characteristics for each collection region. Thus, it is assumed that the biotas of each region, should affect the fatty acids contained in cocoa butter differently, since similar results that were not found in defatted samples.

Effect of the climate and production process, although not directly explicit, cannot be ruled out, considering that they can significantly influence the type of biota and the fermentation process, as well as other possible variables.

ACKNOWLEDGEMENTS

The authors would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Finance Code 01) for the doctoral scholarship of the first author, Programa de Pós-graduação em Ciência e Tecnologia de Alimentos of the Federal University of Pará (PPGCTA / UFPA) for the great support offered.

FUNDING SOURCES

This study was financed by the Instituto Tecnológico Vale (ITV) (Cacau P2) and supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES (Process Number: 1537413).

ETHICAL GUIDELINES

Ethics approval was not required for this research.

STATEMENTS

Research data are not shared. The authors have no conflict of interest.

REFERENCES

- [1] LIMA, LÍ. J. R.; ALMEIDA, M. H.; ROB NOUT, M. J.; ZWIETERING, M. H. *Theobroma cacao* L., “The food of the gods”: Quality determinants of commercial cocoa beans, with particular reference to the impact of fermentation. *Critical Reviews in Food Science and Nutrition*, 51, 731–761, 2011.
- [2] OLEAGA, C.; GARCÍA, M.; SOLÉ, A.; CIUDAD, C.J.; IZQUIERDO-PULIDO, M.; NOÉ, V. CYP1A1 is overexpressed upon incubation of breast cancer cells with a polyphenolic cocoa extract. *European Journal of Nutrition*, 51, 465–476, 2012.
- [3] KHAWAJA, O.; GAZIANO, J. M.; DJOUSSÉ, L. Chocolate and coronary heart disease: A systematic review. *Current Atherosclerosis Reports*, 13, 447–452, 2011.

- [4] HO, V. T. T.; ZHAO, J.; FLEET, G.. The effect of lactic acid bacteria on cocoa bean fermentation. *International Journal of Food Microbiology*, 205, 54–67, 2015.
- [5] KRÄHMER, A.; ENGEL, A.; KADOW, D.; ALI, N.; UMAHARAN, P.; KROH, L. W.; SCHULZ, H. Fast and neat - Determination of biochemical quality parameters in cocoa using near infrared spectroscopy. *Food Chemistry*, 181, 152–159, 2015.
- [6] OYELEKE, S. A.; AJAYI, A. M.; UMUKORO, S.; ADERIBIGBE, A. O.; ADEMOWO, O. G. Anti-inflammatory activity of *Theobroma cacao* L. stem bark ethanol extract and its fractions in experimental models. *Journal of Ethnopharmacology*, 222, 239–248, 2018.
- [7] TEYE, E.; HUANG, X.; DAI, H.; CHEN, Q. Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 114, 183–189, 2013.
- [8] KONGOR, J. E.; HINNEH, M.; WALLE, D.; VAN DE, AFOAKWA, E.O.; BOECKX, P.; &DEWETTINCK, K. Factors influencing quality variation in cocoa (*Theobroma cacao*) bean flavour profile - A review. *Food Research International*, 82, 44–52, 2016.
- [9] CRUZ, J. F. M.; LEITE, P. B.; SOARES, S. E.; BISPO, E. S. Bioactive compounds in different cocoa (*Theobroma cacao*, L) cultivars during fermentation. *Food Science and Technology*, 35, 279–284, 2015.
- [10] CALIGIANI, A.; MARSEGLIA, A.; PRANDI, B.; PALLA, G.; SFORZA, S. Influence of fermentation level and geographical origin on cocoa bean oligopeptide pattern. *Food Chemistry*, 211, 431–439, 2016.
- [11] MEGÍAS-PÉREZ, R.; GRIMBS, S.; D'SOUZA, R. N.; BERNAERT, H.; KUHNERT, N. Profiling, quantification and classification of cocoa beans based on chemometric analysis of carbohydrates using hydrophilic interaction liquid chromatography coupled to mass spectrometry. *Food Chemistry*, 258, 284–294, 2018.
- [12] SIRBU, D.; GRIMBS, A.; CORNO, M.; ULLRICH, M.S.; KUHNERT, N. Variation of triacylglycerol profiles in unfermented and dried fermented cocoa beans of different origins. *Food Research International*, 111, 361–370, 2018.

- [13] AFOAKWA, E., (Eds). Chocolate science and technology. Blackwell, 296 p. Oxford: Wiley, 2010.
- [14] IBGE. Brazilian Institute of Geography and Statistics, Indicators IBGE. Agricultural Production Statistics. Available at: <https://ww2.ibge.gov.br/home/estatistica/indicadores/agropecuaria/>, 2017.
- [15] VESELÁ, A.; BARROS, A. S.; SYNITSYA, A.; DELGADILLO, I.; ČOPÍKOVÁ, J.; COIMBRA, M. A. Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Analytica Chimica Acta*, 601, 77–86, 2007.
- [16] DA COSTA FILHO, P.A. Rapid determination of sucrose in chocolate mass using near infrared spectroscopy. *Analytica Chimica Acta*, 631, 206–211, 2009.
- [17] TEYE, E. & HUANG, X. Novel Prediction of Total Fat Content in Cocoa Beans by FT-NIR Spectroscopy Based on Effective Spectral Selection Multivariate Regression. *Food Analytical Methods*, 8, 945–953, 2015.
- [18] BÜNING-PFAUE, H. NIR Spectroscopy, a Revolution in Analysis. *New Food*, 7, 41-47, 2004.
- [19] WOODCOCK, T.; DOWNEY, G.; KELLY, J.D.; O'DONNELL, C. Geographical classification of honey samples by near-infrared spectroscopy: A feasibility study. *Journal of Agricultural and Food Chemistry*, 55, 9128–9134, 2007.
- [20] URÍČKOVÁ, V.; SÁDECKÁ, J. Determination of geographical origin of alcoholic beverages using ultraviolet, visible and infrared spectroscopy: A review. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 148, 131–137, 2015.
- [21] CHEN, Y.; XIE, M. Y.; YAN, Y.; ZHU, S.; NIE, S. P.; LI, C.; WANG, Y. X.; GONG, X. F. Discrimination of *Ganoderma lucidum* according to geographical origin with near infrared diffuse reflectance spectroscopy and pattern recognition techniques. *Analytica Chimica Acta*, 618, 121–130, 2008.
- [22] CHEN, Q.; ZHAO, J.; LIN, H. Study on discrimination of Roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern

recognition. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 72, 845–850. 2009.

[23] BASSBASI, M.; LUCA, M. DE; IOELE, G., OUSSAMA, A.; RAGNO, G. Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data. *Journal of Food Composition and Analysis*, 33, 210–215, 2014.

[24] COZZOLINO, D.; CYNKAR, W. U.; SHAH, N.; SMITH, P. A. Can spectroscopy geographically classify Sauvignon Blanc wines from Australia and New Zealand? *Food Chemistry*, 126, 673–678, 2011.

[25] DAVIES, A. M. C., FRANKLIN, J. G., GRANT, A., GRIFFITHS, N. M., SHEPHERD, R., FENWICK, G. R. Prediction of chocolate quality from near-infrared spectroscopic measurements of the raw cocoa beans. *Vibrational Spectroscopy*, 2 (1991) 161-172.

[26] ACULEY, P.C.; SNITKJAER, P.; OWUSU, M.; BASSOMPIERE, M.; TAKRAMA, J.; NØRGAARD, L.; NIELSEN, D. S. Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics. *Journal of Food Science*, 75, S300–S307, 2010.

[27] HUE, C.; GUNATA, Z.; BERGOUNHOU, A.; ASSEMAT, S.; BOULANGER, R.; SAUVAGE, F. X.; DAVRIEUX, F. Near infrared spectroscopy as a new tool to determine cocoa fermentation levels through ammonia nitrogen quantification. *Food Chemistry*, 148 (2014) 240–245.

[28] WHITACRE, E.; OLIVER, J.; VAN DEN BROEK, R.; VAN ENGELEN, P.; KREMERS, B.; VAN DER HORST, B.; STEWART, M. JANSENBEUVINK, A. Predictive Analysis of Cocoa Procyanidins Using Near-Infrared Spectroscopy Techniques. *Journal of Food Science*, 68 (9), 2618-2622, 2003.

[29] CAPORASO, N.; WHITWORTH, M. B.; FOWLER, M. S.; FISK, I. D. Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chemistry*, 258, 343–351, 2018.

- [30] HORWITZ, W. & LATIMER, G.W. Official Methods of Analysis of AOAC International. 18th Edition,. Association of Official Analytical Chemistry International, Maryland, 2006.
- [31] SUNOJ, S., IGATHINATHANE, C. & VISVANATHAN, R. Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in Agriculture*, 124, 234–242, 2016.
- [32] BARBIN, D. F., MACIEL, L.F., BAZONI, C.H.V., RIBEIRO, M.S., CARVALHO, R.D.S., BISPO, E.S., MIRANDA, M.P.S., HIROOKA, E.Y. Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses. *J Food Sci Technol*, 55(7) (2018) 2457–2466.
- [33] TEYE, E., HUANG, X., SAM-AMOAH, L.K., TAKRAMA, J., BOISON, D., BOTCHWAY, F., KUMI, F. Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. *Food Chemistry*, 176, 403–410, 2015.
- [34] ANYIDOHO, E.K., TEYE, E., AGBEMAFLE, R. Nondestructive authentication of regional and geographical origin of cocoa beans by using handheld NIR spectroscopy and multivariate algorithm. *Anal. Methods*, 12, 4150-4158, 2020.
- [35] CRUZ-TIRADO, J.P.; PIERNA, J.A.F., Rogez, B., BARBIN, D., BAETEN, Vincent. Authentication of cocoa (*Theobroma cacao*) bean hybrids by NIR-hyperspectral imaging and chemometrics. *Journal Pre-proof*, 118, 2020.
- [36] OLIVEIRA, A.C., MARIEN, A., HULIN, J., MUHOVSKI, Y., BAETEN, V., JANSSEN, E., BERBEN, G., ROGEZ, H., DEBODE, F. Development of real-time PCR methods for cocoa authentication in processed cocoa-derived products. *Food Control* 131, 108414, 2022.
- [37] KADOW, D., BOHLMANN, J., PHILLIPS, W. & LIEBEREI, R. Identification of main fine or flavour components in two genotypes of the cocoa tree (*Theobroma cacao* L.). *Journal of Applied Botany and Food Quality*, 86, 90–98, 2013.
- [38] HO, V.T.T., ZHAO, J. & FLEET, G. Yeasts are essential for cocoa bean fermentation. *International Journal of Food Microbiology*, 174, 72–87, 2014.

[39] AFOAKWA, E.O., KONGOR, J.E., TAKRAMA, J. & BUDU, A.S. Changes in nib acidification and biochemical composition during fermentation of pulp pre-conditioned cocoa (*Theobroma cacao*) beans. *International Food Research Journal*, 20, 1843–1853, 2013.

[40] TORRES-MORENO, M., TORRESCASANA, E., SALAS-SALVADÓ, J. & BLANCH, C. Nutritional composition and fatty acids profile in cocoa beans and chocolates with different geographical origin and processing conditions. *Food Chemistry*, 166, 125–132, 2015.

CONCLUSÃO GERAL

Através das pesquisas realizadas e apresentadas, pode-se comprovar que a técnica NIR é uma ferramenta eficiente, eficaz e sustentável. O que contribui para o aumento da eficiência dos processos industriais sem causar impactos negativos no meio ambiente. O cacau da Amazônia foi avaliado utilizando a técnica NIR. Os resultados obtidos para as predições dos componentes principais como, umidade, nitrogênio total e lipídeos indicam a adequação e robustez dos modelos desenvolvidos e mostram que estes modelos se apresentam como ferramentas analíticas úteis para avaliação de parâmetros de qualidade em amostras de cacau não fermentado. Além disso, também foi possível avaliar a influência da preparação da amostra na discriminação geográfica destas amostras de cacau, usando esta técnica. Observou-se que o processo fermentativo foi crucial para a identificação geográfica, além de fornecer características específicas para cada região avaliada.

ANEXO




ANEXO A – ARTIGO PUBLICADO

Hindawi
Journal of Food Quality
Volume 2022, Article ID 8126810, 9 pages
<https://doi.org/10.1155/2022/8126810>



Research Article

Geographical Discrimination of Ground Amazon Cocoa by Near-Infrared Spectroscopy: Influence of Sample Preparation

Fabielle Negrão Ferreira ¹, Gilson Celso Albuquerque Chagas-Junior,¹
Mozaniel Santana de Oliveira ², Jhonatas Rodrigues Barbosa,³
Marcos Enê Chaves Oliveira,⁴ and Alessandra Santos Lopes ¹

¹Laboratory of Biotechnological Process (LABIOTEC), Graduate Program in Food Science and Technology (PPGCTA), Federal University of Para (UFPA), Belém 66095-100, Pará, Brazil

²Laboratory Adolpho Ducke, Research Field of the Emilio Goeldi Museum Paraense, Perimetral Avenue, 1901, Terra Firme, Belém, Pará, Brazil

³Laboratory for the Extraction of Plant Products (LABEX), Graduate Program in Food Science and Technology (PPGCTA), Federal University of Para (UFPA), Belém 66095-100, Pará, Brazil

⁴Embrapa Eastern Amazon, Belém 66095-100, Pará, Brazil

Correspondence should be addressed to Fabielle Negrão Ferreira; fabielle.negrao@gmail.com and Alessandra Santos Lopes; alessalopes@ufpa.br

Received 2 March 2022; Revised 22 May 2022; Accepted 24 June 2022; Published 31 July 2022

Academic Editor: Daniel Cozzolino

Copyright © 2022 Fabielle Negrão Ferreira et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work presents the application of the NIR technique associated with exploratory analysis of spectral data by main principal components for the discrimination of Amazon cocoa ground seeds. Cocoa samples from different geographic regions of the state of Pará, Brazil (Medicilândia, Tucumã, and Tomé-Açu), were evaluated. The samples collected from each region were divided into four groups distinguished by the treatment applied to the samples, which were fermented (1-with fat and 2-fat-free) and unfermented (3-with moisture and 4-dried). Each set of samples was analyzed separately to identify the influence of moisture, fermentation, and fat on the geographical differentiation of the three regions. From the results obtained, it can be observed that it was not possible to differentiate the samples of seeds not fermented by geographic origin. However, fermentation was crucial for efficient discrimination, providing more defined clusters for each geographic region. The presence of fat in the seeds was a determinant to obtain the best model of geographic discrimination.

1. Introduction

Cocoa (*Theobroma cacao* L.) is one of the most important commercial crops in the world, contributing to the maintenance of inputs to various industries, such as food, pharmaceuticals, and cosmetics, contributing to the capital turnover of millions of dollars per year in addition to significantly influencing the world economy and income generation in several developing countries [1].

It is the primary raw material for the production of chocolate, which contains functional and bioactive groups, such as polyphenols, responsible for anticancer bioactivity

[2], vasodilator [3], in addition to antioxidant bioactivity [4, 5], and anti-inflammatory [6].

The quality of cocoa seeds and various food products is correlated with several aspects, including genotype, crop conditions, and climate. Knowledge of geographical origin is often recognized and appreciated by the food industry and consumers and is an important factor that can add value to cocoa beans [7].

Unfermented cocoa seeds have an estimated chemical composition of 1% organic acids, 1% caffeine, 1-2% theobromine, 2-3% sucrose, 2-3% cellulose, 4-6% pentosans, 4-6% starch, 5 to 6% polyphenols, 10 to 15% proteins, 30 to

32% lipids, and 32 to 39% moisture [1, 8]. These compounds' profiles differ even before fermentation, according to the region where they were cultivated.

The fermentation and drying stages are crucial in cocoa processing, especially for the formation of color, aroma, and flavor of chocolate. These steps lead to profound changes in the chemical structure of cocoa seeds, changing the profile of phenolic, sugars, peptides, and triglycerides, among others [9–12]. Fat is the main constituent of fermented cocoa seeds composition and its fatty acid profile is changed during fermentation and directly affects the texture, viscosity, melting behavior, aroma, and flavor of the food produced from it [13].

According to the report from the Executive Committee of the Cocoa Plantation Plan (Ceplac), the state of Pará reinforced its position as the largest cocoa producer in Brazil in 2017 [14], and this production is dispersed in different localities, which produce fermented cocoa seeds with different sensory characteristics.

Geographical identification can help in the quality control and traceability management of these seeds. This identification and geographical authenticity can be performed through specific laboratory analyses, which are most time-consuming and tedious and require chemical products that are sometimes harmful to the environment [15–17].

The infrared (IR) spectroscopy technique, particularly in the Near-Infrared Region (NIR), is a simple, fast, nonchemical waste method and requires minimal sample handling. This method has been used effectively to determine the origin of various food and nonfood products [18].

Numerous food matrices have already been discriminated by their geographical origins using the NIR technique, among them, products such as honey [19], alcoholic beverages [20], mushrooms [21], green tea [22], butter [23], and wines [24].

The efficiency of the NIR technique has already been tested on cocoa from different regions of the world, and several components of the cocoa have been analyzed and correlated with the NIR method, such as color, volatile, phenolic compounds, antioxidants, fermentation index, and fat [7, 17, 25–29].

In addition, the multivariate classification methods (LDA, KNN, BPANN, and SVM) have also been tested by Teye et al. [7] for the discrimination of cacao originated in different regions. In this study, Teye et al. [7] used 194 samples and evaluated the grain discrimination from seven different regions (Ashanti, Brong-Ahafo, Central, Eastern, Volta, northwest, and southwest), according to the different growing regions in Ghana. PCA was used to test the models, as it provides relevant information about the trend of the samples and the formation of clusters [17, 25–29].

The purpose of this article was to use the NIR technique to evaluate the influence of sample preparation on the geographical discrimination of cocoa fruits in the cities of Medicilândia, Tucumã, and Tomé-Açu, which are the main important producing regions in the state of Pará, Brazil.

2. Material and Methods

2.1. Selecting Collection Regions. The region's selection for the execution of this present work was based on the production volume and the geographical distinction between them. Cities from the three main producing regions of the state of Pará (Medicilândia, Tucumã, and Tomé-Açu) and geographically distant from each other were chosen. In the *Transamazon region*, southwest of Pará, the city of Medicilândia (geographic coordinates 03°26'45" S and 52°53'20" W) was chosen, while in the southeastern and northeastern Pará, the cities of Tucumã (geographic coordinates 06°44'52" S and 51°09'39" W) and Tomé-Açu (geographic coordinates 02°28'41.3" S and 48°16'50.7" W) were chosen, respectively. The production quality was the second criterion, where the region of Medicilândia stands out in the international market for having superior quality.

2.2. Sample Collection and Preparation. A total of 189 samples were used, 117 samples of raw cocoa fruits (RC) and 72 dried and ground fermented cocoa samples (DF). Samples were collected in three rural properties in the cities of Tomé-açu (RC = 40, DF = 20), Tucumã (RC = 40, DF = 21), and Medicilândia (RC = 37, DF = 31), corresponding to three different geographic regions of the state of Pará, Brazil. 15 samples were used for external prediction of the classification model performance.

Raw cocoa samples were collected following the traditional harvesting practice of each region, choosing random fruits at the edges and the center of the plantation.

After the harvest, the fruits were broken and stripped; the cotyledons were separated and milled individually in a multifunctional mill (model A11, IKA, Staufen, Germany). The milled material was then sieved in a 600 μm mesh and stored at -22°C until the analysis.

Fermented cocoa samples were obtained after the fermentation and drying processes, according to the traditional methods of each producer in the three regions mentioned above. These samples were dried to constant weight in an oven (model 80 series, Lucadema, São José do Rio Preto, SP, Brazil) at 80°C for moisture standardization. After the drying process, the dried fermented cocoa cotyledons were milled individually and then sieved in a 600 μm mesh and stored at -22°C until the analysis.

After the NIR analysis, the 117 samples of raw cocoa (RC) were dried to constant weight in an oven under air circulation at 80°C . After drying these samples, the spectra were again obtained.

2.3. Defatting of Cocoa Samples. Dried fermented cocoa samples (56 samples) were exhaustively degreased with petroleum ether (Synth, Diadema, SP, Brazil) at a temperature of $55 \pm 1^{\circ}\text{C}$ (method 963.15, AOAC), for 24 hours, according to the Association of Official Analytical Chemists-AOAC [30]. After removing the fat, DF samples were kept for 24 hours at 80°C in an air circulation oven (model 80 series, Lucadema, Brazil) to remove the solvent. After

solvent removal, dried fermented degreased cocoa samples (DFD) were stored at -22°C until analysis.

Through the procedures described above, it was possible to evaluate cocoa samples in 4 categories: raw cocoa (117), dried unfermented cocoa (117), dried fermented cocoa (72), and dried fermented degreased cocoa (56) (Table 1).

2.4. Obtaining NIR Spectra. NIR spectra were obtained using an MPA FT-NIR spectrometer (Bruker Optics, Ettlingen, Germany). The spectral data were acquired using absorbance mode in the spectral range from 3500 to 12500 cm^{-1} , with 16 cm^{-1} resolution and an average of 32 scans per spectrum. For the samples spectral reading, vials with a volume of 3 ml were used with a capacity for about 1 g of the milled and sieved cotyledon. The spectral reading was performed at $\pm 25^{\circ}\text{C}$.

2.5. Discrimination Model Development by Geographical Origin. The spectroscopy software used to construct discrimination models was OPUS 6.5 (Bruker Optics, Ettlingen, Germany). The spectral data were previously processed in OPUS 6.5, according to the type of sample. For the samples of cocoa RC, DF, and DFD, the vector normalization pretreatment (SNV- standard normal variate) was performed, while for the DU samples, SNV treatment was performed along with the first derivative.

After the application of pretreatments, the data were processed to develop the discrimination model by the geographical origin of each set of samples. The exploratory analysis method, PCA, was applied to multivariate data to make a visual inspection of results more evident.

The spectral region that has best differentiated the geographic regions was chosen by the software operator, based on information related to the sample chemical composition, such as water content, fat, and protein.

3. Results and Discussion

3.1. Spectral Analysis. The complete NIR spectra obtained from RC, DU, DF, and DFD samples are shown in Figure 1. The prominent peaks can be observed around 9000 - 8000 , 8500 - 6000 , 6000 - 5000 , and 5000 - 4500 cm^{-1} within the NIR range. In the RC spectrum (Figure 1(a)), two intense absorption peaks are observed, in 5312 and 7202 cm^{-1} , which are related to the water content in the samples, according to [7], corresponding to the region of the first Overtone of the O-H stretch and O-H deformation. The water absorption bands should be removed to reduce interference with the chemical structures corresponding to the group's CH and CO and combinations of amine groups [31]. With that, we could see in Figures 1(b) and 1(c) that after removing the water, the absorption peaks of other components in the samples became more evident.

The regions corresponding to 9000 - 7500 , 6000 - 6030 , 6030 - 4000 , 4950 - 4500 , and 4500 - 3850 cm^{-1} were used for all the calibration model development.

None of the spectra presented in Figure 1 was possible to verify the differences between fruits from different analyzed

TABLE 1: Spectra collected by region and category: raw cocoa (RC), dried unfermented cocoa (DU), dried fermented cocoa (DF), and dried fermented degreased (DFD).

	RC	DF	DU	DFD
Tomé-Açú	40	20	40	19
Tucumã	40	21	40	19
Medicilândia	37	31	37	18
Total	117	72	117	56

regions, even in the fermented samples. These differences were only noticed after mathematical treatments through the application of statistical analysis techniques for model development.

3.2. Discrimination Models for Raw Cocoa. For the RC cocoa samples, it was impossible that the spectral band (6200 - 4000 cm^{-1}) could discriminate groups of the fruits from different regions, as observed in Figure 2(a).

The graph of the scores presented in Figure 2(b), referring to the spectral range between 4000 and 6200 cm^{-1} , shows the lack of differentiation between the three regions' scores. Bands referring to the phenol group, which can be observed in the regions corresponding to 3562 - 3322 cm^{-1} , aromatic ring related bands, referring to the region of 2925 - 2854 cm^{-1} (attributed to the CH stretch of the aromatic ring) and 1645 - 1544 cm^{-1} (attributed to C in the aromatic ring) [7], were overlapped. This effect is directly related to the amount of water and fats in the RC cocoa samples.

Therefore, high absorption of radiation NIR by water present in samples is a factor that contributes negatively to analysis, since the water spectra overlap with the other spectra of interest, making it challenging to construct an analytical and statistical model of geographic discrimination of the in-nature Amazonian cocoa.

3.3. Discrimination Models for Dried Unfermented Cocoa. After the removal of water by drying, the DU samples have shown a trend toward the formation of groups related to discrimination by geographic region. The best spectral range for the DU discrimination was between 4300 and 9300 cm^{-1} observed in Figure 2(c).

These spectra located between 4300 and 4597 , 4902 - 5199 , 5805 - 6102 , and 6406 - 9300 cm^{-1} in the whole spectrum for DU are associated with the carbonyl group spectral regions, corresponding to stretching combination (CH_2 and CH), first harmonic of CH present in the aromatic ring, combination of C-C and C-N and second harmonic of N-H [15]. These molecular vibrations are caused by functional groups corresponding to polyphenols, alkaloids, vicilin class globulins, proteins, amines, acids, polysaccharides, and other aromatic compounds [7]. These diverse functional groups are the digital identity of each sample from different Amazon regions.

The differentiation between the samples from Tomé-Açu and Tucumã was evident, as observed in Figure 2(d). However, the scores related to the set of Medicilândia samples get mixed with the scores of the samples from the

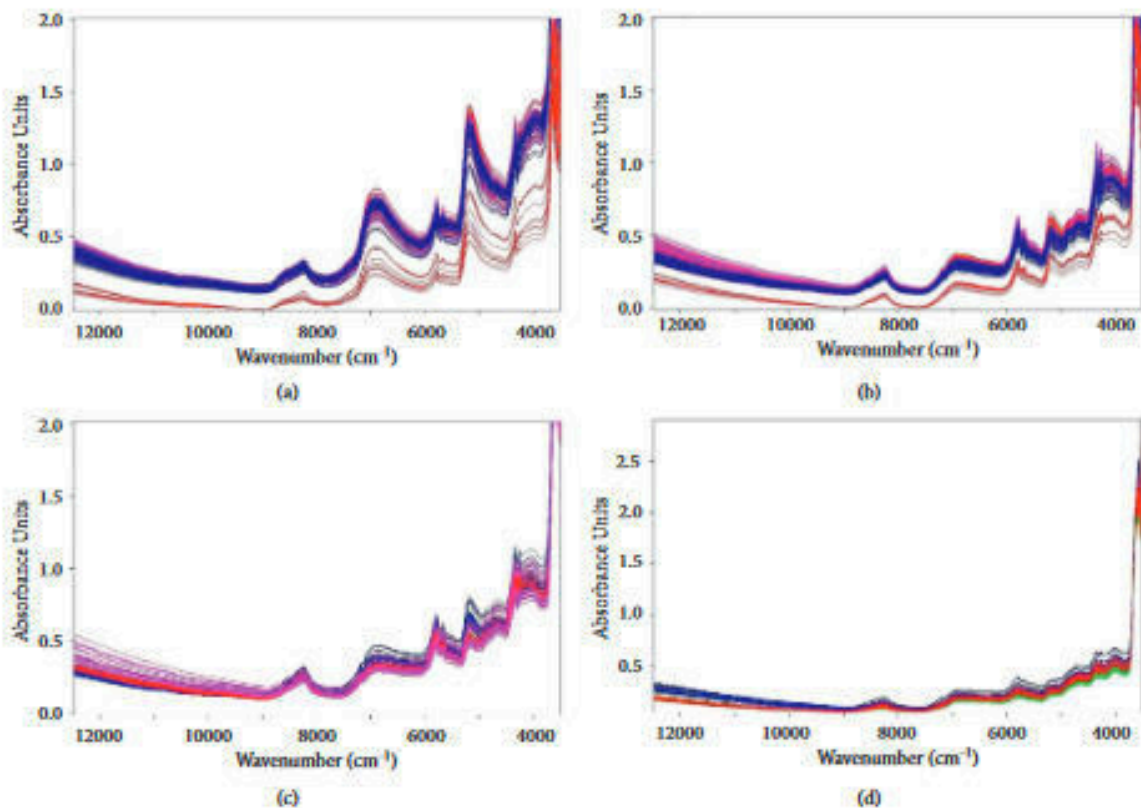


FIGURE 1: Absorption spectrum of (a) RC, (b) DU, (c) DF, and (d) DFD, in the range of 3600 to 12500 cm^{-1} .

other two regions. Several factors can explain the formation of characteristic groups of cocoa, such as chemical composition [32, 33], degree of fermentation [34], and genotype [35], which may be associated with the region of origin of the seeds [36].

According to [37], cocoa, like other food products, has its characteristics influenced by locality. This influence is directly related to the main components of alkaloid polyphenols, proteins, amines, polysaccharides, acids, and diverse aromatic compounds, as previously mentioned [7].

3.4. Discrimination Models for Fermented Cocoa. The fermentation of cocoa beans is essential, due to the diverse microbial processes developed as a consequence of changes in temperature, pH, and oxygen availability, promoting significant biochemical changes in the type and concentration of flavor precursors in cocoa beans [37] which is a crucial step in the formation of quality sensory attributes. Chemical and biochemical complex changes and interactions occur in the cocoa beans during drying, storage, and fermentation, contributing to the complexity and identity of the Amazon cocoa [38].

The fermentation produces flavor and aroma precursors, such as free amino acids and peptides from enzymatic degradation of proteins and sugars that reduce the enzymatic degradation of sucrose, as well as a significant increase

in volatile compounds such as organic acids, esters, alcohols, and aldehydes, after and during the fermentation of cocoa beans. The stoichiometry of their progenitors influences the concentration of these precursors; therefore, this is a response to the nutritional quality of the cultivated soil, from which the nutrients necessary for the synthesis of their parents were collected, revealing the good indicators of the production region [39].

Fermentation was a crucial factor in forming the characteristics of each region, because unlike the dried unfermented samples (DU), the dried fermented samples (DF), whose spectra are shown in Figure 2(e), allowed good discrimination between groups from different regions. The best spectral discrimination range observed was 6000 to 6800 cm^{-1} .

These results allow one to verify that there is a strong influence of the fermentation on the biochemical characteristics of the samples, to allow a clear distinction between them, as shown in Figure 2(f).

Caligiani et al. [10] reported that the progressive fermentation of cocoa seeds causes the hydrolysis of peptides in amino acids, reducing the violet color of the cotyledons at the end of the fermentation. The phenolic compounds are reduced upon drying; this reduction is mainly attributed to the enzymatic action, followed by nonenzymatic reactions due to quinone polymerization, with pH increase and high capitation of O_2 during sun drying [9].

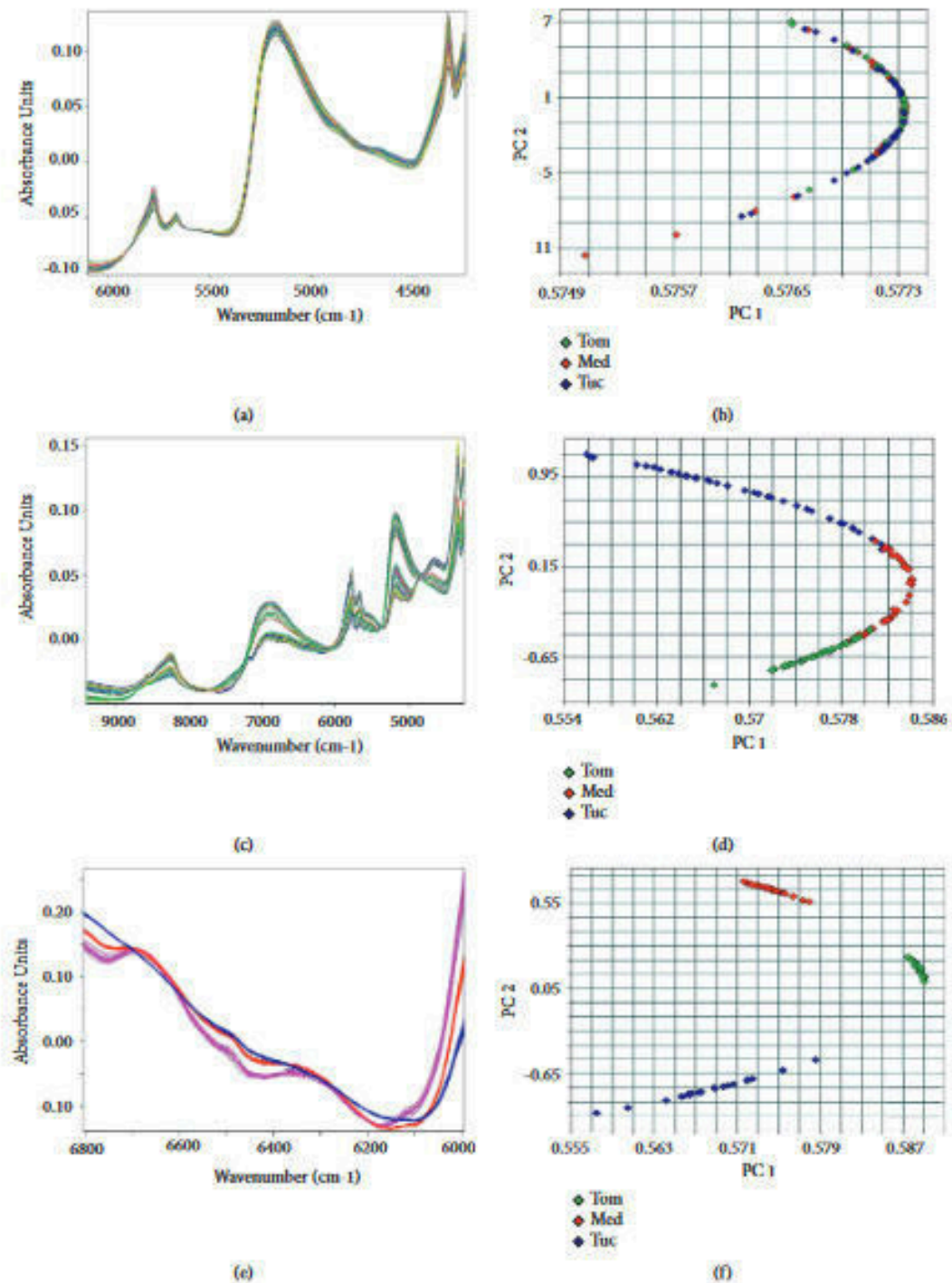


FIGURE 2: The absorption spectrum of the separation regions of (a) RC, (c) DU, and (e) DF, and their respective scores in the PCs graphs (b), (d), and (f).

The study of Sirbu et al. [12] estimated the increase of lipid content during fermentation from 3 to 6% when compared to the value before fermentation. In addition, the

authors observed a change in the triacylglycerol profile caused by fermentation. Some polar triacylglycerols, such as derivatives of hydroxyl allyl fatty acid presented in

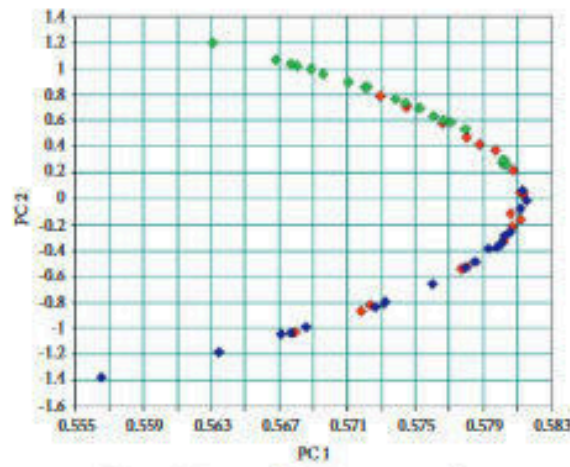


Figure 3: Scores of DFD cocoa samples.

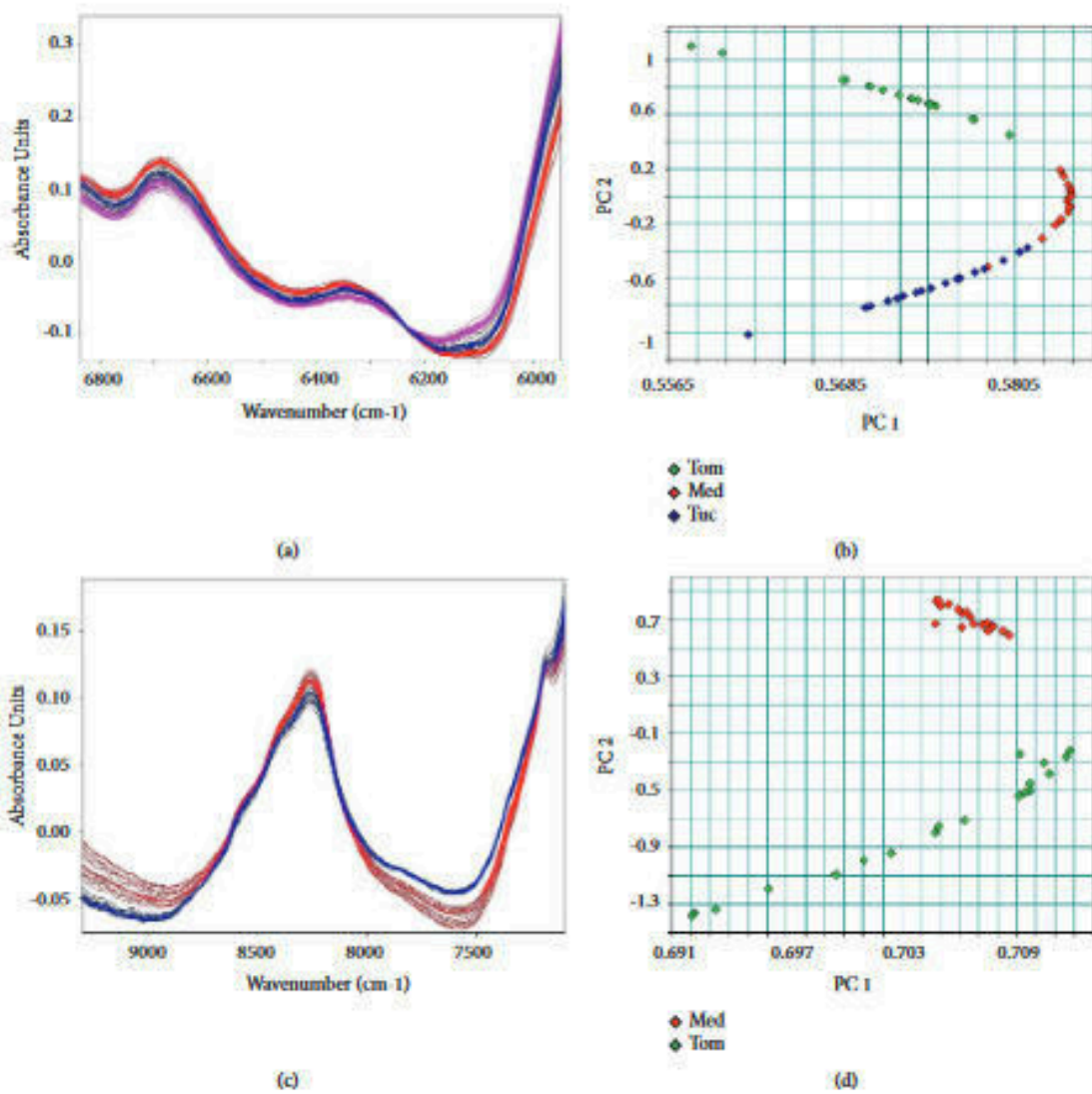


Figure 4: The absorption spectrum of the DFD separation regions (a, c) and their respective scores (b, d), for the two selected optimal separation regions.

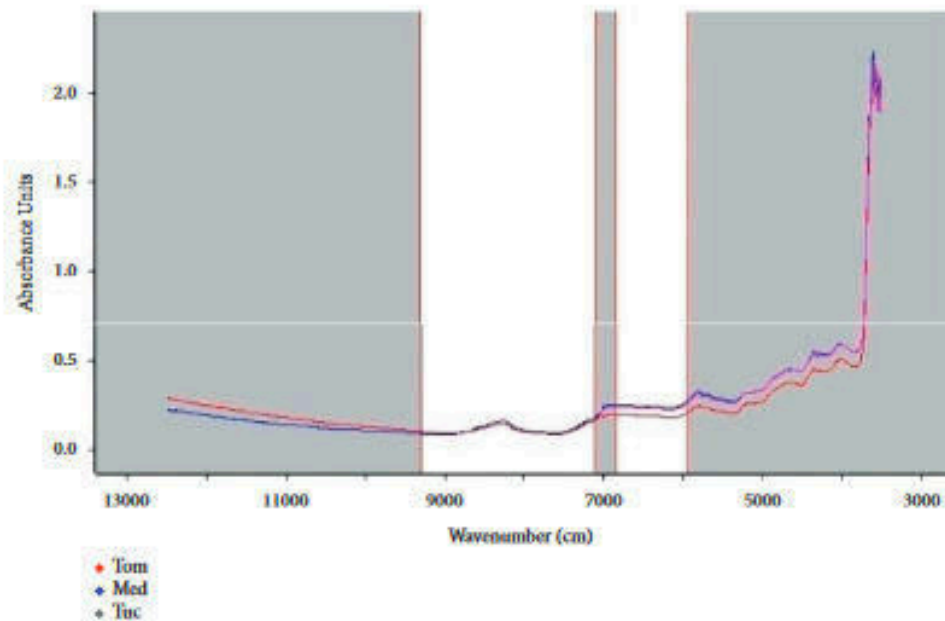


FIGURE 5: Absorption spectrum highlighting the two separation regions of DFD samples (1st and 2nd region).

unfermented cocoa seeds, were not found in dried fermented cocoa seeds, thus allowing chemical differentiation between unfermented and fermented cocoa seeds.

According to the literature, we can see that fermentation causes changes in the profile and concentration of various components of cocoa. Such components or concentrations of components can be characteristic for each region, and because of this, it is possible to carry out geographic identification. The NTR technique associated with chemometrics can previously identify the similarity and differences in various types of samples and classify them, making it possible to visually identify groups that have similar or distinct characteristics.

In this work, we used 15 external samples (not used in the construction of the model) to be able to test the efficiency of the constructed discrimination model (5 samples from Tomé-Açu, 5 from Tucumã, and 5 from Medicilândia). 100% of the samples were correctly classified according to their regions of origin.

3.5. Discrimination Models for Degreased Cocoa. To evaluate the importance of the lipid content of the samples in geographic differentiation, it was made the degreasing of the DF cocoa samples and the spectral evaluation was carried out again, using the same spectral treatment applied to discriminate the nondegreased fermented samples. Scores of the degreased samples are shown in Figure 3.

As observed in Figure 3, even though the mathematical model treatment was used for DF samples (dried fermented), it was not observed the separation between the three groups; however, a partial separation between two regions (Tomé-Açu and Tucumã) was observed, which was the same as visualized in the DU samples. This shows that the fat removal made it difficult to classify the samples in three different geographic

regions using the DFD sample discrimination model, validating partially the results of [40], who have shown a strong correlation of fatty acid composition in the geographic differentiation. In our research, we observed that it was possible to discriminate fermented samples more easily with their original butter constitution.

The lipids and proteins storage cells in cocoa seeds have complex cytology, composed of a compacted cytoplasm with multiple vacuoles, where proteins and lipids can be found, as well as other components such as starch granules, which are important for the definition of specific flavor and aroma characteristics. These biochemical and cytological characteristics of the cells change in percentage and morphological characteristics from region to region, affected by several physical, chemical, and climatic factors [39].

Since there was an indication of partial separation of DFD samples, it was developed a model from two separation steps. In the first separation step, the spectral region of 5950 to 6835 cm^{-1} was used (Figure 4(a)), in which the regions of Tomé-Açu and Tucumã were separated (Figure 4(b)). In the second step of the model, the sample data for the two groups that did not show clear separation in the first step (Medicilândia and Tucumã), the spectral region of 7100 and 9300 cm^{-1} was used (Figure 4(d)). Two spectral regions were used to build the complete separation model of the DFD samples, as shown in Figure 5.

The DFD cocoa seed sample separation model was performed in two stages. For the first stage, it was used the spectral data of the cocoa from the three regions (Medicilândia, Tucumã, and Tomé-Açu) (Figure 4(b)), and for the second stage, it was used the spectral data of the cocoa from Medicilândia and Tucumã, whose separation it was less evident in the first stage (Figure 4(c)). Thus, compared to the previous separation model (of DF seeds), we could observe

that the lipid fraction had considerable influence; however, it was not essential for geographic identification.

According to Sirbu et al. [12], after fermentation, cocoa presents a different profile of triacylglycerol and starch granules, which influence the spectroscopic behavior. And after the fat removal, the other components that changed during fermentation remain there, making geographic differentiation possible [12].

The application of the NIR technique is quite simple; however, the path to acquiring spectral information and analytical data is time-consuming and costly. Once the database of spectral and analytical information is obtained, the mathematical model can be constructed through the use of computational algorithms that plan and/or optimize such experimental procedures. From this, this elaborated model can be used to analyze external samples only with spectral information. The spectral data will be correlated with the data previously obtained and the result is estimated. In this way, the analysis becomes simple, fast, and economical. It is enough that, after the construction of the prediction model, the spectra of the material in question are obtained and the model is applied.

4. Conclusions

The results showed that it was possible to evaluate the influence of sample preparation on the geographic discrimination of cocoa from different geographic regions, in the Pará state, using the NIR technique.

The importance of the fermentation process for geographic identification was observed, and it was noticed provided specific characteristics for each region evaluated. The fermented cocoa lipids were crucial in the formation of groups in the fermented samples, also indicating that fermentation is a crucial step to develop specific characteristics for each collection region. Thus, it is assumed that the biotas of each region should affect the fatty acids contained in cocoa butter differently since similar results were not found in defatted samples.

The effect of the climate and production process, although not directly explicit, cannot be ruled out, considering that they can significantly influence the type of biota and the fermentation process, as well as other possible variables.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Ethical Approval

Ethics approval was not required for this research.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES,

Finance Code 01) for the doctoral scholarship of the first author and Programa de Pós-graduação em Ciência e Tecnologia de Alimentos of the Federal University of Pará (PPGCTA/UFPA) for the great support offered. This study was financed by the Instituto Tecnológico Vale (ITV) (Cacau P2) and supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-CAPES (Process Number: 1537413).

References

- [1] L. J. R. Lima, M. H. Almeida, M. J. Rob Nout, and M. H. Zwietering, "Theobroma cacao L., 'the food of the gods': quality determinants of commercial cocoa beans, with particular reference to the impact of fermentation," *Critical Reviews in Food Science and Nutrition*, vol. 51, pp. 731–761, 2011.
- [2] C. Oleaga, M. Garcia, A. Solé, C. J. Ciudad, M. Izquierdo-Pulido, and V. Noé, "CYP1A1 is overexpressed upon incubation of breast cancer cells with a polyphenolic cocoa extract," *European Journal of Nutrition*, vol. 51, pp. 465–476, 2012.
- [3] O. Khawaja, J. M. Gaziano, and I. Djoussé, "Chocolate and coronary heart disease: a systematic review," *Current Atherosclerosis Reports*, vol. 13, no. 6, pp. 447–452, 2011.
- [4] V. T. T. Ho, J. Zhao, and G. Fleet, "The effect of lactic acid bacteria on cocoa bean fermentation," *International Journal of Food Microbiology*, vol. 205, pp. 54–67, 2015.
- [5] A. Krämer, A. Engel, D. Kadow et al., "Fast and neat—determination of biochemical quality parameters in cocoa using near infrared spectroscopy," *Food Chemistry*, vol. 181, pp. 152–159, 2015.
- [6] S. A. Oyeleke, A. M. Ajayi, S. Umukoro, A. O. Aderibigbe, and O. G. Ademowo, "Anti-inflammatory activity of *Theobroma cacao* L. stem bark ethanol extract and its fractions in experimental models," *Journal of Ethnopharmacology*, vol. 222, pp. 239–248, 2018.
- [7] E. Teye, X. Huang, H. Dai, and Q. Chen, "Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 114, pp. 183–189, 2013.
- [8] J. E. Kongor, M. Hinch, D. V. de Walle, E. O. Afoakwa, P. Boeckx, and K. Dewettinck, "Factors influencing quality variation in cocoa (*Theobroma cacao*) bean flavour profile—a review," *Food Research International*, vol. 82, pp. 44–52, 2016.
- [9] J. F. M. Cruz, P. B. Leite, S. E. Soares, and E. S. Bispo, "Bioactive compounds in different cocoa (*Theobroma cacao*, L) cultivars during fermentation," *Food Science and Technology*, vol. 35, no. 2, pp. 279–284, 2015.
- [10] A. Caligiani, A. Marseglia, B. Prandi, G. Palla, and S. Sforza, "Influence of fermentation level and geographical origin on cocoa bean oligopeptide pattern," *Food Chemistry*, vol. 211, pp. 431–439, 2016.
- [11] R. Megías-Pérez, S. Grimbs, R. N. D'Souza, H. Bernaert, and N. Kuhnert, "Profiling, quantification and classification of cocoa beans based on chemometric analysis of carbohydrates using hydrophilic interaction liquid chromatography coupled to mass spectrometry," *Food Chemistry*, vol. 258, pp. 284–294, 2018.
- [12] D. Sirbu, A. Grimbs, M. Corno, M. S. Ullrich, and N. Kuhnert, "Variation of triacylglycerol profiles in unfermented and dried fermented cocoa beans of different origins," *Food Research International*, vol. 111, pp. 361–370, 2018.

- [13] E. Afoakwa, *Chocolate Science and Technology*, Wiley, Oxford, UK, 2010.
- [14] IBGE, "Brazilian institute of geography and statistics, indicators IBGE. Agricultural production statistics," 2017, <https://ww2.ibge.gov.br/home/estatistica/indicadores/agropecuaria/>.
- [15] A. Veselá, A. S. Barros, A. Synytsya, I. Delgado, J. Čopíková, and M. A. Coimbra, "Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder," *Analytica Chimica Acta*, vol. 601, no. 1, pp. 77–86, 2007.
- [16] P. A. da Costa Filho, "Rapid determination of sucrose in chocolate mass using near infrared spectroscopy," *Analytica Chimica Acta*, vol. 631, no. 2, pp. 206–211, 2009.
- [17] E. Teye and X. Huang, "Novel prediction of total fat content in cocoa beans by FT-NIR spectroscopy based on effective spectral selection multivariate regression," *Food Analytical Methods*, vol. 8, no. 4, pp. 945–953, 2015.
- [18] H. Büning-Pfauc, "NIR spectroscopy, a revolution in analysis," *New Food*, vol. 7, pp. 41–47, 2004.
- [19] T. Woodcock, G. Downey, J. D. Kelly, and C. O'Donnell, "Geographical classification of honey samples by near-infrared spectroscopy: a feasibility study," *Journal of Agricultural and Food Chemistry*, vol. 55, no. 22, pp. 9128–9134, 2007.
- [20] V. Uričková and J. Sádecká, "Determination of geographical origin of alcoholic beverages using ultraviolet, visible and infrared spectroscopy: a review," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 148, pp. 131–137, 2015.
- [21] Y. Chen, M. Y. Xie, Y. Yan et al., "Discrimination of *Ganoderma lucidum* according to geographical origin with near infrared diffuse reflectance spectroscopy and pattern recognition techniques," *Analytica Chimica Acta*, vol. 618, no. 2, pp. 121–130, 2008.
- [22] Q. Chen, J. Zhao, and H. Lin, "Study on discrimination of Roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 72, no. 4, pp. 845–850, 2009.
- [23] M. Bassbasi, M. De Luca, G. Ioele, A. Oussama, and G. Ragno, "Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data," *Journal of Food Composition and Analysis*, vol. 33, no. 2, pp. 210–215, 2014.
- [24] D. Cozzolino, W. U. Cynkar, N. Shah, and P. A. Smith, "Can spectroscopy geographically classify Sauvignon Blanc wines from Australia and New Zealand?" *Food Chemistry*, vol. 126, no. 2, pp. 673–678, 2011.
- [25] A. M. C. Davies, J. G. Franklin, A. Grant, N. M. Griffiths, R. Shepherd, and G. R. Fenwick, "Prediction of chocolate quality from near-infrared spectroscopic measurements of the raw cocoa beans," *Vibrational Spectroscopy*, vol. 2, no. 2-3, pp. 161–172, 1991.
- [26] P. C. Aculey, P. Snitkjaer, M. Owusu et al., "Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics," *Journal of Food Science*, vol. 75, no. 6, pp. S300–S307, 2010.
- [27] C. Hue, Z. Gunata, A. Bergounhou et al., "Near infrared spectroscopy as a new tool to determine cocoa fermentation levels through ammonia nitrogen quantification," *Food Chemistry*, vol. 148, pp. 240–245, 2014.
- [28] E. Whitacre, J. O. Liver, R. Broek et al., "Predictive analysis of cocoa procyanidins using near-infrared spectroscopy techniques," *Journal of Food Science*, vol. 68, no. 9, pp. 2618–2622, 2003.
- [29] N. Caporaso, M. B. Whitworth, M. S. Fowler, and I. D. Fisk, "Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans," *Food Chemistry*, vol. 258, pp. 343–351, 2018.
- [30] W. Horwitz and G. W. Latimer, *Official Methods of Analysis of AOAC International*, Association of Official Analytical Chemistry International, Rockville, MD, USA, 18th edition, 2006.
- [31] S. Sunoj, C. Igathinathane, and R. Visvanathan, "Non-destructive determination of cocoa bean quality using FT-NIR spectroscopy," *Computers and Electronics in Agriculture*, vol. 124, pp. 234–242, 2016.
- [32] D. F. Barbin, L. F. Maciel, C. H. V. Bazoni et al., "Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses," *Journal of Food Science & Technology*, vol. 55, no. 7, pp. 2457–2466, 2018.
- [33] E. Teye, X. Huang, I. K. Sam-Amoah et al., "Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis," *Food Chemistry*, vol. 176, pp. 403–410, 2015.
- [34] E. K. Anyidoho, E. Teye, and R. Agbemaffe, "Nondestructive authentication of the regional and geographical origin of cocoa beans by using a handheld NIR spectrometer and multivariate algorithm," *Analytical Methods*, vol. 12, no. 33, pp. 4150–4158, 2020.
- [35] J. P. Cruz-Tirado, J. A. Fernández Pierna, H. Rogez, D. F. Barbin, and V. Baeten, "Authentication of cocoa (*Theobroma cacao*) bean hybrids by NIR-hyperspectral imaging and chemometrics," *Food Control*, vol. 118, Article ID 107445, 2020.
- [36] A. C. Oliveira, A. Marien, J. Hulin et al., "Development of real-time PCR methods for cocoa authentication in processed cocoa-derived products," *Food Control*, vol. 131, Article ID 108414, 2022.
- [37] D. Kadow, J. Bohlmann, W. Phillips, and R. Lieberei, "Identification of main fine or flavour components in two genotypes of the cocoa tree (*Theobroma cacao* L.)," *Journal of Applied Botany and Food Quality*, vol. 86, pp. 90–98, 2013.
- [38] V. T. T. Ho, J. Zhao, and G. Fleet, "Yeasts are essential for cocoa bean fermentation," *International Journal of Food Microbiology*, vol. 174, pp. 72–87, 2014.
- [39] E. O. Afoakwa, J. E. Kongor, J. Takrama, and A. S. Budu, "Changes in nib acidification and biochemical composition during fermentation of pulp pre-conditioned cocoa (*Theobroma cacao*) beans," *International Food Research Journal*, vol. 20, pp. 1843–1853, 2013.
- [40] M. Torres Moreno, E. Torrescasana, J. Salas Salvadó, and C. Blanch, "Nutritional composition and fatty acids profile in cocoa beans and chocolates with different geographical origin and processing conditions," *Food Chemistry*, vol. 166, pp. 125–132, 2015.