

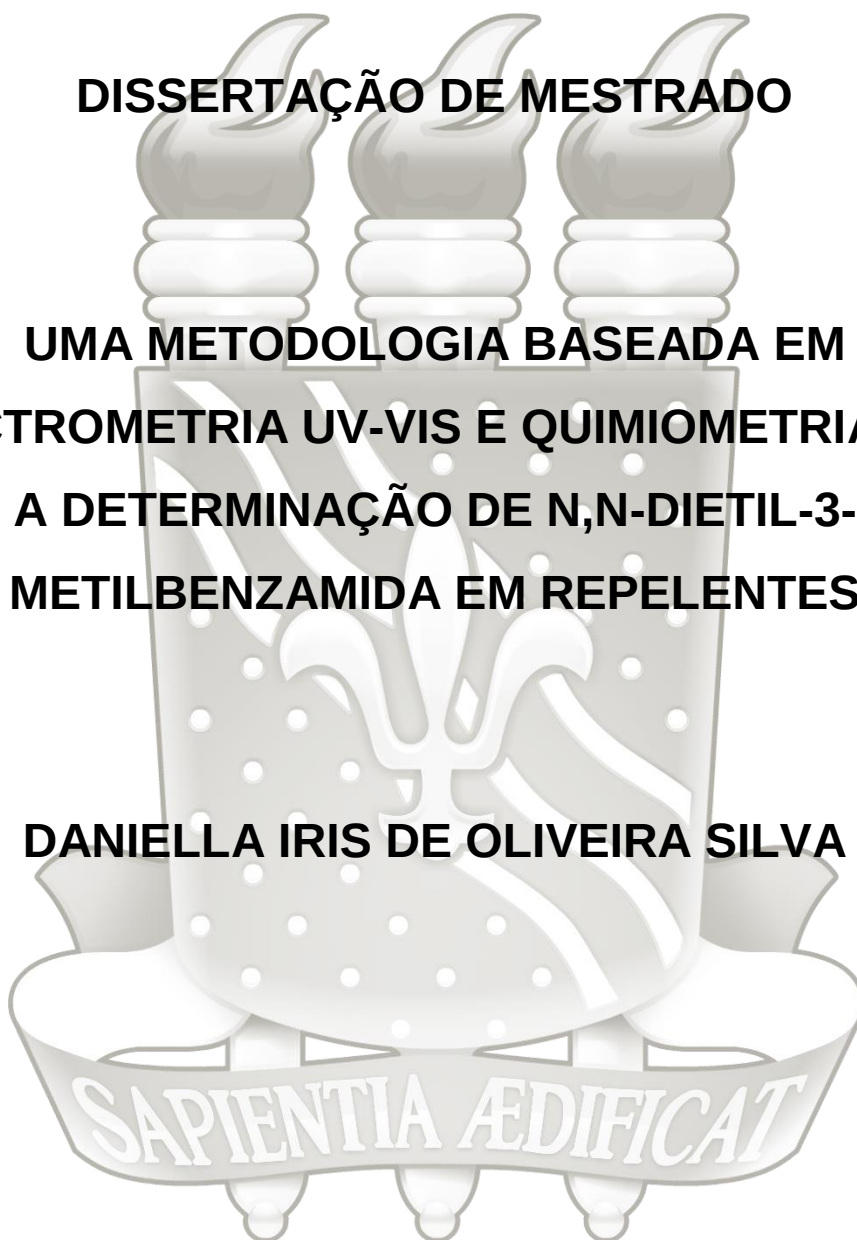


UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

DISSERTAÇÃO DE MESTRADO

**UMA METODOLOGIA BASEADA EM
ESPECTROMETRIA UV-VIS E QUIMIOMETRIA PARA
A DETERMINAÇÃO DE N,N-DIETIL-3-
METILBENZAMIDA EM REPELENTES**

DANIELLA IRIS DE OLIVEIRA SILVA



João Pessoa – PB – Brasil

Julho/2018



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

DISSERTAÇÃO DE MESTRADO

UMA METODOLOGIA BASEADA EM ESPECTROMETRIA UV-VIS E QUIMIOMETRIA PARA A DETERMINAÇÃO DE N,N-DIETIL-3- METILBENZAMIDA EM REPELENTES

DANIELLA IRIS DE OLIVEIRA SILVA*

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal da Paraíba como parte dos requisitos para obtenção do título de Mestre em Química, área de concentração Química Analítica.

Orientador: **Prof. Dr. Márcio José Coelho de Pontes**

* Bolsista: CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

João Pessoa – PB – Brasil

Julho/2018

Catálogo na publicação
Seção de Catalogação e Classificação

S586m Silva, Daniella Iris de Oliveira.
UMA METODOLOGIA BASEADA EM ESPECTROMETRIA UV-VIS E
QUIMIOMETRIA PARA A DETERMINAÇÃO DE
N,N-DIETIL-3-METILBENZAMIDA EM REPELENTES / Daniella
Iris de Oliveira Silva. - João Pessoa, 2018.
72 f. : il.

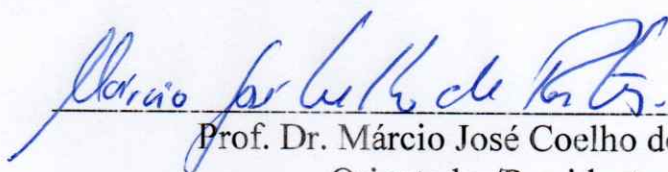
Orientação: MÁRCIO JOSÉ COELHO DE PONTES PONTES.
Dissertação (Mestrado) - UFPB/CCEN.

1. REPELENTES. 2. DEET. 3. ESPECTROMETRIA UV-VIS. 4.
CALIBRAÇÃO MULTIVARIADA. I. PONTES, MÁRCIO JOSÉ COELHO
DE PONTES. II. Título.

UFPB/CCEN

Uma metodologia baseada em espectrometria UV-VIS e quimiometria para determinação de N,N-dietil-3-metilbenzamida em repelentes.

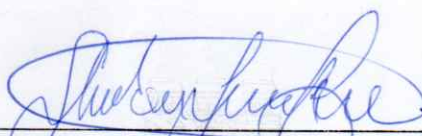
Dissertação de Mestrado apresentada pela aluna Daniella Iris de Oliveira Silva e aprovada pela banca examinadora em 27 de julho de 2018.



Prof. Dr. Márcio José Coelho de Pontes
Orientador/Presidente



Prof. Dr. Everaldo Paulo de Medeiros
Examinador



Prof. Dr. Sherlan Guimarães Lemos
Examinador

*"Cada célula, todo fio de cabelo
Falando assim parece exagero,
Mas se depender de mim eu vou até o fim"*
(Humberto Gessinger)

*Aos meus pais,
Aos meus avós José, Zoca e Severino (in memoriam) e Baia
Eu dedico.*

AGRADECIMENTOS

- *A Deus por me iluminar, me proteger, me dar forças pra superar os momentos difíceis e não me deixar desistir de meus objetivos.*
- *Aos meus pais por todo amor, por terem me criado no caminho do bem e por todo incentivo aos estudos.*
- *Ao meu esposo Thiago Almeida por todo amor, companheirismo, apoio e paciência nos momentos difíceis.*
- *Ao Professor Márcio, pela orientação e grande contribuição para minha formação acadêmica.*
- *A Mayara Ferreira pela amizade de longa data e pelas contribuições para este trabalho e na vida.*
- *A Welma Vilar por toda ajuda, contribuições e por ser minha “referência” neste trabalho.*
- *A Carol Mascarenhas e ao professor Sherlan pela grande ajuda no início dessa jornada.*
- *Aos amigos que incentivaram a seguir em frente sempre com o bom-humor deles: Juliana, Matheus, Kelly, Tamires e a todos que fazem o LAQA e que de alguma forma contribuíram para este trabalho.*
- *A CAPES, pela bolsa concedida.*
- *A UFPB, pela formação acadêmica.*

SUMÁRIO

LISTA DE FIGURAS	iii
LISTA DE TABELAS	v
LISTA DE ABREVIATURAS.....	vi
RESUMO.....	ix
ABSTRACT.....	x
1 INTRODUÇÃO	2
1.1 Caracterização da problemática	2
1.2 Objetivos.....	4
1.2.1 Objetivo geral	4
1.2.2 Objetivos específicos	4
2 FUNDAMENTAÇÃO TEÓRICA	6
2.1 Repelentes	6
2.2 Espectrometria UV-VIS	8
2.2.1 Instrumentação	9
2.3 Quimiometria.....	11
2.3.1 Calibração Multivariada.....	13
2.3.1.1 Regressão em Mínimos Quadrados Parciais (PLS)	14
2.3.1.2 Regressão em Mínimos Quadrados Parciais por Intervalo (<i>i</i> PLS).....	16
2.3.1.3 Regressão Linear Múltipla (MLR)	16
2.3.2 Algoritmos de seleção de variáveis	17
2.3.2.1 Stepwise.....	17
2.3.2.2 Algoritmo Genético	18
2.3.2.3 Algoritmo das Projeções Sucessivas	19
2.3.3 Seleção de amostras	19
2.3.3.1 Particionamento de amostras baseado nas distâncias X-y	20
3 METODOLOGIA	23
3.1 Aquisição e preparo de amostras.....	23
3.2 Análise de referência.....	24
3.3 Aquisição dos espectros.....	24
3.4 Procedimento quimiométrico e software	25
4 RESULTADOS E DISCUSSÃO	28
4.1 Análise espectral.....	28

4.2 Análise exploratória.....	31
4.3 Desenvolvimento dos modelos de calibração.....	32
4.3.1 Modelo PLS.....	32
4.3.2 Modelo <i>i</i> PLS.....	34
4.3.3 Modelos SPA-MLR, SW-MLR e GA-MLR	36
4.4 Análise dos modelos de calibração	39
5 CONCLUSÕES	45
5.1 Perspectivas	47
REFERÊNCIAS.....	48

LISTA DE FIGURAS

Figura 2.1- Figura 1.1 Moléculas de princípios ativos de repelentes no mercado brasileiro. (a) DEET; (b) Icaridida; (c) IR3535.....	6
Figura 2.2- Espectro eletromagnético (adaptado de Skoog [16], 2002).....	8
Figura 2.3- Diagrama geral de um espectrofotômetro de feixe único (adaptado de Skoog [16], 2002).....	10
Figura 2.4- Representação da atuação de radiação eletromagnética em uma solução contendo c mols/L da substância absorvente e caminho óptico b (adaptado de Skoog [16], 2002).....	10
Figura 3.1- Esquema do preparo de amostras.....	24
Figura 4.1- Espectros UV-VIS originais na faixa de 190 a 1100 nm obtidos a partir de 107 amostras.	29
Figura 4.2- Espectros UV-VIS obtidos a partir de 107 amostras: a) espectros brutos com seleção a priori de variáveis; b) espectros pré-processados.....	30
Figura 4.3- Molécula de N,N-Dietil-3- metilbenzamide (DEET) com centros cromóforos em destaque.....	31
Figura 4.4- a) Cromatogramas dos padrões de IR3535 (verde) e DEET (Azul) usados na curva de calibração para o método de referência (HPLC); b) Espectro obtido pelo método de referência na concentração de 8 mg Kg^{-1} do padrão de DEET.....	31
Figura 4.5- Gráfico dos escores PC1 para as 107 amostras de repelentes rotuladas de acordo com a faixa de concentração de DEET (m m^{-1}): (●) menor que 7%; (●) maior ou igual a 7% e menor que 15%; (●) maior ou igual a 15% e menor que 20%;(●) maior ou igual a 20%.....	32

Figura 4.6- Gráfico de pesos de PC1	33
Figura 4.7- Gráficos obtidos a partir de modelo PLS a) valores preditos versus valores de referência para as amostras de calibração (●) e predição (♦) b) Elipse de confiança.....	34
Figura 4.8- Resultados do <i>i</i> PLS para o DEET (%m m ⁻¹) para os 10 intervalos (barras) e para o modelo <i>full-spectrum</i> (---).....	36
Figura 4.9- Gráficos obtidos a partir de modelo <i>i</i> PLS a) valores preditos versus valores de referência para as amostras de calibração (●) e predição (♦) b) Elipse de confiança	37
Figura 4.10- espectros UV-VIS médios com as variáveis selecionadas (●) a) SPA; b) SW; c) GA.....	38
Figura 4.11- Gráficos de valores preditos versus valores de referência para calibração (●) e predição (♦) obtidos com os modelos: a) SPA-MLR e SW-MLR; b) GA-MLR para determinação do teor de DEET em amostras de repelentes. Região da elipse de confiança: d) SPA-MLR; e) SW-MLR; f) GA-MLR (●) Ponto ideal para inclinação e o intercepto.....	40

LISTA DE TABELAS

Tabela 4.1- Resultados obtidos para os modelos de calibração PLS, <i>i</i> PLS, SPA-MLR, SW-MLR e GA-MLR para o teor de DEET.	41
Tabela 4.2- Resultados para o teste <i>F</i> (95% de confiança) obtidos na comparação entre os modelos.....	42
Tabela 4.3- Resultados para o teste <i>t</i> -pareado (95%de confiança) obtidos na comparação dos valores de referências e valores previstos pelos modelos....	43
Tabela 4.4- Valores rotulados, determinados pelo método de referência e estimados pelos modelos SPA-MLR ou SW-MLR..	44

LISTA DE ABREVIATURAS

ANVISA – Agência Nacional de Vigilância Sanitária.

ACN – Acetonitrila

CEC – Contaminants of Emerging Concern (Contaminantes de Preocupação Emergente)

DAD – Detector de Arranjo de Diodos

DDB – Boron-Doped Diamond electrode (eletrodo de diamante dopado com boro)

DEET – N,N- dietil-3-metilbenzamida

EPA – Environmental Protection Agency (Agência de Proteção Ambiental dos Estados Unidos).

FDA – Food and Drugs Administration (Administração de Alimentos e Medicamentos dos Estados Unidos).

FIOCRUZ – Fundação Osvaldo Cruz.

GA – Genetic Algorithm (Algoritmo Genético)

GC – Cromatografia Gasosa

GC/MS – Cromatografia Gasosa acoplada a Espectrômetro de Massa

HOMO – Highest Occupied Molecular Orbital (orbital molecular ocupado de maior energia).

HPLC – High Performance Liquid Chromatography (Cromatografia Líquida de Alta Eficiência)

HPLC/MS – Cromatografia Líquida de Alta Eficiência acoplada a Espectrômetro de Massa

HPLC-DAD – Cromatografia Líquida de Alta Eficiência acoplada a Detector de Arranjo de Diodos

iPLS – Interval Partial Least Squares (Regressão em Mínimos Quadrados Parciais por Intervalo)

INCQS – Instituto Nacional de Controle de Qualidade em Saúde.

IR3535 – Etil 3-[acetil(butil)amino]-propanoato.

IUPAC – International Union of Pure and Applied Chemistry (União Internacional de Química Pura e Aplicada)

KS – Kennard-Stone

LOD – Limit of Detection (Limite de detecção).

LOQ – Limit of Quantification (Limite de quantificação).

LUMO – Lowest Unoccupied Molecular Orbital (orbital molecular desocupado de menor energia)

MLR – Regressão Linear Múltipla (Multiple Linear Regression)

NIR –Near Infrared (Infravermelho próximo)

OMS – Organização Mundial da Saúde.

PLS – Partial Least Squares (Mínimos Quadrados Parciais)

PC – Componente Principal

PCA – Principal Components Analysis (Análise por Componentes Principais)

PPCP – Pharmaceuticals and Personal Care Products (Produtos Farmacêuticos e de Cuidados Pessoais)

RMSE – Root Mean Square Error (Raiz Quadrada do Erro Médio Quadrático)

RMSECV – Root Mean Square Error of Cross Validation (Raiz Quadrada do Erro Médio Quadrático de Validação Cruzada)

RMSEP – Root Mean Square Error of Prediction (Raiz Quadrada do Erro Médio Quadrático de Predição)

RMSEV – Root Mean Square Error of Validation (Raiz Quadrada do Erro Médio Quadrático de Validação)

RS – Random Sampling (Amostragem Randômica)

SNV – Variação Normal Padrão

SPA – Successive Projections Algorithm (Algoritmo das Projeções Sucessivas).

SPE – Solid Phase Extraction (Extração em Fase Sólida)

SPXY– Sample Set Partitioning Based on Joint X-y distances (Particionamento de Amostras Baseado nas distâncias X-y)

SW – Stepwise

TOC – Total Organic Carbon (Carbono Orgânico Total)

UHPLC – Ultra High Performance Liquid Chromatography (Cromatografia Líquida de Ultra Eficiência)

UV-VIS – Ultravioleta-Visível

VL – Variáveis Latentes

RESUMO

SILVA, D. I. O. **Uma metodologia baseada em espectrometria UV-VIS e quimiometria para a determinação de N,N-dietil-3-metilbenzamida em repelentes**. Dissertação de Mestrado (Mestrado em Química) – Laboratório de Automação e Instrumentação em Química Analítica, Universidade Federal da Paraíba, João Pessoa, 2018.

Orientador: Prof. Dr. Márcio José Coelho de Pontes

Nos últimos anos, surtos de doenças transmitidas por vetores tem causado grande apreensão à população. Os repelentes surgem como uma alternativa de prevenção acessível e se torna cada vez mais importante o controle de qualidade destes cosméticos, visto que o teor do princípio ativo está diretamente relacionado à eficiência e ao tempo de proteção proporcionados pelo repelente. Este trabalho propõe o desenvolvimento de uma metodologia para determinação de N,N- dietil-3-metilbenzamida (DEET) em repelentes de insetos em loção, usando espectrometria UV-VIS e calibração multivariada. Para esta finalidade foram desenvolvidos e comparados modelos quimiométricos baseados na regressão pelos mínimos quadrados parciais (PLS), regressão pelos mínimos quadrados parciais por intervalos (iPLS) e regressão linear múltipla com prévia seleção de variáveis pelo algoritmo das projeções sucessivas (SPA-MLR), algoritmo genético (GA-MLR) e formulação stepwise (SW-MLR). Resultados satisfatórios foram encontrados com todos os modelos, com valores de RMPSEP variando de 0,88 a 0,92% m/m. Um Teste *F* realizado com nível de confiança de 95% indicou que não há diferença estatisticamente significativa entre os diferentes modelos desenvolvidos. Adicionalmente, um teste t-pareado (com 95% de confiança) não identificou diferenças entre os métodos propostos e o de referência. Os resultados demonstraram o potencial da espectrometria UV-VIS associada à calibração multivariada para determinar o teor de DEET em repelentes como uma estratégia rápida, simples e com uma boa correlação entre os valores estimados pelo modelo e os valores de referência.

Palavras-chave: repelentes, DEET, espectrometria UV-VIS, calibração multivariada.

ABSTRACT

SILVA, D. I. O. **Uma metodologia baseada em espectrometria UV-VIS e quimiometria para a determinação de N,N-diethyl-3-metilbenzamida em repelentes.** Dissertação de Mestrado (Mestrado em Química) – Laboratório de Automação e Instrumentação em Química Analítica, Universidade Federal da Paraíba, João Pessoa, 2018.

Orientador: Prof. Dr. Márcio José Coelho de Pontes

In recent years outbreaks of vector-borne diseases have caused great concern to the population. The repellents appear as an affordable prevention alternative and it becomes increasingly important the quality control of these cosmetics, since the content of the active ingredient is directly related to the efficiency and the protection time provided by the repellent. This work proposes the development of a methodology to the determination of N, N- Diethyl-3-methylbenzamide (DEET) in insect repellents in lotion using UV-VIS spectrometry and multivariate calibration. For this propose, chemometric models based on Partial Least Squares (PLS) regression, Interval Partial Least Squares (*i*PLS) regression and Multiple Linear Regression (MLR) with previous variable selection by Successive Projections Algorithm (SPA-MLR), Genetic Algorithm (GA-MLR) and Stepwise formulation (SW-MLR) were developed and compared. Satisfactory results were found with all the models, with RMSEP values varying from 0.88 to 0.92% m/m. An *F*- test performed at a confidence level of 95% indicated that there was not statistically significant difference between the different models developed. Additionally, a paired *t*-test (95% of confidence) did not identify differences between the proposed methods and the reference method. The results demonstrated the potential of UV-VIS spectrometry associated with multivariate calibration to determine DEET content in repellent like a fast, simple strategy and with a good correlation between the values estimated by the model and the reference values.

Keywords: repellent, DEET, UV-VIS spectrometry, multivariate calibration.

Capítulo 1

Introdução

1 INTRODUÇÃO

1.1 Caracterização da problemática

Devido ao elevado número de óbitos (700.000 por ano) [1], as doenças transmitidas por vetores tem causado grande preocupação mundial. Os insetos hematófagos são os vetores mais comuns. Os mosquitos ingerem microorganismos de um hospedeiro infectado e injetam em outro hospedeiro transmitindo doenças que já representam 17% de todas as doenças infecciosas. Estas doenças impactam mais fortemente países tropicais e nos últimos anos surtos de dengue, malária, chikungunya, febre amarela e Zika têm afetado estas populações [1]. Em outubro de 2015 o Brasil reportou à Organização Mundial da Saúde (OMS) uma relação alarmante entre a infecção por Zika vírus e a microcefalia. Análises exaustivas de evidências levaram ao consenso científico da associação entre a infecção por Zika vírus durante a gravidez e anomalias cerebrais congênitas, além de o vírus ser um gatilho para a síndrome de Guillain-Barré [2].

Pesquisadores no mundo inteiro vêm trabalhando no desenvolvimento de vacinas para as citadas doenças. No entanto, ainda não há registros de vacinas contra Zika e chikungunya aprovadas pela FDA [3]. Já para o vírus da dengue, a primeira vacina foi registrada no Brasil pela ANVISA em dezembro de 2015, que será fabricada pela companhia francesa Sanofi-Aventis [4].

Assumir medidas práticas agregadas ao uso de um bom repelente tem sido o modo mais acessível para a população de prevenir contra os ataques dos mosquitos e, conseqüentemente, o contágio com diversas doenças [2].

O repelente a ser adotado deve conter um dos três princípios ativos sintéticos que possuem eficácia comprovada e são regulamentados pela ANVISA, que por sua vez adota a regulamentação da Agência de Proteção Ambiental dos Estados Unidos (EPA:Environmental Protection Agency). São eles: N,N-dietil-3-metilbenzamida (DEET); etil-3 [acetil(butil)amino] propanoato (IR3535) e icaridina [5].

O DEET é conhecido como princípio ativo mais empregado em todo o mundo por ser considerado um repelente de amplo espectro. De fato, o mesmo é eficiente em repelência de carrapatos e várias espécies de mosquitos, incluindo o *Aedes aegypti*, responsável pela transmissão da dengue, chikungunya, febre amarela urbana e Zika [6].

É importante destacar que o tempo de ação dos repelentes é influenciado, entre outros fatores, pela concentração dos princípios ativos encontrados nesses produtos. Além disso, a literatura tem reportado que o DEET, quando utilizado em concentrações elevadas, tem provocado diversos efeitos tóxicos em adultos e em crianças [7]. A legislação brasileira restringe o uso de repelentes à base de DEET a crianças a partir de 2 a 12 anos num teor máximo de 10% [8]. Assim, a não conformidade do teor do princípio ativo declarado no rótulo pelo fabricante pode levar o consumidor ao uso de um produto ineficiente ou com maiores riscos toxicológicos [9].

Diante do exposto, torna-se de extrema importância o desenvolvimento de métodos rápidos, seguros e confiáveis, que possam determinar a concentração de DEET em amostras comerciais de repelentes, com aplicabilidade tanto para instituições privadas, atuando no controle de qualidade, quanto para fins de fiscalização (ANVISA).

Os métodos espectrométricos associados à calibração multivariada vêm sendo aplicados na determinação de princípios ativos em formulações cosméticas e farmacêuticas [10-15]. No que diz respeito aos métodos de calibração multivariada, destacam-se a regressão pelos mínimos quadrados parciais (PLS: Partial Least Squares) [16], regressão pelos mínimos quadrados parciais por intervalos (iPLS: Interval Partial Least Squares) [17] e regressão linear múltipla (MLR: Multiple Linear Regression) [18] com prévia seleção de variáveis pelo algoritmo das projeções sucessivas (SPA: Successive Projections Algorithm) [19], algoritmo genético (GA: Genetic Algorithm) [20] e formulação Stepwise (SW) [21].

Até o momento não foram encontrados trabalhos na literatura envolvendo o uso de técnicas espectroanalíticas acopladas aos métodos multivariados para a determinação de ingredientes ativos em repelentes comerciais. Logo, o presente trabalho propõe o desenvolvimento de uma metodologia para determinação de DEET em amostras de repelentes de insetos em loção usando espectrometria UV-VIS e calibração multivariada.

1.2 Objetivos

1.2.1 Objetivo geral

Desenvolver uma metodologia analítica simples, rápida e de baixo custo baseada no uso de espectrometria de absorção molecular UV-VIS e calibração multivariada para a determinação do teor de DEET em amostras comerciais de repelentes.

1.2.2 Objetivos específicos

- ✓ Obter um banco de dados espectrais de amostras (absorção molecular UV-VIS) em diferentes faixas de concentração de DEET;
- ✓ Realizar uma avaliação exploratória dos dados espectrais através da análise de componentes principais (PCA) com intuito de verificar a distribuição das amostras em termos de diferentes faixas de concentração de DEET;
- ✓ Desenvolver modelos de calibração multivariada baseados no uso da faixa espectral completa (PLS *full-spectrum*), de intervalos espectrais (*i*PLS) e de seleção de variáveis (SPA-MLR, GA-MLR e SW-MLR) para a determinação do teor DEET em repelentes;
- ✓ Avaliar os modelos obtidos através de diferentes parâmetros de desempenho (RMSEP, coeficientes de correlação (r) e determinação (R^2), LOD e LOQ).
- ✓ Realizar um estudo comparativo entre os diferentes modelos desenvolvidos baseados em testes estatísticos de significância.

Capítulo 2

Fundamentação

Teórica

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Repelentes

Repelentes são substâncias químicas voláteis que promovem uma barreira de proteção que se estende a cerca de 4 cm da pele, a qual induz o afastamento de artrópodes [6, 22]. São comercializados no mercado brasileiro em várias formas como aerossóis, cremes e loções em spray [22].

Uma substância considerada repelente deve agir na fase de vapor, logo substâncias com pressões de vapor muito baixas e altos pontos de ebulição não são capacitadas para agir como repelentes. Altas pressões de vapor são necessárias, porém quando muito elevadas há volatilização rápida, resultando em um curto período de tempo de proteção [22, 23].

Os repelentes estão inseridos no grupo técnico dos cosméticos visto que são constituídos de substâncias sintéticas, são de uso tópico e tem objetivo de proteção [24]. Devido à regulamentação dos repelentes pela ANVISA como cosméticos, estes acabam demandando uma avaliação menos rigorosa quando comparados com os medicamentos, apesar de sua importância no controle de doenças transmitidas por vetores [25].

As principais substâncias sintéticas usadas como princípios ativos de repelentes no mercado brasileiro são N,N-diethyl-3-metilbenzamida (DEET), etil-3 [acetil(butil)amino]propanoato (IR 3535) e 1-piperidinacarboxilato 2-(2-hidroxi-1-metilpropil)éster (Icaridina) [25] (**Figura 2.1**).

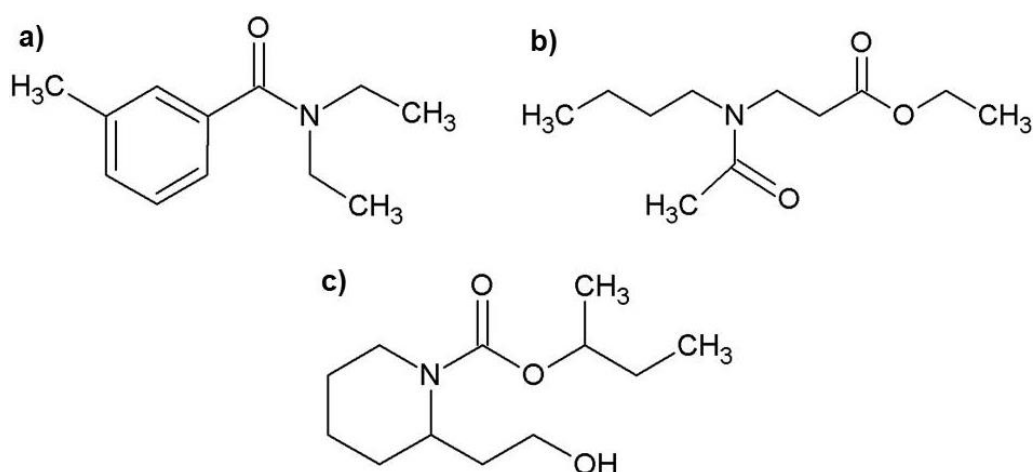


Figura 2.1- Moléculas de princípios ativos de repelentes no mercado brasileiro: a) DEET; b) IR3535; c) Icaridina

Entre os princípios ativos encontrados, o DEET (**Figura 2.1a**) é sem dúvida, o mais empregado por possuir baixo custo, longa duração e ser considerado um repelente de amplo espectro mais efetivo [6, 23]. A palavra DEET é um acrônimo provindo do nome N,N-dietil-meta-toluamida, posteriormente alterado pela IUPAC para N,N-dietil-3-metilbenzamida [23].

A determinação analítica do DEET tem sido realizada principalmente por meio de técnicas cromatográficas. O mesmo foi quantificado simultaneamente com oxibenzona e seus metabólitos por HPLC em amostras biológicas de plasma, urina e fitas de pele de porco, apresentando um LOD de $0,6 \text{ ng } 100 \text{ mL}^{-1}$ [26]. DEET e seus metabólitos também foram determinados por SPE-HPLC/MS on line em urina humana com LOD de $0,1 \text{ ng mL}^{-1}$ [27]. Já em urina e plasma de ratos, sua quantificação foi feita juntamente com piridostigmina (fármaco) e permetrina (inseticida) também por HPLC, com LOD de 50 ng mL^{-1} e LOQ de 150 ng mL^{-1} em plasma e urina respectivamente [28]. N,N-dietil-3-metilbenzamida, bem como a permetrina, foi determinado em plasma humano por CG/MS com LOQ de 1 ng mL^{-1} [29].

Este importante princípio ativo de repelentes é considerado um contaminante emergente (CECs: Contaminants of Emerging Concern) da categoria de Produtos Farmacêuticos e de Cuidados Pessoais (PPCPs) [30, 31]. Tem sido encontrado em diferentes corpos de água em concentrações variando entre 40 e 3.000 ng L^{-1} [30]. SPE-UHPLC/MS foi aplicada para sua detecção e de outros contaminantes orgânicos em águas, apresentando LOD e LOQ de $0,05 \text{ } \mu\text{g L}^{-1}$ e $0,1 \text{ } \mu\text{g L}^{-1}$, respectivamente. [32]. Sua degradação eletroquímica foi estudada em eletrodo de diamante dopado com boro (DDB) frente a outros eletrodos e diferentes parâmetros operacionais (temperatura, densidade de corrente e matriz). Foi alcançado 100% de degradação de DEET além da remoção do carbono orgânico total (TOC) com 1 mol L^{-1} de Na_2SO_4 (eletrólito) em águas residuais de efluente secundário da estação de tratamento [30].

Em pesquisa publicada recentemente, foi feita a quantificação de DEET juntamente com IR3535 por HPLC-DAD em repelentes na forma de loção, apresentando LOQ de $0,099 \text{ mg L}^{-1}$ para o DEET [33]. Apesar dos resultados satisfatórios obtidos e do simples preparo de amostra nesta metodologia, é sabido que a cromatografia líquida de alta eficiência é uma técnica de alto custo que

demanda tempo e mão-de-obra especializada. A metodologia em questão foi utilizada como referência no presente trabalho [33].

2.2 Espectrometria UV-VIS

A espectrometria de absorção molecular na região do UV-VIS usa de fonte de radiação eletromagnética na faixa de comprimento de onda entre 160 e 780 nm (**Figura 2.2**) [34]. O estímulo por esta fonte de radiação pode promover transições eletrônicas na molécula do composto em estudo, sendo limitadas pela presença de um par isolado ou de uma ligação π a partir da qual possa ocorrer a transição eletrônica [35, 36]. Trata-se de uma técnica já bem difundida e de grande aplicação na determinação de várias espécies orgânicas e inorgânicas, inclusive de importância farmacêutica, que possuam grupos absorventes [34].

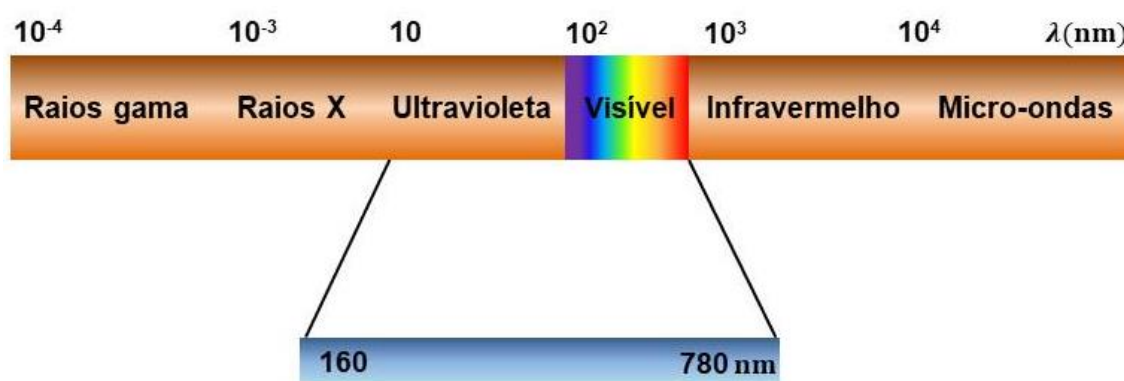


Figura 2.2- Espectro eletromagnético (adaptado de Skoog [34], 2002).

Para que haja a absorção de radiação UV-VIS é necessário que a molécula possibilite transições do tipo $n \rightarrow \pi^*$ e $\pi \rightarrow \pi^*$, o que ocorre com a existência de um grupo funcional insaturado para ceder os orbitais π (ex.: C=O, C=C) [34]. Quanto maior o grau de instauração na molécula absorvente, mais intensas são as bandas de absorção em maiores comprimentos de onda [36]. Estes grupos insaturados que promovem a absorção são denominados cromóforos [36]. Nestes centros de absorção a transição eletrônica ocorre juntamente a transições vibracionais e rotacionais, o que torna compreensível as bandas de absorção alargadas, as quais são características desta técnica [34].

A presença de cromóforos conjugados proporciona um deslocamento das absorções máximas para comprimentos de onda maiores [34]. Este fenômeno é chamado de efeito batocrômico, que pode ser explicado pela deslocalização dos

orbitais π e consequente diminuição do nível de energia do orbital π^* que perde parte de sua característica antiligante, o que diminui a energia entre o HOMO (orbital molecular ocupado de energia mais alta) e o LUMO (orbital molecular desocupado de energia mais baixa) [34, 36]. O efeito batocrômico também pode ocorrer em transições $\pi \rightarrow \pi^*$, desencadeado pelo aumento da polaridade do solvente, visto que os elétrons do solvente podem se rearranjar para promover uma estabilização do estado excitado da molécula, o que também diminui a energia entre os níveis eletrônicos da molécula [34, 36].

O solvente exerce um efeito oposto nos picos associados às transições $n \rightarrow \pi^*$, que são geralmente deslocados para menores comprimentos de onda com o aumento da polaridade do solvente [34]. A interação do solvente com o par solitário no estado fundamental baixa a energia do mesmo e consequentemente aumenta a energia requerida para promover o elétron para o orbital π^* , levando a um deslocamento para comprimentos de onda mais baixos. Este comportamento é denominado efeito hipsocrômico [36].

2.2.1 Instrumentação

Os espectrofotômetros UV-VIS consistem em geral de duas fontes, uma no UV e outra no visível, uma célula apropriada (cubeta) por onde passa a radiação e um detector para medir a quantidade de luz que passa pela célula. Atualmente são computadorizados e podem medir a absorbância em um determinado comprimento de onda ou podem varrer toda a faixa UV-VIS [36].

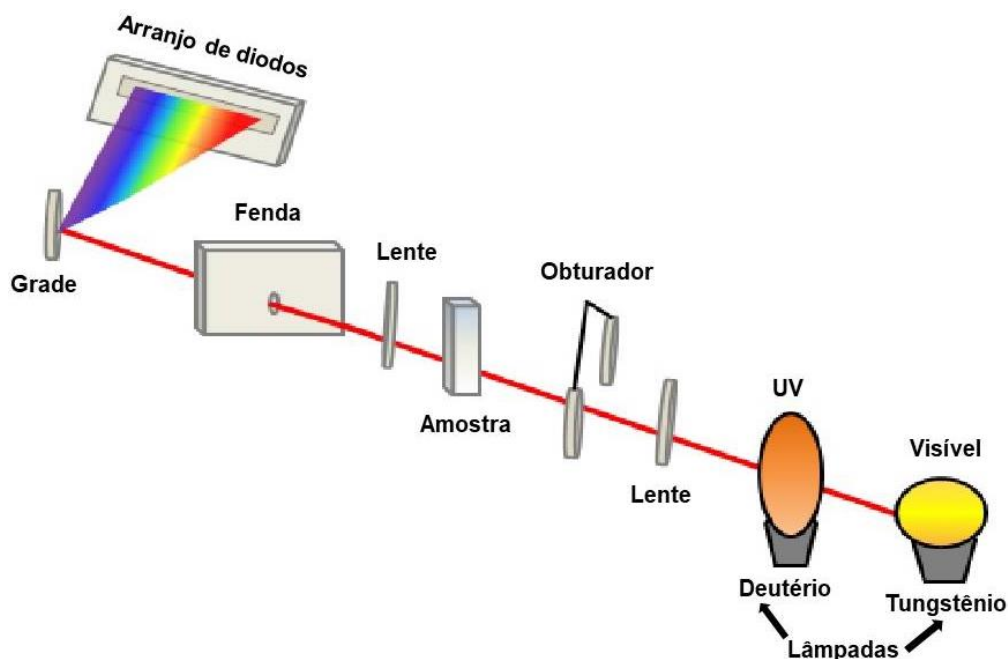


Figura 2.3- Diagrama geral de um espectrofotômetro de feixe único (adaptado de Skoog [34], 2002).

Na espectrometria de absorção molecular UV-VIS, uma radiação de potência P_0 incide sobre a amostra contida na cubeta e parte dessa radiação é absorvida pela amostra, resultando numa potência P menor que P_0 [34].

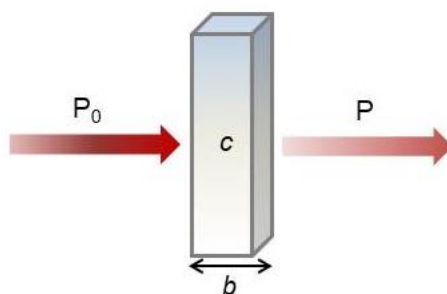


Figura 2.4- Representação da atuação de radiação eletromagnética em uma solução contendo c mol L^{-1} da substância absorvente e caminho óptico b (adaptado de Skoog [34], 2002).

Com o intuito de compensar as perdas por reflexão e espalhamento é feita a comparação da potência do feixe transmitido pela célula contendo a solução em análise (P) com a potência do feixe transmitido por uma célula contendo somente o solvente (P_0). A absorbância é obtida por meio da equação 2.1 [34, 36].

$$A = \text{Log} \frac{P_0}{P}$$

A absorvância está associada com a concentração do analito pela lei de Lambert-Beer e é diretamente proporcional ao caminho ótico b (comprimento em cm da cubeta contendo o analito) e a concentração, c , da molécula absorvente (mol L^{-1}). Esta importante relação está representada na **equação 2.2**, em que ϵ representa um parâmetro característico da molécula absorvente chamado de absortividade molar [34, 36].

$$A = \epsilon cb$$

2.2

Esta lei tem êxito em representar a absorção em meios com concentrações baixas de analito ($<0,01 \text{ mol L}^{-1}$), porém em concentrações elevadas há um desvio desta lei proporcionado pela menor distância entre as moléculas absorventes em solução, o que pode alterar a capacidade de absorção da molécula. Há então uma limitação ao uso de soluções diluídas [34].

A espectrometria UV-VIS tem sido utilizada na quantificação de vários analitos [10, 37-41], porém, as bandas alargadas obtidas na técnica, provenientes de sobreposições espectrais, tornam difícil a aplicação de métodos univariados por conter informações de outros componentes da fórmula além do analito. Em consequência deste comportamento são requeridas técnicas quimiométricas como recurso para a interpretação do espectro de forma mais ampla [10].

2.3 Quimiometria

A quimiometria é uma área que emprega técnicas estatísticas e matemáticas para o planejamento e otimização de experimentos, bem como para extrair o máximo de informações químicas relevantes por meio da análise de dados [42-44]. É considerada parte do campo da quimioinformática, que trata do uso de técnicas de informática para a resolução de problemas químicos, utilizando-se também de matemática e estatística [42]. Mais do que isto, é considerada uma ciência autônoma que traz uma linguagem formal para a quimioinformática [45].

O desenvolvimento da quimiometria tem proporcionado uma ampla aplicação da mesma, que já é bem difundida em diversos campos da química. Apesar de sua utilização em diversas áreas, sua aplicação se destaca na química analítica com um elevado número de artigos e atividades associadas [46]. O que é plausível visto que há uma grande demanda de análises da composição química,

muitas vezes em matrizes complexas, bem como a necessidade de determinações simultâneas de espécies químicas em misturas complexas [42, 44].

Entre os métodos multivariados encontrados na literatura, a calibração multivariada é uma das mais importantes metodologias aplicadas a problemas analíticos [47]. Antes mesmo da realização de um método quantitativo multivariado, a realização de uma análise exploratória dos dados pode ser uma estratégia interessante para a verificação da distribuição das amostras e variáveis no espaço multivariado. Neste contexto, destaca-se a análise de componentes principais (PCA), visto que é uma técnica eficiente para extrair e visualizar as principais informações de dados multivariados, como a formação de agrupamentos entre as amostras; identificação das variáveis que possuem maior contribuição para diferenciá-las, assim como detectar amostras anômalas [42, 48].

Em PCA, a dimensionalidade dos dados é reduzida pela transformação das variáveis em $\mathbf{X}$, altamente correlacionadas, em uma quantidade menor de variáveis latentes ortogonais entre si denominadas Componentes Principais (PCs) e que podem descrever, aproximadamente, toda a informação contida nos dados originais [48, 49]. As PCs são combinações lineares das variáveis originais e são geradas em ordem decrescente de máxima variância. Desta forma uma grande quantidade de informações é alcançada a partir de uma quantidade reduzida de componentes [48].

Matematicamente, a PCA é realizada a partir da decomposição da matriz $\mathbf{X}$ ($I \times J$) de dados em duas matrizes, uma de escores $\mathbf{T}$ e uma de pesos $\mathbf{L}$ (Equação 2.3) [49].

$$\mathbf{X} = \mathbf{TL}' + \mathbf{E} \quad 2.3$$

Para mais de uma PC:

$$\mathbf{X} = \mathbf{t}_1\mathbf{l}'_1 + \mathbf{t}_2\mathbf{l}'_2 + \mathbf{t}_a\mathbf{l}'_a + \mathbf{E}_a \quad 2.4$$

A matriz $\mathbf{T}$ é formada pelas coordenadas de cada amostra na nova variável (PC) enquanto $\mathbf{L}$ compreende a quantidade de informação que cada variável original contribuiu (peso) na formação desta PC. Os pesos correspondem aos cossenos dos ângulos entre a PC e os eixos das variáveis originais, $\mathbf{E}$ corresponde à matriz de

resíduos que contém toda a informação original não descrita e será usada para calcular a PC seguinte. [49, 50].

2.3.1 Calibração Multivariada

A calibração multivariada consiste no processo de associar múltiplas respostas instrumentais a uma ou mais propriedades desconhecidas da amostra, podendo ser aplicada em determinações simultâneas, em sistemas complexos mesmo na presença de alta correlação e diferenças espectrais [44, 51]. Busca-se uma correspondência entre um conjunto de medidas com outro conjunto adquirido por um método de referência. Esta relação permite que as medidas possam ser previstas a partir de medidas obtidas geralmente, de formas mais simples diminuindo assim o custo e o tempo de análise [43, 50].

Para estes fins são construídos modelos matemáticos que permitem prever uma propriedade y a partir das variáveis x ($x_1, x_2, \dots, x_n$) [42]. Vários modelos de calibração multivariada têm sido empregados como Regressão Linear Múltipla (MLR), Regressão por Componentes Principais (PCR) e Regressão em Mínimos Quadrados Parciais (PLS) [52].

Inicialmente é feita a modelagem para estabelecer uma relação entre $\mathbf{X}$ e $\mathbf{y}$. Na etapa de validação o valor das respostas das amostras de validação é então estimado a partir do modelo construído e deve-se calcular o erro desta predição [44, 47, 53]. Na validação por série de teste há a divisão das amostras de calibração em subgrupos de treinamento, para construção do modelo, e a partir do grupo de teste a Raiz Quadrada do Erro Médio Quadrático de Validação (RMSEV) é obtida [47].

A validação cruzada (validação interna) é a mais utilizada em quimiometria, nela apenas o conjunto de calibração é utilizado. Segmentos de amostras são retirados do conjunto, o modelo é então construído e em seguida são feitas as predições com as amostras que não participaram do modelo. Este procedimento de construção do modelo e predição é realizado repetidamente. O processo de validação é avaliado, entre outros parâmetros, pela Raiz Quadrada do Erro Médio Quadrático de Validação Cruzada (RMSECV) obtido a partir da **equação 2.5** [42, 47, 50].

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^I (y - \hat{y}_{\text{cv},i})^2}{I}}$$

2.5

Em que I é o número de amostras do conjunto de calibração, y e $\hat{y}_{cv}$ são respectivamente os valores medido e previsto para o parâmetro de interesse na i -ésima amostra de calibração.

Finalmente, o modelo de regressão obtido e otimizado pode ser utilizado para a predição de amostras externas ao modelo e sua habilidade é avaliada pelo RMSEP (Raiz Quadrada do Erro Médio Quadrático de Predição), calculado a partir da **equação 2.6** em que N_p corresponde ao número de amostras de predição [43, 50].

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{N_p} (y - \hat{y})^2}{N_p}} \quad 2.6$$

2.3.1.1 Regressão em Mínimos Quadrados Parciais (PLS)

A Regressão em Mínimos Quadrados Parciais é o método de regressão mais utilizado em se tratando de dados espectrais de grandes dimensões e com variáveis correlacionadas, pois o PLS é indiferente à colinearidade das variáveis regressoras e normalmente todas as variáveis espectrais podem ser empregadas (*full-spectrum*) [46, 50].

Neste método de regressão, há a transformação das variáveis $\mathbf{X}$ em variáveis latentes (VLs) ou fatores que são utilizadas para a regressão com a variável $\mathbf{y}$ [42]. As VLs relacionam $\mathbf{X}$ e $\mathbf{y}$ e são obtidas a partir da maximização da covariância entre os escores da matriz $\mathbf{X}$ ($\mathbf{t}$) e a variável dependente $\mathbf{y}$ [54, 50]. A covariância irá relacionar tanto a alta variância de $\mathbf{X}$ (estabilidade) quanto a alta correlação com $\mathbf{y}$, sendo a variável latente que prediz $\mathbf{y}$ uma combinação linear das variáveis originais [42].

Para a modelagem PLS aplicam-se as equações fundamentais 2.7 e 2.8.

$$\mathbf{X} = \mathbf{TP} + \mathbf{E} \quad 2.7$$

$$\mathbf{y} = \mathbf{Tq} + \mathbf{f} \quad 2.8$$

Inicialmente, o vetor de pesos $\mathbf{w}$ é calculado (**equação 2.9**) e os escores $\mathbf{t}$ ao longo deste eixo são obtidos a partir da **equação 2.10** [43, 50, 55].

$$\mathbf{w} = \mathbf{X}'\mathbf{y} \quad 2.9$$

$$\mathbf{t} = \frac{\mathbf{X}\mathbf{w}}{\sum \mathbf{w}^2} \quad 2.10$$

Os componentes em $\mathbf{X}$ são então regredidos em $\mathbf{t}$ a fim de obter o vetor de loadings $\mathbf{p}$ (Equação 2.11) [43, 55]. Os loadings em $\mathbf{y}$ ($\mathbf{q}$) são fornecidos partindo da regressão de $\mathbf{y}$ em $\mathbf{t}$ (Equação 2.12).

$$\mathbf{p} = \frac{\mathbf{t}'\mathbf{X}}{\sum \mathbf{t}^2} \quad 2.11$$

$$\mathbf{q} = \frac{\mathbf{y}'\mathbf{t}}{\sum \mathbf{t}^2} \quad 2.12$$

Os produtos $\mathbf{tp}$ e $\mathbf{tq}$ são subtraídos de $\mathbf{X}$ e $\mathbf{y}$ respectivamente (equações 2.13 e 2.14) para obter a matriz de dados residual pela subtração do efeito do novo fator da matriz de dados [43, 55].

$$\mathbf{X}_{\text{res}} = \mathbf{X} - \mathbf{tp} \quad 2.13$$

$$\mathbf{y}_{\text{res}} = \mathbf{y} - \mathbf{tq} \quad 2.14$$

Este procedimento se reinicia caso sejam necessárias mais variáveis latentes, diferindo apenas na substituição de $\mathbf{X}$ e $\mathbf{y}$ por seus resíduos antes de aplicar a equação 2.9. O processo continua até a quantidade de componentes desejada seja extraída, podendo ser, no máximo, igual à quantidade de variáveis espectrais.

A relação entre $\mathbf{X}$ e $\mathbf{y}$ é ilustrada na equação 2.15. Construído o modelo, a propriedade em $\mathbf{y}$ pode ser determinada pela equação 2.16 [55].

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{f} = \mathbf{T}\mathbf{q} + \mathbf{f} \quad 2.15$$

$$\mathbf{y} = \mathbf{x}\mathbf{b} \quad 2.16$$

O coeficiente de regressão b é obtido a partir da **equação 2.17** [55].

$$\mathbf{b} = \mathbf{W}(\mathbf{P}\mathbf{W})^{-1}\mathbf{q} \quad 2.17$$

2.3.1.2 Regressão em Mínimos Quadrados Parciais por Intervalo (iPLS)

A regressão em mínimos quadrados parciais por intervalo trata-se de uma regressão local de dados com a finalidade de melhorar a capacidade preditiva de modelos PLS, bem como auxiliar na interpretação da relação entre o modelo e a composição da amostra, a partir do desenvolvimento de modelos locais de PLS em subintervalos equidistantes do espectro. Baseando-se nestas subdivisões, o algoritmo viabiliza uma visão geral da relevância das informações, possibilitando a eliminação de interferências e o foco nas regiões importantes do espectro [17, 56].

No iPLS são construídos modelos PLS locais (intervalos de mesma largura) e seus desempenhos de previsão são comparados com o modelo global, principalmente com base no RMSECV, porém parâmetros como R^2 , inclinação e offset também podem ser avaliados [17]. O intervalo de variáveis é selecionado se o RMSECV for menor que RMSECV médio de todos os subintervalos [56].

O número de VLs para os intervalos geralmente é diferente daquele utilizado para o modelo PLS construído com espectro completo, pois é influenciado pela quantidade de informações e pelo ruído/interferência carregados pelo intervalo de variáveis em questão. Esta quantidade de VLs é selecionada de acordo com o menor erro produzido para o modelo [17].

2.3.1.3 Regressão Linear Múltipla (MLR)

A regressão linear múltipla MLR é a abordagem mais simples entre os métodos de regressão para a construção de modelos de calibração e para resolução de problemas lineares [57]. A relação estabelecida entre x (resposta instrumental) e y (parâmetro de interesse) por este método obedece a **equação 2.18**, em que $\mathbf{y}$ corresponde ao vetor do parâmetro a ser determinado, $\mathbf{X}$ é a matriz das variáveis regressoras, $\mathbf{b}$ é o vetor dos coeficientes de regressão e $\mathbf{e}$ corresponde ao vetor de resíduos [50].

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad 2.18$$

A **equação 2.19** mostra o modelo de regressão em que se fundamenta a Regressão Linear Múltipla [43, 50].

$$y_i = b_0 + \sum_{j=1}^J x_{ij}b_j + e_i \quad \text{para } i=1,2,3,\dots,I \quad 2.19$$

Os coeficientes de regressão, b_j s, são estimados pelo método dos mínimos quadrados, a partir da **equação 2.20**

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad 2.20$$

Nota-se que a matriz $\mathbf{X}'\mathbf{X}$ precisa ser invertida e a colinearidade entre as variáveis causa complicações [42, 50]. Outra restrição, é que o número amostras deve ser maior que o número de variáveis [45, 50]. Porém, como o que ocorre geralmente é a obtenção de uma grande quantidade de respostas e uma quantidade menor de amostras, pode-se utilizar de técnicas de seleção de variáveis para contornar estas restrições [51]. No presente trabalho foram utilizados alguns algoritmos já bem estabelecidos na literatura, a saber: o Stepwise (SW) [21]; o algoritmo genético (GA) [20]; o algoritmo das projeções sucessivas (SPA) [19] e o *i*PLS [17].

2.3.2 Algoritmos de seleção de variáveis

2.3.2.1 Stepwise

O SW seleciona o subconjunto ótimo de variáveis por busca exaustiva a partir do critério de melhoria dos termos de regressão [43, 46]. O algoritmo pode adicionar (*forward selection*) ou remover (*backward elimination*) uma variável por vez ou ainda, utilizar de uma mistura dos dois procedimentos em busca do melhor modelo. Este último procedimento é o mais utilizado e consiste de uma fase de inclusão seguida de uma exclusão [58, 45].

Na fase de inclusão, cada variável ainda não modelada é submetida a um teste F e se o valor obtido for maior que o valor para o critério de entrada, determinada variável é incluída no modelo. Na fase de exclusão é feito um teste F

com cada variável modelada e se o valor obtido for menor que o valor crítico para eliminação haverá a exclusão da variável do modelo e esta ficará novamente disponível para seleção. O processo de seleção cessa a partir do momento em que variáveis não podem mais ser adicionadas ou removidas pelos critérios de teste F [45].

2.3.2.2 Algoritmo Genético

O Algoritmo Genético é baseado na teoria evolutiva das espécies de Darwin, nele a seleção de variáveis é feita por meio de uma simulação computacional do que acontece no processo biológico da evolução das espécies [45, 59].

Um subconjunto de variáveis selecionadas pode ser representado por um vetor de componentes randômicos binários em que “0” refere-se a uma variável não selecionada e “1” uma selecionada. O número de variáveis deste vetor é representado pelo comprimento m e o vetor determina um subconjunto provável de variáveis. Este representaria um cromossomo e as m variáveis os genes. A população é composta de um conjunto de cromossomos e seu tamanho inicial, p , está geralmente entre 20-500 indivíduos [42, 45]. Deve-se associar a cada cromossomo um valor relacionado à aptidão, que se trata de uma medida de desempenho do modelo desenvolvido a partir deste subconjunto de variáveis [42, 59]. Geralmente esta resposta é analisada pela diferença entre o valor esperado e o valor previsto com a finalidade de encontrar o menor erro possível, a partir do qual a decisão de vida ou morte dos indivíduos é tomada. Também a partir deste erro, em alguns algoritmos, é estipulada a probabilidade de cruzamento de determinado cromossomo [59].

A tática do GA é fornecer as melhores populações, promovendo a evolução da população. A população seguinte é adquirida por ações de competição, cruzamento e mutação. Na competição, cromossomos piores são excluídos para dar lugar a novos cromossomos. No cruzamento, cromossomos são originados de pares de bons cromossomos, em busca de adquirir um cromossomo melhorado [42]. Já a mutação ocorre aleatoriamente com uma pequena percentagem (1-2%) dos genes ocasionando uma troca entre alguns “0” e “1” [42, 45]. Esta alteração pode agir de modo benigno ou maligno. Se houver a geração de indivíduos mais adaptados, as

características desenvolvidas serão transferidas para indivíduos das próximas gerações [59].

Após a formação da população inicial, as etapas acima são repetidas continuamente até que haja a satisfação do critério de parada, que pode ser um número máximo de gerações ou o alcance de valor alvo de resposta, que pode ser um erro mínimo, por exemplo [45, 59].

2.3.2.3 Algoritmo das Projeções Sucessivas

O SPA seleciona subconjuntos de variáveis com conteúdo infimamente redundante, a partir de uma sucessão de projeções compreendendo a coluna de variáveis instrumentais, para corrigir problemas de colinearidade, o que é particularmente interessante para MLR [19, 60]. O algoritmo inicia com uma variável e vai introduzindo variáveis com maior projeção ortogonal uma a uma até alcançar um número N (*forward selection*) [45, 19, 60, 61].

O algoritmo é processado em três etapas: a geração das cadeias de variáveis; a verificação das cadeias geradas e a eliminação de variáveis. Inicialmente, na geração de cadeias há a formação de subconjuntos de variáveis com a menor colinearidade possível entre os elementos. Na segunda fase, as cadeias formadas anteriormente são analisadas com base na capacidade preditiva do modelo, que pode ser avaliada pelo RMSEV ou RMSECV, dependendo do tipo de validação utilizada [61]. Finalmente, na etapa de eliminação, há a exclusão de variáveis com contribuição insignificante à capacidade de predição do modelo MLR. Para este fim, é feita a comparação do RMSE (Raiz Quadrada do Erro Médio Quadrático) de cada variável com o RMSE mínimo alcançado por meio de um teste F . Feito isto, são selecionadas variáveis cujos valores de RMSE não sejam significativamente maiores que o RMSE mínimo [60, 61].

2.3.3 Seleção de amostras

Equações adequadas em calibração multivariada somente são obtidas com bons modelos estatísticos, procedimentos de estimação eficazes e dados de calibração satisfatórios. Geralmente, a capacidade preditiva do modelo melhora à medida que a quantidade de amostras aumenta. No entanto, a qualidade dos dados de calibração não é determinada somente pela quantidade de amostras, a forma de

seleção também é relevante já que a capacidade de detectar problemas do modelo é um aspecto importante [43]. Logo, a seleção de amostras é de grande importância não só para a seleção de um conjunto de calibração representativo, mas também para a validação e a predição [62].

A amostragem randômica (RS) é bastante utilizada para a seleção de amostras de calibração devido à simplicidade da técnica ao fato de que um conjunto de dados extraídos aleatoriamente de um conjunto maior segue a distribuição estatística do todo. Contudo, a representatividade do conjunto não é garantida e nem a inclusão das amostras nos limites do conjunto é assegurada [62, 63]. Dito isto, há uma necessidade de métodos alternativos ao RS, como os algoritmos Kennard-Stone (KS) [64] e o de particionamento de amostras baseados nas distância de X-y (SPXY: Sample Set Partitioning based on Joint X-y Distances) [62].

2.3.3.1 Particionamento de amostras baseado nas distâncias X-y

O algoritmo KS clássico é utilizado para selecionar um subconjunto representativo de amostras uniformemente, de modo que as distâncias euclidianas entre os vetores de resposta instrumental sejam maximizadas [62-64]. O algoritmo trabalha em etapas em que as seleções são feitas em regiões espaciais distantes das amostras selecionadas anteriormente [64]. Para esta finalidade são adotadas as distâncias euclidianas $d_x(p, q)$ entre os vetores $\mathbf{x}$ de cada par pq conforme **equação 2.21** [62].

$$d_x(p, q) = \sqrt{\sum_{j=1}^J [x_p(j) - x_q(j)]^2} \quad p, q \in [1, N] \quad 2.21$$

Em que $x_p(j)$ e $x_q(j)$ são respostas instrumentais no j -ésimo comprimento de onda das amostras p e q . A princípio é tomado o par de amostras p_1 e p_2 com maior distância d_x . Posteriormente, o algoritmo seleciona amostras que possuam a maior distância com relação a qualquer uma das amostras selecionadas anteriormente. Este processo se repete até atingir a quantidade de amostras designada pelo analista [62]. Desta forma, o algoritmo propicia a seleção de amostras de todo espaço de dados com a vantagem de evitar extrapolações [45, 63].

O fato da variável y não ser levada em consideração pode ter como consequência que valores de y não contemplem a faixa de calibração inteira [45].

Para considerar as distâncias em $\mathbf{X}$ e $\mathbf{y}$, uma versão modificada do KS, denominada SPXY, foi proposta por Galvão et al. [62, 65].

Neste algoritmo, a distância definida na **equação 2.21** é acrescida pela distância em $\mathbf{y}$ representada pela **equação 2.22** [62].

$$d_y(p, q) = \sqrt{(y_p - y_q)^2} = |y_p - y_q|; \quad p, q \in [1, N] \quad 2.22$$

As distâncias $d_x(p, q)$ e $d_y(p, q)$ são divididas pelos seus valores máximos no conjunto de dados, a fim de atribuir importância equivalente para distribuição das amostras em $\mathbf{X}$ e $\mathbf{y}$. Logo, as distâncias X-y são normalizadas conforme a **equação 2.23** [62].

$$d_{xy}(p, q) = \frac{d_x(p, q)}{\max_{p, q \in [1, N]} d_x(p, q)} + \frac{d_y(p, q)}{\max_{p, q \in [1, N]} d_y(p, q)} \quad p, q \in [1, N] \quad 2.23$$

Capítulo 3

Metodologia

3 METODOLOGIA

3.1 Aquisição e preparo de amostras

Para este trabalho foram utilizadas 107 amostras comerciais de repelentes em loção de 10 diferentes fabricantes e lotes. As amostras foram adquiridas em várias regiões do Brasil e suas concentrações variaram entre 5% e 30% m m⁻¹ do princípio ativo DEET. Dentre os excipientes destes compostos, os mais comuns são álcool, água e fragrâncias (*parfum*) [66]

As amostras comerciais de repelentes foram acondicionadas em temperatura ambiente e posteriormente diluídas cerca de 60.000 vezes em solução 10% de acetonitrila (CH₃CN) em água ultrapura (ACN/H₂O). Inicialmente foi medido um volume de 25 µL de repelente comercial, aferido para 10 mL com a solução ACN/H₂O (10% v v⁻¹) e em seguida agitada em vortex por cerca de 10s. Após a homogeneização, uma alíquota de 54 µL desta primeira diluição foi diluída a 8 mL e homogeneizada como citado anteriormente (**Figura 3.1**). Para aquisição das amostras que representam o controle negativo ou prova em branco, foi utilizada apenas a solução de ACN/H₂O 10% (v v⁻¹) em água. A **figura 3.1** detalha o esquema adotado para o preparo das amostras utilizadas neste trabalho.



Figura 3.1- Esquema do preparo de amostras

3.2 Análise de referência

Os valores de referência foram obtidos em um cromatógrafo Ultimate 3000 Dionex constituído de bomba quaternária, injetor manual com um loop fixo de 20 µL e um detector de arranjo de diodos UV-VIS (HPLC-DAD). Foi empregada uma coluna cromatográfica Dionex Acclaim® 120 C18 de 50 mm x 2,1 mm, tamanho de partículas de 3 µm e 300 Å de tamanho de poro. As amostras foram diluídas 20.000 vezes em solução de ACN/H₂O (10% v v⁻¹) e eluídas em triplicatas [33]

3.3 Aquisição dos espectros

As amostras foram diluídas conforme esquema da **figura 3.1** e em seguida foram conduzidas as medidas para obtenção dos espectros de forma aleatória e em triplicatas na faixa de 190-1.100 nm. Foi utilizado para as medidas um espectrofotômetro de feixe único Hewlett-Packard modelo 8453 com resolução de 1 nm, equipado de cubeta de quartzo com 10mm de caminho óptico.

3.4 Procedimento quimiométrico e software

Para a determinação do teor de DEET em amostras comerciais de repelentes foram desenvolvidos modelos de calibração multivariada utilizando os espectros de absorção molecular no UV-VIS e os valores de referência adquiridos por HPLC.

Inicialmente, uma avaliação espectral por meio de uma inspeção visual foi realizada com intuito de descartar regiões não informativas e reter aquelas de maior qualidade de sinal. Após a escolha da região de trabalho, a correção da linha de base pelo método baseline offset foi empregada como técnica de pré-processamento dos dados.

Para a construção dos modelos, o conjunto de dados foi dividido em subconjuntos de calibração (64) e de predição (43) aplicando o algoritmo SPXY [62]. O conjunto de calibração foi empregado em um procedimento de validação cruzada (*leave-one-out*), enquanto o conjunto de predição foi utilizado para avaliar a capacidade preditiva dos modelos desenvolvidos.

O limite de detecção (LOD) e o limite de quantificação (LOQ) foram determinados a partir dos desvios padrão obtidos através da predição de 30 medidas do branco nos respectivos modelos [67]. Os modelos PLS, iPLS, SW-MLR, SPA-MLR e GA-MLR construídos tiveram seus desempenhos avaliados de acordo com os valores de RMSEP, coeficientes de correlação (r) e determinação (R^2), LOD, LOQ e número de variáveis/fatores selecionados.

A significância estatística das diferenças entre os valores de RMSEP foi avaliada usando teste F em um nível de confiança de 95%. Os valores de F foram calculados pela proporção entre os quadrados dos maiores e menores valores de RMSEP (**Equação 3.1**). Essa relação foi comparada com o valor crítico F ($n-1$, $n-1$, 0,95) [68- 70]. No presente trabalho o n se iguala ao número de amostras no conjunto de predição ($n = 43$).

$$F = \frac{(RMSEP_1)^2}{(RMSEP_2)^2} \quad 3.1$$

O teste t -pareado, ao nível de confiança de 95%, foi empregado para avaliar a existência de diferenças significativas entre o método proposto e o método de referência. Os valores de t -calculado (**Equação 3.2**) foram comparados com os valores de t -crítico (0,95; n) tabelado.

$$t = \frac{\bar{d}}{s_d/\sqrt{N}} \quad 3.2$$

Em que $\bar{d}$ é a diferença média entre os valores de referência e os estimados pelo modelo, S_d é o desvio padrão da diferença e N o número de amostras utilizadas para a predição.

O pré-processamento dos dados e a construção dos modelos PLS e MLR foram realizados utilizando o software The Unscrambler® 9.7 (CAMO S.A.). O SPXY e os algoritmos de seleção de variáveis (SPA, Stepwise, GA e *i*PLS) foram realizados com Matlab® R2010a 7.10.0.499.

Capítulo 4

Resultados

e

Discussão

4 RESULTADOS E DISCUSSÃO

4.1 Análise espectral

A **Figura 4.1** corresponde aos espectros UV-VIS médios originais (sem pré-processamento) na faixa entre 190 nm e 1.100 nm, referente às 107 amostras utilizadas para calibração e predição. Analisando os espectros pode-se observar uma ampla região (280-950 nm) não informativa e outra com elevado ruído (saturação do detector), na região superior a 950 nm. Estas duas regiões foram removidas e o tratamento quimiométrico dos dados foi realizado na faixa espectral de 190 a 279 nm, como apresentado na **figura 4.2a**.

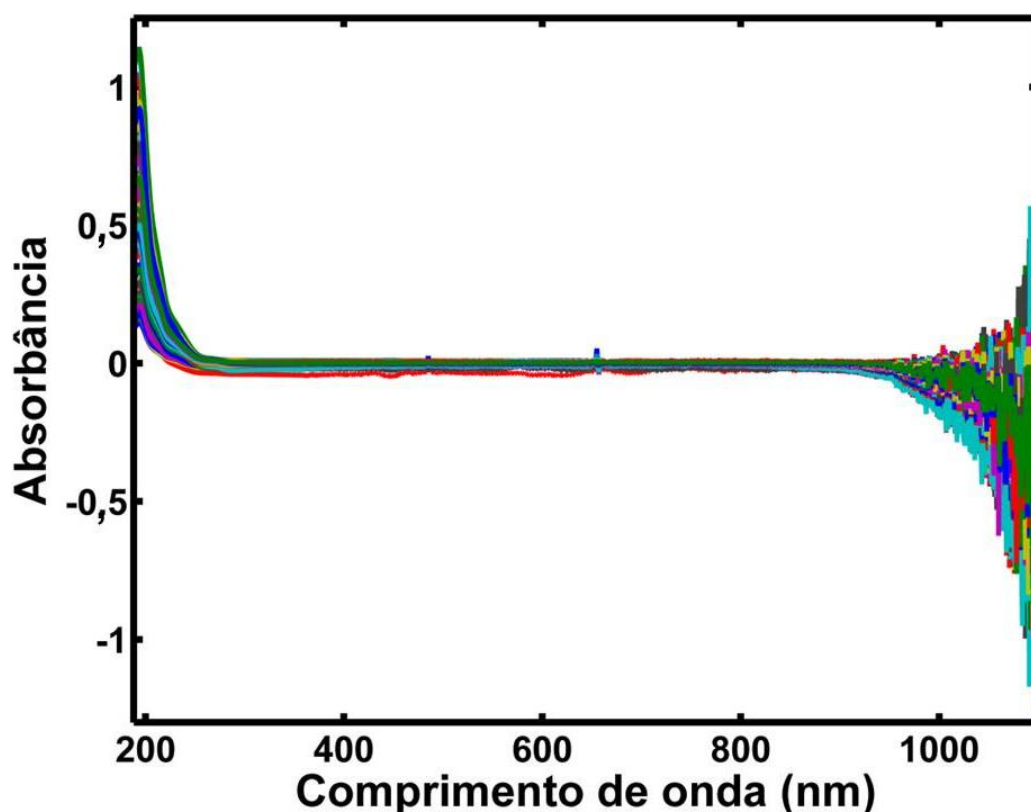


Figura 4.1- Espectros UV-VIS originais na faixa de 190 a 1100 nm obtidos a partir de 107 amostras.

Os espectros apresentam deslocamento de linha de base provavelmente devido a própria variabilidade na composição das amostras. Uma correção da linha de base (*baseline offset*) dos espectros foi realizada no domínio dos dados originais (**Figura 4.2a**) e os espectros processados são apresentados na **figura 4.2b**.

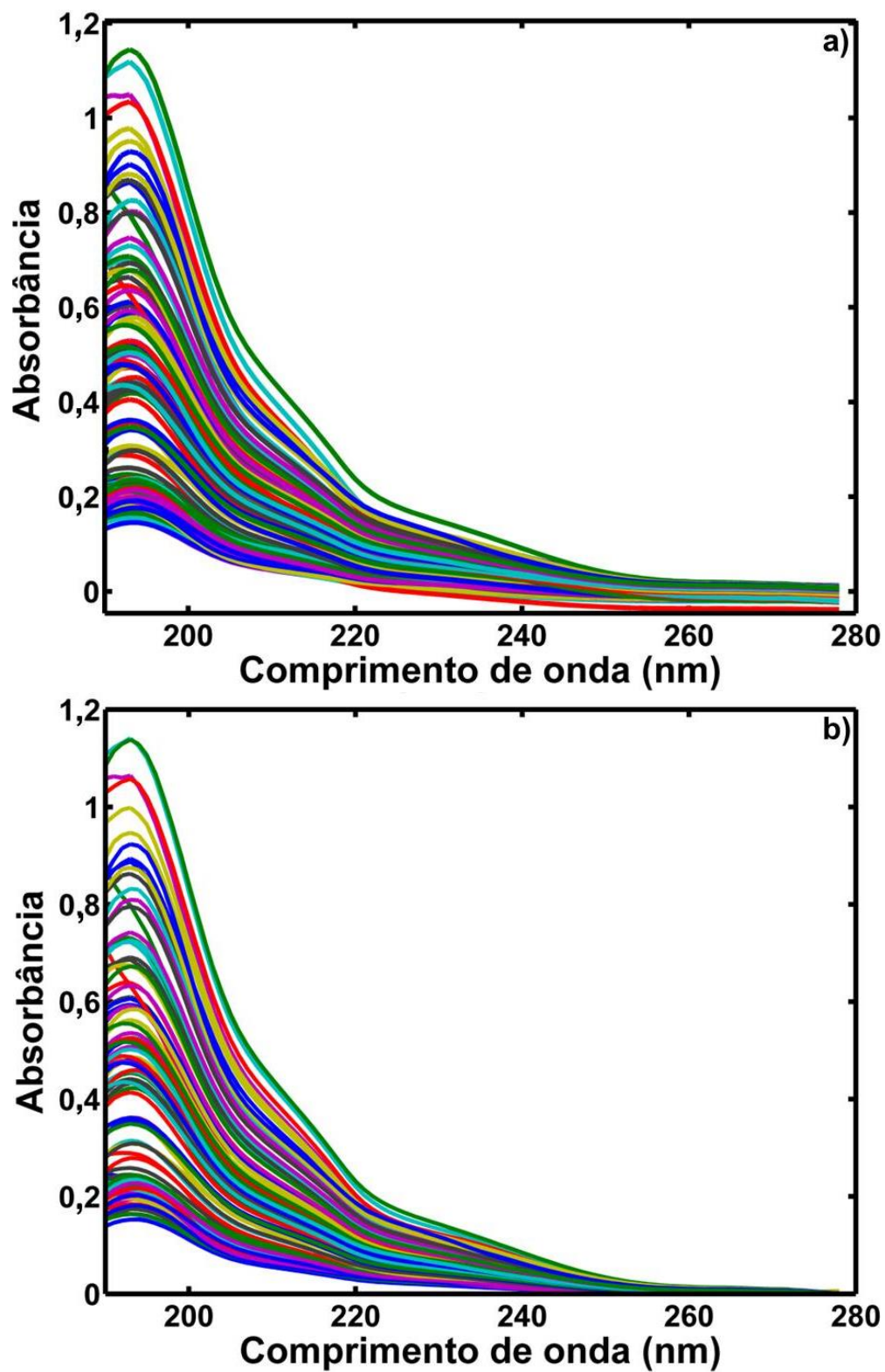


Figura 4.2- Espectros UV-VIS obtidos a partir de 107 amostras: a) originais com seleção a priori de variáveis; b) pré-processados.

A região adotada está relacionada com as transições eletrônicas $n \rightarrow \pi^*$ e $\pi \rightarrow \pi^*$ possibilitadas pela molécula de DEET. Em particular, a banda em torno de 193 nm está associada à absorção dos cromóforos C=O e C=C na molécula deste princípio ativo (**Figura 4.3**) [36].

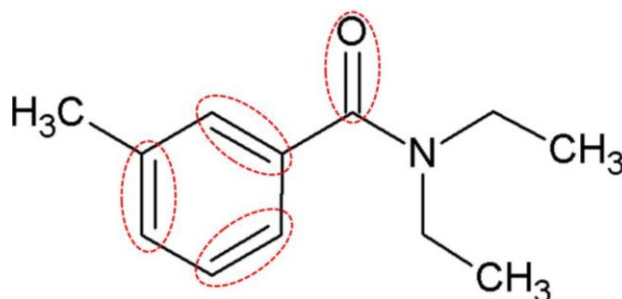


Figura 4.3- Molécula de N,N-Dietil-3- metilbenzamida (DEET) com centros cromóforos em destaque.

A **figura 4.4a** mostra os cromatogramas de padrões usados na construção da curva de calibração do método de referência. Em verde está representado o pico do IR3535 e em azul o pico do DEET. O espectro do padrão de DEET (**Figura 4.4b**), também obtido pelo método de referência, tem aspecto bastante semelhante aos obtidos por UV-VIS (**Figura 4.2**). Isto corrobora a utilização de um método baseado em absorção molecular UV-VIS como uma boa alternativa à HPLC, uma vez que o método de referência apresenta alto custo e moderada complexidade.

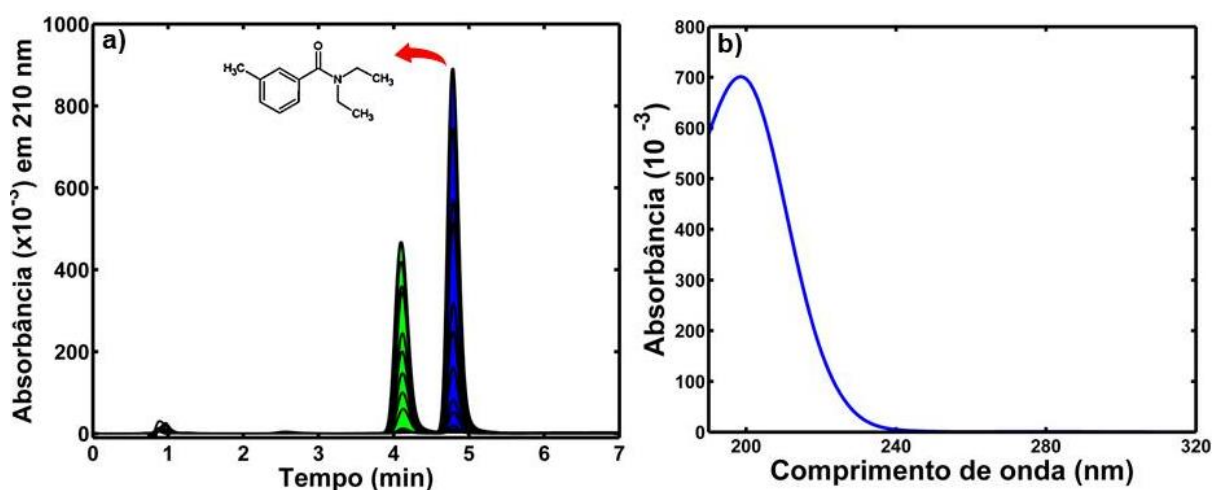


Figura 4.4- a) Cromatogramas dos padrões de IR3535 (verde) e DEET (Azul) usados na curva de calibração para o método de referência (HPLC); b) Espectro do padrão DEET na concentração de 8 mg Kg⁻¹. (adaptado de Vilar [33], 2018).

4.2 Análise exploratória

A fim de realizar uma avaliação exploratória do comportamento dos dados, uma PCA (Análise por Componentes Principais) foi aplicada aos espectros UV-VIS com intuito de verificar tendências das amostras de repelentes com respeito às faixas de concentrações de DEET encontradas nas amostras dos principais fabricantes.

Na **figura 4.5** visualiza-se o gráfico dos escores obtido pela aplicação da PCA aos dados pré-processados. A PC1 e PC2 explicam 99,8% e 0,15% da variância dos dados, respectivamente. A PC1 é responsável por distinguir as diferentes faixas de concentração do DEET nas amostras analisadas. De fato, em valores mais positivos de escores em PC1 encontram-se as amostras de maior teor do princípio ativo.

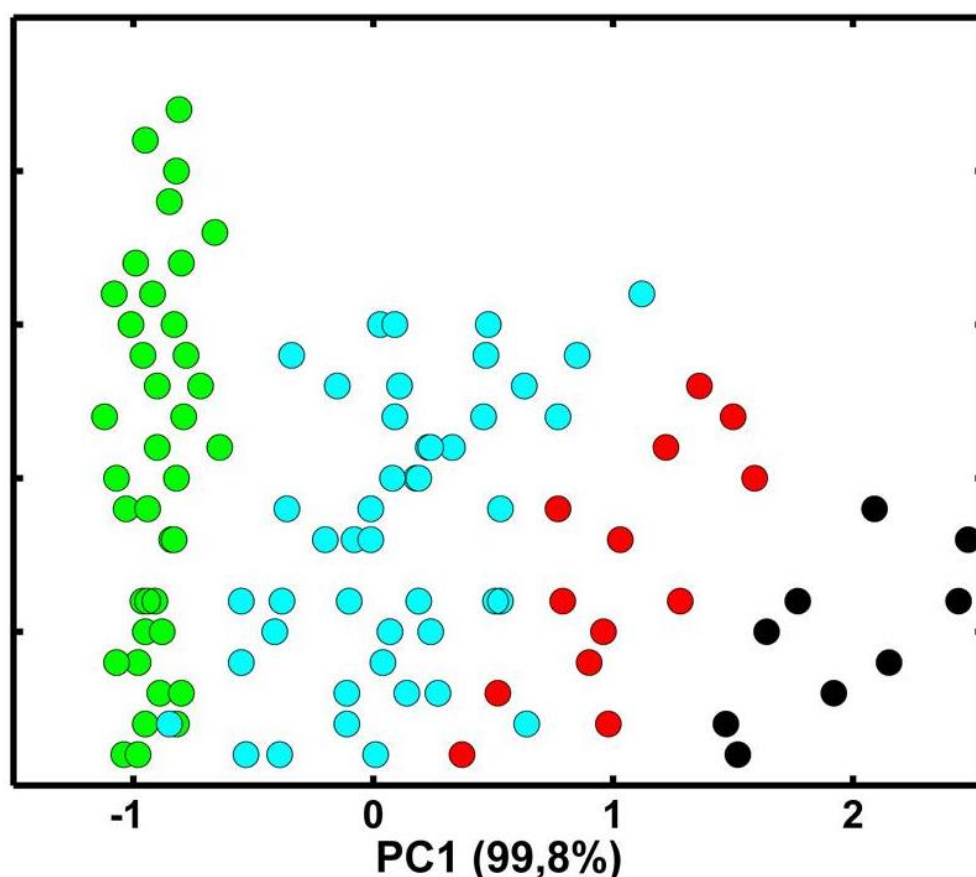


Figura 4.5- Gráfico dos escores de PC1 para as 107 amostras de repelentes rotuladas de acordo com a faixa de concentração de DEET ($\% \text{ m m}^{-1}$): (●) menor 7%; (●) maior ou igual a 7% e menor que 15%; (●) maior ou igual a 15% e menor que 20%; (●) maior ou igual a 20%.

No gráfico dos pesos obtido pela PCA (**Figura 4.6**), é possível observar que os valores de pesos são positivos em PC1, assemelhando-se ao espectro UV-VIS das amostras de repelentes. Em especial, valores mais positivos de PC1 encontram-se em comprimentos de onda em torno de 190-198 nm, região característica do DEET.

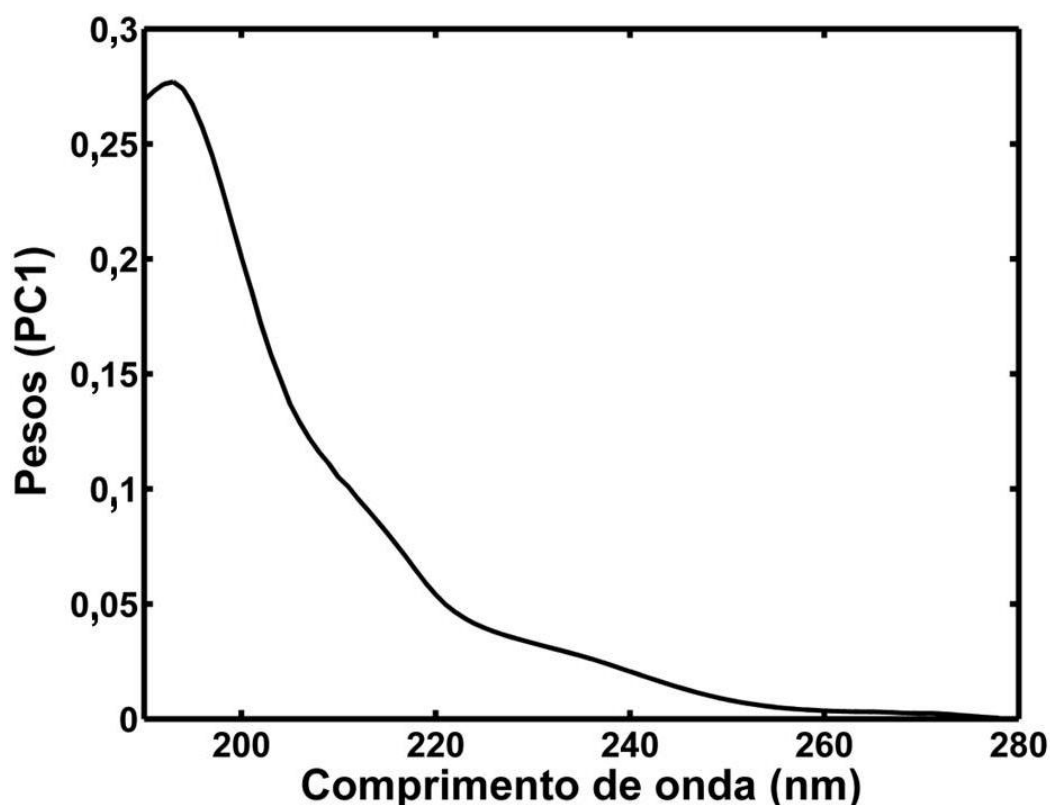


Figura 4.6- Gráfico de pesos de PC1.

4.3 Desenvolvimento dos modelos de calibração

4.3.1 Modelo PLS

Foi construído o modelo de regressão multivariada por mínimos quadrados parciais (PLS) contendo como conjunto de calibração amostras de repelentes no intervalo de 5% a 30% m m^{-1} de DEET. Apenas uma variável latente (VL) explicou 100% e 94% da variância de **X** e **y**, respectivamente. A **figura 4.7a** apresenta o gráfico de valores preditos versus valores de referência obtidos a partir do modelo PLS. Observa-se que as amostras estão distribuídas randomicamente em torno da bissetriz ao longo de toda a extensão de valores **y**, evidenciando a inexistência de erros sistemáticos. A exatidão do modelo foi expressa em termos de RMSECV

(1,42% m m⁻¹) e RMSEP (0,92% m m⁻¹). Os coeficientes de correlação (r) e de determinação (R^2) foram de 0,97 e 0,94, respectivamente.

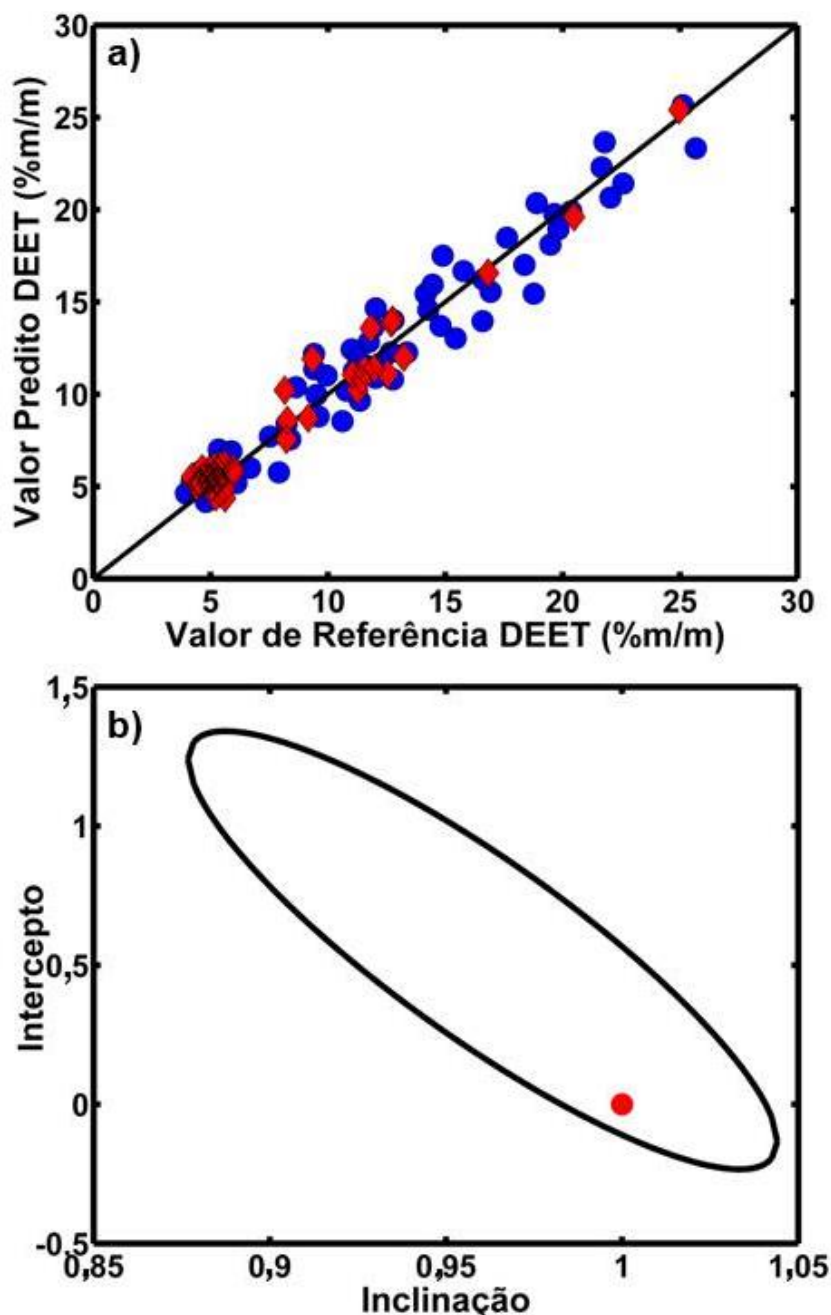


Figura 4.7- Gráficos obtidos a partir de modelo PLS: a) valores preditos versus valores de referência para as amostras de calibração (●) e predição (♦); b) Elipse de confiança.

Como medida da exatidão dos resultados do método proposto frente ao usado como referência, fez-se uma comparação dos valores obtidos para a inclinação e o intercepto de uma reta ajustada entre os valores de referência e os estimados para cada modelo por meio de uma elipse de confiança [71]. Na **figura**

4.7b está representada a elipse de confiança para o intercepto e inclinação da regressão entre as concentrações de DEET previstas pelo modelo e os valores de referência determinados por HPLC-DAD. A elipse de confiança contém o ponto ideal (1,0) indicando que os valores determinados pelo modelo PLS não apresentam diferença significativa dos valores determinados por HPLC com 95% de confiança [71, 72].

4.3.2 Modelo *i*PLS

A **figura 4.8** ilustra o procedimento de seleção de intervalo de variáveis pelo *i*PLS. Os comprimentos de onda no lado esquerdo de cada barra de erro no eixo horizontal correspondem aos limites dos 10 intervalos espectrais equidistantes considerados. Cada barra vertical indica o valor de RMSECV obtido pelo modelo PLS local para o respectivo intervalo e na parte inferior das barras é indicado o número de variáveis latentes do modelo em questão. A linha horizontal em vermelho equivale ao valor de RMSECV para o modelo PLS *full-spectrum*.

O critério de seleção baseado nos valores de RMSECV permite observar que o intervalo com o menor valor de RMSECV corresponde ao primeiro intervalo (com apenas 1 VL). Esta região escolhida pelo *i*PLS é, de fato, apropriada por corresponder à banda característica do DEET. O valor de RMSECV deste intervalo não apresentou melhora significativa quando comparado ao modelo *full-spectrum*, porém houve uma melhoria em termos de outros parâmetros avaliados, que serão discutidos posteriormente. Além disso, é possível observar uma piora no desempenho dos modelos em intervalos com comprimentos de onda superior a 208 nm. Isto é particularmente constatado quando se observam regiões com baixa relação sinal ruído (9º e 10º intervalos).

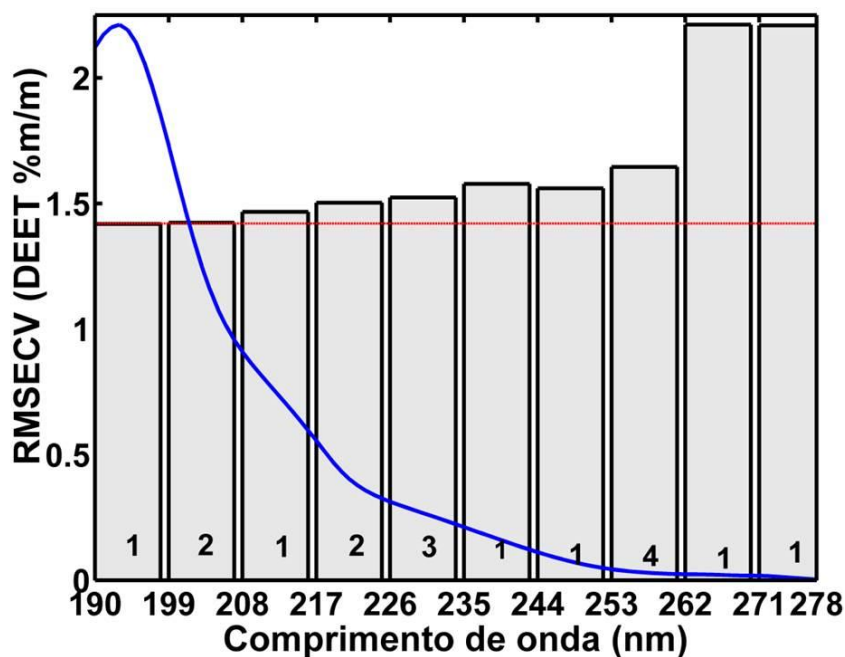


Figura 4.8- Resultados do *i*PLS para o DEET (%m m⁻¹) para os 10 intervalos (barras) e para o modelo *full-spectrum* (---).

Na **figura 4.9a** está disposto o gráfico de valores preditos versus valores de referência obtidos a partir do modelo de Regressão em Mínimos Quadrados Parciais por Intervalo, *i*PLS. O modelo apresentou valores baixos de RMSECV (1,42% m m⁻¹) e RMSEP (0,93% m m⁻¹), este último indicando uma boa capacidade de predição do modelo. Os coeficientes de correlação (*r*) e determinação (*R*²) foram 0,97 e 0,94, respectivamente. De modo geral, o gráfico da **figura 4.9a** apresentou comportamento similar aquele desenvolvido com o modelo *full-spectrum*.

A **Figura 4.9b** mostra a elipse de confiança contendo o ponto ideal, demonstrando que os valores estimados pelo modelo *i*PLS e os valores de referência não apresentam diferença significativa com 95% de confiança.

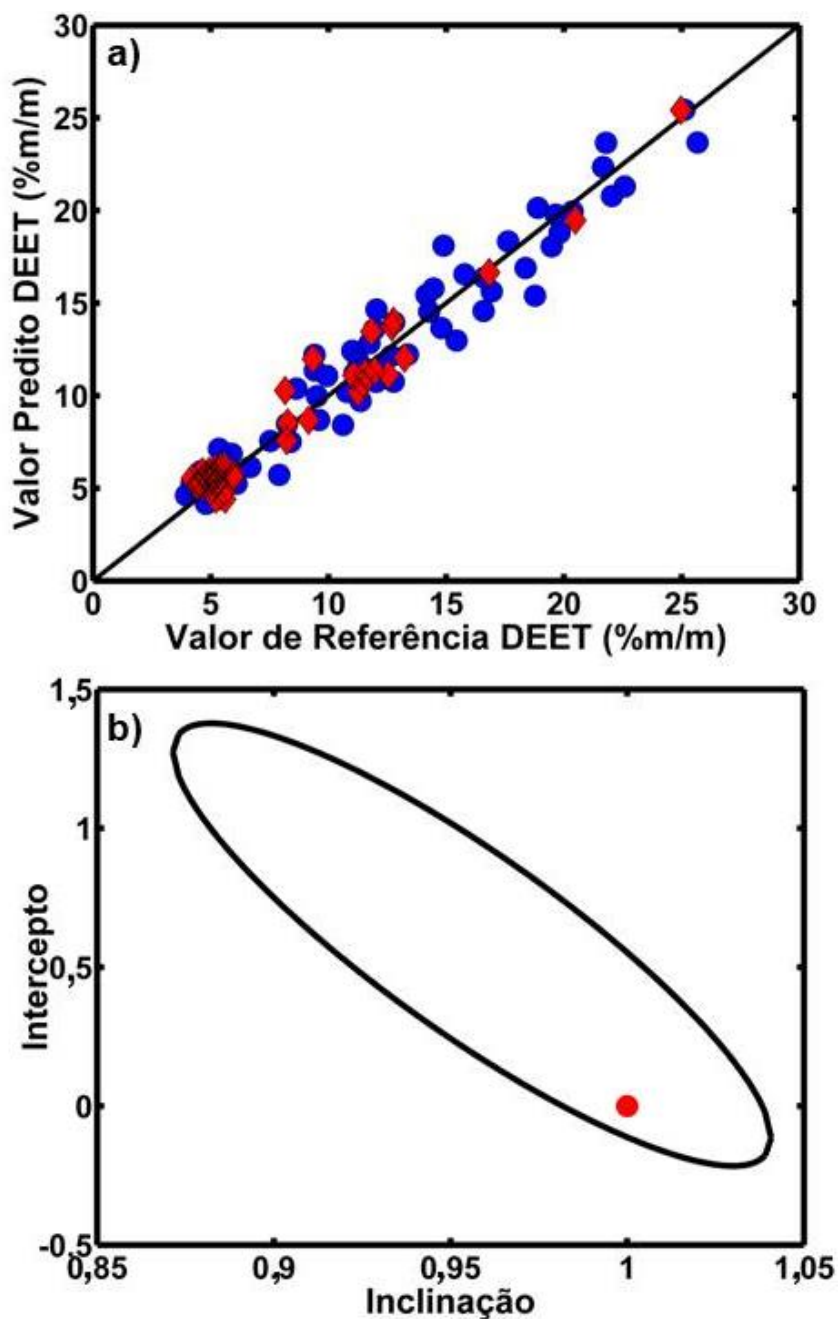


Figura 4.9- Gráficos obtidos a partir de modelo *i*PLS: a) valores preditos versus valores de referência para as amostras de calibração (●) e predição (◆); b) Elipse de confiança

4.3.3 Modelos SPA-MLR, SW-MLR e GA-MLR

A fim de diminuir a dimensionalidade dos dados e encontrar variáveis mais relevantes para os modelos MLR, as técnicas de seleção de variáveis (seções 2.3.2.1 a 2.3.2.3) foram aplicadas. Na **figura 4.10** estão representados os gráficos das variáveis selecionadas empregadas na construção dos modelos SPA-MLR, SW-MLR e GA-MLR.

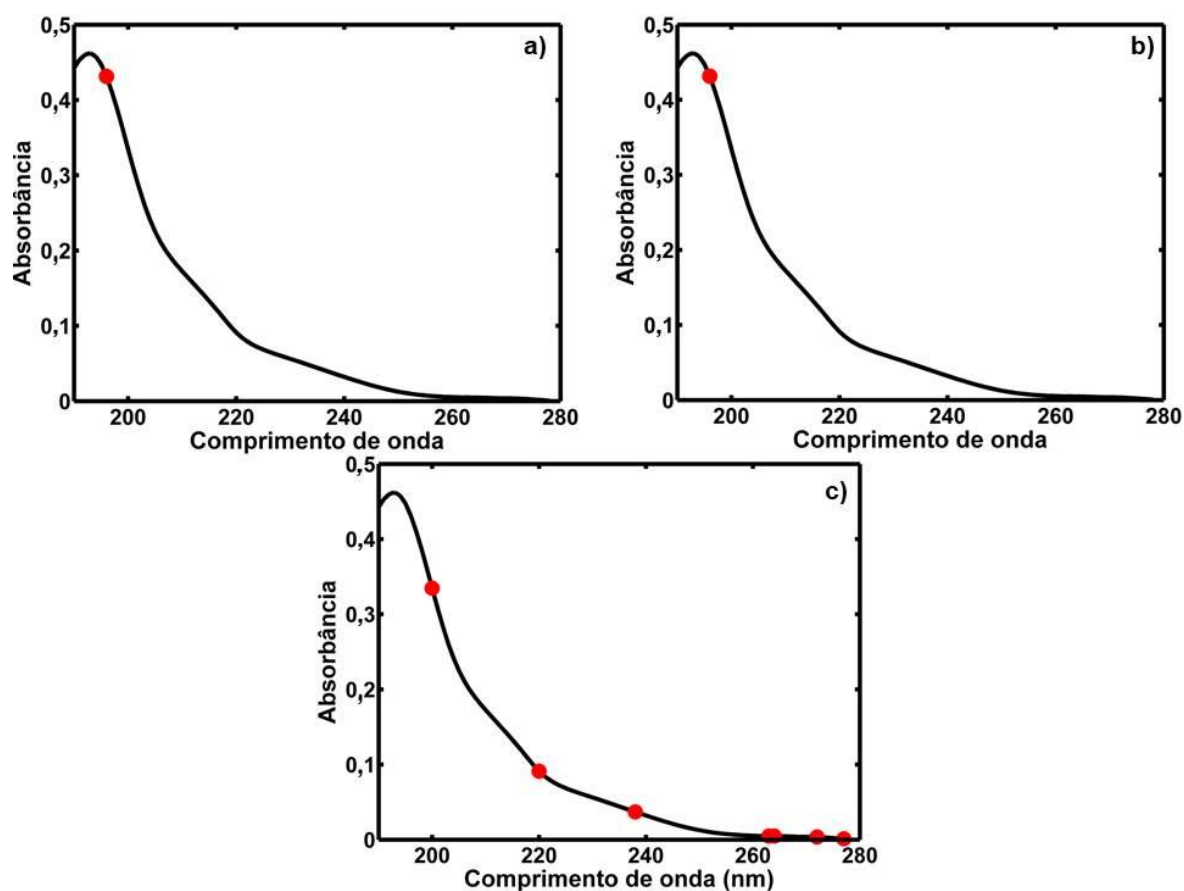


Figura 4.10- Espectros UV-VIS médios com as variáveis selecionadas (●): a) SPA, b) SW, c) GA.

Observa-se que tanto o SPA, quanto o SW selecionaram a mesma variável (**Figura 4.10a** e **Figura 4.10b**), correspondente ao comprimento de onda 196 nm, próximo ao ponto de maior intensidade do espectro (193 nm). Já o GA selecionou um total de sete variáveis (**Figura 4.10c**) distribuídas pelo espectro, incluindo pontos em regiões de baixa relação sinal ruído (260 a 280 nm).

Como os algoritmos SPA-MLR e SW-MLR selecionaram a mesma variável, os modelos desenvolvidos com estes dois algoritmos são idênticos. (**Figura 4.11a**). As amostras apresentam distribuição aleatória em torno da bissetriz em toda extensão de y e de modo não tendencioso. Os valores de RMSECV e RMSEP foram de 1,40% m m⁻¹ e 0,93% m m⁻¹, respectivamente. O coeficiente de correlação (r) apresentou valor de 0,97 e o coeficiente de determinação (R^2) um valor de 0,94. A **figura 4.11c** representa a elipse de confiança para os modelos. A inclusão do ponto ideal (1,0) para a inclinação e o intercepto evidencia que os valores estimados pelos modelos não apresentam diferença significativa com 95% de confiança.

Diferentemente dos outros modelos anteriormente apresentados, o GA-MLR resultou em uma maior dispersão das amostras ao longo da bissetriz. Algumas amostras, inclusive, apresentaram uma elevada diferença entre os valores verdadeiros e estimados pelo modelo (**Figura 4.11b**). Este fato pode estar associado à seleção de variáveis realizada a partir do GA, que indicou alguns comprimentos de onda na região com menor informação química do espectro e, conseqüentemente, com menor relação sinal-ruído. As amostras em questão revelaram comportamentos anômalos apenas neste modelo.

O modelo GA-MLR apresentou, no geral, parâmetros menos satisfatórios que os avaliados anteriormente. Uma queda nos coeficientes de correlação ($r=0,91$) e determinação ($R^2=0,81$) indicou uma baixa linearidade entre os valores de referência e os estimados pelo modelo GA-MLR, bem como uma baixa proporção da variável **y** explicada pela regressão. Os valores de RMSECV e RMSEP foram de 2,42% m m^{-1} e 0,88% m m^{-1} , respectivamente, demonstrando melhora na capacidade preditiva do modelo, apesar do valor elevado de RMSECV. De acordo com a elipse de confiança (**Figura 4.11d**), valores estimados pelo modelo não apresentam diferença significativa em relação aos resultados de referência ao nível de 95% de confiança.

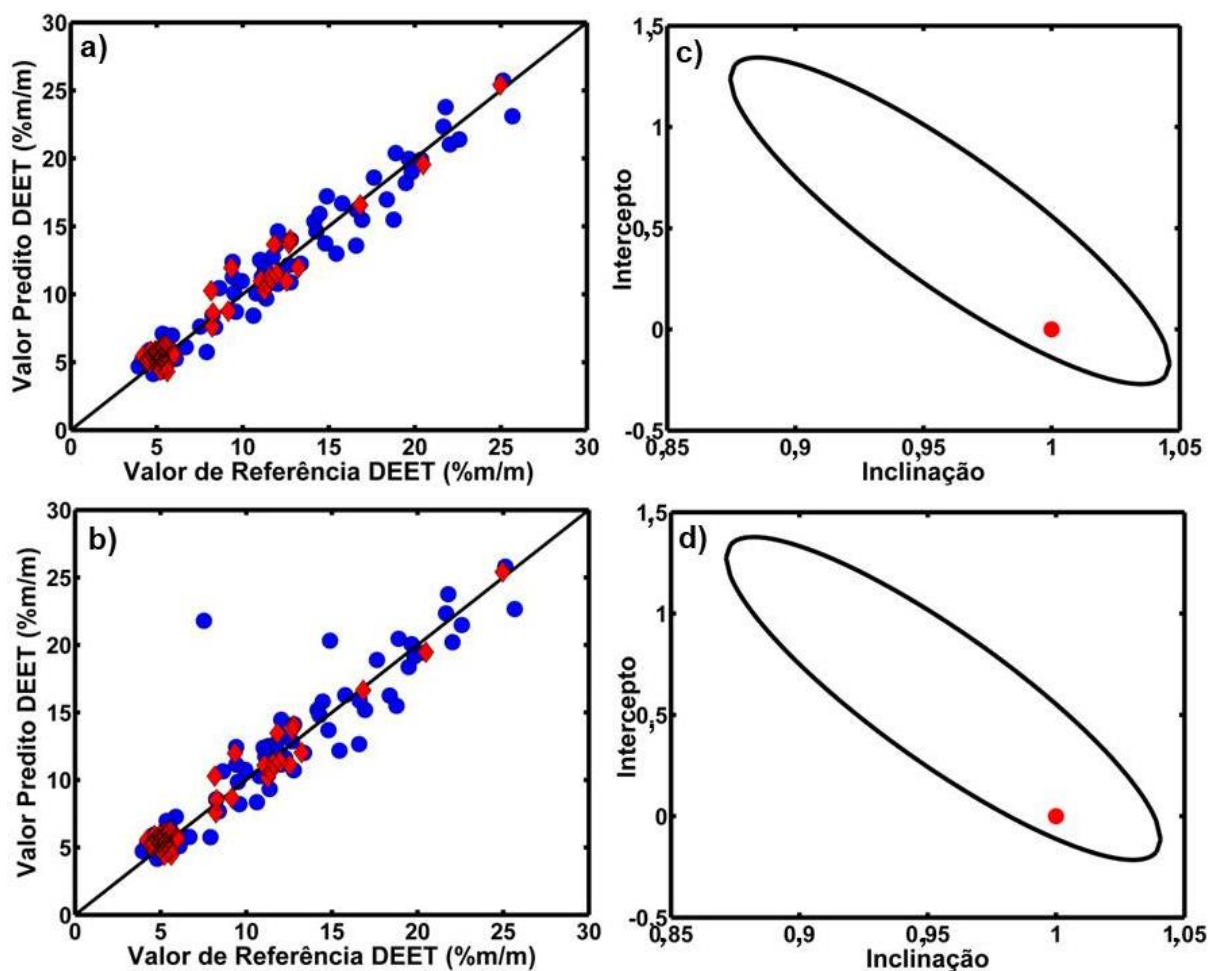


Figura 4.11- Gráficos de valores preditos versus valores de referência para calibração (●) e predição (♦) obtidos com os modelos: a) SPA-MLR ou SW-MLR; b) GA-MLR para determinação do teor de DEET em amostras de repelentes. Região da elipse de confiança: c) SPA-MLR e SW-MLR; d) GA-MLR (●) Ponto ideal para inclinação e o intercepto.

4.4 Análise dos modelos de calibração

A **tabela 4.1** resume os resultados dos parâmetros para avaliação e comparação dos modelos.

Tabela 4.1- Resultados obtidos para os modelos de calibração PLS, *i*PLS, SPA-MLR, SW-MLR e GA-MLR para o teor de DEET.

Parâmetro	Modelo	RMSECV (%m m ⁻¹)	r _{cv}	R ²	RMSEP (%m m ⁻¹)	r _{pred}	LOD (%)	LOQ (%)
DEET %m m ⁻¹	PLS	1,42	0,97	0,94	0,92	0,98	0,13	0,38
	<i>i</i> PLS	1,42	0,97	0,94	0,93	0,98	0,03	0,09
	SW- MLR	1,40	0,97	0,94	0,93	0,98	0,06	0,20
	SPA- LR	1,40	0,97	0,94	0,93	0,98	0,06	0,20
	GA- MLR	2,42	0,91	0,82	0,88	0,98	1,55	4,70

Na comparação entre os modelos PLS e *i*PLS não houve grandes alterações na maioria dos parâmetros de desempenho, exceto para os valores de LOD e LOQ que, para o método por intervalo, resultados mais adequados foram alcançados. Provavelmente, pelo fato de a seleção do intervalo mais informativo ter descartado grande quantidade de ruído.

Os modelos MLR com seleção de variáveis pelo SPA ou SW apresentaram, em sua maioria, desempenho similar àquele obtido pelo modelo PLS (*full-spectrum* e por intervalos). A vantagem, contudo, dos modelos MLR (SPA ou SW) em relação aos modelos PLS é que apenas uma variável espectral é selecionada (196 nm), o que torna a (s) estratégia (a) mais simples.

O modelo GA- MLR apresentou valores mais elevados de RMSECV, LOD e LOQ e valores de r , R^2 inferiores quando comparados com aqueles obtidos pelos modelos PLS, *i*PLS, MLR (SPA ou SW). Todavia, o RMSEP (0,88% m m⁻¹) alcançado pelo modelo foi, de fato, similar àqueles encontrados pelas demais estratégias. Possivelmente, este fato pode ser atribuído à seleção de algumas variáveis pouco informativas (baixa relação sinal ruído) para a construção do modelo (GA-MLR). Apesar dos valores de LOD e LOQ serem adequados, visto que o valor mínimo de teor encontrado no mercado brasileiro é de 5% m m⁻¹ (rotulado), a concentração real pode estar muito abaixo disto.

Para avaliar a significância estatística entre os valores de RMSEP de todos os modelos desenvolvidos, um teste F ao nível de confiança de 95% foi aplicado. (Tabela 4.2). Os valores de F calculado obtidos permaneceram menores que o tabelado ($F=1,67$), o que demonstrou que não há diferenças estatisticamente significativas entre os valores de RMSEP obtidos com os modelos. Logo, estatisticamente, não há diferença na precisão dos modelos elaborados.

Tabela 4.2- Resultados para o teste F (95% de confiança) obtidos na comparação entre os modelos.

Modelo	$F_{\text{calculado}}$	$F_{\text{crítico}}$
PLS/ <i>i</i> PLS	1,02	1,67
PLS/SPA (SW)-MLR	1,02	
PLS/GA-MLR	1,09	
<i>i</i> PLS/SPA (SW)-MLR	1,00	
<i>i</i> PLS /GA-MLR	1,12	
SPA (SW)-MLR/GA-MLR	1,12	

Um teste t -pareado (**Tabela 4.3**) foi feito ao nível de confiança de 95% e não indicou diferenças estatisticamente significativas entre os valores previstos pelos modelos e os valores de referência. Porém, o PLS apresentou valor muito próximo do t crítico.

Tabela 4.3- Resultados para o teste t -pareado (95%de confiança) obtidos na comparação dos valores de referências e valores previstos pelos modelos.

Modelo	$t_{\text{calculado}}$	$t_{\text{crítico}}$
PLS	1,675	1,683
<i>i</i> PLS	1,588	
SPA-MLR	0,743	
SW-MLR	0,743	
G A-MLR	0,153	

Na **tabela 4.4** estão dispostos os valores de concentração estimados pelo modelo SW-MLR ou SPA-MLR bem como os valores de referência e o rotulado pelos fabricantes.

Tabela 4.4- Valores rotulados, determinados pelo método de referência e estimados pelos modelos SPA-MLR ou SW-MLR.

Valor rotulado (%m m ⁻¹)	Valor de referência (%m m ⁻¹)	Valor predito pelo modelo (%m m ⁻¹)
14,6	9,16	8,75
10,0	8,16	10,2
10,0	9,34	11,9
7,00	4,78	5,44
7,00	5,62	4,30
6,79	4,58	4,98
6,79	5,58	4,67
6,79	5,98	5,47
6,79	4,65	5,79
6,79	5,31	5,31
6,79	5,19	6,04
6,79	5,56	6,14
6,79	4,94	5,85
6,79	4,45	5,13
6,79	4,20	5,46
15,0	12,8	14,0
15,0	11,4	10,7
15,0	11,6	11,2
15,0	12,7	13,7
5,00	5,98	5,66
5,00	4,94	5,30
5,00	5,06	5,13
5,00	5,22	4,41
5,00	4,99	4,84
5,00	5,28	5,39
6,65	4,59	5,27
6,65	4,43	5,04
6,65	5,22	5,78
6,65	4,93	5,64
6,65	5,28	6,07
16,0	11,9	11,5
16,0	11,3	10,3
7,60	5,23	5,29
7,60	5,41	6,16
10,0	8,23	7,60

9,50	8,27	8,62
21,0	16,8	16,6
29,0	24,9	25,4
23,0	20,5	19,6
16,0	11,8	13,6
15,0	13,2	11,9
15,0	12,6	10,9
12,0	11,1	10,9

Capítulo 5

Conclusões

5 CONCLUSÕES

Este trabalho propôs uma metodologia analítica baseada em espectrometria de absorção molecular UV-VIS e análise multivariada para determinação de N, N-dietil-3-metilbenzamida (DEET) em amostras de repelentes comerciais.

Inicialmente uma análise exploratória dos dados foi realizada através da aplicação da PCA aos dados pré-processados com intuito de verificar tendência das amostras em relação às faixas de concentração de DEET encontradas em amostras dos principais fabricantes brasileiros. Com base no gráfico dos escores obtido pela PCA, foi possível concluir que a PC1, com 99,8 % da variância explicada, responde ao teor de DEET das amostras de repelentes. Nos valores mais positivos de escores em PC1 encontram-se aquelas amostras de maior teor do princípio ativo envolvido. O gráfico dos pesos evidenciou, também, que a região de maior peso é justamente aquela responsável pela absorção do DEET (entre 190 e 198 nm)

Os modelos de calibração multivariada baseados em PLS, *i*PLS, SPA-MLR, SW-MLR e GA-MLR foram desenvolvidos, validados e comparados entre si. Os modelos PLS e *i*PLS apresentaram um desempenho muito parecido, exceto para os valores de LOD e LOQ que, quando comparados com todos os modelos testados, alcançaram 0,03 % e 0,09%, respectivamente, quando o *i*PLS é aplicado.

Entre as técnicas que empregam seleção de variáveis espectrais originais, os melhores resultados foram alcançados pelos modelos SPA-MLR e SW-MLR. Tais modelos selecionaram a mesma variável espectral (196 nm), alcançando valores próximos daqueles obtidos pelos métodos PLS (*full-spectrum* e por intervalos). O GA-MLR, em sua otimização, precisou de 7 variáveis espectrais e, no que diz respeito aos parâmetros destacados no procedimento de validação, conseguiu um desempenho inferior às demais estratégias testadas. Contudo, quando o modelo GA-MLR é avaliado em relação ao conjunto externo de amostra, resultados similares são alcançados em termos de RMSEP e coeficiente de correlação, mas com valores de LOD e LOQ mais elevados.

Um teste *F* realizado com nível de confiança de 95% indicou que não há diferença estatisticamente significativa entre os diferentes modelos desenvolvidos. Um teste *t*-pareado (com 95% de confiança) também não identificou diferenças entre os métodos propostos com o de referência.

A inexistência de diferenças entre os modelos que utilizam muitas variáveis (PLS, iPLS e GA-MLR) com aqueles que usam apenas uma única variável (SPA-MLR ou SR-MLR) conduz este estudo para estratégia univariada, por sua simplicidade e valores adequados.

A metodologia proposta tem como vantagens ser simples, rápida sensível, necessitar de quase nenhum preparo de amostra, requerer baixo consumo de reagentes, ser de baixo custo e com resultados concordantes com os obtidos pelo método de referência. Tais características a tornam interessante, tanto para auxiliar no controle de qualidade por parte da indústria de cosméticos, quanto para a utilização por órgãos fiscalizadores dada a necessidade de métodos alternativos eficientes que proporcionem agilidade e redução de custos de produção.

5.1 Perspectivas

- ✓ Aplicar a metodologia desenvolvida no controle de qualidade de outros cosméticos que contenham DEET em sua composição;
- ✓ Construir um fotômetro portátil no comprimento de onda selecionado (196 nm);
- ✓ Realizar um estudo de classificação com os demais princípios ativos encontrados no mercado;
- ✓ Utilizar espectrometria UV-VIS e calibração multivariada para a determinação de IR3535 em repelentes;

REFERÊNCIAS

1. WHO- World Health Organization. Vector-borne Diseases. Disponível em: <http://www.who.int/mediacentre/factsheets/fs387/en/>. Acessado em: Out. 2017.
2. WHO- World Health Organization. Zika virus. Disponível em: <http://www.who.int/mediacentre/factsheets/zika/en/>. Acessado em: Jan. 2018.
3. FDA- Food and Drugs Administration. Zika Virus Response Updates from FDA. Disponível em <https://www.fda.gov/emergencypreparedness/counterterrorism/medicalcountermeasures/mcmissues/ucm485199.htm>. Acessado em: Jan. 2018.
4. ANVISA- Agência Nacional de Vigilância Sanitária. ANVISA Registra Primeira Vacina Contra Dengue no Brasil. Disponível em: http://portal.anvisa.gov.br/noticias/-/asset_publisher/FXrpx9qY7FbU/content/anvisa-registra-primeira-vacina-contradengue-no-brasil/219201/. Acessado em: Jan. 2018.
5. ANVISA- Agência Nacional de Vigilância Sanitária. ANVISA não vê restrições no uso de repelentes por gestantes. Disponível em: http://portal.anvisa.gov.br/noticias/-/asset_publisher/FXrpx9qY7FbU/content/anvisa-nao-ve-restricoes-no-uso-de-repelentes-por-gestantes/219201/pop_up?inheritRedirect=false. Acessado em Jan. 2018.
6. E. Lupi, C. Hatz, P. Schlagenhauf, The efficacy of repellents against Aedes, Anopheles, Culex and Ixodes spp.-A literature review. Travel Med. and Infectious Dis.11 (2013) 374 – 411.
7. D. Wiles, J. Yee, U. Castillo, J. Russel, H. Spiller, M. Casavant, A Lethal Case of DEET Toxicity Due to Intentional Ingestion. J. Anal. Toxicol. 38 (2014) 696 – 698.
8. FIOCRUZ- Fundação Osvaldo Cruz. Repelentes contra insetos serão testados pela Fiocruz. Disponível em: <https://rededengue.fiocruz.br/noticias/322-repelentes-contrainsetos-serao-testados-pela-fiocruz>. Acessado em Jan. 2018.
9. Brasil Ministério da Saúde. Agência Nacional. de Vigilância Sanitária. Resolução RDC nº 19 Dispõe sobre os requisitos técnicos para a concessão de registro de produtos cosméticos repelentes de insetos e dá outras providências. Brasília, DF (2013)
- 10.C. J. Correr, G. Cordeiro, J. Gasparetto, P. Peralta-Zamora, R. Pontarolo, Determinação de Ácido Kójico em Produtos Farmacêuticos por Espectroscopia

- UV-Vis e Processo de Calibração Multivariada. *Acta Farm. Bonaerense*. 24 (2005) 416 – 420.
11. M. Blanco, M. Alcalá, J. Planells, R. Mulero, Quality control of cosmetic mixtures by NIR spectroscopy. *Anal. Bioanal. Chem.* 389 (2007) 1577–1583.
 12. I. M. Palabiyik, E. Dinç, F. Onur, Simultaneous spectrophotometric determination of pseudoephedrine hydrochloride and ibuprofen in a pharmaceutical preparation using ratio spectra derivative spectrophotometry and multivariate calibration techniques. *J. Pharm. and Biomed. Anal.* 34 (2004) 473 – 483.
 13. A. H. M. Sarrafi, E. Konoç, M. Ghiyasvand, Simultaneous determination of Atorvastatin Calcium and Amlodipine Besylate by Spectrophotometry and Multivariate Calibration Methods in Pharmaceutical Formulations. *E-J. Chem.* 8 (2011) 1670 – 1679.
 14. F. A. C. Sanches, R. B. Abreu, M. J. C. Pontes, F. C. Leite, D. J. E. Costa, R. K. H. Galvão, M. C. U. Araújo, Near-Infrared Spectrometric Determination of Dipyrone in Closed Ampoules. *Talanta* 92 (2012) 84 – 86.
 15. J. M. Marques Jr, A. L. H. Muller, E. L. Foletto, A. B. Costa, C. A. Bizzi, E. I. Muller, Determination of Propranolol Hydrochloride in Pharmaceutical Preparations Using Near Infrared Spectrometry with Fiber Optic Probe and Multivariate Calibration Methods. *J Anal. Methods Chem.* 2015 (2015) 1 – 6.
 16. M. Sjöström, S. Wold, W. Lindberg, J. Persson, H. Martens, A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables. *Anal. Chim. Acta* 150 (1983) 61 – 70.
 17. L. Norgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck, S. B. Engelsen, Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Appl. Spectrosc.* 54 (2000) 413 – 419.
 18. G. K. Uyanik, N. Güler, A Study on Multiple Linear Regression. *Procedia Soc. Behav. Sci.* 106 (2013) 234 – 240.
 19. M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvão, T. Yoneyama, H. C. Chame, V. Visani, The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemom. Intell. Lab. Syst.* 57 (2001) 65 – 73.
 20. C. B. Lucasius, G. Kateman, Genetic Algorithms for Large-scale Optimization in Chemometrics: an Application. *Trends Anal. Chem.* 10 (1991) 254 – 258.

21. M. Forina, S. Lanteri, M. Casale, M. C. C. Oliveros, Stepwise Orthogonalization of Predictors in Classification and Regression Techniques: An “Old” Technique Revisited. *Chemom. Intell. Lab. Syst.* 87 (2007) 252 – 261.
22. M. Brown, A. A. Herbert. Insect repellents: An overview. *J. Am. Acad. Dermatol.* 36 (1997) 243 – 249.
23. W. S. Leal, The enigmatic reception of DEET – The gold standard of insect repellents. *Curr. Opin. in Insect. Sci.* 6 (2014) 93 – 98.
24. INCQS- Instituto Nacional de Controle de Qualidade em Saúde. GT/COS- Grupo Técnico de Cosméticos. Disponível em: https://www.incqs.fiocruz.br/index.php?option=com_content&view=article&id=85&Itemid=93. Acesso em Jan. 2018.
25. F. J. R. Paumgartenl, I. F. Delgado, Repelentes de mosquitos, eficácia para prevenção de doenças e segurança do uso na gravidez. *Visa em debate.* 4 (2016) 98 – 104.
26. S. Kasichayanul, J. D. House, T. Wang, X. Gu, Simultaneous analysis of insect repellent DEET, sunscreen oxybenzone and five relevant metabolites by reversed-phase HPLC with UV detection: Application to an in vivo study in a piglet model. *J. Chromatogr. B* 822 (2005) 271 – 277.
27. P. Kuklenyik, S. E. Baker, A. M. Bishop, P. Morales-A, A. M. Calafat, On-line solid phase extraction-high performance liquid chromatography–isotope dilution–tandem mass spectrometry approach to quantify N,N-diethyl-m-toluamide and oxidative metabolites in urine. *Anal. Chim. Acta* 787 (2013) 267 – 273.
28. A. W. Abu-Qare, Mo. B. Abou-Donia, Simultaneous determination of Pyridostigmine Bromide, N,N-diethyl-m-toluamide, Permethrin, and their Metabolites in Rat Plasma and Urine by High-Performance Liquid Chromatography. *J. Chromatogr. B.* 749 (2000) 171 – 178.
29. S. A. Cherstniakova, G. E. Garcia, J. Strong, D. Bi, J. Weitz, M. J. Roy, L. R. Cantilena, Rapid Determination of N,N-Diethyl-m-Toluamide and Permethrin in Human Plasma by Gas Chromatography-Mass Spectrometry and Pyridostigmine Bromide by High-Performance Liquid Chromatography. *J. Anal. Toxicol.*, 30 (2006) 21 – 26.
30. T-S. Chen, P-H. Chen, K-L. Huang, Electrochemical degradation of N,N-diethyl-m-toluamide on a boron-doped diamond electrode. *J. Taiwan Inst. Chem. Eng.* 45 (2014) 2615 – 2621.

31. S. Merel, S. A. Snyder, Critical assessment of the ubiquitous occurrence and fate of the insect repellent N,N-diethyl-m-toluamide in water. *Environ. Int.* 96 (2016) 98 – 117.
32. T. Anumol, S. Merel, B. O. Clarke, S. A. Snyder, Ultra High Performance Liquid Chromatography Tandem Mass Spectrometry for Rapid Analysis of Trace Organic Contaminants in Water. *Chem. Cent. J.* 7 (2013) 1 – 14.
33. W. T. S. Vilar, E. S. Sousa, L. Pinto, M. C. U. Araújo, M. J. C. Pontes, Development and Validation of a HPLC Method to quantify DEET and IR3535 in insect repellents. *Anal. Methods.* 10 (2018) 1911 – 1917.
34. D. A. Skoog, F. F. Holler, T. A. Nieman, *Princípios de análise instrumental*, Bookman: São Paulo, 2002.
35. E. C. Silva, *Apostila de Métodos Espectroanalíticos*. João Pessoa- PB, 2010.
36. R. J. Anderson, D. J. Bendell, P. W. Groundwater, *Organic spectroscopy analysis*, The Royal Society of Chemistry: Cambridge, 2004.
37. M.J. Martelo-Vidal, M. Vázquez, Determination of Polyphenolic Compounds of red Wines by UV–VIS–NIR Spectroscopy and Chemometrics Tools. *Food Chem.* 158 (2014) 28 – 34.
38. D. Xu, W. Fan, H. Lv, Y. Liang, Y. Shan, G. Li, Z. Yang, L. Yu, Simultaneous Determination of Traces Amounts of Cadmium, Zinc and cobalt based on UV–Vis Spectrometry Combined with Wavelength Selection and Partial Least Squares regression. *Spectrochim. Acta A Mol Biomol. Spectrosc.* 123 (2014) 430 – 435.
39. G. G. Shimamoto, M. Tubino, Alternative Methods to Quantify Biodiesel in Standard Diesel-biodiesel blends and samples adulterated with vegetable oil through UV–Visible Spectrometry. *Fuel* 186 (2016) 199 – 203.
40. T. Eickner, T. Reske, S. Illner, K-P. Schmitz, D. P. Martin, N. Grabow, Fast parallel Quantification of Dual Drug Systems Via UV Spectrometry – a Case Study with Antibiotic Substances. *Curr. Directions in Biomed. Eng.* 3.(2017) 375 – 378
41. S. Dobrinhas, A. Soceanu, V. Popescu, G. Stanciu, S. Smalberger, Optimization of a UV-VIS Spectrometric Method for Caffeine Analysis. *Sci. Stud. & Res. Chemistry & Chem. Eng., Biotechnol. Food Ind.* 14 (2013) 71 – 78.
42. K. Varmuza, P. Filzmoser, *Introduction to Multivariate statistical Analysis in chemometrics*, Taylor & Francis Group: Boca Raton, 2008.

43. T. Naes, T. Isaksson, T. Fearn, T. Davies, A User-friendly guide to Multivariate Calibration and classification , NIR Publications: Chichester, 2002.
44. M. M. C. Ferreira, A. M. Antunes, M. S. Melgo, P. L. O. Volpe, Quimiometria I: calibração multivariada, um tutorial. Quím. Nova. 22 (1999) 724 – 731.
45. S.D. Brown, R. Tauler, B. Walkzak, Comprehensive Chemometrics-Chemical and biochemical data analysis, Elsevier: Broomfield, 2009.
46. F-T. Chau, Y-Z. Liang, J. Gao, X-G. Shao, Chemometrics From Basics to Wavelet Transform, John Wiley & Sons, Inc., Publication: New Jersey, 2004.
47. M. Forina, S. Lanteri, M. Casale, Multivariate Calibration. J. Chromatogr. A 1158 (2007) 61 – 93.
48. J. M. M. Neto, G. C. Moita, Uma Introdução à Análise Exploratória de Dados Multivariados. Quim. Nova. 21 (1998) 467 – 469.
49. P. R. M. Correia, M. M. C. Ferreira Reconhecimento de Padrões por Métodos não Supervisionados: Explorando Procedimentos Quimiométricos para Tratamento de Dados Analíticos. Quim. Nova. 30 (2007) 481 – 487.
50. M. M. C. Ferreira, Quimiometria Conceitos métodos e aplicações, Editora da Unicamp: Campinas, 2015.
51. K. R. Beebe, R. J. Pell, M. B. Seasholtz, Chemometrics: a practical guide, John Wiley and Sons Inc.: New York, 1998.
52. J. W. B. Braga, R. J. Poppi Validação de Modelos de Calibração Multivariada: Uma Aplicação na Determinação de Pureza Polimórfica de Carbamazepina por Espectroscopia no Infravermelho Próximo. Quim. Nova. 27 (2004) 1004 – 1011.
53. M. I. S. Gonçalves. Aplicações de Imagens Digitais e Análise Multivariada para Classificação e Determinação de Parâmetros de Qualidade em Plumas de Algodão. João Pessoa-PB, Programa de Pós Graduação em Química, ufpb, 2013, Dissertação de Mestrado.
54. M. Sjogström, S. Wold, W. Lindberg, J-K. Persson, H. Martens, A Multivariate Calibration Problem in Analytical Chemistry Solved by Partial Least-Squares Models in Latent Variables. Analytica Chim. Acta. 150 (1983) 61 – 70.
55. R. G. Brereton, G. R. Lloyd, Partial Least Squares Discriminant Analysis: taking the magic away. J. Chemometrics. 28 (2014) 213 – 225.
56. Y. Liu, X. Sun, H. Zhang, O. Aiguo, Nondestructive measurement of internal quality of Nanfeng mandarin fruit by charge coupled device near infrared spectroscopy. Comput. Electron. Agric. 71S (2010) S10 – S14.

57. C. Zhang, N. Xu, L. Luo, F. Liu, W. Kong, L. Feng, Y. He, Detection of Aspartic Acid in Fermented Cordyceps Powder Using Near Infrared Spectroscopy Based on Variable Selection Algorithms and Multivariate Calibration Methods. *Food Bioprocess Technol.* 7 (2014) 598 – 604.
58. R. R. Hocking, The analysis and selection of variables in linear regression. *Biometrics* 32 (1976) 1 – 49.
59. P. A. da C. Filho e R. J. Poppi, Algoritmo Genético em Química. *Quím. nova*, 22 (1999) 405 – 411.
60. R. K. H. Galvão, M. C. U. Araújo, W. D. Fragoso, E. C. Silva, G. E. José, S. F. C. Soares, H. M. A Paiva, Variable Elimination Method to Improve the Parsimony of MLR Models Using the Successive Projections Algorithm. *Chem Intell Lab Sys* 92 (2008) 83 – 91.
61. R. K. H Galvão, M. C. U Araújo, E. C. Silva, G. E. José, S. F. C. Soares, H. M. Paiva, Cross-validation for the Selection of Spectral Variables Using the Successive Projections Algorithm. *J. Braz. Chem. Soc.* 18, (2007). 1580 – 1584.
62. R. K. H. Galvão, M. C. U. Araújo, G. E. José, M. J. C. Pontes, E. C. Silva, T. C. B. Saldanha, A method for Calibration and Validation Subset Partitioning. *Talanta*, 67 (2005) 736 – 740.
63. Z. He, M. Li, Z. Ma, Design of a reference value-based sample-selection method and evaluation of its prediction capability. *Chemometr. Intell. Lab. Syst.* 148 (2015) 72 – 76.
64. R. W. Kennard, L. A. Stone, Computer Aided Design of Experiments. *Technometrics*. 11 (1969) 137– 148.
65. H. M. Paiva, S. F. C. Soares, R. K. H. Galvão, M. C. U. Araújo, A graphical User Interface for Variable Selection employing the Successive Projections Algorithm. *Chemometr. Intell. Lab. Syst.* 118 (2012) 260 – 266.
66. OJEU. Official Journal of the European Union. Decision 96/335/EC establishing an inventory and a common nomenclature of ingredients employed in cosmetic products. 528p. 2006.
67. A. C. Olivieri, Analytical Figures of Merit: From Univariate to Multiway Calibration. *Chemical Reviews* 114 (2014) 5358 – 5378.
68. F. V. C. Vasconcelos, P. F. B. Souza Jr., M. F. Pimentel, M. J. C. Pontes, C. F. Pereira, Using near-infrared overtone regions to determine biodiesel content and

- adulteration of diesel/biodiesel blends with vegetable oils. *Anal. Chim. Acta* 716 (2012) 101 – 107.
69. M. J. C. Pontes, A. M. J. Rocha, M. F. Pimentel, C. F. Pereira, Determining the Quality of Insulating Oils Using Near Infrared Spectroscopy and Wavelength Selection. *Microchem. J.* 98 (2011) 254 – 259.
70. D. A. Skoog, D. M. West, F. J. Holler, S. R. Crouch, *Fundamentos de Química Analítica*, Thomson: São Paulo, 2006.
71. P. Valderrama, J. W. B. Braga, R. J. Poppi, Estado da Arte de Figuras de Mérito em Calibração Multivariada. *Quim. Nova.* 32 (2009) 1278 – 1287.
72. D. A. Santos, K. P. Lima, P. H. Marçó, P. Valderrama, Espectroscopia UV e calibração multivariada multiproduto na determinação da acidez total em néctar de frutas. *REBRAPA.* 6 (2015) 1– 8.