Rosane Beatriz Oliveira Severo

**Identificação de Planta Medicinal baseada em
Espectroscopia e Lógica Fuzzy**

São Paulo

2010

Rosane Beatriz Oliveira Severo

**Identificação de Planta Medicinal baseada em
Espectroscopia e Lógica Fuzzy**

Tese apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Doutor em Engenharia Elétrica.

Área de concentração: Sistemas Eletrônicos.

Orientador: Prof. Dr. Sérgio Takeo Kofuji

São Paulo

2010

Este exemplar foi revisado e alterado em relação a versão original, sob responsabilidade única do autor e com a anuência do seu orientador.

São Paulo, 08 de fevereiro de 2010.

Assinatura do autor

Assinatura do orientador

Ficha catalográfica

Severo, Rosane Beatriz Oliveira

Identificação de Planta Medicinal baseada em Espectroscopia e Lógica Fuzzy

163 páginas

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos.

1. Fitoterapia(Controle de qualidade)
2. Lógica *fuzzy*
3. Espectroscopia

I.Universidade de São Paulo Escola Politécnica. Departamento de Engenharia de Sistemas Eletrônicos II.t.

Rosane Beatriz Oliveira Severo

**Identificação de Planta Medicinal baseada em
Espectroscopia e Lógica Fuzzy**

Tese apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Doutor em Engenharia Elétrica.

Área de concentração: Sistemas Eletrônicos.

Orientador: Prof. Dr. Sérgio Takeo Kofuji

São Paulo

2010

Agradecimentos Dedicados

Aos professores doutores: Sérgio Takeo Kofuji, Reinaldo Giudici, Marilza Antunes de Lemos, Lindolpho Capellari Junior, Wilhelmus Von Noije, Marcelo Knörich Zuffo, Terezinha de Jesus Andreoli Pinto, Nilsa Sumie Yamashita Wadt, Pedro Aladar Tonelli, Roberto Kenji Hiramatsu, Paulo Victor Albuquerque Bergo, Marina Silveira, Liane Margarida Rockembach Tarouco, Nilson Cristino da Cruz, Elidiane Cipriano Rangel, Octávio Ferreira Afonso, Desiré Nguessan, José Alberto Quintanilha; Dr. Dennis Chicoma Lara; Doutorandos Adir José Moreira, Gustavo Adolfo Cerezo Vásquez, (e professora) Jussara Marandola Kofuji; Farmacêutica especialista da Bruker Optics do Brasil, Andréa Andrade; a Empresa Bruker Optics do Brasil; Roberto Serson, diretor da empresa SourceTech Química Ltda, à Empresa SourceTech Química Ltda; Professor Dr. Nobel Penteado Freitas e a graduanda Josiana Aparecida Prestes, do Viveiro da Universidade de Sorocaba; A jornalista (e graduanda) Maria Teresa Oliveira Severo e a advogada Gecy Oliveira Severo; Tradutora Maria Cristina Vidal Borba, mestre Marcio Almeida Mendes, graduandos Miguel Basili Dias, Bruno Feitosa, graduando (e técnico) Cícero Conceição. Agradeço à minha família, Gecy Oliveira Severo, Odará Severo, Maria Teresa Oliveira Severo, Paulo César Oliveira Severo (*in memoriam*) e sua família, Marialva da Costa Severo, Paula Severo, Nicole Severo, Isabele Severo, Jeandra Severo e Felipe Severo. E, a toda a minha grandiosa família, lado Longoni Oliveira, que reside em Porto Alegre-RS. Agradeço, a todas as pessoas, que contribuíram, ajudando para a realização deste trabalho, como os colegas amigos da USP, da Engenharia Elétrica, do Lab. de Sistemas Integráveis, Programa de Sistemas Eletrônicos (LSI-PSI), do Progr. de Sistemas Digitais (PSD), do Progr. de Energia e Automação (PEA), da Engenharia Química, Lab. de Simulação e Controle de Processos (LSCP-PEQ-EP-USP), do Instituto de Matemática e Estatística (IME-USP), do Lab. de Microscopia Eletrônica, Instituto de Física (LME-IF-USP), da ESALQ-USP Piracicaba, da FZEA-USP Pirassununga, da UNESP campus Sorocaba, aos colegas amigos professores, aos pós-graduandos, aos graduandos, aos moradores e amigos das cidades de São Paulo e Sorocaba, a Bee e a Maninha, evocando meu ser Zen.

Resumo

A presente tese de doutorado discorre sobre o método da espectroscopia na região do Infravermelho Próximo, usando Transformada de Fourier (IVP-TF), aplicada à identificação de plantas medicinais, método de análise química qualitativa, que se propõe para o controle de qualidade de fitoterápicos. É apresentado, também, um método computacional, implementado para realizar esta identificação, utilizando modelo e lógica nebulosos (*fuzzy*). Remete-se a discussão da espectroscopia IVP-TF como um método rápido e de baixo custo, para taxonomia de plantas medicinais brasileiras, bem como promove-se a comparação dos resultados obtidos nos diferentes métodos utilizados.

Palavras-chave: Fitoterapia(Controle de qualidade), Lógica fuzzy, Espectroscopia.

Abstract

This doctoral thesis is about the method of Near Infrared spectroscopy using Fourier transform (FT-NIR), applied to identification of medicinal plants, method of the chemical analysis qualitative, which is proposed to the quality control of herbal medicines. It is presented, also, a computational method, implemented to make this identification using fuzzy logic model. Refers to discussion of FT-NIR spectroscopy as a rapid method with low cost for taxonomy of Brazilian medicinal plants, and makes comparison of results obtained in the different methods used.

Keywords: Quality Control of Herbal Medicines, Fuzzy logic, Spectroscopy.

Lista de Figuras

1.1	28 dos 60 fitoterápicos analisados no curso de farmagnosia da Faculdade de Farmácia e Bioquímica, São Paulo	20
1.2	32 dos 60 fitoterápicos analisados. Fitoterápico anunciado (V ou F)? . .	21
1.3	Tipos de espectros ópticos. Extraído de Gonzalez, Rafael C., Woods, Richard E. 2002. Digital Image Processing. Upple Saddle River Prentice Hall Inc New Jersey.	26
2.1	Representação esquemática de seção vertical de uma folha normal, incluindo a veia principal e tecidos circundantes. (<i>Extraído de Oi Wah Liew, Pek Ching Jenny Chong, Bingqing Li and Anand K. Asundi, Signature Optical Cues: Emerging Technologies for Monitoring Plant Health, Sensors 2008, 8, 3205-3239</i>)	33
3.1	Jaborandi teste identificado no <i>cluster</i> de Jaborandi	57
3.2	Relatório do Id	59
3.3	Fatorização - Cálculo da distância espectral.	60
4.1	Procedimentos de <i>Fuzzy</i> ificação.	81
4.2	Arquivo de Referência.	82
4.3	Arquivo de Teste.	82
4.4	Ordem Reference	82
4.5	Ordem Teste	82
4.6	Arquivo Picos.	82
4.7	Grau de pertinência da planta teste.	84

4.8	Pertinência de uma diferença de $0.3 * \lambda$ para as funções de pertinência quadrática (à esquerda) e modular (à direita).	98
4.9	Pertinência de uma diferença de $0.75 * \lambda$ para as funções de pertinência quadrática (à esquerda) e modular (à direita).	99
4.10	Avaliação de AbacateiroMarcio contra Abacateiro0.	102
4.11	Avaliação de AbacateiroMarcio contra Abacateiro5.	104
4.12	Avaliação de AbacateiroMarcio contra Alcachofra0.	105
4.13	Avaliação de AbacateiroMarcio contra Alecrim0.	106
5.1	Agrupamento de média dos testes.	121
A.1	Modelo mecânico da molécula diatômica vibrando.	127
A.2	Três picos de absorção para constantes de forças diferentes.	128
A.3	Três picos de absorção para massas atômicas diferentes.	128
A.4	Curva do potencial de energia para um oscilador harmônico.	130
A.5	Curva do potencial de energia para um oscilador não harmônico.	130
A.6	Separação em três faixas espectrais do Infravermelho.	131
A.7	Primeiras tecnologias IV monocromáticas.	131
A.8	Evolução das tecnologias monocromáticas.	131
A.9	AOTF e técnica de modulação usando FT	132
C.1	Caixa de Especimens	143
C.2	Drogas Vegetais	143
C.3	Pilocarpus microphyllus SourceTech	143
C.4	P. microphyllus SourceTech	143
C.5	Poejo exsicata	144
C.6	Guaco exsicata	144
C.7	Pata de vaca exsicata	145

C.8	<i>Pilocarpus pennatifolius</i> Praça Sé	145
C.9	<i>Pilocarpus pennatifolius</i> Praça Sé	145
C.10	<i>Pilocarpus jaborandi</i> Praca da Sé	145
C.11	Drogas Vegetais	145
C.12	Drogas Vegetais	145
C.13	Drogas Vegetais	146
C.14	<i>Hibiscus sabdariffa</i> L.	146
C.15	<i>Hibiscus sabdariffa</i> L.	146
C.16	Drogas Vegetais	146
C.17	<i>Alecrim exsicata</i>	147
C.18	<i>Bardana exsicata</i>	148
C.19	Drogas Vegetais	149
C.20	<i>Espinheira santa exsicata</i>	149
C.21	<i>Maytenus ilicifolia</i> Reissek	149
C.22	<i>Maytenus ilicifolia</i> Reissek	149
C.23	<i>Maytenus ilicifolia</i> Reissek	149
C.24	<i>Pilocarpus microphyllus</i> SourceTech	149
C.25	<i>Urucum exsicata</i>	150
C.26	ESALQ Universidade de São Paulo	150

Lista de Tabelas

1.1	Bandas Temáticas Satélite LANDSAT da NASA. Expressas em comprimento de onda μ m igual a 10^{-6} m	26
3.1	Plantas medicinais catalogadas pelo herbário ESA-ESALQ-USP.	47
3.2	Espécimes caracterizados espectralmente - IVP e Raman - na Primeira Fase.	50
3.3	Plantas medicinais catalogadas pelo herbário ESA-ESALQ-USP.	53
3.4	Resultados possíveis codificados para os testes de identidade.	70
3.5	Resultados possíveis para os testes de classe e seus respectivos números codificados.	72
4.1	Espectros que formam a Base de Referência, formam a primeira aquisição espectral com reprodutibilidade espectral de 6 a 7.	85
4.2	Espectros que formam a Base de Referência de Testes com Jaborandis.	92
5.1	Valores médios agrupados de testes entre as plantas testes obtidas da reprodutibilidade 5 e da respectiva planta referência de reprodutibilidade zero.	110
5.2	Valores médios agrupados de testes entre as plantas testes obtidas da reprodutibilidade 5 e da respectiva planta referência de reprodutibilidade zero.	113
5.3	Resultado dos testes entre maiores números <i>fuzzy</i> das plantas de reprodutibilidade 5 e qualquer planta de referência diferente da planta teste.	114
5.4	Resultado dos testes entre maiores números <i>fuzzy</i> das plantas de reprodutibilidade 5 e qualquer planta de referência diferente da planta teste.	115

5.5	Conforme valores das tabelas 5.1, 5.2, 5.3, 5.4 tem-se os resultados considerados válidos:	115
5.6	Conforme valores das tabelas 5.1, 5.2, 5.3, 5.4 tem-se os resultados considerados válidos:	116
5.7	Método das diferenças e resultados, como as correlações para planta testexplanta de referência com reprodutibilidade.	117

Lista de Siglas

ANVISA Agência Nacional de Vigilância Sanitária

AOTF Acousto-Optic Tunable Filters

CP Componente Principal

DEM Distância Euclideana Média

DR Diffuse Reflection - Reflexão Difusa

DRIFT Diffuse Reflectance Infrared Fourier Transform

EIVP Espectroscopia no Infravermelho Próximo

ESA Herbário ESALQ-USP

FAR Far Infrared

FPA Focal Plane Array

HPLC High Performance Liquid Chromatography

MIR Middle Infrared

MPA Multipurpose Analysis

NIRS Near Infrared Reflectance Spectroscopy

PAT Process Analytical Technology

PCA Principal Component Analysis - Análise de Componentes Principais

PE Pelkin Elmer

RDC Resolução da Direção Colegiada

RMS Root Mean Square

ROI Regions of Interest

SD Standard Deviation - Desvio Padrão

SIM Spectral Identity Mapping

SNR Signal to Noise Ratio - Relação Sinal Ruído

TF Transformada de Fourier

UV Ultravioleta, região do espectro eletromagnético

VIS Visível, região do espectro eletromagnético

WEFA Window Evolving Factor Analysis

Sumário

1	Introdução	16
1.1	Motivação, Relevância, Problemas	17
1.2	Objetivos	30
1.2.1	Objetivo Principal	30
1.2.2	Objetivos Secundários	30
1.3	Descrição do trabalho	30
2	Aspectos Teóricos	32
2.1	Métodos utilizados para análise de plantas medicinais	32
3	Assinatura Espectral no Infravermelho Próximo	41
3.1	Aquisição dos Espécimes	43
3.1.1	Drogas Vegetais	43
3.2	Material e Métodos de Identificação	52
3.3	Resultados do Id	74
4	Assinatura Espectral <i>Fuzzy</i>	78
4.1	Implementação segundo o modelo de Bafflert	80
4.2	Aplicação do Método das Diferenças com a Lógica <i>Fuzzy</i>	95
5	Discussão dos Resultados	108
5.1	Assinatura Espectral Id	108

5.2	Assinatura Espectral Fuzzy	109
6	Conclusões	122
6.1	Trabalhos Futuros	124
	Apêndice A	126
A.1	Introdução	126
A.2	Espectroscopia IV	126
A.3	IVP	136
	Apêndice B	139
B.1	Modelo e Lógica Fuzzy	139
B.1.1	Modelo Fuzzy	140
	Apêndice C	143
C.1	Exsicatas e conjunto de amostras	143
	Apêndice D	151
D.1	Detalhes das Implementações	151

1 *Introdução*

A motivação para a realização deste trabalho é inerente a relevante área de saúde. Os estudos e as observações, com a identificação dos problemas e as respectivas intervenções, culminaram com a formulação dos objetivos para o desenvolvimento deste trabalho.

Neste sentido, este trabalho se estenderá sobre a forma de implementações futuras, decorrentes, isto é, seus objetivos vão além desta tese. São detalhadamente expostas estas considerações, na presente descrição que segue, dividida em partes, com coerência funcional e premiando também, suas recorrências.

Este *envolvimento* tem sua parte inicial, centrada nos conhecimentos adquiridos em atividades de pesquisa exercidas junto aos colegas da área de saúde, notoriamente, aqui, da área de Farmácia e Bioquímica.

Os relatos seguintes mostram a busca de idéias, que levem à elucidação das questões postas, rumando ao campo de outras áreas, como o processamento digital de imagens, notoriamente, o processamento e análise de imagens de satélite de sensoriamento remoto e orbital, culminando com a espectroscopia, que compõe o ferramental dos quimiometristas, nas análises químicas qualitativas e quantitativas (incluindo a fitoquímica), neste caso, também, além da farmacognosia, o objeto dos taxonomistas, as plantas de interesse econômico - medicinais, e ornamentais.

É apresentado neste trabalho um histórico de tentativas implementadas com sucesso. Algumas destas serão melhor exploradas em trabalhos futuros, como o uso da espectroscopia na região do infravermelho médio, a espectroscopia Raman e outras, como a avaliação de umidade com uso de microondas.

Podem ser observados relatos de diversas outras áreas do conhecimento, para se chegar aos objetivos eleitos para esta pesquisa, sempre visualizando os resultados,

solução ou soluções dos problemas particionados, levando a mais de uma solução sistêmica. Com a presente contribuição, é apresentado o processo de identificação de aproximadamente cem plantas medicinais ou drogas vegetais, fitoterápicos, com uso de espectroscopia, na região do infravermelho próximo e o desenvolvimento de programas que implementam soluções *fuzzy*. Com a pertinência *fuzzy* (82 é este número fuzzy) pode-se afirmar que, de acordo com as informações de nosso conhecimento (da base de referência), trata-se da correta identificação, no caso, de *Pilocarpus microphyllus* (Jaborandi), a corrente droga vegetal testada. Este trabalho a partir de suas publicações, pretende expor e discutir o tema, bem como dos resultados obtidos (para que serve, é útil esta pertinência (número) *fuzzy*, em um dado (que) contexto?), no meio científico local e remoto.

1.1 Motivação, Relevância, Problemas

No ano de 1999, esta autora participou de cinco disciplinas da pós-graduação em engenharia elétrica, ministradas na Escola Politécnica da Universidade de São Paulo, versando sobre assinaturas digitais de doenças, usando CT (Tomografia Computadorizada), MR (Ressonância Magnética), NM (Medicina Nuclear), US (Ultra-Som), inerentes ao campo do processamento digital de imagens avançado, no caso para aplicações médicas.

Tive ainda a oportunidade de cursar a disciplina de processamento de imagens de ME (Microscopia Eletrônica), junto ao Instituto de Física, usando microscópio eletrônico de transmissão, culminando com a disciplina de processamento de imagens de sensoramento remoto, onde foi estudado e pesquisado, o tema espectrometria, sendo enfatizada a região do visível (VIS), com experiências práticas, acadêmicas, realizadas em laboratório no Instituto de Física da USP.

As plantas visualizadas por satélite (sensoriamento remoto) apresentam uma assinatura espectral, sendo por exemplo caracterizadas, datadas a sua idade, conforme a cor (VIS), de modo análogo, para citar, as águas profundas ou rasas, azul escuro ou azul claro, respectivamente.

No ano de 2004, durante um trabalho de conclusão de curso da ciência da computação, iniciou-se uma pesquisa cujo objetivo era o controle de qualidade de fitoterápi-

cos. Na ocasião, tomou-se conhecimento de diversos temas serão apresentados, envolvendo o controle de qualidade de fitoterápicos.

Fitoterápico é um medicamento obtido empregando-se exclusivamente matérias primas ativas vegetais. É caracterizado pelo conhecimento da eficácia e dos riscos de seu uso, assim como pela reprodutibilidade e constância de sua qualidade. Sua eficácia e segurança é validada através de levantamentos etnofarmacológicos de utilização, documentos tecnocientíficos em publicações ou ensaios clínicos, numa terceira fase. Não se considera medicamento qualquer fitoterápico aquele que, na sua composição, inclua substâncias ativas isoladas de qualquer origem, nem as associações destas com extratos vegetais, (AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2004a).

A idéia primordial na indicação do uso de fitoterápicos na medicina humana não é substituir medicamentos registrados e comercializados, mas aumentar a opção terapêutica dos profissionais de saúde, indicando medicamentos também registrados e talvez mais baratos, podendo ser indicado também como um medicamento complementar às medicações existentes, (SIMÕES, 2004). Até a década de 1950 os remédios vegetais tinham uso exclusivo na medicina terapêutica. Devido ao difícil controle de qualidade (químico, físico-químico, farmacológico ou toxicológico dos extratos vegetais utilizados) as drogas vegetais foram substituídas aos poucos nas farmácias por medicamentos contendo as substâncias ativas deles extraídas ou seus derivados sintéticos. Consequentemente poucas foram as plantas medicinais estudadas segundo protocolos mais modernos. A maioria das informações disponíveis referem-se à década de 1950, portanto obsoletas frente ao estágio atual do conhecimento científico. Entre os adeptos de produtos fitoterápicos, é comum o pensamento de que as plantas medicinais já foram testadas e homologadas pelo uso prolongado na própria espécie humana e de que são remédios eficazes e seguros, naturalmente balanceados, sem os efeitos comuns aos produtos sintéticos, não necessitando da avaliação exigida para este tipo de medicamento. É inegável que a maioria da população de baixa renda recorre às plantas medicinais para a cura de seus males, porque o comércio de medicamentos, no Brasil, atende mais a faixa da população economicamente ativa. A validação científica desses produtos é essencial ao seu uso como medicamentos alternativos. Se a intenção é utilizar uma planta medicinal como medicamento, ela deve ser previamente validada, isto é, ter sua ação comprovada e sua toxicidade potencial avaliada cientificamente na espécie humana, como qualquer outro medicamento (SIMÕES, 2004).

Como fonte para coleta de dados foram utilizadas as monografias anuais da disciplina de farmacognosia da Faculdade de Farmácia e Bioquímica de duas instituições da cidade de São Paulo. Assim os componentes das monografias serviram de entrada de dados para o estudo. Uma das hipóteses formuladas, que serviu como motivação, é que 70% dos fitoterápicos estudados são verdadeiros, isto é, correspondem ao produto anunciado, sendo comercializado na cidade de São Paulo. Também, parte destes estariam impróprios para o consumo, considerando um grau de pureza inferior a 95%. As monografias realizadas pelos alunos do curso de graduação em Farmácia e Bioquímica, no período de cinco anos, são compostas dos seguintes itens:

1. Descrição Macroscópica: Esta análise é feita à vista desarmada ou com o auxílio de lupa de pequeno aumento. Na análise das folhas, as principais características macroscópicas são as seguintes: aspecto geral, consistência, cor, forma, sabor, odor, superfície, tamanho e transparência (OLIVEIRA, AKISUE, 1998).
2. Descrição Microscópica: Do ponto de vista da microscopia são importantes a forma da célula vegetal, seu conteúdo, a natureza e o espessamento de sua parede, os tipos de tecidos presentes e suas características, bem como as estruturas.
3. Fitoquímica: A pesquisa fitoquímica visa conhecer os constituintes químicos de espécies vegetais ou avaliar a sua presença.
4. Ação Farmacológica: Está relacionada ao preparo da forma farmacêutica para administração, com a garantia da qualidade e uniformidade da amostra.

Para garantir a uniformidade de um fitoterápico, é necessário que os diferentes fitoterápicos intermediários sejam caracterizados através de seus constituintes químicos, ou de sua(s) atividade(s) farmacológica(s) (OLIVEIRA, AKISUE, 1998).

5. Ação Toxicológica: Deve indicar qual o grau de confiança a ser depositado em um medicamento a ser administrado à espécie humana. Os testes de toxicidade em geral devem guardar relação dose-efeito satisfatória e permitir estabelecer relação causa-efeito.
6. Ensaios Clínicos e Testes de Segurança: A experiência da equipe executora, as condições dos laboratórios, os animais utilizados e a definição clara dos objetivos são condições mais importantes do que o cumprimento automático das normas aparentemente rígidas dos protocolos oficiais.

7. Grau de Pureza: A pureza da droga pode ser modificada em função de contaminação ou fraude. A contaminação ocorre acidentalmente durante as etapas de coleta de vegetal, preparo, conservação ou armazenamento. Esta contaminação pode se dar com outros vegetais ou outros órgãos, do mesmo vegetal, ou ainda, por microrganismos, insetos e roedores, (OLIVEIRA, AKISUE, 1998).

NOME BOTÂNICO	NOME POPULAR	V	F	% PUREZA
<i>Cichorium intybus</i> L.	CHICÓRIA-DO-CAFÉ	X		92.73
<i>Cichorium intybus</i> L.	CHICÓRIA-DO-CAFÉ		X	
<i>Achyrocline satureioides</i> (Lam.) DC.	MARCELA-DO-CAMPO	X		70.4
<i>Achyrocline satureioides</i> (Lam.) DC.	MARCELA-DO-CAMPO	X		8.96
<i>Hedera helix</i> L.	HERA		X	
<i>Allium sativum</i> L.	Alho	X		
<i>Hamamelis virginiana</i> L.	HAMAMELIS	X		3.5
<i>Quassia amara</i> L.	QUINA		X	
<i>Quassia amara</i> L.	QUINA	X		92
<i>Quassia amara</i> L.	QUINA	X		87
<i>Anemopaegma arvense</i> (Vell.)*	CATUABA		X	
<i>Centella asiatica</i> (L.) Urban.	CENTELA	X		94.8
<i>Centella asiatica</i> (L.) Urban.	CENTELA	X		
<i>Arnica montana</i> L.*	ARNICA	X		
<i>Brugmansia suaveolens</i> (Willd.) Bercht.	TROMBETEIRA	X		
<i>Citrus aurantium</i> L.	LARANJA-AMARGA	X		
<i>Solanum paniculatum</i> L.	JURUBEBA-VERDADEIRA	X		33.3
<i>Casearia sylvestris</i> Sw.	GUAÇATONGA	X		
<i>Casearia sylvestris</i> Sw.	GUAÇATONGA	X		
<i>Cassia angustifolia</i> Vahl.	SENE	X		
<i>Cassia angustifolia</i> Vahl.	SENE	X		
<i>Lavandula officinalis</i> Chaix & Kitt. *	ALFAZEMA	X		
<i>Centella asiatica</i> (L.) Urban	CENTELA	X		75
<i>Rheum tanguticum</i> var. <i>tanguticum</i>	RUIBARBO	X		
<i>Cynara scolymus</i> L. *	ALCACHOFRA	X		97.3
<i>Erythroxylum vacciniifolium</i> Mart.	CATUABA	X		87.7
<i>Eucalyptus globulus</i> Labill.	EUCALIPTO	X		
<i>Eucalyptus globulus</i> Labill.	EUCALIPTO	X		

Figura 1.1: 28 dos 60 fitoterápicos analisados no curso de farmagnosia da Faculdade de Farmácia e Bioquímica, São Paulo

Além dos itens já referidos remetendo-se ao conhecimento popular, nas monografias estão descritos os usos populares conhecidos da droga. Foram apurados os resultados, conforme figuras 1.1 e 1.2, que basicamente respondem a duas perguntas:

- Sendo o fitoterápico proveniente do comércio em geral, trata-se verdadeiramente

NOME BOTÂNICO	NOME POPULAR	V	F	% PUREZA
<i>Glycyrrhiza glabra</i> L.	ALÇAÇUZ	X		
<i>Hedera helix</i> L.	HERA		X	
<i>Piper umbellatum</i> L. *	CAPEBA	X		
<i>Centella asiatica</i> (L.) Urban.	PATA DE CAVALO	X		85
<i>Illicium verum</i> Hook. f.	ANIS-ESTRELADO	X		
<i>Illicium verum</i> Hook. f.	ANIS-ESTRELADO	X		
<i>Jacaranda caroba</i> (Vell.) A. DC.	CAROBINHA	X		75.2
<i>Malva sylvestris</i> L.	MALVA		X	
<i>Matricaria chamomilla</i> L.	CAMOMILA	X		
<i>Mentha piperita</i> L.	HORTELÃ-PIMENTA	X		
<i>Mentha piperita</i> L.	HORTELÃ-PIMENTA		X	
<i>Mikania glomerata</i> Spreng.	GUACO	X		100
<i>Paullinia cupana</i> Kunth.	GUARANÁ		X	
<i>Paullinia cupana</i> Kunth.	GUARANÁ	X		
<i>Petiveria alliacea</i> L.	GUINÉ	X		86.8
<i>Peumus boldus</i> Molina	BOLDO-DO-CHILE	X		
<i>Peumus boldus</i> Molina	BOLDO-DO-CHILE	X		92
<i>Pilocarpus microphyllus</i> Stapf	JABORANDI	X		58
<i>Pilocarpus microphyllus</i> Stapf	JABORANDI	X		
<i>Pimpinella anisum</i>	ERVA-DOCE	?	?	67
<i>Piper umbellatum</i> L. *	CAPEBA		X	
<i>Punica granatum</i> L.	ROMÃ	X		
<i>Punica granatum</i> L.	ROMÃ	X		99.5
<i>Punica granatum</i> L.	ROMÃ	X		
<i>Quillaja saponaria</i> Molina	QUILAIA	X		100
<i>Rosmarinus officinalis</i> L.	ALECRIM	X		98
<i>Ruta graveolens</i> L.	ARRUDA	X		100
<i>Smilax papyracea</i> Poir.	SAL SAPARRILHA	X		
<i>Salvia officinalis</i> L.	SÁLVIA	X		
<i>Siparuna apiosyce</i> (Mart. ex Tul.) A. DC.	LIMÃO-BRAVO	X		99.5
<i>Stryphnodendron barbatiman</i> Mart.	BARBATIMÃO	X		20

Figura 1.2: 32 dos 60 fitoterápicos analisados. Fitoterápico anunciado (V ou F)?

do fitoterápico anunciado? (V ou F)

- Qual o grau de pureza da droga vegetal?

No sentido da discussão dos resultados, estes foram comparados com a literatura e normas vigentes. A seguir, apresentam-se os resultados obtidos com os dados processados das monografias, tendo sido comprovados os resultados das hipóteses observadas. Os resultados computados estão apresentados. Após, foi procedida a discussão destes resultados encontrados versus as normas brasileiras da ANVISA, bem como foi esta literatura estudada.

- Depreende-se que o processo de controle de qualidade de fitoterápicos ensinado,

é composto, basicamente das seguintes fases:

- análise macroscópica: o especialista em farmacognosia identifica a planta medicinal macroscopicamente por experiência adquirida, conhecimento prévio e/ou comparando-a com o catálogo de herbário, com a respectiva exsicata;
 - análise microscópica: usando microscópio óptico, o especialista identifica microscopicamente a planta medicinal, de modo análogo aos procedimentos da fase macroscópica; procedimentos similares são adotados para a microscopia do pó;
 - prospecção fitoquímica: o especialista em farmacognosia procede uma série de análises fitoquímicas, utilizando diferentes métodos e/ou equipamentos como, para citar, espectroscopia VIS (Visível), UV (Ultra-Violeta), HPLC ¹, GC-MS ², gravimetria, CCD ³, TLC ⁴, doseamento de princípios ativos, teste de pureza, etc. Este processo é executado em laboratório, seguindo os protocolos de processamento das respectivas drogas vegetais, a princípio descritos nas farmacopéias em conformidade com a legislação, que controla processos do ponto de vista de qualidade.
- A Agência Nacional de Vigilância Sanitária (ANVISA) publica suas resoluções que normatizam o uso de fitoterápicos. A Resolução de Diretoria Colegiada (RDC) 48/04 (AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2004a), classifica:
 - Droga vegetal é a planta medicinal ou suas partes, após processos de coleta, estabilização e secagem, podendo ser íntegra, rasurada, triturada ou pulverizada.
 - Derivado de droga vegetal são os produtos de extração da matéria prima vegetal: extrato, tintura, óleo, cera, exsudato, suco, e outros.

¹Cromatografia líquida de alta eficiência (CLAE) (alta pressão ou cromatografia líquida) é uma forma de coluna cromatográfica utilizada frequentemente em bioquímica e química analítica para separar, identificar e quantificar compostos. CLAE utiliza uma coluna que detém cromatograficamente embalagens (fase estacionária), uma bomba que move a fase móvel através da coluna, e um detector que mostra os tempos de retenção das moléculas. Tempo de retenção varia de acordo com as interações entre a fase estacionária, as moléculas sendo analisadas, e do solvente utilizado.

²Gas Chromatography Mass Spectrometry

³Cromatografia em Camada Delgada

⁴Thin layer chromatography, Cromatografia em Camada Fina. Técnica cromatográfica que é usada em muitos laboratórios de pesquisa (fito)farmacêutica para a separação e identificação por coloração (croma=grego para cor) de compostos orgânicos numa mistura de centenas de compostos.

- Fitoterápico, o medicamento obtido empregando-se exclusivamente matérias-primas ativas vegetais.
- Princípio ativo de medicamento fitoterápico é a substância, ou classes químicas (ex: alcalóides, flavonóides, taninos, ácidos graxos, etc.), quimicamente caracterizada, cuja ação farmacológica é conhecida e responsável, total ou parcialmente, pelos efeitos terapêuticos do medicamento fitoterápico.
- Marcador é o componente ou classe de compostos químicos (ex: alcalóides, flavonóides, ácidos graxos, etc.) presente na matéria-prima vegetal, idealmente o próprio princípio ativo, e preferencialmente que tenha correlação com o efeito terapêutico, que é utilizado como referência no controle de qualidade da matéria-prima vegetal e dos medicamentos fitoterápicos.
- A RDC 48/04 informa em seu item “abrangência” que os medicamentos cujos princípios ativos sejam exclusivamente derivados de drogas vegetais serão objetos de registro como fitoterápicos.
- Relatório de controle de qualidade: Informações referentes à droga vegetal, laudo de identificação próprio ou emitido por profissional habilitado, quando não existirem especificações em farmacognosia que permitam a confirmação da identidade botânica. Referência bibliográfica da Farmacopéia consultada e reconhecida pela ANVISA, de acordo com a legislação vigente.
- Informações referentes aos derivados de droga vegetal: Nomenclatura botânica oficial (gênero, espécie e autor), a nomenclatura farmacopéica e/ou tradicional. Parte da planta utilizada, Solventes, excipientes e/ou veículos utilizados na extração do derivado.
- Estas informações devem vir acompanhadas do laudo de análise do fornecedor. Testes de autenticidade (caracterização organoléptica, identificação macroscópica e microscópica); Testes de pureza e integridade, incluindo: cinzas, cinzas insolúveis em ácido clorídrico, umidade, pesquisa de matérias estranhas, pesquisa de contaminantes microbiológicos e de metais pesados. Em caso de utilização de métodos para eliminação de contaminantes, descrever o método e a pesquisa de eventuais alterações da matéria-prima. Análise qualitativa e quantitativa dos princípios ativos e/ou marcadores, quando conhecidos, ou classes de compostos químicos característicos da espécie.

- Controle de qualidade do produto acabado: Apresentar a descrição detalhada de todas as metodologias utilizadas no controle de qualidade, com os métodos analíticos devidamente validados para o medicamento, de acordo com a Guia para Validação de Métodos Analíticos e Bioanalíticos, indicando a sua fonte bibliográfica ou de desenvolvimento. Resultado da prospecção (*screening*) fitoquímica, ou perfil cromatográfico (*fingerprint*) por cromatografia líquida de alta eficiência - CLAE ou cromatografia gasosa - CG, quando cabível. Apresentar tradução quando o idioma não for inglês ou espanhol; Especificações do material de embalagem primária. Certificado de Boas Práticas de Fabricação e Controle (BPFC) emitido pela ANVISA, para a linha de produção na qual o produto classificado como medicamento fitoterápico será fabricado, ou ainda, cópia do protocolo de solicitação de inspeção para fins de emissão do certificado de BPFC. Este protocolo é válido desde que a linha de produção pretendida esteja satisfatória na última inspeção para fins de verificação do cumprimento de BPFC realizada.
- Desde a publicação da RDC 48/04, fitoterápicos não são mais cadastrados como isentos de registro. Sendo assim, tais produtos devem se adequar a esta resolução, sendo registrados como fitoterápicos.

O (BRASIL MINISTÉRIO DA SAÚDE, 2006) no sentido do interesse popular e institucional vem crescendo no sentido de fortalecer a fitoterapia no Sistema Único de Saúde (SUS). A partir da década de 80, diversos documentos foram elaborados, enfatizando a introdução de plantas medicinais e fitoterápicos na atenção básica no sistema público, entre os quais se destaca:

- a Resolução Ciplan no. 8/88, que regulamenta a implantação da fitoterapia nos serviços de saúde e cria procedimentos e rotinas relativas a sua prática nas unidades assistenciais médicas;

Problemas específicos que tem sido relatados, segundo (JORGE, 2000) no que se refere a qualidade da droga vegetal:

- droga vegetal consiste em mistura de vários constituintes;

- princípio(s) ativo(s) nem sempre é (são) conhecido(s);
- métodos analíticos seletivos podem não existir;
- substâncias químicas de referência (marcadores) podem não estar disponíveis comercialmente;
- variabilidade química do material vegetal sofre influência das condições de coleta, secagem e estocagem;
- influência do processo de extração (polaridade do solvente, modo de extração, instabilidade dos constituintes);

A primeira idéia avaliada foi a identificação de plantas medicinais, considerando a primeira fase do processo em farmacognosia, pretendendo-se implementar a identificação macroscópica através de técnicas de processamento digital de imagens, como uso de reconhecimento de padrões. Esta proposta em estudo evolui para a proposta atual envolvendo a terceira etapa da análise fitoquímica. Pela experimentação vivenciada neste trabalho, percebe-se que as análises químicas qualitativas e quantitativas, disjuntas, são permeadas pelas técnicas de análise de fatores, especialmente no que se refere às caracterizações procedidas na região do IVP, logo, no presente trabalho, percebeu-se que o imageamento de drogas vegetais envolve necessariamente a fitoquímica, a quimiometria. Prova disto são os trabalhos referidos adiante, que vão desde modelos de absorção óptica por folhas (vegetais). Parece que a misteriosa lacuna aberta pela farmacognosia, já referi da, com estes métodos avança na sua exploração de novas técnicas para “facilitar” o controle de qualidade de fitoterápicos, esforço este, internacional, vide Process Analytical Technology (PAT).

Segundo Gonzalez (GONZALEZ,2002), a principal fonte de energia para imagens em uso nos dias de hoje é o espectro eletromagnético. Entre outras fontes pode-se citar a eletrônica (na forma de feixes de elétrons, usados na microscopia eletrônica). Imagens baseadas na radiação do espectro eletromagnético são mais familiares, especialmente as imagens de R-X e das bandas visíveis do espectro. Ondas eletromagnéticas podem ser conceituadas como propagação de ondas senoidais de comprimentos de ondas diversos, ou podem ser pensadas como um fluxo de partículas de menor massa, movimentando-se numa onda padrão, na velocidade da luz. Cada partícula de pequena massa contém uma certa quantidade de energia (ou feixe) desta. Cada feixe é chamado fóton. Sendo

Figura 1.3: Tipos de espectros ópticos. Extraído de Gonzalez, Rafael C., Woods, Richard E. 2002. Digital Image Processing. Upple Saddle River Prentice Hall Inc New Jersey.

as bandas espectrais agrupadas de acordo com a energia por fóton, tem-se o espectro da figura 1.3. Imagens na banda do infravermelho são utilizadas em microinspeção para a caracterização de materiais, sensoriamento remoto, entre outras aplicações.

A tabela 1.1 apresenta as bandas temáticas do LANDSAT da NASA, incluindo bandas nas regiões do visível e infravermelho.

Tabela 1.1: Bandas Temáticas Satélite LANDSAT da NASA. Expressas em comprimento de onda μ m igual a 10^{-6} m

Banda	Nome	Compr. de onda μ	Características e usos
1	Azul (VIS)	0.45—0.52	Penetração máx. na água
2	Verde (VIS)	0.52—0.60	Bom para medir a idade de plantas
3	Vermelho (VIS)	0.63—0.69	Discriminação de vegetais
4	IV Próximo	0.76—0.90	Mapeamento de biomassa e da costa
5	IV Médio	1.55—1.75	Quantia umidade vegetal e solar
6	IV Térmico	10.4—12.5	Umidade solar; mapeam. térmico
7	IV Médio	2.08—2.35	Mapeamento mineral

Os fundamentos do processamento digital de imagens residem na transformada de Fourier do sinal adquirido. Isto remete as observações para além do casamento ou reconhecimento de padrões de imagens digitais de plantas medicinais lidas, e neste sentido, vindo dos conceitos de processamento digital de imagens, chega-se ao fundamento deste tema, o espectro do sinal e o espectro característico da amostra, sendo usado nas demais etapas do processo em farmacognosia, especialmente na terceira fase, da análise fitoquímica.

Segundo (SALA, 2002), várias técnicas possibilitam a captura de informações sobre estrutura molecular, níveis de energia e ligações químicas, no caso a espectroscopia Raman e o infravermelho. Pode-se restringir o estudo aos aspectos vibracionais, rotacionais (e de rotação-vibração) de moléculas dinâmicas, passando para moléculas poliatômicas. A espectroscopia estuda a interação da radiação eletromagnética com a matéria, sendo um de seus objetivos principais a determinação dos níveis de energia dos átomos ou moléculas. Diretamente obtém-se as diferenças (transições) entre estes e a partir destas medidas, determinam-se as posições relativas dos níveis energéticos. No caso de moléculas, a região espectral depende do tipo de níveis envolvidos, sendo as transições eletrônicas situadas na região do ultravioleta ou visível, as vibracionais na região do infravermelho e as rotacionais na região de micro-ondas (casos particulares, também no infravermelho longínquo). Cada tipo de espectrômetro tem uma tecnologia adequada. Desconsiderando a energia devida aos movimentos translacionais, a energia total de uma molécula é a soma da energia eletrônica (maior que a vibracional), da energia vibracional (maior que a energia rotacional) e a energia rotacional (caso de moléculas gasosas). Cada tipo de espectro pode ser estudado independentemente do tipo de interação entre os níveis. A separação entre os movimentos dos núcleos e dos elétrons, conhecida como aproximação Born-Oppenheimer, resulta da grande diferença de massa destes. Pode-se considerar a posição dos núcleos fixada, durante a transição eletrônica (movimento dos elétrons é muito mais rápido). Durante o movimento dos núcleos, pode-se considerar uma distribuição média dos elétrons.

A interação da radiação eletromagnética com o movimento vibracional dos núcleos origina o espectro vibracional no infravermelho ou o espalhamento Raman. Os espectros no infravermelho são usualmente observados por absorção, também, por emissão. Indiretamente, podem ser observados os espectros infravermelhos, transferindo-os para a região do visível, através do espalhamento Raman. As frequências são obtidas pelas

diferenças entre frequências das radiações espalhadas e da radiação incidente. A absorção no infravermelho ocorre, quando a frequência de radiação multiplicada pela constante de Planck, tem o mesmo valor da diferença de energia entre dois estados vibracionais, trata-se de uma ressonância entre a diferença de níveis de energia da molécula e a radiação eletromagnética. No espalhamento Raman, uma radiação, geralmente na região do visível ou ultravioleta, interage com a molécula e é espalhada com frequência ligeiramente modificada. Trata-se da diferença de energia entre dois estados vibracionais. Nos mesmos estados vibracionais, ambas as frequências são iguais. Para um modo vibracional ser ativo no infravermelho é necessário haver variação no momento dipolar durante essa vibração.

A atividade no Raman difere no sentido de que o momento de dipolo considerado é o induzido pela radiação eletromagnética, deve haver variação da polaridade da molécula durante a vibração. As medidas de intensidade das bandas envolvem mecânica quântica. A análise vibracional envolvendo somente os valores das frequências pode ser feita com noção das propriedades de simetria e pode-se fazer previsões dos espectros, quanto ao número de modos vibracionais ativos no Raman e infravermelho, com uso da teoria de grupos, com base matemática. Na espectroscopia Raman, a introdução da radiação de laser pulsado de rubi, como fonte de excitação, deu-se em 1962; em 1963, utilizou-se pela primeira vez o laser contínuo de HeNe (632,8nm); e, no final da década de 1960, o desenvolvimento dos lasers de Ar^* ⁵ e de Kr^* ⁶, associados com duplo ou triplo monocromador, detecção com fotomultiplicadoras, efeito Raman ressonante, efeito Raman intensificado por superfície, efeito Raman inverso, espalhamento Raman anti-Stokes, etc.

Atualmente detectores multicanais permitem a obtenção de espectros em frações de segundos. Os espectrômetros modernos no infravermelho avançaram com o desenvolvimento dos instrumentos interferométricos, usando transformada de Fourier, em meados da década de 1960, sendo sua faixa estendida a $10cm^{-1}$, permitindo obter-se considerável região espectral de uma única vez, em curto tempo. Desde a caracterização de compostos extremamente instáveis (vida muito curta) a moléculas complicadas,

⁵O laser de argônio é de uma família de lasers de ions utilizando gás nobre. Este laser emite em vários comprimentos de onda através do espectro visível e ultravioleta: 351 nm, 454,6 nm, 457,9 nm, 465,8 nm, 476,5 nm, 488,0 nm, 496,5 nm, 501,7 nm, 514,5 nm, 528,7 nm.

⁶Laser de ions de gás criptônio. Quando o criptonio é misturado com argônio, criam-se lasers tipo "white-light". Emitem em vários onda através do espectro visível: em 406,7 nm, 413,1 nm, 415,4 nm, 468,0 nm, 476,2 nm, 482,5 nm, 520,8 nm, 530,9 nm, 568,2 nm, 647,1 nm, 676,4 nm.

podem ser estudadas, usando computador, com rapidez e grande refinamento.

Ora, identificar plantas medicinais é então um processo semelhante a análise de uma droga sintética? Observa-se o seguinte:

- A composição química de uma droga sintética é previamente esperada (conhecida), com o uso da espectroscopia no infravermelho próximo. Por exemplo, pode-se verificar o resultado esperado tendo sido obtido ou não na droga fabricada, industrializada; a composição química de uma planta medicinal é bastante variável, pois “mãe natureza” determina uma grande variabilidade, ainda que tratando-se da mesma espécie, por exemplo, um Jaborandi plantado no interior do estado de São Paulo, pode apresentar diferentes comportamentos sazonais, devido a mais ou menos umidade, ainda, a mesma planta medicinal encontrada na Amazônia, apresenta outras especificidades devido ao clima da região, enfim, não obstante as distinções dentro da mesma espécie, o Jaborandi ⁷, possui pelo menos 3 espécies catalogadas, neste trabalho.
- Considerando-se a questão de catalogar plantas em herbário, se constata que esta é uma tarefa muito especializada, requerendo conhecimento especializado de taxonomia ou sistematização de plantas, inerentes a especialistas em fitoquímica, junto a herbários e/ou especialistas, pesquisadores na academia. Pergunta-se, poderia haver um método que auxiliasse nesta tarefa, sendo um facilitador?
- Ainda, observando-se o processo em farmacognosia, poderia haver um método que responde-se a pergunta - É Jaborandi? Rapidamente, em comparação com o dito processo utilizado em farmacognosia? De modo análogo, as perguntas e respostas poderiam ser obtidas para o processo de catalogação em herbários (processo praticado por taxonomistas, principalmente)?

⁷Família: Rutaceae, Nome Botânico: *Pilocarpus Jaborandi* Holmes, Nomes Populares: Jaborandi, Jaborandi-de-pernambuco, jaborandi-verdadeiro; Parte utilizada: folhas. Constituintes químicos: alcalóides (Classe de substâncias muito encontrada nos vegetais, de propriedades básicas, e de grande interesse medicinal. Só substâncias químicas presentes numa variedade de plantas medicinais; possuem uma estrutura nitrogenada, de reação alcalina, com um ou mais átomos de carbono, normalmente em estrutura cíclica. Apresentam atividades farmacológicas marcantes e muito diversificadas.) (1%): pilocarpina (0,5%), isopilocarpina, pilosina, pilocarpidina; óleo essencial (0,25%): cariofileno, 2-trideconona, outros terpenos; taninos; ácido jabórico; ácido policárpico; resinas. Indicação Terapêuticas: Tônica, diaforetica, sialagoga, contra afecção catarral (via respiratória, bronquite, asma úmida, edema, pleurisia, pneumonia, difteria, nefrite aguda, “mal de bright”(f.congestivas), hidropsia renal, intoxicação, diabete pancreática, doença de pele, hiperidroses gerais, hiperidroses locais, oftálmica. In <http://www.plantamed.com.br/plantaservas/generos/Pilocarpus.htm>

- Dá-se início ao estudo desta proposta a partir da Farmacopéia Brasileira, sendo eleito o Jaborandi um objeto de estudo. Esta proposta, criativamente, evolui para a necessidade de uma base de referência robusta que contém pelo menos 70 drogas vegetais catalogadas pelo herbário ESA, pertencente a ESALQ-USP.

1.2 Objetivos

1.2.1 Objetivo Principal

Propor um método qualitativo (rápido), pensado para ser de baixo custo, para identificar plantas medicinais, que auxilie no seu processamento em ambientes como o de farmácias de manipulação.⁸

1.2.2 Objetivos Secundários

- Instrumentar, com sistema computacional, programas, em sua primeira versão, chamada Beta, o controle de qualidade de fitoterápicos, de modo que sua prática seja propiciada (facilitada) na rotina das farmácias de manipulação;
- Proceder a análise comparativa (investigativa) da contribuição entre os diferentes métodos pesquisados, como análise de fatores e lógica fuzzy.
- Criar uma base de espectros de plantas medicinais, em torno de cem espécies;
- Criar um método rápido alternativo para taxonomia de plantas, que seja um auxiliar dos taxonomistas.

1.3 Descrição do trabalho

A descrição da presente tese foi desenvolvida premiando os seguintes ações teóricas e práticas:

- Descrever os passos metodológicos desta tese de doutorado no sentido dos resultados obtidos, situando desde o início do processo criativo com a constatação das

⁸Ferramental para controle de qualidade de fitoterápicos.

carências no tema, que motivaram o trabalho, passando pelas idéias (seu estudo) para sua implementação;

- De acordo com os problemas constatados, introduzir, sem o objetivo de esgotar, senão remeter ao contínuo aprofundamento, os conceitos (fundamentos) teóricos que envolvem as grandes linhas de conhecimento abrangidas por este projeto, interdisciplinar: a espectroscopia na região do infravermelho próximo, bem como as técnicas quimiométricas aplicadas neste trabalho, a lógica *fuzzy*. Do mesmo modo, estudo da flora, especialmente a brasileira, no que tange as plantas medicinais e seus processos em farmacognosia (parcialmente), exemplificados na extratificação do Jaborandi, para citar. Passando, também, pelo trabalho especializado de catalogação em herbários.
- Descrever o desenvolvimento da prática laboratorial para obtenção da identificação de plantas medicinais caracterizadas, cujos espectros foram capturados usando espectrômetro Raman e infravermelho próximo da Bruker; Do mesmo modo, usando do software OPUS, da Bruker, pacote Ident, proceder a identificação da planta com grau de certeza na faixa mínima de 95% de acertos.
- Descrever o desenvolvimento de um sistema baseado em lógica fuzzy para identificação de plantas medicinais, usando o software Matlab e a linguagem de programação C. Submissão do sistema a testes com a massa de dados adquiridos num volume tal que possam indicar a robustez da solução testada. Os testes foram implementados de forma analítica com esgotamento de todas as possíveis combinações de variáveis (conjunto completo das possibilidades de cruzamentos) para a identificação de drogas vegetais, especialmente para validar o sistema de programas *fuzzy* desenvolvidos.
- Descrever de modo analítico e sintético os resultados obtidos nos diferentes métodos comparados.

No presente capítulo e no decorrer dos capítulos seguintes, esta ordenação das idéias, teórico/práticas, estão descritas.

2 Aspectos Teóricos

2.1 Métodos utilizados para análise de plantas medicinais

Modelos foram desenvolvidos para representar a interação da luz com diferentes camadas celulares, em relação as características anatômicas e bioquímicas, estando destacada nas figura 2.1, que representa uma folha. Stéphane Jacquemoud, em sua página *web*¹ intitulada *Leaf Optical Properties Database* descreve um dos modelos para planta vascular, sendo formado pelo tecido vascular da veia principal, delineado a partir do mesófilo esponjoso por uma camada compacta de células do parênquima, chamada bainha. Células envoltentes podem ou não conter uns poucos cloroplastos. Fechado, dentro da bainha do feixe, estão o xilema adaxial, floema abaxial e algumas células esclerênquima de suporte. Uma rede de conexão com vertentes vasculares, constitui um sistema contínuo, ao longo da folha com ramos de dimensões cada vez mais finas, provenientes da veia principal. Devem ser evitados pontos de coleta de dados espectrais ao longo das nervuras principais das folhas, para reduzir as variações nas medidas espectrais. E, apresenta outros modelos. Diversos artigos científicos relatam o uso de IVP para diferentes fins que culminam com diferentes tipos de controle de qualidade de diversas substâncias.

O uso experimentado de espectroscopia IVP para identificação de drogas derivadas de plantas, com ênfase em análise quantitativa, também, relatada em (CORTI, 1990).

- Conforme (OSBORNE, 1988) é apresentado um novo método para análise discriminante ou qualitativa de chá preto comercial (método rápido de análise de alimentos) usando espectroscopia IVP, com a técnica de amostragem da reflectân-

¹<http://teledetection.ipgp.jussieu.fr/opticleaf/sommaire.php3>, [10/29/09], Inst. Physique du Globe da Université Paris 7 Denis Diderot.

Figura 2.1: Representação esquemática de seção vertical de uma folha normal, incluindo a veia principal e tecidos circundantes. (Extraído de Oi Wah Liew, Pek Ching Jenny Chong, Bingqing Li and Anand K. Asundi, *Signature Optical Cues: Emerging Technologies for Monitoring Plant Health*, **Sensors** 2008, 8, 3205-3239)

cia. Neste caso são eleitos os valores de 1660, 1720, 2050 e 2300 nm, correspondendo às bandas de absorção dos chás pretos. Foi usada a distância estatística (Mahalanobis) para o cálculo. Na indústria de chás a qualidade é assegurada por testadores que consideram a sua aparência e outras características, sendo um procedimento subjetivo e laborioso. O uso de IVP tem sido limitado devido a natureza imprecisa da calibração dos dados.

A análise estatística é procedida sobre os dados gravados usando $\log(1/R)$ sendo R a reflectância, com resolução com intervalo de 2nm na faixa de 1200 a 2400nm. (R é definida como sendo $\frac{P_s}{P_o}$ onde P_s é a potência da radiação refletida pela amostra (*Power*) e P_o o mesmo da amostra padrão (cerâmica padrão), obtida à priori para cada amostra). O seu experimento executou 50 *scans* em 30s e as amostras foram lidas ao acaso. É usada a segunda derivada $\frac{d^2 \log(1/R)}{d\lambda^2}$ como técnica de pré-processamento e o desvio padrão é apresentado graficamente. Os autores acreditam que a segunda derivada consegue minimizar as diferenças oriundas do tamanho das partículas, granularidade das diferentes amostras e supõem ter resolvido o problema de diferenças de espalhamento, tendo obtido sinais de absorção. Finalmente afirmam as bandas citadas como sendo as bandas de espectro IVP para chás pretos. Esclarecem que as bandas 1440 e 1920nm correspondem a absorção da água, bandas de umidade, rejeitadas para a presente análise. Especificamente os comprimentos de onda/respectivos grupos: 1660nm/aromático, 1720, 2300/metileno, e 2050nm/hidroxila. Os autores usam a distância Mahalanobis para investigar os intervalos muito próximos que contemplam a variabilidade das amostras encontradas no mercado. Em seu procedimento, contaminam as amostras com um constituinte extra numa quantidade conhecida para verificar se o sensor consegue detectá-la. Usando os dados da reflectância com $p = 4$ comprimentos de onda selecionados e uma métrica definida em $1/p$ vezes a matriz de variância do conjunto A , a distância estatística é calculada. Os autores usam *cut-off* de observações anômalas ou dados discrepantes ou aberrantes (*outliers*) e obtém taxa de sucesso de 91% e prometem estudar um filtro, porque sentem que o uso de só quatro comprimentos de onda os impede de obter um desempenho melhor e prometem aumentar o número de variáveis (a complexidade). Também, afirmam *It is also considered important that a chemical basis be established to justify a relationship between the reported absorption bands and black tea quality.*

- Conforme (BAKALLI, PESTI, ETHERIDGE, 2000) refere o uso da espectroscopia IVP(EIVP) com um método rápido e não destrutivo da amostra, para a predição de composição química de alimentos e ingredientes de alimentos, incluindo úmidade, proteína, carboidratos, gordura, etc. Os procedimentos convencionais de laboratório consomem tempo e custo. Para os procedimentos da EIVP não envolvem reagentes, nem geração de produtos residuais. Um dos fatores mais importantes que influenciam a acurácia e a precisão da análise com o equipamento EIVP é o tamanho médio de partícula das amostras. O tamanho médio de partícula, por sua vez, foi influenciado pelo método de moagem. Isto foi observado em determinações de proteínas, óleo e a úmidade usando EIVP. Assim sendo, foram procedidos testes comparativos entre os métodos convencionais, técnicas de laboratório padronizadas (pelo órgão de controle de qualidade *AOAC-The Association of American Feed Control Officials*) e a EIVP no sentido da determinação do tamanho da amostra, ensaios de úmidade, de proteína, de gordura, de fibra bruta.

O equipamento de EIVP utilizado neste estudo era um espectrofotômetro (NIRS IS1-2 da Intrasoftware International), cujo *scanner* óptico contém uma fonte de luz (tungstênio), com esfera de integração (para captura da luz refletida da amostra) com dois fotodetectores de sulfeto de chumbo. A luz que colide com a amostra é difusamente refletida. Um dos detectores a captura e o sinal é amplificado usando o amplificador de resposta logarítmica e gravado como um $\log(1/R)$ onde R é a reflectância. Com os dados foram executadas: análise de variância, coeficientes de correlação, diâmetro geométrico médio e desvio padrão geométrico do tamanho da partícula.

Os autores revelam vários resultados onde a EIVP foi superior aos resultados obtidos por processos químicos, tanto para sementes de soja, proteína, do mesmo modo para a fibra bruta, etc, todavia nas suas conclusões são cautelosos afirmando que os métodos se equivalem e os resultados para fibra dependem do método de moagem e finalmente da dependência do fabricante no sentido dos ajustes finos das curvas de calibração para se obter maior acurácia.

- De acordo com (WENNING; SEILER; SCHERER, 2002) foi procedido um estudo para identificar leveduras. As células crescem em microcolônias de 70 a 250 μ m em diâmetro. Os espectros IV foram registrados usando microscópio IV

acoplado em um espectrômetro IV e a identificação foi procedida por comparação ao espectro de referência. Os registros foram equivalentes, com a vantagem para a referida como microespectroscopia, de requerer menor tempo de preparação da amostra. A microespectroscopia é referida como uma ferramenta de análise de populações mista, rápida. Com a ressalva de que quatro espécies de microrganismos, apresentaram grau de confiança de 91% a 92% de um total de 45 cepas. Interessante neste estudo é observar que ferramentas como a análise de *clusters* também foram empregadas, tendo sido apresentados os dendogramas por espécie com reprodutibilidade sendo destacada a sua SD (distância espectral). No pré-processamento foi empregada a segunda derivada e diferentes janelas espectrais.

- Conforme (YOON et al., 2004), foi utilizada a EIVP com reflectância, técnica rápida e preservadora da amostra, sendo mostrada a sua sensibilidade às propriedades físicas e químicas da amostra em combinação com vários métodos da quimiometria (análise de componentes principais, distância Euclideana, sistema de qualificação polar, coeficiente de correlação).
- Schulz entende que a fim de garantir a qualidade dos medicamentos utilizados como matérias-primas para a produção de cosméticos, extratos de materiais ou aroma dos produtos fitofarmacêuticos, geralmente a GC e o HPLC combinado com sistemas de detecção de diferentes (por exemplo, FID, MS, UV-Vis, fluorescência) são utilizados. Uma vez que estes métodos são mais demorados e caros comparativamente, algumas tentativas têm sido feitas para encontrar opções alternativas de análise. Neste contexto, vários novos métodos de espectroscopia vibracional, em combinação com sofisticados algoritmos quimiométricos foram introduzidos com sucesso nos últimos anos (por exemplo, usando IVP, ATR-IV e técnicas de espectroscopia Raman) para uma determinação eficaz dos componentes individuais de óleo essencial (SCHULZ, 2004). Considerando que os dados NIRS (EIVP) podem ser interpretados apenas pela aplicação de métodos estatísticos, os espectros Raman na maioria dos casos apresentam bandas características fundamentais e, portanto, em princípio, permitem com mais facilidade a discriminação de diferentes espécies de vegetais ou quimiotipos. A espectroscopia Raman também tem provado ser um método rápido e confiável para o controle de qualidade de vários extratos isolados que requerem tanto a preparação da amostra mínima e quantidade mínima do analito. Além disso, esta técnica

permite, principalmente, o identificar de modo não-destrutivo de vários componentes em materiais vegetais frescos se as bandas características das moléculas podem ser encontradas no espectro (SCHULZ; QUILITZSCH; KRUGER, 2003), (SCHULZ; BARANSKA; BELZ, 2004B). Neste contexto, especialmente IVPTF e Raman tem sido descrito como uma ferramenta valiosa para as investigações, porque a fluorescência e decomposição térmica dos tecidos vegetais pode ser reduzida a um mínimo. Mapeamentos usando Raman permitem que se obtenha imagens espectroscópicas de 2 dimensões das amostras investigadas, que podem ser diretamente comparados com a imagem correspondente visual. Além disso, o mapeamento fornece informações detalhadas sobre a distribuição de substâncias vegetais específicas que ocorrem na camada superficial (BARANSKA, 2004), (SCHULZ, 2004C). Com base nestes dados, não apenas qualitativa, mas também interpretações semi-quantitativas das medições espectroscópicas são possíveis. O mapeamento Raman não precisa de nenhuma preparação da amostra e também pode ser realizado em material como planta viva.

- Conforme (MAURER, 2008) o Food and Drug Administration (FDA) promoveu a iniciativa Process Analytical Technology (PAT) cujo objetivo era o entendimento dos avanços nos processos e controle através de diferentes ferramentas, encorajando o desenvolvimento e a implementação de sistema inovadores no desenvolvimento da farmácia, manufatura e garantia da qualidade. Tecnologias analíticas fazem parte destes sistemas; outras partes são, o projeto estatístico de experimentos ou estratégias para monitoramento de processos. Neste trabalho foram avaliadas diferentes técnicas como Terahertz Pulsed Spectroscopy (TPS) e imageamento (TPI e EIVP) na manufatura em escala completa de produtos farmacêuticos sólidos. A EIVP serviu para identificação de matéria-prima e determinação da umidade durante a secagem. Entende-se que o número de aplicações é crescente e a EIVP pode ser uma ferramenta PAT valiosa. Também, é procedida a comparação qualitativa de imagens obtidas da superfície. Entretanto análise quantitativa, neste caso, não foi possível, pois o método precisaria ser adequado para uso rotineiro num ambiente industrial. Outro uso tem sido para estimar o tamanho da amostra analisada, a profundidade de penetração da radiação IVP, em tabletes de revestidos com filmes. A vantagem do IVP é a sua resolução alta em comparação com TPI.

- (EBEL, 1992) apresenta métodos *fuzzy* para a análise de espectros IVP.

Um cientista se esforça para trabalhar exatamente e para obter resultados claros. Neste contexto, é importante dar um passo atrás do significado normal da palavra *fuzzy* e para ver como os matemáticos consideram um modelo que é representado pelo termo *fuzzy*.

- Em matemática, a expressão *fuzzy* foi introduzida por Zadeh. Em um conjunto não-nebuloso (nítido, *crisp*) de dados, cada elemento pode assumir apenas um determinado valor. Um dado conjunto *fuzzy* pode ser descrito por uma função de pertinência, que se caracteriza pela largura γ uma medida de sua pouca nitidez (*fuzzyness*) e pode assumir valores entre 0 e 1. Os valores negativos podem ser consideradas como um vetor de dados $a(\lambda)$. Uma relação quadrática é escolhida e aplicada como uma função de pertinência apropriada. Um espectro de IVP pode ser considerado como um vetor de dados $x(\lambda)$. Considerando-se um espectro como um conjunto difuso, pode ser representado por uma função de pertinência. Esta função é dependente do comprimento de onda, onde a variável $x(\lambda)$ esta sendo definida na dimensão da absorbância. Dois algoritmos para comparação espectral, baseada em métodos Fuzzy, foram testados quanto à sua adequação na região IVP, descritos a seguir.

Ebel apresenta o algoritmo para o reconhecimento de padrões de acordo com Blaffert, (BLAFLERT, 1984).

O padrão de espectro de referência é tomado como *fuzzy*, o espectro da amostra como *crisp*. No primeiro caso estudado a localização dos máximos de absorbância característica no espectro de referência são descritas por funções de pertinência. O padrão da amostra ($\lambda_{s,max}$) é colocado sobre o padrão de referência ($\lambda_{r,max}$). Os cruzamentos podem ser interpretados como valores de adesão $msf(\lambda_{s,max})$. Uma medida de acordância N, resulta da soma das interseções, dividida pelo maior número de amostra, ou máximo de referência n_{max} . No segundo caso estudado o padrão de amostra não nebuloso ($\lambda_{s,max}$ e $A_{\lambda,s}$) é colocado sobre o padrão de referência *fuzzy* ($\lambda_{r,max}$ e $A_{\lambda,r}$), como no primeiro caso. A medida de acordância é calculada utilizando posições e os valores da absorbância máxima.

Ebel, também, apresenta o algoritmo para o chamado método das diferenças.

Ambos os espectros de referência e da amostra são considerados como conjuntos

fuzzy. Uma função de pertinência para a diferença de dois números *fuzzy* a e b é usada. A diferença de dois vetores de dados *fuzzy* (espectros) utiliza um vetor de dados (valores de absorvância) $(a_s(\lambda))$ do espectro da amostra e o vetor de dados $(a_r(\lambda))$ do espectro de referência. De acordo com Otto e Bandemer (OTTO; BANDEMER, 1986) um critério de desvio N_d pode ser utilizado. Ao inserir os dados de referência pode ser utilizado o critério de desvio menor possível. Para calcular o $N_{d,max}$ maior, só os dados de referência são utilizados. Um valor γ_d é tomado como a largura apropriada. A medida de acordância M precisa ser normalizada com valores entre 0 e 1. Caso aconteça de $(N_d - N_{d,min} > N_{d,max})$, então M é definido como 0. M se aproxima de 1, para espectros correspondentes. Ebel apresenta exemplos de resultados práticos da aplicação, dos métodos *fuzzy* sugeridos, aos dados de IVP. São fornecidos dois exemplos de métodos *fuzzy* aplicado a dados IVP. Em ambos os casos espectros foram coletados usando um espectrômetro Pelkin-Elmer Lambda 9 UV/VIS/NIR, com esfera de integração de 60 milímetros de ouro laminado. O pacote de software PECSS para UV / VIS / IVP foi utilizado para processar os dados. Para o primeiro exemplo, os espectros de Cetiol foram coletados no modo de transmissão entre 862 e 2250nm, dois tipos de Cetiol: Cetiol V e Cetiol HE. Estes foram comparados. Diferentes concentrações de misturas de Cetiol V no Cetiol HE, também, foram comparados com Cetiol HE. Para o segundo, propafenona-HCL de diferentes fabricantes, e os diferentes tamanhos de partículas, foram medidas no modo de refletância entre 862 e 2500nm. Em ambos os casos, os algoritmos citados foram aplicados aos espectros. Com os resultados para os espectros de Cetiol e os resultados para os espectros de propafenona HCL, foram arroladas as seguintes conclusões.

- A influência de um ou mais deslocamentos podem impedir a correta identificação das amostras. Essa dificuldade pode ser minimizada por meio de reconhecimento de padrões propostos no caso 1.
- Os espectros são apenas deslocados na dimensão da absorvância, as posições dos máximos permanecem as mesmas. Se, por exemplo, um deslocamento causado por diferentes tamanhos de partículas é um fator importante, o caso 2 leva a bons resultados, porque os valores de absorvância influenciam o grau de concordância.
- Impurezas podem influenciar os valores de absorção de alguns comprimentos

de onda, sem mudar as posições dos máximos. Neste caso, a detecção de impurezas, que não é possível com o processo de reconhecimento de padrões caso 1, mas é possível com o caso 2, no exemplo prático, tratando-se de uma concentração menor que 2,5%.

- Com o método das diferenças, todo o espectro é considerado. Cada diferença entre referência e espectros da amostra são registradas. Isso evita o risco de que informações espectrais sejam inutilizadas. A aplicabilidade deste método é considerada boa.

3 *Assinatura Espectral no Infravermelho Próximo*

Pretende-se aqui apresentar o trabalho desenvolvido que culminou com a identificação de plantas medicinais catalogadas, conforme as tabelas 3.1 e 3.3, a serem apresentadas no decorrer deste capítulo. Estas setenta plantas medicinais catalogadas da referida tabela compreendem a entrada para o sistema de assinatura espectral IVP, iniciando com a aquisição, seguida de processamento e/ou tratamento destas, resultando na sua identificação com o uso do conjunto de programas OPUS da Bruker Optics do Brasil.

O sistema é formado, em nível inicial ou zero, das seguintes partes abaixo apresentadas. Antes de enumerá-las cabe aqui uma ressalva muito importante. É necessário que seja visualizado que os processos são aplicados na formação da base de referência e parcialmente, recorrem, na fase que uma amostra teste é checada contra a base de referência. E, aproveitando o ensejo, tudo isto pode ser visualizado, também, em um dendograma da análise de *clusters*.

- (1) Entradas: Basicamente, é a matéria prima, para um sistema auxiliar no controle de qualidade de fitoterápicos, plantas medicinais catalogadas, isto é, drogas vegetais, que passaram apenas pelo processo de secagem e/ou moagem, sem qualquer outro tipo de tratamento, como a remoção de contaminantes. Do mesmo modo que as plantas medicinais catalogadas formam a amostra ou espécimen referência para o sistema, ocorre a aquisição de drogas vegetais provenientes do comércio, em geral, formando as amostras teste do sistema, amostras que são identificadas, processadas pelo sistema, contra uma determinada amostra de referência ou versus a base de referência completa.
- (2) Processamento: Basicamente, no sistema, durante o seu desenvolvimento

(fase anterior ao seu uso rotineiro ou fase de produção (uso) em laboratório) foi procedida a divisão do processamento em três grandes módulos ou fases (optou-se pela análise ou modelagem funcional do processo);

- (2.1) primeira fase - aquisição e validação (como a reprodutibilidade por espécimen - 6 a 7) das amostras de referência, fase repetida, de modo similar, para as amostras teste;
 - (2.2) segunda fase - todo e qualquer pré-processamentos dos espectros adquiridos, inclusive a seleção de regiões de interesse (ROI);¹
 - (2.3) esta fase é o processamento propriamente dito, de identificação da droga vegetal teste ao término da execução de diversos (sub)processos, com diferentes possibilidades ou fluxos de execução. Para citar pode se vislumbrar procedimentos de inicialização ou configuração de cenários quaisquer, com base, para citar na prévia decisão da ROI. São realizadas contagens, cálculos, etc. Por exemplo, pode ser utilizado o algoritmo padrão.
- (3) Com os resultados da fase anterior, prepara-se o *ranking* das respostas geradas pelo processamento e diversos modos de sua apresentação. Também são escritos os registros de dados inerentes a cada processamento conforme um dado fluxo de execução, inclusive os registros dos resultados apurados, o que permite a persistência (dos dados), conforme escolha do usuário, o registro de de qualquer tipo de processos executados, bem como de suas informações resultantes. O sistema mostra (possui “inteligência” para mostrar). também, metadados, como registros completos dos procedimentos na fase de aquisição de espécimen (espécimes), isto é, registro completo do(s) experimento(s), inclusive com mecanismos essenciais de controle de acesso, de tal forma utilizável para controle de qualidade em ambiente de produção em fábrica. Por exemplo, de acordo com a inicialização ocorrida

¹Nesta fase o processamento pode ou não assumir aspecto supervisionado, isto é, de acordo com o interesse do usuário (inteligência humana, *mindscape*) decisões são tomadas com base em conhecimentos prévios do comportamento do sistema, obtidos ao acaso. Por exemplo, o arquiteto do sistema informa ao usuário que o mesmo se acomodará para processar qualquer tipo de amostra teste. O usuário resolve testar para perceber tendências no comportamento do sistema, sendo testado (o sistema) com os procedimentos de teste *piloto*. Isto analiticamente pode corresponder a processamento numérico considerável, portanto, adequado para computadores, onde filtros do usuário, *queries*(consultas) são processadas em máquinas cliente/servidor, além dos fundamentais relatórios sintéticos que podem ser gerados.

na fase (2.1) são ranqueados dez relevantes (parâmetro configurável), correspondentes sucessos na lista de identificação de assinatura spectral.

A seguir, nas seções seguintes estão descritos todos os passos do desenvolvimento da identificação da população (N) de 100 drogas vegetais adquiridas pelo sistema.

3.1 Aquisição dos Espécimes

3.1.1 Drogas Vegetais

Parte das amostras ou espécimes podem ser apresentados através de imagens apresentando plantas medicinais embaladas e/ou exsiccatas, apresentadas no anexo C.².

Jorge (JORGE, 2000) afirma que sem a observação da legislação e a correção de suas lacunas, todo avanço que as pesquisas científicas realizem fica desvalorizado, pois permanece o descrédito e o desprestígio dessa medicação(os fitoterápicos) em função da insegurança que seu uso apresenta.

A farmacopéia Brasileira (BRASIL. Ministério da Saúde, 1988), primeira edição, registra as seguintes proporções relativas das diversas partes vegetais oficializadas como drogas: raízes e rizomas (31,6%), folhas (22,4%), cascas caulinares (20,4%), flores e frutos (19%), sementes (6,6%).

A coleta predatória pode levar à extinção de espécies, principalmente quando os órgãos subterrâneos são o objeto de interesse medicinal. A ipecacuanha e o jaborandi são exemplos fortes de vegetais importantes para a fitoterapia, que, mesmo assim, correm risco de extinção. O que dizer então das espécies ainda desconhecidas ou mal conhecidas taxonomicamente?

Quanto ao aspecto morfológico das folhas, estas podem apresentar-se amarrotadas, inteiras, fragmentadas ou pulverizadas. Elas podem ainda estar contaminadas ou mesmo modificadas por misturas com outros órgãos da mesma planta ou de outra, constituindo adulteração, fraude ou falsificação. Jorge exemplifica: funcho em lugar de erva-doce (espécies diferentes da mesma família) constitui fraude; orégano misturado

²Todas estas imagens de drogas vegetais foram adquiridas pela própria autora.

com manjerona (espécies diferentes do mesmo gênero) constitui adulteração; mate com grande quantidade de pecíolos constitui falsificação; camomila com insetos constitui produto deteriorado, corrompido ou avariado. A consistência da folha guarda relação com a resistência que ela apresenta a certas ações mecânicas, como flexão e pressão. Sob esse aspecto, a folha pode ser: dura, mole, friável e flexível. As Farmacopéias Brasileiras I e II apresentam descrições, tais como: consistência de couro, consistência de papel, consistência de membrana e carnosa ou succulenta (com parênquima aquífero).

Segundo Capellari, (CAPELLARI, 2009), ao estudar plantas medicinais e aromáticas, o principal desafio consiste em se conhecer quais as espécies, ou quais são os grupos de plantas, desta categoria? Ainda, que a maioria dos vegetais empregados como aromáticos possuem também efeito fitoterápico. Praticamente todas as espécies vegetais poderiam ser utilizadas em tratamentos alternativos de diversas doenças uma vez conhecidos seus princípios ativos, sua atuação no organismo humano e a forma de seu emprego para cada caso, isto é, a forma de extração dos princípios ativos, a estabilização dos mesmos, sua veiculação, a dosagem recomendada, a combinação com outros princípios, etc.

A fitoquímica se encarrega de estudar estas substâncias ativas, sua estrutura, sua distribuição na planta, suas modificações e os processos de transformação que se produzem no decurso de vida da planta, durante a preparação do remédio vegetal e no período de armazenagem.

As substâncias ativas das plantas medicinais são de dois tipos: os produtos do metabolismo primário (essencialmente sacarídeos), substâncias indispensáveis à vida das plantas que se formam em todos os vegetais verdes graças a fotossíntese; o segundo tipo de substâncias é composto pelos produtos do metabolismo secundário, ou seja, processos que resultam essencialmente da assimilação do nitrogênio. Estes produtos parecem frequentemente serem inúteis à planta, mas os seus efeitos terapêuticos, em contrapartida são notáveis. Na verdade estes produtos auxiliam a planta na sua defesa química contra ataques de herbivória.

Tratam-se designadamente de óleos essenciais (ou essências naturais), resinas, alcalóides como os da cravagem (um fungo do trigo) ou do ópio (da papoula). Geralmente estas substâncias se encontram na planta sob a forma de complexos, cujos diferentes componentes se completam e reforçam na sua ação sobre o organismo. Mesmo quando

a planta medicinal só contém uma substância ativa, esta tem sobre o organismo humano um efeito mais benéfico que o produzido pela mesma substância obtida pela síntese química. A substância ativa apresenta também um equilíbrio fisiológico. Pode-se citar como exemplo o ópio (*Papaver somniferum L.*) onde cada alcalóide isolado tem ação totalmente diferente do ópio no seu conjunto e provocam no organismo humano efeitos específicos, típicos e originais (efeitos farmacológicos). O mesmo se passa com os glucosídeos da dedaleira (*Digitalis purpurea L.*)

A natureza química da droga é determinada pelo seu teor em substâncias pertencentes aos seguintes grupos principais: alcalóides, glucosídeos, saponinas, princípios amargos, taninos, substâncias aromáticas, óleos essenciais e terpenos, óleos graxos, glucoquininas, mucilagens vegetais, hormônios e anti-sépticos vegetais.

Segundo Capellari, a taxonomia é a ciência que trata da identificação, nomenclatura e classificação dos objetos, dizendo respeito, em geral, apenas a objetos de natureza biológica; quando restrita às plantas, é denominada, frequentemente Botânica Sistemática.

Um sistema de identificação é formado pelas unidades chamadas taxon ou tâxones, no plural. Estas unidades são escritas em latim e seguem regras, determinações e recomendações contidas no Código Internacional de Nomenclatura Botânica. As unidades de um sistema estão divididas em unidades fundamentais e sub-unidades. As fundamentais estão presentes para a classificação de qualquer indivíduo, as sub-unidades apresentam-se apenas em certos casos, quando um taxon fundamental for demasiadamente grande. Capellari apresenta a seguinte lista:

Reino: Plantae

Divisão: radical+terminação phyta Subdivisão:phytina

Classe:radical+terminação opsida ou atae Subclasse:(idae),Coorte(iidae)

Ordem:radical+terminação ales Subordem:(inales), Superfamília:(ineales)

Família:radical+terminação aceae Subfamília:(oideae), Tribo:(eae),Subtribo:(inae)

Genêro:substantivo latinizado com inicial maiúscula Subgênero, Secção, Subsecção, Série

Espécie: gênero+designação ou epíteto específico (adjetivo latinizado) com ini-

cial maiúscula, seguido pelo nome abreviado de seu autor) Subespécie, Variedade, Subvariedade, Forma, Linha, Clone (estas denominadas categorias infra-específicas)

A bibliografia, de um modo geral, para plantas medicinais e aromáticas está baseada no sistemas de classificação Engler ou Conquist, sendo o sistema APG II (The Angiosperm Phylogeny Group, de 2003) também utilizado.

O reino vegetal pode ser dividido em dois grandes grupos de acordo com a reprodução das espécies: plantas inferiores (sem flores) ou Criptógamas e plantas superiores (com flores) ou Fanerógamas. O primeiro grupo são as algas, os líquens e musgos e as pteridófitas. As plantas com flores podem ser colocadas em duas categorias distintas principalmente de acordo com a presença ou ausência de um ovário envolvendo os óvulos e consequentemente formando um fruto. Entre as principais famílias de plantas medicinais e aromáticas estão as Pteridófitas, Gimnospermas e Angiospermas. As Pteridófitas, aparecem desmembradas em sete grupos, um exemplo de Gênero e espécies principais é a cavalinha, *Equisetum arvense* L. da família Equisetaceae, um exemplo de Gênero e espécies principais de Gimnospermas é o Ginkgo, *Ginkgo biloba* L. da família Ginkgoaceae, e um exemplo de Gênero e espécies principais de Angiospermas é o Jaborandi *Pilocarpus jaborandi* Holmes da família Rutaceae. Nesta família também podem ser encontrados: laranjas-azedas, limão-galego, lima tipo doce e arruda, nomes comuns. Ainda, Capellari elege para seguir o sistema de classificação de Cronquist (1981), citando também os sistemas APG II (2003), o de Glifford e Foster (1998) e o de Scagel et al (1965), para casos específicos.

Segundo Lorenzi (LORENZI, 2002), as plantas medicinais podem ser listadas de acordo com os respectivos usos etnomédicos. No presente trabalho foram objeto de captura inicialmente para fins de geração do espectro e posterior análise de resultados em torno de 100 amostras de plantas medicinais, com algumas repetições. Quando este trabalho de aquisição de espectros e testes de amostragem começou, isto ocorreu. Após, foram catalogadas pelo taxonomista, no herbário ESA (ESALQ-USP) 70 plantas conforme as tabelas 3.1 e 3.3

A seguir são apresentados detalhes do desenrolar da aquisição dos dados de entrada do sistema:

- Testes de amostragem e análises são procedidas para avaliar o potencial de resposta espectral das amostras; as amostras vão desde tintura de Jaborandi, pas-

Tabela 3.1: Plantas medicinais catalogadas pelo herbário ESA-ESALQ-USP.

Nome popular ¹	Família	Nome específico
panacéia	SOLANACEAE	Solanum cernuum Vell.
sete-sangrias	LYTHRACEAE	Cuphea cartaginensis (Jacq.) Macbr.
erva-doce	APIACEAE (UMBELLIFERAE)	Pimpinella anisum L.
funcho	APIACEAE (UMBELLIFERAE)	Foeniculum vulgare Mill.
laranja	RUTACEAE	Citrus sinensis (L.) Osbeck.
maracujá	PASSIFLORACEAE	Passiflora edulis Sims.
alfazema	LAMIACEAE (LABIATAE)	Lavandula stoechas L.
erva-de-sta-maria	AMARANTHACEAE	Chenopodium ambrosioides L.
chá-branco	THEACEAE	Camellia sinensis (L.) Kuntze.
tanchagem	PLANTAGINACEAE	Plantago major L.
pinhão-roxo	EUPHORBIACEAE	Jatropha gossypifolia L.
mentruz	BRASSICACEAE (CRUCIFERAE)	Coronopus didymus (L.) Sm.
poejo	LAMIACEAE (LABIATAE)	Mentha pulegium L.
melissa	LAMIACEAE (LABIATAE)	Melissa officinalis L.
erva-de-são-joão	ASTERACEAE (COMPOSITAE)	Ageratum conyzoides L.
unha-de-gato	RUBIACEAE	Uncaria tomentosa (Willd. ex Schult.) DC.
pulmonária	BORAGINACEAE	Pulmonaria officinalis L.
pau-tenente	SIMAROUBACEAE	Quassia amara L.
cipó-cabeludo	ASTERACEAE (COMPOSITAE)	Mikania hirsutissima DC.
malva	MALVACEAE	Malva sylvestris L.
verbena	VERBENACEAE	Verbena officinalis L.
chá-de-bugre**	BORAGINACEAE	Cordia salicifolia Cham.
douradinha	MALVACEAE	Waltheria douradinha A.St.-Hill.
stévia	ASTERACEAE (COMPOSITAE)	Stevia rebaudiana (Bert.) Bertoni
erva-de-bicho	POLYGONACEAE	Polygonum hydropiperoides Michx.
assa-peixe	ASTERACEAE (COMPOSITAE)	Vernonia polyanthes Less.
hortelã	LAMIACEAE (LABIATAE)	Mentha arvensis L.
porangaba**	BORAGINACEAE	Cordia discolor Cham.
menta	LAMIACEAE (LABIATAE)	Mentha x piperita var. citrata (Ehrh.) Briq.
chapéu-de-couro	ALISMATACEAE	Echinodorus grandiflorus (Cham. e Schltdl.) Micheli
sene	FABACEAE (LEGUMINOSAE)	Cassia angustifolia Vahl.
carobinha	BIGNONIACEAE	Jacaranda decurrens Cham.
jaborandi	RUTACEAE	Pilocarpus microphyllus Stapf ex Holm.
jaborandi*	RUTACEAE	Pilocarpus trachylopus Holm.
jaborandi	RUTACEAE	Pilocarpus pennatifolius Lem.
jaborandi	RUTACEAE	Pilocarpus jaborandi Holm.
confrei	BORAGINACEAE	Symphytum officinale L.

* jaborandi de identificação duvidosa, a ser comprovada. ** porangaba e chá-de-bugre são nomes que geram confusões, pois muitas vezes

sando por diluições. Neste período foi definido a rota do trabalho no sentido da aquisição e análise espectral de droga vegetal sólida. Para tanto foram executados estudos, descritos no decorrer.

- O Jaborandi foi escolhido como a planta que serviria de objeto de avaliação mais detida exemplar por se tratar de uma droga vegetal de amplo estudo e com suas especificidades descritas na Farmacopéia Brasileira, 4a Edição, além de fazer parte da lista de fitoterápicos controlados pela ANVISA.
- Farmacognosia é a ciência que tem por objetivo o conhecimento dos fármacos, limitados dos provinientes, dos dois reinos vivos, vegetal e animal, em particular, aos de origem vegetal. Pretende realizar o seu estudo nos diversos aspectos, da origem, culturas, colheita, preparação e conservação, identificação pelos seus caracteres macroscópicos e microscópicos, composição química, pesquisa de falsificações, atividade farmacêutica relacionada com métodos de aferição, usos diversos, por vezes, também, sob alguns outros aspectos, o histórico, o econômico, etc, ainda na separação, dos seus constituintes ativos, isolamento de secreções ou grupos de princípios imediatos e estabelecimentos das relações bioquímicas (corpos gordos, óleos essenciais, resinas, etc), mas entre todos, interessa-lhe, particularmente, os problemas relacionados com a identificação e análise. Várias ciências como a Botânica, a Química e Farmacodinâmica, são reputadas fundamentais (BRUNETON, 2001).
- A Farmacopéia Brasileira, na sua 4a edição, apresenta os métodos de análise com os procedimentos técnicos aplicados à medicamentos, métodos físicos e físico-químicos, químicos, de Farmacognosia, Biológicos, bem como os procedimentos aplicáveis aos ensaios biológicos, os radiofármacos, metodologias para teste de sensibilidade aos antibacterianos, recipientes e materiais empregados na sua fabricação, métodos de fabricação, substâncias corantes, reagentes, nomes simbólicos e massas atômicas, microorganismos empregados em testes e ensaios, etc. Portanto, os experimentos do ponto de vista farmacológico tem como embasamento a Farmacopéia Brasileira e outras como as Farmacopéias européias e americana. Como já, foi referido, as normas da ANVISA, notoriamente a RC048.
- A habilidade da espectroscopia de reflectância no IVP para rápida determinação de muitas propriedades químicas de plantas tem sido amplamente investigada.

Pode-se citar seu uso na oxigenação, como na quantificação de saturação de oxigênio nos tecidos obtida usando um método de múltiplas fontes, no domínio da frequência, baseado em diodos emissores de luz, em dois comprimentos de onda, inclusive são referidos os índices da saturação cerebral, parâmetros investigados por (KALRA, 1998). De modo análogo, foi procedido para saturação em músculo braquiorradial, tendo sido observadas declinações durante a isquemia do antebraço e a contração voluntária máxima, (KALRA, 1998).

- Na primeira fase do trabalho, as amostras foram preparadas para coleta do espectro corresponde a uma certa variável por exemplo a concentração em amostra líquida ou o tamanho da partícula em amostra sólida; A avaliação começa com o uso do método da transfectância com a diluição da amostra, em álcool, no caso a tintura (extrato padronizado) do Jaborandi e o mesmo para as amostras da Pilocarpina, seu princípio ativo. São utilizadas diferentes concentrações. No caso da Pilocarpina³ estas diferentes concentrações têm seus espectros comparados com o marcador da Sigma-Aldrich e com o marcador proveniente da empresa Sourcetechn. Quer-se caracterizar, usando IVP - quantitativamente as amostras coletadas. Para a obtenção de uma curva de calibração pretende-se utilizar, também, os métodos descritos (BRASIL. Ministério da Saúde, 1988) de espectrofotometria, gravimetria e cromatografia. A captura de espectros foi realizada utilizando o espectrômetro da Bruker Optics, IFS 28/N FRA 106/S (sendo utilizada a espectroscopia Raman, também). A resposta da transfectância IVP foi 20%, 80% da informação IVP da amostra líquida foi considerada com informação de difícil interpretação, com muito ruído. Passou-se a fazer amostragens sólidas utilizando espectroscopia Raman. Deste primeiro estudo com IVP foi rumado para a espectroscopia de “referência” Raman, resultou na caracterização quantitativa da Pilocarpina nas folhas estudadas, conforme tabela 3.2, trabalho a ser publicado no próximo ano. Esta tabela apresenta as diferentes concentrações testadas com uso da transfectância em IVP, bem como para as amostragens caracterizadas com espectroscopia Raman,

³Substância Química de Referência (marcador) para Análise da Sigma-Aldrich, número do produto: P6503, Nome do produto: Pilocarpine hydrochloride, > 98%(*trituration*), pó, Sigma, Número CAS 54-71-7, Fórmula molecular $C_{11}H_{16}N_2O_2.HCl$, Peso molecular 244,72, Re-teste recomendado em 5 anos. QC Data de liberação junho 2007. Aparência: pó branco, *titration*: mínima de 99%, [http://www.sigmaaldrich.com/catalog/LotNo ... 08/09/2008](http://www.sigmaaldrich.com/catalog/LotNo...08/09/2008)

Tabela 3.2: Espécimes caracterizados espectralmente - IVP e Raman - na Primeira Fase.

Planta	concentrações
Jaborandi <i>Pilocarpus microphyllus</i> (SourceTech)	folhas
Jaborandi	extrato folha redonda USP
Jaborandi	extrato folha comprida USP
Jaborandi	extrato mercado público
Jaborandi Panizza	folha pó 0,236 mm
Jaborandi Panizza	folha pó 0,150 mm
Jaborandi Sourcetek	folha 0,236mm
Jaborandi Sourcetek	folha 0,150mm
Jaborandi Folha comprida USP	folha 0,236mm
Jaborandi Folha comprida USP	folha 0,150mm
Jaborandi Folha redonda USP	folha 0,236mm
Jaborandi Folha redonda USP	folha 0,150mm
Princípio ativo - Padrões	concentrações
Pilocarpina Sigma	2,0g MeOH HCl 10%
Pilocarpina Sigma	0,2g 10 ml Metanol
Pilocarpina Sigma	0,1g 10 ml Metanol
Pilocarpina Sigma	0,1g 10 ml Etanol
Pilocarpina Sigma	0,1g H2SO4 0,25N
Pilocarpina Sigma	0,1g HCl 10%
Pilocarpina Sigma	0,2g HCl 10%
Pilocarpina Sigma	0,3g HCl 10%
Pilocarpina Sigma	0,4g HCl 10%
Pilocarpina Sigma	0,5g HCl 10%
Pilocarpina SourceTech	0,75g HCl 10%
Pilocarpina SourceTech	1,0g HCl 10%
Pilocarpina SourceTech	2,0g HCl 10%
Outros	concentrações
Metanol	
Sulfato quinino	
Quina	pó 0,150mm
HCl	10%
Álcool	70%
Boldo	pó 0,168mm
Boldo	pó 0,150mm
Goiaba	extrato hidroalcoólico

- O Jaborandi encontra-se na natureza sob várias espécies. Jaborandi do Maranhão, *Pilocarpus microphyllus* Stapf, família das Rutáceas. Jaborandi de Pernambuco, *P. jaborandi* Holmes, Jaborandi do Rio ou Paraguai, *P. pennatifolius* Lem., para citar. Parte usada são os folíolos. Os alcalóides são compostos cíclicos que

contém um átomo de nitrogênio básico, formam um grupo heterogêneo de substâncias orgânicas, em particular de origem vegetal, definido pela função da amina, que empresta seus constituintes com propriedades químicas próprias, às quais se aliam a uma toxicidade elevada e de muitas vezes também, atividade farmacológica notável. São reconhecidos por formarem sais com ácidos, à semelhança dos álcalis e de precipitarem destas soluções ou adquirirem cores características com os denominados reagentes gerais, Os alcalóides aparecem sobretudo nas plantas vasculares. Sua utilização pelas plantas é desconhecida. Inicialmente atribuiu-se o papel de proteção. A base orgânica predominante e de maior importância terapêutica é a pilocarpina, com estrutura determinada por análise, depois por síntese (COSTA, 1987).

- As amostras adquiridas e caracterizadas usando MPA FT-NIR da Bruker Optics. Conforme lista nas tabelas 3.1 e 3.3 de referência das 70 plantas, além disto, aproximadamente outras 30 plantas, também, tiveram seus espectros característicos obtidos, porém com a catalogação comprobatória, por taxonomista, pendente atualmente.
- Os experimentos fazem parte do *log*, registro de dados (metadados) dos arquivos do Opus. A seguir são apresentados alguns exemplos para ilustrar a captura de espectros no Opus.

3.2 Material e Métodos de Identificação

O pacote de *software* chamado Ident (Id), parte do sistema OPUS, permite a criação de uma biblioteca de referência, definida pelo usuário, bem como a geração de um método de identificação. A análise da identificação e os relatórios resultantes são explicados a seguir, bem como as teorias das rotinas de teste deste *software*.

A análise de *clusters* (conjuntos, aglomerados ou grupos) são descritas, também. São apresentadas todas as funções do Id, questões sobre as funcionalidades ou os problemas ao usar o Id. O Id é um pacote de *software*, projetado para identificar substâncias por seus espectros infravermelho. Um espectro desconhecido (espectro de teste) é comparado diretamente aos espectros de referência de uma biblioteca. O Id identifica aqueles espectros de referência que estão mais próximos do espectro de teste e determina os desvios padrão entre este espectro e o espectro de teste. Isto faz com que o Id identifique as substâncias desconhecidas, por exemplo, polímeros e avalie o grau de conformidade de uma substância com um padrão de referência. Esta última aplicação é uma tarefa tipicamente encontrada em controle de qualidade.

Para executar um teste de identidade primeiro é necessário ter uma biblioteca da referência a qual se compare com o espectro de teste. Se nenhuma biblioteca apropriada existir, é medido um conjunto de espectros de referência, isto é, pelo menos um espectro por substância. Entretanto, recomenda-se medir diversos lotes da mesma substância para habilitar o programa com obtenção de mais informações sobre possíveis variações existentes. Antes, as amostras têm que ser classificadas com um modo de análise de referência comum. Pode ser prosseguida a medição de um espectro a ser identificado, considerando que a etapa seguinte será gerar um método de identificação. Para executar um teste de identidade pode se selecionar diversos algoritmos e definir parâmetros da identificação usando o Id. Os resultados de um teste de identidade são armazenados em um relatório que inclui o resultado da análise, o método usado assim como os parâmetros definidos.

1. Primeiro - ajuste de um método de teste do Id - incorpora-se pelo menos um espectro de referência por substância em uma biblioteca, definindo uma escala espectral apropriada para a identificação, selecionando um método de processo dos dados, medição e análise do espectro de teste.

Tabela 3.3: Plantas medicinais catalogadas pelo herbário ESA-ESALQ-USP.

Nome popular	Família	Nome específico
camomila	ASTERACEAE (COMPOSITAE)	Chamomilla recutita (L.) Rauschert.
abacateiro	LAURACEAE	Persea americana Mill.
guaco	ASTERACEAE (COMPOSITAE)	Mikania glomerata Spreng.
baleeira	BORAGINACEAE	Varronia verbenacea (DC.) Borhidi.
ginkgo	GINKGOACEAE	Ginkgo biloba L.
alcachofra	ASTERACEAE (COMPOSITAE)	Cynara scolymus L.
cavalinha	EQUISETACEAE	Equisetum hyemale L.
jambolão	MYRTACEAE	Syzygium cuminni (L.) Skeels
cardo-santo	PAPAVERACEAE	Argemone mexicana L.
capim-cidreira	POACEAE (GRAMINEAE)	Cymbopogon citratus (DC.) Stapf.
centela-asiática	APIACEAE (UMBELLIFERAE)	Centella asiatica (L.) Urban.
arnica	ASTERACEAE (COMPOSITAE)	Porophyllum ruderale (Jacq.) Cass.
catinga-de-mulata	ASTERACEAE (COMPOSITAE)	Tanacetum vulgare L.
louro	LAURACEAE	Laurus nobilis L.
carqueja	ASTERACEAE (COMPOSITAE)	Baccharis trimera (Less.) DC.
guine	PHYTOLACCACEAE	Petiveria alliacea L.
espinheira-santa	CELASTRACEAE	Maytenus ilicifolia Reissek
amora	MORACEAE	Morus nigra L.
alfafa	FABACEAE (LEGUMINOSAE)	Medicago sativa L.
guassatonga	SALICACEAE	Casearia sylvestris Sw.
hibisco	MALVACEAE	Hibiscus sabdariffa L.
orégano	LAMIACEAE (LABIATAE)	Origanum vulgare L.
losna	ASTERACEAE (COMPOSITAE)	Artemisia absinthium L.
sálvia	LAMIACEAE (LABIATAE)	Salvia officinallis L.
pata-de-vaca	FABACEAE (LEGUMINOSAE)	Bauhinia forficata Link.
alecrim	LAMIACEAE (LABIATAE)	Rosmarinus officinalis L.
graviola	ANNONACEAE	Annona muricata L.
avenca	ADIANTACEAE	Adiantum pentadactylon Langsd. e Fish.
artemisia	ASTERACEAE (COMPOSITAE)	Artemisia alba Turra
dente-de-leão	ASTERACEAE (COMPOSITAE)	Taraxacum officinale Weber ex F.H.Wigg.
barbatimão	FABACEAE (LEGUMINOSAE)	Stryphnodendron adstringens (Mart.)Coville
boldo-do-chile	MONIMIACEAE	Peumus boldus Molina
tília	MALVACEAE	Tilia cordata Mill.

2. Uma biblioteca principal dos espectros (IP1) é criada. Todas as bibliotecas secundárias relacionadas a esta biblioteca de referência serão armazenadas no mesmo arquivo do método que usa *FAA como extensão.

A qualidade de uma análise do Id depende substancialmente do método de pré-processamento dos dados e das regiões espectrais de cada espectro (*Region of Interest* ROI), ambos selecionados para as amostras bem como o método da identificação. Para a biblioteca principal (IP1) recomenda-se usar o método padrão e definir uma ROI grande, porque o ruído espectral será “alisado”(torna a “textura” fina, *smoothed*) substancialmente. Em caso das sub-bibliotecas (IP2, IP3 etc.) pode ser melhor usar o método de Fatorização e limitar a ROI, que, naturalmente, faz com que o ruído espectral seja “alisado” de modo insuficiente, mas o espectro mostra muitos detalhes significativos. Deve ser observado que caso sejam atualizados os dados da biblioteca, os espectros novos devem ser re-calculados em cada nível da biblioteca. Trata-se do processo clássico de manutenção de bases de dados, conhecido como reorg(anização da base). A normalização vetorial é selecionada frequentemente como o método escolhido para o pré-processamento dos dados. Às vezes, entretanto, resultados melhores são obtidos com a seleção da primeira e segunda derivada. Selecionando, por exemplo, a normalização vetorial da lista de pré-processamento, pode imediatamente ser acionado o início do cálculo para obtenção das distâncias espectrais. Tem sido comumente utilizado neste trabalho uma combinação que tem apresentado um comportamento robusto, a normalização vetorial e a primeira derivada. Sem dúvida, o comportamento do sistema pode variar conforme o pré-processamento, significando o sucesso na identificação do sinal, ou a indicação clara de que os problemas deveriam ter sido melhor filtrados previamente. No caso de terem sido mantidos, para serem resolvidos pelos métodos de processamento (posteriores ao pré-processamento), podem representar a parte árdua do insucesso, obtido nas fases seguintes como a de separabilidade.

O limite ou limiar do teste de identidade contém os limites do teste de Id indicados. Este dado limiar de um espectro de referência é a soma calculada a partir da distância máxima listada no relatório de média, mais o produto resultando do desvio padrão (SDev) e de um dado fator de x. Para cada grupo os valores de limiar são listados na coluna do limiar. Pode-se incorporar qualquer fator no campo de entrada, 0,25 é ajustado por *default*. O fator configurado é válido para todos os espectros da referência, e a coluna do limiar é atualizada adequadamente. De modo similar pode-se ajustar o

valor para o nível da confiança. Pode ser incorporado qualquer fator entre 95 e 99.9999% no campo de entrada, 99.99% é ajustado por *default*. Ao confirmar sua entrada, a coluna do limiar é atualizada. A coluna *outlier* indica o número de espectros que estão fora do limiar. Com o ajuste da confiança em nível de 95%, 5% do número total dos espectros estará identificado como *outlier*. Estes importantes conceitos que envolvem o limiar (*threshold*), formulados, serão detalhados e apresentados pictoricamente, com exemplos aplicados, a frente e discutidos nos resultados.

Na validação ou verificação da biblioteca, primeiro é observado se os parâmetros do Id selecionados para o método estão otimizados para todos os espectros de referência, e se uma atribuição não ambígua de espectros de teste pode ser assegurada a um grupo. Isto é feito por um procedimento de validação que compare cada espectro original com os espectros médios de todos os grupos. O resultado da validação é indicado no formulário de um relatório e armazenado em um arquivo que use a extensão *.VAL. Os relatórios da validação podem também ser criados para uma única biblioteca da referência, sub-biblioteca ou nivelar para a estrutura de dados inteira da biblioteca, começando no nível de onde a validação inicia. Ao executar mais de uma validação, os arquivos dos resultados estarão numerados consecutivamente (*.v00, *.v01,) Pode-se comparar resultados de diferentes validações, um com o outro, selecionando o arquivo respectivo. Os valores de limiar são considerados como a região de confiança, durante a validação. Os resultados são classificados em três categorias:

- identificado excepcionalmente; os resultados que pertencem a esta categoria tem sua distância espectral, entre o espectro original e o médio, dentro do valor limítrofe.
- não identificado; a distância espectral é mais elevada do que o valor de limiar.
- pode ser confundido com; esta categoria indica que a distância espectral de um espectro original é menor do que o nível de confiança, comparado pelo menos a um espectro médio diferente.

A execução da análise do Id, compara o espectro de teste com todos os espectros da referência. O resultado de uma comparação entre os espectros A e B resulta na distância espectral D, que é chamada sucesso, encontrado, na qualidade (*hit quality*). Os dois espectros que melhor combinam, tem menor distância espectral. O sucesso

qualitativo para espectros idênticos é 0 (isto é, um espectro da referência é comparado consigo mesmo). Os resultados de um teste de identidade e o cálculo da média dos espectros originais, gerando arquivos da referência para uma biblioteca do Id, são armazenados em blocos de relatório. Os resultados da comparação de espectros, entre o espectro de teste e os espectros da referência, são escritos em um relatório do Id. Este relatório é armazenado no arquivo do espectro de teste. O índice do relatório depende dos parâmetros e dos algoritmos selecionados para executar o teste da identidade. O relatório do Id contém informação detalhada no método e em uma lista de distâncias espectrais, entre o espectro de teste e os espectros de referência. Esta lista inclui distâncias em ordem ascendente, isto é, sucesso No. 1 é o espectro da referência que é o mais similar ao espectro de teste. O número ou a quantia de distâncias listadas pode ser definido ao criar o método do Id.

O método de Fatorização, no seu algoritmo, utiliza autovalores e autovetores. Os valores dos fatores podem ser recuperados a partir do arquivo de relatório, do arquivo de método do Id (extensão *.FAA), no OPUS, facilmente recuperado abrindo o bloco dos dados de relatório (*report data block*) e abrindo o sub-diretório do método da busca de identidade. Neste sub-diretório pode-se selecionar autovetores ou autovalores, estes valores estarão indicados na janela do relatório. Os valores de T (Método de Jacobi para cálculo de autovalores e autovetores) são listados no sub-diretório dos autovetores. Conhecido o uso prático da análise de fatores adiante serão aprofundados seus conceitos.

A análise de *clusters* de uma planta teste, pode ser observada na figura 3.1, onde foi obtido o sucesso na localização da planta teste Jaborandi.

A análise de aglomerados ou agrupamentos (*clusters*) é a análise do conjunto de espectros teste de IVP-TF (Infravermelho Próximo-transformada de Fourier), no que se refere a sua similaridade, contrastando com o teste da identidade, porque nenhuma informação de entrada é requerida. A análise de *clusters* divide espectros similares em grupos. Estes grupos são chamados classes ou conjuntos. Aglomerações podem ser indicadas em um dendograma. Na teoria, a distância espectral indica o grau de similaridade espectral. Dois espectros com uma distância espectral de 0 são inteiramente idênticos (dentro das escalas de frequência testadas). Quanto mais elevada a diferença

Figura 3.1: Jaborandi teste identificado no *cluster* de Jaborandi

entre dois espectros, mais elevada é a distância espectral. Os algoritmos hierárquicos de um *cluster* executam as seguintes tarefas:

- As distâncias espectrais entre todos os espectros são calculadas;
- Os dois espectros com a similaridade mais elevada (isto é, os espectros com a distância espectral menor) são fundidos em um conjunto;
- As distâncias entre este conjunto e todos os espectros restantes são calculadas. Diversos métodos (enlace único(*single linkage*), enlace completo(*complete linkage*),) estão disponíveis para calcular as distâncias;
- Os dois espectros (espectro/espectro ou espectro/*cluster*) com a distância menor são fundidos outra vez, em um conjunto novo; As distâncias entre este conjunto novo e todos os espectros restantes (conjunto dos espectros) são fundidas em um conjunto novo; Este procedimento é repetido somente até que um *cluster* grande ser obtido. As distâncias espectrais entre todos os dois espectros, de um conjunto de n espectros, podem ser representadas em uma matriz. Esta matriz é simétrica e os elementos diagonais principais são 0s. Subsequentemente, é suficiente testar somente um triângulo desta matriz. O eixo y do dendrograma mostra as distâncias espectrais entre conjuntos(*clusters*) diferentes. As linhas horizontais indicam os níveis da fusão, que são as distâncias espectrais dos conjuntos diferentes e o espectro antes de se aglomerar novamente.

O alvo da análise com o Id é determinar as diferenças entre um espectro de teste e os espectros da referência, em uma biblioteca. Deve ser definido um método para testar a similaridade espectral e o limiar (*threshold*). Este limiar determina se um espectro é similar somente ou mesmo, idêntico ao espectro da (biblioteca de) referência.

Basicamente, são dois algoritmos para executar a análise do Id:

1. o método do padrão(*standard*);
2. o método da fatorização ou análise de fatores.

Durante a análise, ambos, os métodos comparam o espectro de teste com todos os espectros da referência. O resultado de uma comparação entre dois espectros é a qualidade (grau) do sucesso(*hit quality*), também, chamada distância espectral D . Os dois espectros que combinam melhor, tem menor a distância espectral. A qualidade do

sucesso para espectros idênticos é 0 (isto é, se um espectro da referência for comparado consigo mesmo).

A seguinte fórmula, chamada distância Euclideana, é usada no método padrão:

$$D = \sqrt{\sum_k (a(k) - b(k))^2} \quad (3.1)$$

onde $a(k)$ e $b(k)$ são os valores da ordenada dos espectros a e b . A soma incorpora todos os k pontos de dados selecionados.

Figura 3.2: Relatório do Id

Como mostra o exemplo, no relatório do Id 3.2, a menor distância espectral, que foi determinada pela comparação espectral, é 0,098. Os dois primeiros sucessos são bem separados entre si. A distância espectral para o sucesso no. 2 é bastante grande.

A Figura 3.3 mostra um esquema que representa o método Fatorização - cálculo da distância espectral.

Figura 3.3: Fatorização - Cálculo da distância espectral.

O método da fatorização representa o espectro como combinações lineares dos chamados fatores espectrais ou componentes principais (*loadings*):

$$a = T_{1a} \cdot f1 + T_{2a} \cdot f2 + T_{3a} \cdot f3 + \dots \quad (3.2)$$

O vetor a mostra o espectro a e os fatores espectrais são denotados $f1, f2, f3$ etc. T indica os coeficientes (contagens, *scores*) necessários para reconstruir o espectro original a . Para calcular a distância espectral D entre os dois espectros a e b , os coeficientes T são usados no método de fatorização:

$$D = \sqrt{\sum_i (T_{ia} - T_{ib})^2} \quad (3.3)$$

O somatório é executado para um determinado número de coeficientes. Estes coeficientes T são chamados, também, contagens (*scores*). As diferenças entre os espectros,

o original e o reconstruído são conhecidas como resíduos espectrais.

O método padrão usa diretamente as intensidades espectrais para calcular a distância espectral e o somatório é executado para todos os pontos de dados, dentro das regiões frequenciais especificadas (que poderiam ser mais de 100 pontos). O número de fatores espectrais ou coeficientes de contagem têm que ser incluídos em uma biblioteca do Id é um aspecto muito importante, explicado a seguir. Ao fatorizar a biblioteca do Id, os espectros médios s são transformados em fatores espectrais s . Estes fatores espectrais são ortogonais entre si. O efeito que um determinado fator tem na reprodução do espectro de referência é indicado pelo respectivo autovalor. Os fatores espectrais são classificados de acordo com estes autovalores. O primeiro fator espectral é o mais importante e tem o autovalor mais elevado. Quanto mais o autovalor diminui, mais baixas são as intensidades espectrais (valores da ordenada), destes fatores espectrais, e mais intenso é o ruído. Os fatores espectrais que consistem principalmente de ruído, não devem ser usados em um método do Id. Os fatores espectrais são armazenados no Opus como qualquer outro espectro, no diretório do método do Id, usando o formato de arquivo do Opus e com o mesmo nome do arquivo do método. Entretanto os espectros têm uma extensão de arquivo numérica adicional, começando com o 0 para o primeiro fator espectral. Estes espectros podem ser carregados no Opus como qualquer outro espectro. Pode-se facilmente verificar a ortogonalidade dos fatores espectrais, multiplicando dois fatores espectrais usando a calculadora de espectro do Opus. Isto é seguido por uma integração através da escala de frequência inteira do espectro resultante. O resultado da integração será 0 (ou aproximadamente 0 devido a exatidão computacional finita).

Importantes aspectos teóricos da análise de fatores a seguir são apresentados no sentido da demonstração matricial de seu uso com o programa Id.

Assumindo que cada espectro de referência s , consiste de pontos de dados d . Os espectros da referência são representados pelos os vetores de coluna $d_1, d_2, d_3, \dots$ que formam a matriz de dados D (com dimensões $d \times s$):

$$D = [d_1, d_2, d_3, \dots, d_s] \quad (3.4)$$

Ao trocar as linhas pelas colunas desta matriz, obtém-se a transposta da matriz,

D^T de dimensão $s \times d$. Multiplicando a matriz transposta D^T pela matriz original D obtém-se a matriz de covariância Z :

$$Z = D^T \times D \quad (3.5)$$

A diagonalização e transformação ortogonal de Z produz os autovetores e os autovalores de Z :

$$\Lambda = L_T \times Z \times L \quad (3.6)$$

Os vetores coluna da matriz L de dimensão $(s \times s)$ são $l_1, l_2, l_3, \dots$ autovetores da matriz Z . Esses autovetores são ortonormais, isto é, as seguintes equações são válidas para o produto escalar de dois autovetores:

$$l_i \times l_j = 0 \quad i \neq j \quad (3.7)$$

$$l_i \times l_i = 1 \quad (3.8)$$

A matriz Λ de dimensão $(s \times s)$ contém $\lambda_1, \lambda_2, \lambda_3, \dots$ autovalores da matriz Z como sendo a diagonal principal, todos os outros elementos da matriz são 0. Isto significa:

$$Z_i = \lambda_i \times l_i \quad (3.9)$$

A matriz de dados D é fatorizada pela matriz do autovetor L , usando a multiplicação:

$$F = D \quad (3.10)$$

A matriz F tem as mesmas dimensões da matriz de dados D ($d \times s$) e inclui os vetores $f_1, f_1, f_1, \dots$ fatores espectrais como colunas. Multiplicando a matriz transposta F^T por F produz:

$$F^T \times F = (D \times L)^T \times (D \times L) = L^T \times D^T \times D \times L = L^T \times Z \times L = \Lambda \quad (3.11)$$

Os elementos da matriz quadrática $F^T F$ de dimensão $(s \times s)$ são os produtos escalares que podem ser criados aos pares, juntamente com os fatores espectrais. A equação (3.11) faz com que os fatores espectrais sejam ortogonais entre si:

$$f_i \times f_j = 0 \quad i \neq j \quad (3.12)$$

$$f_i \cdot f_j = \lambda_i \quad (3.13)$$

A norma do vetor de um fator espectral é igual a raiz quadrada do correspondente autovalor. Usando a ortogonalidade L , a matriz de dados D pode ser a seguinte:

$$D = D \times 1 = D \times L \times L^T = F \times L^T \quad (3.14)$$

Os espectros de referência são representados como combinações lineares dos fatores espectrais, e os coeficientes estão contidos nas colunas da matriz L^T . Baseado na equação (3.14), aplica-se ao primeiro espectro de referência, o seguinte:

$$d_1 = L_{1,1}^T \cdot f_1 + L_{2,1}^T \cdot f_2 + L_{3,1}^T \cdot f_3 + \dots + L_{s,1}^T \cdot f_s \quad (3.15)$$

Os coeficientes de pontuação (*scores*) são as coordenadas dos espectros de referência no sistema de fatores espectrais. Todo o espectro u pode ser representado como combinação linear dos fatores espectrais:

$$u = F \cdot k \cdot e \quad (3.16)$$

O vector coluna k , desconhecido, corresponde aos elementos da coluna da matriz L^T . O espectro de erro e é a diferença entre o espectro de u e espectro reconstruído. A solução de quadrados mínimos para k , que minimiza o erro, é, como segue:

$$k = (F^T \cdot F)^{-1} \cdot F^T \cdot u = \Lambda^1 \cdot F^T \cdot u \quad (3.17)$$

Se apenas os primeiro fator espectral r é considerado, a distância espectral D entre um espectro de u e um espectro de referência D_a é:

$$D = \sqrt{(k_1 - L_{1a}^T)^2 + (k_2 - L_{2a}^T)^2 + \dots + (k_r - L_{ra}^T)^2} \quad (3.18)$$

Em vez de usar os vetores coluna da matriz L^T , pode-se usar os vetores linha L . A equação para a distância espectral D do fator no espaço de dimensão r é então:

$$D = \sqrt{(k_1 - L_{a1})^2 + (k_2 - L_{a2})^2 + \dots + (k_r - L_{ar})^2} \quad (3.19)$$

Ao utilizar o método Padrão, pode-se especificar um intervalo de valores para as distâncias D . Selecionando o pré-processamento: normalização vetorial. A escala vai de 0 (espectros idênticos) a 2 (diferença espectral máxima). Isto não se aplica a D , quando utiliza-se o método de fatoração. Se todos os fatores espectrais são utilizados, as distâncias espectrais entre os espectros de referência são constantes, ou seja, $D = \sqrt{2}$. Para calcular as distâncias espectrais, os elementos da matriz autovetor L são utilizados. Essa matriz consiste de vetores unitários ortogonais no espaço dimensional s . A distância entre dois vetores unitários ortogonais sempre tem que ser $\sqrt{2}$. No entanto, quando se utiliza o método de fatorização nem todos os fatores espectrais são utilizados, pois os espectros de maior fator espectral, principalmente, aumentam o nível de ruído do espectro reconstruído. O método de Fatorização também permite calcular distâncias espectrais usando resíduos. O resíduo espectral é calculado pela diferença entre o espectro original e o reconstruído. Para calcular o resíduo espectral, sendo u um espectro qualquer, a equação $SpecResu_u$ é a seguinte:

$$SpecResu_u = \sqrt{k(u(k) - (k_1 \cdot f_1(k) + k_2 \cdot f_2(k) + \dots + (k_r - f_r(k)))^2} \quad (3.20)$$

O somatório é realizado para todos os k pontos de dados selecionados. A distância espectral D entre o espectro u e o espectro da referência é calculada da seguinte forma:

$$D = \sqrt{(k_1 - L_{a1})^2 + (k_2 - L_{a2})^2 + \dots + (k_r - L_{ar})^2 + (SpecRes_u - SpecRes_a)^2} \quad (3.21)$$

Os algoritmos *Scaling to 1st Range* e *Normalize to Reprolevel* podem ser usados para identificar microorganismos. Ao contrário dos métodos padrão e de análise de fatores, sobreposição de intervalos espectrais não são mescladas em um só intervalo para fins de cálculo de distâncias. O pré-processamento usando a normalização vetorial é calculada automaticamente, separadamente para cada intervalo espectral. Os valores resultantes são usados para determinar uma distância média. A normalização vetorial considera todos os pontos dos dados selecionados, quando se utiliza o padrão e algoritmo de fatorização.

Para o cálculo das distâncias espectrais são considerados:

1. Para calcular as distâncias espectrais têm de ser em primeiro lugar normalizado o vetor, para cada faixa de frequência.
2. Então, o coeficiente de correlação r , de Pearson, é calculado. O tipo de coeficiente define a correlação entre um espectro a e um espectro b:

$$r = \sum a_n(k) \times b_n(k) \quad (3.22)$$

Essa correlação é calculada separadamente para cada faixa de frequência. O somatório abrange todos os pontos k de dados de uma faixa de frequência: a_n b_n são as intensidades espectrais normalizados. A normalização resulta:

$$r = \frac{\sum (a_n(k) - a_m) \cdot (b_n(k) - b_m)}{\sqrt{\sum (a_n(k) - a_m)^2} \cdot \sqrt{\sum (b_n(k) - b_m)^2}} \quad (3.23)$$

a_m b_m são as intensidades espectrais médias dentro da faixa espectral, enquanto $a(k)$ e $b(k)$ são as intensidades espectrais originais. O intervalo de valores do coeficiente de correlação r atinge de -1 (espectros inversos) a +1 (espectros idênticos).

3. O coeficiente de correlação é transformado na distância espectral D através da seguinte equação:

$$D_r = (1 - r) \times 1000 \quad (3.24)$$

A distância espectral D pode estar entre 0 (espectros idênticos) e 2000 (espectros inversos).

4. a. O *Scaling to 1st Range* determina o valor mínimo e máximo das distâncias espectrais para o primeiro intervalo espectral. Em seguida, as distâncias de todos os outros intervalos espectrais são calculadas e dimensionadas para o mesmo intervalo de valores, ou seja, os mesmos mínimos e máximos, como o primeiro intervalo espectral. Exemplo: As distâncias espectrais na faixa espectral primeiro estão entre 2 e 10, no segundo, entre 6 e 22. As distâncias do segundo intervalo espectral são transformados como segue:

$$D \rightarrow 0,5 \times D - 1 \quad (3.25)$$

Após essa transformação das distâncias espectrais da segunda faixa espectral, estas têm os mesmos valores que as distâncias que pertencem ao primeiro intervalo espectral. Como a escala (*scaling*) foi submetida ao primeiro intervalo espectral, não importa qual faixa espectral é selecionada em primeiro lugar. É necessário selecionar o intervalo espectral correto como primeiro intervalo. Se as distâncias espectrais são classificados de acordo com valores ascendentes, essa ordem resulta diretamente a partir dos intervalos espectrais selecionadas. Por exemplo, quando se compara um espectro de teste com um espectro de referência, utilizando um teste Ident, a distância espectral pode ter o menor valor para o quarto espectro de referência (melhor *hit*), baseado no primeiro intervalo espectral. Considerando-se a segunda faixa espectral, a distância espectral pode ter o menor valor para o décimo primeiro espectro de referência.

- 4b. Ao usar o método *Normalize to Replevel* deve ser definido um nível de reprodução para cada faixa espectral. As distâncias espectrais serão divididas por esse nível de reprodução. Esta é a razão porque as distâncias espectrais são indicadas como unidades de reprodução de nível, o que significa que se pode definir um limite para o teste de identidade. Por exemplo, se a qualidade (*hit*) é inferior a 1, no caso de um espectro de teste, a amostra é considerada como sendo identificada. Se, no entanto, a distância espectral é superior a 1, o espectro

não pode ser atribuído a qualquer espectro de referência.

5. Independentemente do método utilizado, as distâncias espectrais podem ser ponderadas para cada intervalo espectral único, de acordo com a seguinte equação:

$$D = \frac{\sum w_j \cdot D_j}{\sum w_j} \quad (3.26)$$

Distâncias espectrais calculadas pelo algoritmo *Normalize to Replevel* podem ser superior a 2000, se os níveis de reprodução são diferente de 1. Quando se utiliza o método *Scaling to 1st Range*, os valores têm que estar entre 0 e 2000.

Vários métodos de pré-processamento podem ser utilizados sendo comum a normalização vetorial. O valor máximo da qualidade do sucesso tem que ser definido apenas se a normalização do vetor foi usada para pré-processar os dados. Uso de um método de pré-processamento que não seja normalização vetorial, não requer definição do limite máximo para a qualidade do sucesso, ou seja, pode se usar qualquer valor numérico. A distância máxima espectral é 2 (diferença máxima dos espectros) no caso da normalização vetorial, desde que tenha sido selecionado o método padrão. A normalização vetorial primeiro calcula o valor y médio dos espectros e utiliza apenas os pontos de dados dentro dos intervalos espectrais selecionados. O valor médio calculado será, então, subtraído do espectro, o que faz com que o espectro seja centrado em torno de $y = 0$. Isto é seguido pelo cálculo da soma dos quadrados de todos os valores de y , e o respectivo espectro é dividido pela raiz quadrada dessa soma calculada. O vetor normalizado do espectro resultante é 1:

$$a_m = \frac{\sum_k a(k)}{N} \quad (3.27)$$

$$a'(k) = a(k) - a_m \quad (3.28)$$

$$a''(k) = \frac{a'(k)}{\sqrt{\sum_k (a'(k))^2}} \quad (3.29)$$

$$\sum_k (a''(k))^2 = 1 \quad (3.30)$$

A normalização vetorial também reduz as diferenças entre cada medição da mesma amostra. Espectros adquiridos de uma única amostra, tal como uma substância em pó diferem substancialmente entre si. Estas diferenças podem ser consideravelmente reduzidas usando a normalização vetorial.

Também, espectros derivados da mesma amostra apresentam as diferenças muito pequenas. A normalização vetorial é altamente recomendado nestes casos. Outros métodos de pré-processamento de dados são a primeira e segunda derivadas.

Em geral, uma questão é até que ponto a distância espectral (*hit Quality*) pode aumentar e estar ainda dentro de um limite aceitável para identificar corretamente o espectro de teste? Para definir esse limiar, não é suficiente medir apenas um espectro único por substância de referência. Devem ser medidos muitos espectros e ser determinado esse limiar a partir das flutuações espectrais (diferenças espectrais). O espectro médio é calculado a partir de cada medição e então é adicionado à biblioteca do Id como espectro de referência.

Para determinação do valor limite para o teste de identidade ou para definir o (limiar) valor limite para um grupo do Id, existem as possibilidades:

- O algoritmo fixo determina (máximo *hit* + x SDev). O limite é calculado a partir do pior (*worst hit*) caso (isto é, o maior valor de sucesso(*hit*) na qualidade, no relatório média) e o desvio padrão S_0 :

$$\text{Threshold } D_T = D_{Max} + S_0 \cdot x \quad (3.31)$$

onde o valor padrão(*default*) é de x é 0.25.

O limiar é escolhido de forma que todos os espectros originais usados para criar o espectro de referência (espectro médio) têm uma distância menor, superior a esse limite do espectro de referência. Se a análise de um espectro da amostra produz uma distância espectral, que é maior do que este limiar, o espectro da amostra será definido como não sendo amostras idênticas.

- O nível de confiança utiliza dois parâmetros são derivados de distâncias espectrais para definir a região de confiança para o espectro médio: A distância média D_M

$$D_M = \sum_i \frac{D(i)}{n} \quad (3.32)$$

O desvio padrão S_0

$$S_0 = \sqrt{\sum_i \frac{D(i)^2}{n-1}} \quad (3.33)$$

com n sendo o número de espectros originais. Note-se que S_0 é o desvio padrão de zero e não o desvio-padrão S_m proveniente do valor médio. O desvio padrão da distância média S_M pode ser calculado a partir D_M e S_0 .

$$S_M = \sqrt{\sum_i \frac{(D(i) - D_M)^2}{n-1}} = \sqrt{S_0^2 - \frac{D_M^2 \cdot n}{n-1}} \quad (3.34)$$

O limite é calculado multiplicando o desvio padrão S_M com um fator de f e acrescentando a distância média D_M :

$$D_T = D_M + f \cdot S_M \quad (3.35)$$

O fator f é calculado a partir da probabilidade Φ que pode ser escolhida entre 95% e 99,9999%. Exemplo: Escolhendo $\Phi = 97,7\%$, o fator f será 2. As distâncias espectrais são supostamente distribuídas de acordo com uma distribuição normal. Note que o valor de probabilidade é calculada para o limite de um lado só. Sendo selecionado $\Phi = 95\%$, em seguida, 5% dos espectros originais estão fora da região de confiança (a distância espectral para o espectro médio é maior que o limite em questão). Se isso é realmente correto ou não pode ser testado na fase de Validação, cuja descrição é apresentada após as descrições seguintes do Teste de Identidade e do Teste de Classe.

- A opção limiar absoluto aceita que se defina o limiar para cada espectro de referência. Este limiar, porém, só terá efeito se o grupo é composto por diversos espectros. Sendo definido o limiar, devem ser certificadas as considerações das diferenças espectrais do espectro original.

Teste de Identidade

A rotina de teste de identidade gera uma lista de ocorrências que é armazenada no relatório do Id, os testes são ordenados em ordem crescente de suas distâncias espectrais. Caso uma referência esperada tenha sido selecionada (opção oferecida pelo sistema de testar contra uma referência selecionada da base) automaticamente a partir do nome da amostra, ou tenha sido definido pelo usuário, às três categorias de resultados de testes de identidade seguintes, são possíveis: “idênticos”; “não idênticos”; ou, “podem ser confundidos com”. Se um grupo for constituído por apenas um espectro, o resulta do do teste de identidade será: Identidade não checada. Se nenhuma referência esperada foi definida, os resultados do teste de identidade serão: Identificada como; não identificadas; não é possível a identificação única.

Tabela 3.4: Resultados possíveis codificados para os testes de identidade.

Resultado dos testes de Identidade	Códigos
Identidade não checada (sem limiar)	0
Idêntico a ou Identificado como	1
Pode ser confundido com ou Identificacao possível não é única	-1
Não idêntico ou Não identificado Ok	-2
Pode ser confundido com $< N >$ outros <i>hits</i>	2

Os possíveis resultados do teste de identidade pelo Id são apresentados na tabela 3.4 e descritos a seguir.

- Possibilidades de resposta obtida pelo usuário dependem da opção escolhida, o usuário pode testar sua amostra teste contra qualquer substância da base ou ter optado por um ensaio “contra” determinada (escolhida) substância. Primeiro são apresentadas as possibilidades de resultados versus qualquer substância da base e depois a amostra de teste versus uma amostra de referência que teve seu nome indicado entre as substâncias da base de referência.
- “Identificado como” (no caso de qualquer referência definida):

Todos os valores limite são considerados, o teste será procedido contra a base completa. Em caso de identificadas como, o *hit* da qualidade da primeira substância (ou seja, a distância espectral entre um espectro de teste e no espectro de referência, no relatório) tem que ser menor do que o limiar correspondente.

Mas o *hit* de Qualidade de todas as outras substâncias tem de ser maior do que o limite correspondente.

- “Não Identificado” (no caso de referência qualquer tendo sido definida):

Neste caso, o sucesso na qualidade de todas as substâncias de referência tem que ser maior que o limiar.

- “Não é possível a identificação única” (no caso de referência qualquer definida):

The sucesso (*hit*) na qualidade de mais de uma substância é menor que o limite correspondente. Os números de código são atribuídos aos resultados individuais.

- “Idêntico à” (no caso de referência esperada):

O primeiro sucesso(*hit*) deve ser a referência que era a esperada. O usuário escolheu, tenho Jaborandi e vou ver se é o Jaborandi da base de referência.

O *hit* da qualidade da primeira substância (ou seja, a distância espectral entre um espectro de teste e o espectro de referência, no relatório) tem de ser inferior ao limiar da referência esperada, e as distâncias espectral entre o espectro de teste e todos os outros espectros médios são sempre maiores do que os limiares correspondentes.

- “Não idêntico à” (no caso de referência esperada):

A distância espectral entre o espectro de teste e espectro de referência esperado é maior que o valor limiar.

- “Pode ser confundida com $< N >$ Outros *hits*” (no caso de referência esperada)

O número de código 2 é utilizado se todos os espectros de referência com sucesso de qualidade inferior ao limiar tem o mesmo nome da amostra (ou sub nome da amostra) do espectro de teste. O valor de 2 só pode ocorrer se a opção selecionado automaticamente(referência esperada a partir do nome da amostra de opção Nome tenha sido ativada.

Idealmente, todos os espectros de referência são bem distintos uns dos outros e os limiares são tão pequenas que as suas regiões de confiança não se sobreponham. Os espectros de referência são os pontos (em um sentido matemático) no espaço dimensional n , sendo N o número de pontos de dados selecionados. As regiões de confiança

podem ser representadas por esferas que representam os centros de espectros de referência. No entanto, pode ocorrer que as regiões de confiança de alguns espectros de referência se sobrepõem. Nesta observação qualquer das regiões de confiança do A, B, C e espectros de referência claramente se sobrepõem. É possível definir vários espectros de referência, que são **membros de uma classe**, e realizar um teste de classe durante a análise com o Id. Tanto no caso de uma referência prevista, bem como em caso de uma análise sem referência definida, o teste de classe vai determinar se todos os espectros de referência, com o seu sucesso na qualidade, abaixo do limiar correspondente, são membros da mesma classe. Se assim for, o resultado vai ser da classe teste OK. Caso contrário, o resultado será Test NÃO OK. Caso o espectro de referência esperado não faz parte de qualquer classe, o relatório de ID vai dizer teste de classe não realizado. É necessário certificar-se de definir apenas um nome de classe para um grupo no método Id. Sendo carregado um método utilizado anteriormente, incluindo membros extraordinários, somente o nome da primeira classe será considerado. Todos os membros da classe utilizam os seus limiares individuais. A seguir é sumariada a tabela 3.5 com os códigos numéricos resultantes possíveis resultados do teste da classe.

Tabela 3.5: Resultados possíveis para os testes de classe e seus respectivos números codificados.

Resultado dos testes de classe	Códigos
Teste de classe Ok	1
Teste de classe não executado	0
Teste de classe não Ok	-2

Quando da criação de uma biblioteca Id, é preciso verificar se os parâmetros do método Id são otimizados para todos os espectros de referência em uma biblioteca. Isso pode ser feito através da **validação**, a qual compara com os espectros originais os espectros de média. Todos os espectros originais são comparados com os espectros médios da biblioteca. O limiares absolutos são utilizados como região de confiança no processo de validação. Os resultados são identificados de forma exclusiva(único), não identificado e pode ser confundido com, de modo semelhante ao teste de identidade. O número de espectros que pertencem à classe respectiva é indicado na parte final do relatório de validação. O total de todos os espectros tem que ser igual ao total dos espectros originais testados. Em caso de sobreposições o relatório de Resultados Detalhados contém informação adicional sobre quais grupos devem ser atribuídos a uma nova sub-biblioteca comum. Um espectro original é identificado exclusivamente se as

distâncias espectrais entre este espectro e o espectro médio correspondente for menor do que o limite, enquanto as distâncias espectral entre o espectro original e todos os outros espectros médios são maiores do que a região de confiança correspondente. Os espectros originais, que são identificados exclusivamente não estão listadas no relatório. Se um espectro não é identificado, a distância espectral entre o espectro original e o espectro médio é maior que o limiar. Neste caso, o relatório indica o espectro médio correspondente, o nome de amostra e limite (região de confiança) especificado para esta substância. Os nomes de arquivo dos espectros originais que não tenham sido identificados, e as distâncias espectrais entre os espectros original eo espectro médio (*hit*) são listados como Espectro original fora da Região de Confiança definida. O resultado “Pode ser confundido com”, indica que a distância espectral de um espectro original para o espectro médio correspondente é menor do que a região de confiança, enquanto uma ou mais distâncias espectrais entre o espectro original e espectros de média quaisquer são ainda menores do que as regiões de confiança correspondentes. Caso um espectro original é testado e “Pode ser confundido com” outras referências, em primeiro lugar, o seu espectro médio, o nome da amostra e região de confiança estão listados no relatório, seguido do nome do espectro original e do limiar. Além disso, o nome do espectro médio, o nome da amostra e da distância espectral (*hit*) entre o espectro médio e do espectro original são listadas com a informação “Sobreposição com”. A distância espectral é menor do que o limite (região de confiança). Se este espectro original sobrepõe a média de vários espectros, todos eles serão listados no relatório. Se todos os espectros de referência são selecionados, cada espectro original será comparado com todos os espectros de média. Mesmo sendo selecionados apenas alguns dos espectros de referência, os espectros originais pertencentes a esses espectros de referência são testadas versus todos os espectros médios. Esta opção é extremamente útil se uma biblioteca existente, deve ser estendida por novos espectros de referência e apenas estes espectros novos têm de ser testados. Se tiver sido ativado o parâmetro Sempre usar nível IP mais baixo, o relatório detalhado lista apenas os resultados do mais baixo nível de IP para todos os espectros de cada grupo.

3.3 Resultados do Id

Amplos resultados foram conseguidos com uso do equipamento da Bruker, MPA, com o software OPUS e com o software Id. Com o equipamento da Bruker Optics Multipurpose Analyser Espectrômetro na região do infravermelho próximo usando transformada de Fourier, foram obtidos mais de mil espectros de plantas medicinais ou drogas vegetais. Estes espectros variaram em tipo do experimento e “preparação” ou formato da amostra. Desde diluições com experimentos líquidos foram realizados, inclusive com espectroscopia Raman, usando *vials*. Com amostra sólida foram obtidos os espectros em pó, 0,150 mm e com 0,236mm, além da massiva obtenção de espectros direto em copo grande e pequeno usando esfera de integração. Isto possibilitou testes intensivos com diferentes concentrações de diferentes soluções, desde extrato de jaborandi, bem como a sua diluição, com HCl, por exemplo. Concluiu-se que as observações e análises com infravermelho usando amostras líquidas apresentam uma perda de informação da ordem de 80%, ficando pois, no presente momento com os recursos que se dispunha, postergado este tipo de análise. Optou-se pela análise comparativa utilizando espectroscopia Raman, desde as amostras líquidas até sólidas.

Com uso do MPA foram testadas diferentes granularidades (tamanho de partícula sólida) com diferentes experimentos. Conclui-se que as questões que envolvem os diferentes experimentos, como uso de copo grande, ou pequeno e a esfera de integração, tratam-se certamente de questões que envolvem o “processamento” da amostra, o que de qualquer modo deve levar a sua identificação correta, limitada pelas questões “tecnológicas” de processamento utilizando o MPA e o *software* OPUS.

Especificamente o *software* Id, a exemplo do OPUS, quando da aquisição das amostras, pode apresentar oferecer possibilidades de sofisticadas análises dos resultados, no sentido da busca de melhores evidências e inclusive do aprendizado no uso de suas potencialidades parcialmente exploradas por diferentes tipos de usuários/aplicações,

Para citar tem-se o caso dos algoritmos de análise de picos e seu comportamento com as diferentes granularidades das amostras e/ou experimentos. A seguinte tabela apresenta o seu comportamento com diferentes parametrizações utilizadas na execução dos algoritmos.

Os resultados do Id em termos de identificação de amostras são os esperados. Na

criação da base de referência, todas as amostras foram identicamente reconhecidas, isto é, qualquer espectro, considerando a reprodutibilidade variando de seis a sete, foi reconhecido quando testado contra ele mesmo, sendo idêntico o reconhecimento, resultado esperado, que comprova o funcionamento correto do sistema.

De forma semelhante comportou-se a análise de *clusters*, fig. 3.1, indo além do reconhecimento idêntico, isto é, tratando-se de uma amostra teste sendo rastreada nos *clusters* de referência, com sucesso a amostra foi identificada, e, considerando que foram usados sem uso de média, os espectros com a respectiva reprodutibilidade, a amostra teste teve afinidade por uma determinada amostra dentro do conjunto das respectivas amostras obtidas com reprodutibilidade, conforme esperado.

Os resultados de rastreamento das amostras teste contra a base de referências, apresentaram resultados esperados no seguinte sentido a ser demonstrado: primeiro, as amostras teste de espécies que existiam na base de referência apresentaram resultados positivos de reconhecimento e apresentaram, também, resultados que chamamos de falso negativo, pois embora o sistema tenha dito desconhecer a amostra teste, em seu *ranking* de qualidade de sucesso apresentando os dez espectros, que tiveram afinidade com a amostra teste, em primeiro lugar sempre estava a respectiva espécie da amostra teste. Isto conduz a uma série de reflexões sobre este comportamento do sistema que vem desde a questão do caminho ótico “particular” que resulta em cada espectro, durante as diferentes aquisições com reprodutibilidade, passando até as conhecidas questões que envolvem ainda detalhes do experimento como a resolução espectral, uso de algoritmos de pré-processamento, variações na temperatura, etc. Também concorrem, neste caso, as clássicas questões que são o tamanho dos *clusters* envolvidos, a separabilidade destes, a “especialização” com uso da análise de fatores, além dos já referidos algoritmos de pré-processamento, por exemplo, com uso exclusivo da normalização vetorial os resultados requerem maior atividade no sentido da supervisão da separabilidade. Com o uso da primeira derivada combinada com a normalização vetorial os resultados da separabilidade são mais significativos, ficando um percentual menor de amostras para a análise de fatores. Do mesmo modo, na análise de fatores o uso dos algoritmos de primeira e ou segunda derivada associados a normalização vetorial apresentam resultados significativos caindo a um número mínimo a separabilidade onde o sistema afirma estar confuso sobre determinados espectros. Muitas conjecturas e testes inclusive do comportamento dos componentes principais (os fatores) podem ser

e inclusive são recomendáveis para a busca da “estabilidade” no comportamento do reconhecimento dos espectros, como o caso da eleição do(s) fator(es) mais representativos da amostra, uso ou não de resíduos, uso ou de médias, etc. Reiterando o já citado, tudo culmina com as questões que envolvem também a amostra propriamente dita. O sistema precisa ser robusto e flexível para abarcar a variabilidade da amostra. Assim um resultado falso negativo pode ser atribuído a falta de abrigo, pelo sistema, de uma maior variabilidade, por exemplo, um *cluster* de Jaborandi, na verdade teria de abrigar uma série de pelo menos dez *clusters* que representam as diferentes espécies, sem considerar que ocorrem casos de a mesma espécie estar coberta em diferentes áreas do país, como é o caso do *Pilocarpus microphyllus* Stapf, no interior do estado de São Paulo e ser espontâneo do nordeste do Brasil, e exportado pelos portos do Maranhão, Pará e Paraíba. Outra observação que sem dúvida vai impactar a análise quantitativa são as questões que envolvem a umidade dos diferentes solos, portanto uma mesma espécie pode ter um diferente grau de pureza, conforme esperado, este caso, já foi mencionado. Parece ficar claro que duas questões devem ser consideradas para a estabilidade da base de referência, o número de espécies e/ou amostras abrigadas em um *cluster* e o nível de especialização, uso da análise de fatores, dos referidos grupos. Do mesmo modo, ficou claro que a identificação (nível qualitativo) de drogas vegetais, neste trabalho, é factível. Ressalta aos procedimentos no sentido de facilitar a estabilidade da base de referência deve ser dita que é a necessidade de supervisão por parte do usuário do sistema, isto é, o sistema pode vir a ser dotado de ferramentas “mais inteligentes”, sem a necessidade de supervisão do processo de separabilidade e especialização usando análise de fatores, por outro lado, uma vez que este passo seja supervisionado, a estabilidade da base de referência com seus *clusters* especializados poderá ser desenvolvida com o abrigo das variabilidade dentro deste mesmo conceito. Com pode-se afirmar tal questão? Pelo motivo inerente ao próprio sistema, o Id possui robustez para decidir a exemplo do taxonomista, isto é, “fácil” de separar, trata-se de *Chamomilla recutita* (L.), inclusive com pouca variedade, etc. Por outro lado, o sistema avisa quando está confuso e tem dificuldade para distinguir quer sejam amostras que precisam ser supervisionadas e reagrupadas, usando análise de fatores ou quando se trata de amostras de espécies diferentes que também requerem o mesmo procedimento. A robustez do sistema está em que uma vez tendo sido supervisionada a sua “especialização”, novamente com a análise de fatores ele decide, com um índice de confusão de pouco significado. E, ainda permite, que sejam variados os componentes principais(fatores) para uma estabilidade

apropriada à aplicação do momento. Pode-se dizer que o sistema Id tenta comportar-se de modo análogo a *Solanum cernuum* Vell, família *Solanaceae*, nome popular panacéia, que tenta ser “solução” para diversos problemas.

4 *Assinatura Espectral Fuzzy*

Visão Geral

O uso da lógica *fuzzy* permite um julgamento mais flexível dos dados, próximo do raciocínio humano, tornando-o mais abrangente e menos susceptível a preconceitos, comumente encontrados na lógica booleana.

A lógica *fuzzy* se baseia nos princípios teóricos de que uma proposição pode possuir diversos valores lógicos, variando entre 0 e 1, em contraste com a lógica booleana, que assume a uma proposição um valor lógico que pode ser somente 1 (ou verdadeiro) ou 0 (ou falso).

Apesar de parecer um absurdo relacionar a lógica, que é um conceito baseado na certeza, com a incerteza que *fuzzy* carrega, é exatamente esta flexibilidade que torna a lógica *fuzzy* mais precisa.

Em se tratando este trabalho de um projeto multidisciplinar parece razoável o uso de um exemplo simples ou comum para se demonstrar isto. Trata-se de um radar de velocidade em uma rodovia. Se o limite de velocidade nesta rodovia for 80 km/h, um carro a 83 km/h e outro a 95 km/h serão multados (valor lógico 1), mas um carro a 77 km/h e outro a 65 km/h não serão (valor lógico 0). Ou seja, para o radar de velocidade, acima de 80 km/h é *rápido* (valor lógico 1) e abaixo disso é *lento* (valor lógico 0).

Porém, ao se julgar a velocidade dos carros, a pouca diferença entre o carro que está a 83 km/h e o que está a 77 km/h é visível, quando comparada com a velocidade do carro que está a 95 km/h ou ao que está a 65 km/h.

Portanto, poderíamos julgar, com uma velocidade média de 80 Km/h, o carro que está a 65 km/h como “lento” (valor lógico 0,1), o que está a 77 km/h como “razoavelmente lento” (valor lógico 0,45), o que está a 83 km/h como “razoavelmente rápido”

(valor lógico 0,55) e o que está a 95 km/h como “rápido” (valor lógico 0,9), por exemplo.

Associar esta vaguidão diretamente relacionada à lógica *fuzzy* à avaliação da diferença entre as curvas, torna este trabalho mais simples, já que as classificações de distância entre as curvas passam a incluir a avaliação do quão distante ou próximo estas estão.

A idéia de conjunto *fuzzy* se inicia com o fato de que um certo elemento pode pertencer ou não a um conjunto. Por exemplo, na situação anterior, um carro a 79 km/h não pertenceria ao grupo dos carros que serão multados, mas um carro a 81 km/h pertenceria a este conjunto. A essa característica que um elemento possui se dá o nome de pertinência. Na teoria de conjuntos simples (ou *crisp sets*), ou um elemento pertence ao conjunto (pertinência igual a 1) ou não pertence (pertinência igual a 0). Ainda usando o exemplo do radar de trânsito, ou o carro será multado, ou não será multado.

Todavia, para o julgamento de uma característica mais abstrata, como dizer se um carro está rápido ou não, essa definição se torna incompleta. Por exemplo, define-se que um carro acima de 80 Km/h está rápido e abaixo disso, estará lento. Portanto, um carro que está a 79 km/h é julgado lento, assim como um a 65 km/h. Porém, um carro a 81 km/h é julgado como rápido, assim como um carro a 95 km/h. Nesse ponto de vista, uma diferença de 2 km/h (79 km/h e 81 km/h) pode mudar completamente o julgamento da velocidade do carro, assim como a diferença de 30 Km/h (65 km/h e 95 km/h), que é 15 vezes maior que a primeira.

Para resolver essa ambiguidade, podemos definir os conjuntos *fuzzy* (ou *fuzzy sets*), nos quais a pertinência de certo elemento pode ser um valor real entre 0 e 1. Isso significa que um certo elemento pode pertencer ao conjunto (pertinência igual a 1), pertencer apenas parcialmente ao conjunto (pertinência entre 0 e 1, sendo estes não inclusos) ou não pertencer ao conjunto (pertinência igual a 0).

Julgando novamente, pode-se dizer que:

- O carro que está a 65 km/h estaria *ligeiramente rápido* (valor lógico 0,1);
- O que está a 77 km/h estaria *quase rápido* (valor lógico 0,45);
- O que está a 83 km/h como *razoavelmente rápido* (valor lógico 0,55);

- e o que está a 95 km/h como *rápido* (valor lógico 0,9).

Comparando os dois julgamentos, é possível observar que, o segundo está mais de acordo com a realidade observada pelo homem, pois admite qualificadores ao julgamento, tornando-o mais flexível e menos propenso a preconceitos.

As implementações seguintes são baseadas no artigo apresentado no capítulo 2, conforme (EBEL, 1992).

4.1 Implementação segundo o modelo de Bafflert

O arquitetura de sistemas aqui utiliza a lógica *fuzzy* e a programação utilizando a ferramenta *software* MatLab como ambiente de desenvolvimento, no sentido do objetivo de implementar a identificação de espécies de plantas medicinais ou drogas vegetais, através da comparação de uma dada planta teste com a base das plantas de referência.

Para isto são utilizados os espectros (em tabelas) da planta teste e das plantas referências caracterizadas (adquiridas) com uso do *software* Opus, além das frequências características dos picos eleitos, da planta teste, com o emprego do mesmo *software*. A comparação é feita utilizando as intensidades dos picos. Primeiro, são encontradas as intensidades dos picos das plantas referências para que seja montada uma função de pertinência triangular para cada referência.

O pico do triângulo (grau 1) é a intensidade da planta, representada na função de pertinência e a base (grau 0) é formada pelas intensidades das duas plantas, que estão próximas da planta, representada na função de pertinência.

Para o caso da referência ter a menor ou a maior intensidade, a base do triângulo, onde não há planta “vizinha” é dada pela mesma distância da planta, da função de pertinência com a outra “vizinha”. A figura 4.1 exemplifica este procedimento.

No passo seguinte, encontra-se a intensidade do pico da planta teste e encontra-se, também, o grau de pertinência para cada referência. A programação contém o programa principal (Principal.m) e duas funções *intensidadepico.m* e *fuzzificacao.m*.

O programa principal (Principal.m) é o programa, o qual o usuário deverá rodar.

Para o correto funcionamento do programa, devem ser montados os arquivos de

Figura 4.1: Procedimentos de *Fuzzificação*.

dados, o usuário deve carregar o arquivo Referencias, extensão xls, com as referências (deve-se colocar na coluna 1 as frequências e na coluna 2 as intensidades, estando em cada planilha a tabela de uma planta de referência). Carrega-se o arquivo Teste, extensão xls, com as plantas testes (disposto da mesma maneira que o arquivo das referências). Também, carregar os arquivos OrdemReferencia, extensão xls, com a ordem de como as referências estão colocadas nas planilhas (é colocado o nome sequencialmente em cada linha da primeira coluna), o OrdemTeste, extensão xls, com a ordem de como os testes estão colocadas nas planilhas.

Finalmente, deve-se carregar o arquivo Picos, extensão xls, onde são colocados os picos dos espectros das plantas testes, obtidos na fase de aquisição de espectros,

com o *software* Opus (as planilhas são colocados na mesma ordem das planilhas do arquivo Teste extensão xls). Todos os arquivos devem estar na mesma pasta do arquivo Principal.m.

	A	B
1	12489,4	0,441155
2	12481,69	0,438779
3	12473,98	0,436725
4	12466,26	0,436194
5	12458,55	0,436714
6	12450,83	0,436912
7	12443,12	0,436356
8	12435,4	0,435367
9	12427,69	0,433858
10	12419,98	0,432344
11	12412,26	0,432084
12	12404,55	0,432504
13	12396,83	0,432206
14	12389,12	0,431575
15	12381,4	0,431265
16	12373,69	0,431312
17	12365,98	0,431819
18	12358,26	0,432385
19	12350,55	0,43179
20	12342,83	0,429701
21	12335,12	0,427915
22	12327,4	0,427425
23	12319,69	0,42743
24	12311,98	0,427416
25	12304,26	0,426853
26	12296,55	0,427904

Figura 4.2: Arquivo de Referência.

	A	B
1	12489,4	0,439091
2	12481,69	0,437306
3	12473,98	0,437511
4	12466,26	0,438941
5	12458,55	0,438559
6	12450,83	0,436687
7	12443,12	0,435625
8	12435,4	0,435095
9	12427,69	0,434061
10	12419,98	0,433113
11	12412,26	0,433476
12	12404,55	0,434622
13	12396,83	0,434394
14	12389,12	0,432504
15	12381,4	0,430138
16	12373,69	0,428098
17	12365,98	0,426806
18	12358,26	0,426195
19	12350,55	0,426324
20	12342,83	0,426779
21	12335,12	0,426441
22	12327,4	0,425185
23	12319,69	0,424239
24	12311,98	0,423879
25	12304,26	0,422966
26	12296,55	0,421867

Figura 4.3: Arquivo de Teste.

	A	B
1	Referência_1	
2	Referência_2	
3	Referência_3	
4	Referência_4	

Figura 4.4: Ordem Referência

	A	B
1	Teste_1	
2	Teste_2	
3	Teste_3	

Figura 4.5: Ordem Teste

	A	B
1	8336,0593	
2	6820,738	
3	5789,954	
4	5172,253	
5	4668,493	
6	4337,303	
7	4259,195	

Figura 4.6: Arquivo Picos.

Os arquivos estão ilustrados nas figuras 4.2, 4.3, 4.4, 4.5 e 4.6.

Ao rodar o programa principal, primeiro, é solicitado ao usuário que digite o número de planilhas contidas no arquivo Referencia.xls e então o número de planilhas do Teste.xls. Então é lido o arquivo Ordem Referencia e o arquivo Referências, dentro de um *looping for* limitado pelo número de planilhas (é lida uma planilha em

cada *loop*). Em seguida é feito, respectivamente, para os arquivos Ordem Teste e Teste, o mesmo procedimento descrito anteriormente. São lidas, ainda, as frequências dos picos eleitos da planta teste, que estão dispostas no arquivo Picos. Os espectros obtidos do *software* Opus têm o mesmo número de linhas (por exemplo, 1154), informação é essencial para o bom funcionamento do programa. Então os espectros são colocados em uma matriz (de espectros), para que esta seja passada como parâmetro para as funções. É necessário colocar a tabela do espectro da planta teste, como sendo as duas últimas colunas da matriz. Foi utilizado um *looping for* para obter as intensidades dos picos das plantas referências e da planta teste (função *intensidadepico*) e a *fuzzy*ificação (função *fuzzy*ificação), para cada pico eleito. Com isso é procedida a *defuzzy*ificação dos resultados, usando a média e o mínimo. A média é calculada somando os graus de pertinências de cada pico e dividindo pelo número de picos. O mínimo é obtido ordenando os graus de pertinências de cada pico e elegendo o menor deles.

A função *intensidadepico.m* recebe como parâmetro, do programa principal, uma matriz contendo os espectros (tabela com as frequências e as respectivas intensidades) e a frequência do pico eleito. Então a função devolve um vetor com as intensidades de todas as plantas para aquele pico (devolve na ordem em que os espectros estão na matriz passada como parâmetro). Inicialmente acha-se o número de espectros de plantas contidas na matriz e então para cada planta são feitos os limites de frequência, que circundam a frequência do pico eleito. Foram eleitos 15 pontos de maior frequência e 15 de menor frequência. Com isso são ordenadas as intensidades, que estão dentro deste limite e escolhe-se a maior destas, como sendo a máxima intensidade para este pico.

A função *fuzzificacao.m* recebe como parâmetro, do programa principal, o vetor contendo as intensidades dos picos e devolve um vetor com o grau de pertinência da planta teste (última intensidade do vetor, uma vez que é, obrigatoriamente, o último espectro da matriz de espectros), com relação a cada planta de referência. Primeiramente as intensidades das plantas da base de referências são ordenadas, para que seja encontrada a maior e a menor intensidade e para que seja calculado o *range* máximo e mínimo valor. Isto se faz necessário, para que seja criada a base do triângulo da função de pertinência, da planta de maior e menor intensidade, como foi referido. Então são ordenadas as intensidades das plantas referências, da planta teste e o *range* máximo e mínimo, para que sejam feitos os cálculos dos graus de pertinência. O grau de pertinên-

cia, da planta teste, com relação as funções de pertinências das plantas de referências vizinhas, é calculado por semelhança de triângulo. Este processo está ilustrado na figura 4.7.

Figura 4.7: Grau de pertinência da planta teste.

Resultados

Primeiramente foi montada a base de referência com as seguintes plantas, conforme mostra a tabela 4.2. Trata-se dos espectros de plantas consideradas como referências (catalogadas, reconhecidas pelo taxonomista), obtidos da reprodutibilidade zero (primeira aquisição espectral, desta mesma amostra, de um total de 6 a 7). Primeiro, foram feitos testes submetendo os espectros da própria base de referência, e o resultado obtido, para o espectro da planta teste igual o da planta referência, foi de grau 1 de pertinência para todos os picos, assim como esperado. Como pode ser observado a seguir.

Também são apresentados casos onde o número apresenta o comportamento *fuzzy*, com valores inferiores a um. Casos em que são testados espectros da mesma espécie e diferente reprodutibilidade, entre outros. E, os casos de amostra teste versus a base

de referência. Espera-se que esta versão Beta pretenda levantar esta discussão em nossa comunidade.

Tabela 4.1: Espectros que formam a Base de Referência, formam a primeira aquisição espectral com reprodutibilidade espectral de 6 a 7.

Número	Nome da amostra	Número	Nome da amostra
1	Jab_penna_Nutri_0	34	Alfafa_0
2	Abacateiro_0	35	AssaPeixe_0
3	Baleeira_0	36	Carobinha_0
4	Avenca_0	37	ErvadeBicho_0
5	Artimisia_0	38	ErvStaMaria_0
6	CardoSanto_0	39	Sene_0
7	CatingaMulata_0	40	Porongaba_0
8	ChapeudeCouro_0	41	Hortela_0
9	CipoCabeludo_0	42	ChaBugre_0
10	DenteLeao_0	43	Malva_0
11	FolhadeLaranja_0	44	UnhadeGato_0
12	Funcho_0	45	ErvSaoJoao_0
13	Doradinha_0	46	Melissa_0
14	Graviola_0	47	Mentruz_0
15	Guassatonga_0	48	Tanchagem_0
16	Guine_0	49	ChaBranco_0
17	Losna_0	50	Alfazema_0
18	Menta_0	51	Maracuja_0
19	PatadeVaca_0	52	Ervadoce_0
20	PauTenente_0	53	SeteSangria_0
21	PinhaoRoxo_0	54	BoldoChile_0
22	Verbena_0	55	Barbatimao_0
23	Cidreira_0	56	Alecrim_0
24	Espinheira_0	57	Oregano_0
25	Guaco_0	58	Hibisco_0
26	Jab_micro_ST_0	59	Amoreira_0
27	Jambolao_0	60	Carqueja_0
28	Panaceaia_0	61	Arnica_0
29	Poejo_0	62	CentellaAsiatica_0
30	Pulmonaria_0	63	Cavalinha_0
31	Salvia_0	64	Guinkgo_0
32	Stevia_0	65	Camomila_0
33	Tilia_0	66	Alcachofra_0

1 Teste =

'Jab_micro_ST_0' 'X' 'Jab_micro_ST_0'

Respostas =

'Pico'	[1]	[1]
'Pico'	[2]	[1]
'Pico'	[3]	[1]
'Pico'	[4]	[1]
'Pico'	[5]	[1]
'Pico'	[6]	[1]
'Pico'	[7]	[1]

Media = 1 Minimo = 1

2 Teste =

'Jab_penna_Nutri_5'	'X'	'Jab_penna_Nutri_0'
---------------------	-----	---------------------

Respostas =

'Pico'	[1]	[0.6030]
'Pico'	[2]	[0.6523]
'Pico'	[3]	[0.9041]
'Pico'	[4]	[0.1655]
'Pico'	[5]	[0.7048]
'Pico'	[6]	[0.9216]
'Pico'	[7]	[0.9216]

Media =

0.6961

Minimo =

0.1655

3 Teste =

'Jab_penna_Nutri_0' 'X' 'Jab_penna_Nutri_0'

Respostas =

'Pico' [1] [1]

'Pico' [2] [1]

'Pico' [3] [1]

'Pico' [4] [1]

'Pico' [5] [1]

'Pico' [6] [1]

'Pico' [7] [1]

Media = 1 Minimo = 1

4 Teste =

'ChapeudeCouro_5' 'X' 'ChapeudeCouro_0'

Respostas =

'Pico' [1] [0.9180]

'Pico' [2] [0.7282]

'Pico' [3] [0.7794]

'Pico' [4] [0]

'Pico' [5] [0.8009]

Media = 0.6453

Minimo = 0

5 Teste =

'FolhadeLaranja_5' 'X' 'FolhadeLaranja_0'

Respostas =

'Pico'	[1]	[0.6569]
'Pico'	[2]	[0]
'Pico'	[3]	[0.8075]
'Pico'	[4]	[0]
'Pico'	[5]	[0.4423]
'Pico'	[6]	[0.8425]

Media = 0.4582

Minimo = 0

6 Teste =

'Guine_5' 'X' 'Guine_0'

Respostas =

'Pico'	[1]	[0.9949]
'Pico'	[2]	[0.8589]
'Pico'	[3]	[0.8627]
'Pico'	[4]	[0.5949]
'Pico'	[5]	[0]

Media = 0.6623

Minimo = 0

7 Teste =

'PauTenente_5' 'X' 'PauTenente_0'

Respostas =

'Pico'	[1]	[0.4613]
'Pico'	[2]	[0]
'Pico'	[3]	[0]
'Pico'	[4]	[0]
'Pico'	[5]	[0]
'Pico'	[6]	[0.8236]

Media = 0.2142

Minimo = 0.

8 Teste =

'Verbena_5'	'X'	'Verbena_0'
-------------	-----	-------------

Respostas =

'Pico'	[1]	[0.7362]
'Pico'	[2]	[0.9783]
'Pico'	[3]	[0.9892]
'Pico'	[4]	[0.0796]
'Pico'	[5]	[0.9462]

Media = 0.7459

Minimo = 0.0796

9 Teste =

'Cidreira_5'	'X'	'Cidreira_0'
--------------	-----	--------------

Respostas =

'Pico'	[1]	[0.9526]
'Pico'	[2]	[0.4213]
'Pico'	[3]	[0]

```
'Pico'      [4]      [      0]
'Pico'      [5]      [0.9200]
```

Media = 0.4588 Minimo = 0

10 Teste =

```
'Alfafa_5'   'X'      'Alfafa_0'
```

Respostas =

```
'Pico'      [1]      [      0]
'Pico'      [2]      [      0]
'Pico'      [3]      [      0]
'Pico'      [4]      [      0]
'Pico'      [5]      [0.7372]
'Pico'      [6]      [0.3983]
```

Media = 0.1892 Minimo = 0

11 Teste =

```
'Jab_micro_ST_5' 'X'      'Abacateiro_0'
```

Respostas =

```
'Pico'      [1]      [      0]
'Pico'      [2]      [      0]
'Pico'      [3]      [0.9071]
'Pico'      [4]      [      0]
'Pico'      [5]      [      0]
'Pico'      [6]      [      0]
'Pico'      [7]      [      0]
```

Media = 0.1296 Minimo = 0

12 Teste =

'BoldoChile_5' 'X' 'Jab_micro_ST_0'

Respostas =

'Pico' [1] [0]

'Pico' [2] [0]

'Pico' [3] [0]

'Pico' [4] [0.1732]

'Pico' [5] [0]

Media = 0.0346 Minimo = 0

13 Teste =

'Jab_penna_Nutri_5' 'X' 'Tilia_0'

Respostas =

'Pico' [1] [0]

'Pico' [2] [0.3477]

'Pico' [3] [0]

'Pico' [4] [0]

'Pico' [5] [0]

'Pico' [6] [0]

'Pico' [7] [0]

Media = 0.0497 Minimo = 0

14 Teste =

'BoldoChile_5' 'X' 'BoldoChile_0'

Respostas =

'Pico'	[1]	[0.9698]
'Pico'	[2]	[0.8722]
'Pico'	[3]	[0.7367]
'Pico'	[4]	[0]
'Pico'	[5]	[0.7516]

Media = 0.6660 Minimo = 0

É fácil de deduzir, as possibilidades de testes são inúmeras, isto considerando apenas esta pequena base de referência Beta. Sem dúvida, trata-se da necessidade de compartilhar estes resultados desta versão Beta, na rede mundial para sua crítica e aperfeiçoamento. Ainda somando-se os espectros adquiridos.

Para fins de observação, do comportamento do sistema, foi montado uma base de referência contendo as seis amostras reproduzidas do Jaborandi *Pilocarpus microphyllus* e do Jaborandi *Pilocarpus pennatifolius*. E foram testadas as seis reprodutibilidades de cada um dos Jaborandis teste, listados abaixo. Apresentaremos alguns casos, pois esta listagem implicaria em 100 páginas. Em mídia eletrônica estes resultados serão disponibilizados. A tabela seguinte apresenta a lista das amostras teste:

Tabela 4.2: Espectros que formam a Base de Referência de Testes com Jaborandis.

Número	Nome da amostra
1	FolBoaVista
2	FolCumprCeag
3	FolCumprUsp
4	FolLojaIndio
5	FolMercPub
6	Pilo_micro_Panizza
7	FolRed_STech
8	FolRedUsp
9	FolSehPrim
10	FolSehSeg
11	FolSuely
12	FolZeFranco

Nos resultados obtidos houve, pelo menos, um pico com grau de pertinência diferente de zero, alguns estão mostrados a seguir, e demonstram claramente que há indícios de que o Jaborandi Folha Comprida USP seja da espécie *Pilocarpus pennatifolius* e

certamente pelas médias abaixo de 0.0 trata-se de uma espécie diferente do *Pilocarpus microphyllus*.

Teste =

'FolCumprUsp_0' 'X' 'Pilo_Penna_Nutri_3'

Respostas =

'Pico'	[1]	[0]
'Pico'	[2]	[0.8458]
'Pico'	[3]	[0.9700]
'Pico'	[4]	[0.8591]
'Pico'	[5]	[0.8847]
'Pico'	[6]	[0.9757]
'Pico'	[7]	[0.9757]

Media = 0.7873 Minimo = 0

Teste =

'FolCumprUsp_1' 'X' 'Pilo_Penna_Nutri_3'

Respostas =

'Pico'	[1]	[0]
'Pico'	[2]	[0.8628]
'Pico'	[3]	[0.9885]
'Pico'	[4]	[0.8757]
'Pico'	[5]	[0.9044]
'Pico'	[6]	[0.9957]
'Pico'	[7]	[0.9957]

Media = 0.8033 Minimo = 0

Teste =

'FolCumprUsp_3' 'X' 'Pilo_Penna_Nutri_3'

Respostas =

'Pico'	[1]	[0]
'Pico'	[2]	[0.8362]
'Pico'	[3]	[0.9647]
'Pico'	[4]	[0.8496]
'Pico'	[5]	[0.8824]
'Pico'	[6]	[0.9757]
'Pico'	[7]	[0.9757]

Media = 0.7835 Minimo = 0

Teste =

'FolCumprUsp_4' 'X' 'Pilo_Penna_Nutri_3'

Respostas =

'Pico'	[1]	[0]
'Pico'	[2]	[0.8606]
'Pico'	[3]	[0.9882]
'Pico'	[4]	[0.8644]
'Pico'	[5]	[0.8991]
'Pico'	[6]	[0.9919]
'Pico'	[7]	[0.9919]

Media = 0.799 Minimo = 0

4.2 Aplicação do Método das Diferenças com a Lógica *Fuzzy*

Introdução

Nesta segunda implementação de acordo com o artigo do Ebel (EBEL, 1992), com o objetivo de se diferenciar plantas medicinais através de seus espectros NIR, foi estudado o uso de Lógica *Fuzzy* em combinação com o Método das Diferenças para se realizar esta avaliação, chamaremos de Beta2. Foram obtidas diversas amostras de espectros para servirem de referências e testes, sendo escolhidas como referências aquelas que possuíam autenticação pelo taxonomista. Após o estudo do Método das Diferenças e da Lógica *Fuzzy*, foi desenvolvido, no ambiente de programação do MatLab, um programa que implementa a comparação entre espectros introduzidos pelo usuário contra uma base de dados interna e retorna os resultados sobre a semelhança entre a amostra de teste e cada amostra de referência no base de referência. A seguir será discutido o método de avaliação abordado e sua implementação, assim como os resultados obtidos com a implementação do método para diversas entradas. Ao final, serão discutidas perspectivas futuras para o aprimoramento do código do programa e sua execução.

Descrevendo método

Com o objetivo de identificar plantas, um dos meios de se realizar essa diferenciação é a avaliação e comparação do espectro IVP da planta desconhecida com os espectros das plantas conhecidas usando-as como referências. Dentro deste meio de identificação, o método direto consiste em achar a diferença entre as duas curvas dos espectros subtraindo-os, assim avaliando a nova curva obtida, que representa esta diferença, sendo este o Método das Diferenças, conforme Ebel (EBEL, 1992). Ainda, para avaliar esta diferença, uma aproximação possível é a conversão desta diferença em um número *fuzzy*, que avalia de ponto-a-ponto a distância entre a referência e a amostra, convertendo essa distância em um número que pode variar entre 0 (muito distante) e 1 (estão no mesmo ponto). A vantagem no uso do número *fuzzy* para esta avaliação está no fato de que o número *fuzzy* pode assumir qualquer valor entre 0 e 1, podendo assim, a avaliação da distância entre dois pontos ser definida com 0,5 (distância média) ou também 0,75 (relativamente próximo).

O Método das Diferenças pode ser introduzido pela constatação de que consiste em

avaliar a distância entre as curvas subtraindo-as diretamente e classificar esta diferença como grande ou pequena de acordo com uma certa margem de referência. O princípio deste método é amplamente utilizado em aproximações numéricas, onde diferenças de ordem abaixo de 3 ou 6 casas decimais (dependendo da aproximação requerida) entre dois valores podem ser desconsideradas, julgando-se estes valores como iguais. Este método requer uma margem de referência para que sirva de base para o julgamento da pertinência deste valor à decisão a ser tomada. Para aproximações numéricas, isto significa que os valores abaixo de certa casa decimal passam a ser ignorados. Por exemplo: para uma avaliação da terceira casa decimal, os números 3,4151 e 3,4152 são iguais a 3,415 (a diferença de 0,0001 é desconsiderada). Como o valor desta margem de avaliação depende diretamente da diferença entre as absorvâncias dos espectros de duas plantas esta margem foi obtida neste trabalho por meio de dados estatísticos relativos ao grupo de amostras de referência, a será chamada de λ (lambda). A avaliação entre espectros de plantas não possui margens definidas para dizer que esta diferença possa ser desprezada. Fez-se necessário o uso da Lógica *Fuzzy* para julgar a diferença, aproximando estas diferenças para números *fuzzy*, com o uso do λ como parâmetro referencial.

Aplicação do Método das Diferenças com a Lógica *Fuzzy* A aplicação do método das diferenças, segundo Ebel, com a Lógica *Fuzzy* consiste em transformar a curva da amostra de referência em um conjunto *fuzzy* e calcular a intersecção entre esse conjunto e a curva da amostra de teste. Como a curva da amostra de teste pode ser representada diretamente como um conjunto simples, e não como um conjunto *fuzzy*, essa intersecção se torna mais simples de ser calculada, não requisitando as regras de intersecção de conjuntos *fuzzy*. Definiu-se as funções de pertinência da curva da amostra de referência como sendo a triangular (Equação 1) e a parabólica (Equação 2), onde k é o valor do número de onda, $v_{ref}(k)$ é o valor da absorvância da curva de referência em função do número de onda, o $v_{teste}(k)$ é o valor da absorvância da curva de teste em função do número de onda e o λ é o parâmetro comparativo que define se a diferença é grande ou pequena. Como a curva da amostra de teste é representada por um conjunto simples, ela pode ser vista como sendo uma reta, significando que apenas aquele valor de absorvância pertence à curva. Dessa forma, cada valor de número de onda tem associado a si uma diferença entre as curvas, e, portanto, um valor de pertinência relacionado a essa diferença. Esse valor de pertinência representa o quão pequena é a diferença entre as curvas (se $f(k)$ está próximo de 1) ou o quão grande é a distância (se $f(k)$ está próximo de 0).

$$f_{triang}(k) = \max(0, 1 - |\frac{v_{ref}(k) - v_{teste}(k)}{\lambda}|)$$

Esta equação apresenta a função de pertinência triangular para calcular a intersecção entre as curvas de Referência e de Teste.

$$f_{quad}(k) = \max(0, 1 - (\frac{v_{ref}(k) - v_{teste}(k)}{\lambda})^2)$$

Esta equação apresenta a função de pertinência quadrática para calcular a intersecção entre as curvas de Referência e de Teste. A seguir pode ser observado o comportamento das curvas da amostra de referência e de amostra de teste, sendo vistas em um único ponto da curva, onde o eixo x representa a distância entre as curvas em função do λ , e o y a pertinência resultante. As retas perpendiculares ao eixo x representam as curvas das amostras de teste (sendo a pontilhada a pertinência da curva e a reta a intersecção entre as curvas de referência e teste) e as curvas, as funções de pertinência da curva de referência. Vale ressaltar que nessas comparações, que a pertinência quadrática retorna valores de pertinência maiores para um mesmo valor de distância, se comparada com a função de pertinência triangular.

É possível observar a partir destes gráficos, que diferenças maiores que o parâmetro λ retornam valores de pertinência nulos, indicando que a curva de teste não tem nenhuma semelhança com a curva de referência, e que diferenças próximas de 0, retornam valores de pertinência próximos de 1, indicando alta semelhança entre as curvas. Como entradas do programa foram reunidos 213 espectros IVP de 70 plantas para servirem de referências para as avaliações, montando-se assim a base de referência. Também foi definida uma Região de Interesse (*Region Of Interest ROI*), que tinha como objetivo limitar os valores de número de onda a serem analisados. Esta pode ser escolhida, é definida pelo usuário ao início da execução do programa, estando esta intervalo dentro do intervalo dos dados, *range* das frequências IVP. Após definida a ROI, os espectros do base de referência foram avaliados estatisticamente em relação à distância média entre os mesmos (calculadas por média comum e por rms *Root Mean Square*) e as variâncias dessas diferenças. As médias das diferenças permitiram saber o quão distantes as curvas estavam entre si, e as variâncias o quanto essa diferença mudava ao longo da curva. Foi observado que algumas diferenças possuíam médias pequenas, porém

Figura 4.8: Pertinência de uma diferença de $0.3 * \lambda$ para as funções de pertinência quadrática (à esquerda) e modular (à direita).

variâncias grandes, indicando que as curvas eram diferentes entre si, mas essa diferença se equilibrava, pois em certas partes da curva esta diferença era negativa e em outros era positiva, nivelando a diferença média. Portanto, passou a ser usado apenas a média RMS para se observar as diferenças médias, visto que esta despreza os sinais das diferenças. Também deve ser observado que esta diferença média entre as curvas, calculadas a partir da média em RMS pode ser chamada de Distância Euclidiana Média, sendo referida neste documento a partir deste ponto como DEM. Tal pode ser obtida através da seguinte equação, a seguir descrita.

$$DEM = \sqrt{\sum_{k=1}^n \frac{(v_{ref}(k) - v_{teste}(k))^2}{n}}$$

Esta equação apresenta a Distância Euclidiana Média.

O λ foi determinado a partir da média das DEM entre os pares de curvas da base de referência e foi usado para avaliar as amostras de teste contra as amostras de referência. Primeiramente usou-se o λ para avaliar a diferença entre os espectros do próprio

Figura 4.9: Pertinência de uma diferença de $0.75 * \lambda$ para as funções de pertinência quadrática (à esquerda) e modular (à direita).

base de referência contra eles mesmos, tendo esta redundância o objetivo de testar o Método das Diferenças contra ele mesmo. Os resultados mostraram que as curvas do base de referência eram todas diferentes entre si, demonstrando a confiabilidade do método. A implementação foi dividida majoritariamente na parte da organização dos dados e processamento dos dados. Os detalhes dessas duas divisões e de cada uma de suas sub-etapas é definida e explicada em profundidade a seguir. A avaliação da semelhança entre as curvas se dá a partir da análise do usuário dos resultados demonstrados pelo programa, a respeito das funções de pertinência, DEM, variância e correlação, calculados e exibidos. A correlação é uma informação estatística processada a partir dos espectros de teste e referência que indica a similaridade entre os formatos das curvas. Foi utilizado neste trabalho o coeficiente de correlação r de Pearson, função embutida no programa Matlab, versão 7.0.1.

Etapas de Implementação

Os Dados foram organizados de modo a adquirir as curvas de absorbância das amostras de referência organizadas e também de forma que fosse criada uma tabela

contendo os nomes de cada espectro. Os espectros já foram previamente manipulados, eliminando-se o eixo contendo os número de onda das curvas, por ser um eixo em comum entre todas, e, portanto, só representariam um aumento desnecessário na quantidade de dados.

A criação da base de referências se deu através da carga de uma tabela contendo todos os espectros concatenados em ordem de acordo com a lista criada anteriormente. Desta forma, nessa nova tabela cada linha da mesma corresponde ao nome guardado na linha de mesmo valor na lista. Isso simplifica o acesso e manipulação dos dados. O primeiro passo na manipulação da base de referência é a definição a ROI a partir de uma entrada do usuário. Esta entrada é crucial, pois a escolha de uma ROI pode levar a uma interpretação errada dos resultados, visto que certos números de onda carregam informações desnecessárias para a análise. Para os resultados demonstrados, foi escolhido como a ROI (*default*), o intervalo entre 8750 e 4200 cm^{-1} . Após a entrada da ROI pelo usuário, o programa “corta” o eixo contendo os números de onda e o base de referência, eliminando todos os valores fora da ROI. Esta escolha também pode permitir a análise de regiões chave, como picos das curvas. O processamento dos dados das Referências começou com a obtenção do λ , onde foi necessário um processamento das amostras de referência, como citado anteriormente, a partir dos cálculos da DEM de cada par de curvas do base de referência. Para tal, são obtidas algumas informações sobre o grupo de amostras de referência, tal como número de amostras no base de referência. Esses dados são usados para se criar de antemão as matrizes que receberão os dados estatísticos para servirem como auxiliares nas avaliações. O processamento das Referências para a obtenção do λ ou (*Fuzzyficação*, tem início com a obtenção do λ , onde é realizada a diferença entre todas as curvas do base de referência (par a par e sem repetição), e após cada diferença é calculada a DEM e a variância dessa diferença. A variância é a informação sobre a forma desta diferença (quanto mais próxima de zero, mais constante é esta diferença, e mais semelhantes são as formas dessas curvas, mesmo que sejam distantes uma da outra). O λ passa a ser definido como a maior diferença média entre duas curvas (escolha do pior caso possível). Avaliação das Referências usando o λ (*Defuzzyficação*) Com o λ obtido, são avaliadas as diferenças entre as curvas do base de referência para testar o método de avaliação contra o próprio base de referência, como descrito anteriormente, através do uso das Equações 1 e 2. Os resultados são demonstrados mostrando-se os gráficos das duas curvas sendo avaliadas,

o gráfico de sua diferença, os gráficos das pertinências das curvas para cada função, a DEM, a variância da diferença, os valores médios das funções de pertinência e a correlação entre as duas curvas. A avaliação triangular é mais restrita do que a quadrática, como se pode observar e fora discutido nas figuras mostradas a seguir, onde para um mesmo valor de diferença, a quadrática retorna um valor de pertinência do que a modular. A entrada das amostras de teste é processada, quando requisita-se do usuário a entrada de um espectro para ser testado contra o base de referência. Por simplicidade, os espectros foram previamente manipulados para reduzir a quantidade de informação, assim como feito com os espectros de referência.

O processamento da Entrada de Teste (ou *Fuzzyficação*) tem início com entrada sendo *cortada* de acordo com a ROI e testada contra todo o base de referência, assim como a base de referência foi testado contra ele mesma, ou seja, é calculada a diferença entre a entrada e todo a base de referência um a um, os gráficos das pertinências das curvas para cada função, a DEM, a variância da diferença, os valores médios das funções de pertinência e a correlação entre as duas curvas. Estes valores são salvos pelo programa, para caso o usuário queira, após a execução do programa, observar os dados. Os resultados são salvos pelo MatLab na forma de arquivos de texto contendo somente os valores de cada avaliação, na ordem que foram avaliados, requerindo que o usuário compare a linha que está localizado o dado com a lista das amostras do base de referência. A avaliação da entrada de teste ou (*Defuzzyficação*) se dá após a avaliação da entrada contra o base de referência, o usuário é questionado sobre quais resultados serão mostrados e como serão mostrados. Os resultados são mostrados de forma similar à avaliação das amostras da base de referência. Devido a grande quantidade de métodos de avaliação e resultados presentes para o usuário, foi deixado ao critério deste definir a avaliação como satisfatória ou não. O programa de Avaliação, segue em anexo, o programa foi totalmente comentado, após a sua escrita no ambiente de programação do MatLab. Para a execução do programa, é necessário uma tabela de referência, as funções de corte dos eixos e amostras para serem testadas contra o grupo de amostra de referências.

Os resultados e discussões deste programa podem ser acompanhados em anexo, onde são mostrados os valores da avaliação do base de referência contra ele próprio, constando os nomes de cada comparação, os valores da DEM, variância, valor médio das funções de pertinência e correlação entre as curvas para cada comparação. São

Figura 4.10: Avaliação de AbacateiroMarcio contra Abacateiro0.

mostradas a seguir para ilustrar o funcionamento do programa, algumas avaliações realizadas considerando a base de referência versus uma dada amostra de teste. Devido

ao grande número de amostras de referência na base de dados, são mostradas apenas algumas avaliações para cada amostra de teste. Avaliação de AbacateiroMarcio contra Abacateiro0, conforme figura 4.10

```
AbacateiroMarcio contra Abacateiro0
DEM 0.0000000e+000
Correlacao 1.0000000e+000
Avaliacao Quadratica 1.0000000e+000
Avaliacao Triangular 1.0000000e+000
Variancia das Diferencas 0.0000000e+000
```

Essa avaliação é um caso de perfeita semelhança entre as curvas, como pode ser visto pela nula DEM e variância da diferença, e valor unitário para a correlação e avaliações quadrática e triangular, lembrando que estas avaliações em *fuzzy* significa a exata semelhança entre as curvas.

Avaliação de AbacateiroMarcio contra Abacateiro5, conforme figura 4.11

```
AbacateiroMarcio contra Abacateiro5
DEM 1.0964919e-004
Correlacao 1.0000000e+000
Avaliacao Quadratica 9.9995794e-001
Avaliacao Triangular 9.9437749e-001
Variancia das Diferencas 3.9667630e-007
```

Essa avaliação é um caso de altíssima semelhança entre as curvas, como pode ser visto pela baixa DEM e variância da diferença, e valor quase unitário para a correlação e avaliações quadrática e triangular.

Levando-se em conta a avaliação anterior, e que esta nova referência é apenas uma reprodutibilidade da mesma amostra, sendo ainda a mesma amostra.

Avaliação de AbacateiroMarcio contra Alcachofra0, conforme figura 4.12

```
AbacateiroMarcio contra Alcachofra0
DEM 7.1113462e-003
```

Figura 4.11: Avaliação de AbacateiroMarcio contra Abacateiro5.

Correlacao 9.9337595e-001

Avaliacao Quadratica 8.2310438e-001

Figura 4.12: Avaliação de AbacateiroMarcio contra Alcachofra0.

Avaliacao Triangular 6.5059439e-001

Variancia das Diferencas 6.6899309e-004

Figura 4.13: Avaliação de AbacateiroMarcio contra Alecrim0.

Essa avaliação é um caso de baixa semelhança entre as curvas, como pode ser visto pela alta DEM e variância da diferença, valores medianos (próximos de 0,5) para as

avaliações quadrática e triangular, porém com alta semelhança entre os formatos das curvas, demonstrado pela alta correlação.

Avaliação de AbacateiroMarcio contra Alecrim0, conforme figura 4.13

```
AbacateiroMarcio contra Alecrim0
DEM 1.4605993e-002
Correlacao 9.9294790e-001
Avaliacao Quadratica 4.7552487e-001
Avaliacao Triangular 3.8834140e-001
Variancia das Diferencas 2.8752713e-003
```

Levando-se em conta a avaliação anterior, e que esta nova referência é apenas uma reprodutibilidade da mesma amostra, sendo ainda a mesma amostra. Essa avaliação é um caso de muito baixa semelhança entre as curvas, como pode ser visto pela alta DEM e variância da diferença, valores pequenos (abaixo de 0,5) para as avaliações quadrática e triangular, porém com alta semelhança entre os formatos das curvas, demonstrado pela alta correlação.

Perspectivas Futuras

O método demonstrado possui flexibilidade em sua mudança e adição de novos métodos avaliativos, podendo ser amplamente utilizado para comparações de curvas. Com a adição de mais amostras no base de referência, o método de organização demonstrado aqui se tornará ineficaz e de difícil manutenção, sendo necessária sua alteração com o acréscimo de estruturas de dados (structs) para persistência dos mesmos. Este método organizacional também permitirá a diferenciação das amostras de referência em espécies e subespécies, ainda possibilitando um procedimento de criação de λ 's específicos para diferenciação dentro destes subconjuntos. As variâncias demonstraram que algumas curvas apesar de um pouco distantes, podem ter formas semelhantes, o que poderia significar que ainda são a mesma curva, apenas estando levemente deslocada horizontalmente. Nestes casos, pode-se criar uma rotina de reavaliação da diferença após a realização de uma normalização dessa diferença. Devido a importância dos picos em gráficos de espectro IVP, ainda se poderia adicionar à avaliação um aumento da importância das regiões dos picos nas avaliações através de fatores multiplicativos, por exemplo.

5 *Discussão dos Resultados*

A discussão dos resultados deste trabalho, da chamada versão Beta implementada, remete diretamente à continuidade do trabalho, inclusive em ambiente compartilhado com a comunidade científica na rede mundial, Internet.

5.1 *Assinatura Espectral Id*

O índice de acerto do Id depende de vários fatores, uma vez, estabelecida esta “configuração” de teste, o sistema apresenta os seguintes resultados:

- Estado bem comportado, em termos conclusivos, existe uma banda de confiança adequada, a partir do desvio padrão, adequada permitindo o acerto de praticamente todos os casos, e a base de referência comporta esta variabilidade.

No presente estudo por falta desta maior adequação, ocorrem casos assim chamados falso negativos, na prática a droga está identificada, mas o sistema informa desconhecer.

No restante, todas as amostras são reconhecidas conforme o esperado.

- Estado variável existem erros de avaliação que são resolvidos mediante uma nova configuração a ser estudada, nestes casos, já os problemas de separabilidade são apresentados (referência ao capítulo 3).

Com a fatorização são reduzidos drasticamente, ficando praticamente desprezíveis os resultados *confusos* do sistema, novamente na separabilidade (com o novo método - especializado - análise de fatores).

Outro fator importante neste caso é a retirada dos *outliers*, com o aumento ainda maior do nível de reconhecimento, até o limite (estatístico) de confiança do sistema.

Uma solução utilizada que equaciona de modo melhor ainda as questões deste estado da amostra é a utilização de bandas de confiança, para melhor abrigar as variabilidades.

- Análise de *clusters* dado que a sua entrada é a mesma do Id, os espectros e dadas as restrições deste tipo de análise, seu comportamento é bastante estável.

Cabe aqui ressaltar o potencial encontrado no dendograma, pois abriga com facilidade, adequadamente, as reprodutibilidades das amostras de referência.

Um espectro é identificado corretamente, dada a sua similaridade a algum dos espectros do dendograma, sendo corretamente *indexado*.

5.2 Assinatura Espectral Fuzzy

O modelo **Blafflert** apresentou os seguintes resultados conclusivos para o momento.

Considerando a base de referência com 66 drogas vegetais, foram obtidas os seguintes resultados usando número fuzzy:

- O primeiro grupo completo foi testado de acordo com os modos listados no capítulo anterior. De um total de 492 testes, 97 Identificados foram encontrados considerando um número fuzzy maior que 0,5, aproximadamente 20% de Identificados.

Cabe aqui a discussão da contribuição destes resultados.

O que é a pertinência (número) *fuzzy*?

Numa análise empírica foi feito este *cut* utilizando 0,5 para separar as amostras em um arquivo de Identificados e NãoIdentificados, a partir das médias calculadas.

Sabe-se claramente que dentro destes 20% estão contemplados os casos idênticos(amostra de referência versus amostra de referência)¹, portanto a princípio, a implementação, com erro desprezível, visualizado na bateria de testes, funciona.

Todavia neste intervalo da vizinhança de 0,5, existem casos e casos, que precisam ser analisados para ser procedido o *tuning* do sistema Beta.

¹o sistema acerta **sempre** que uma amostra bem comportada lhe é oferecida e é “fácil” de separar, isto é, o sistema fuzzy acerta em 1, quando a amostra é idêntica, naturalmente.

Tabela 5.1: Valores médios agrupados de testes entre as plantas testes obtidas da reprodutibilidade 5 e da respectiva planta referência de reprodutibilidade zero.

PlantaTeste versus Planta Referência	Número <i>Fuzzy</i>
Ervadoce5XErvadoce0	0,9695
Sene5XSene0	0,933
Espinheira5XEspinheira0	0,9223
MentaXMenta0	0,8839
Barbatimao5XBarbatimao0	0,882
Malva5XMalva0	0,8351
Guassatonga5XGuassatonga0	0,8102
Guaco5XGuaco0	0,7543
Verbena5XVerbena0	0,7459
CatingaMulata5XCatingaMulata0	0,7457
Abacateiro5XAbacateiro0	0,699
JabpennaNutri5XJabpennaNutri0	0,6961
Graviola5XGraviola0	0,6864
BoldoChile5XBoldoChile0	0,666
Guine5XGuine0	0,6623
ChapeudeCouro5XChapeudeCouro0	0,6453
Mentruz5XMentruz0	0,6362
ErvStaMaria5XErvStaMaria0	0,6025
Poejo5XPoejo0	0,6004
Alcachofra5XAlcachofra0	0,5882
CardoSanto5XCardoSanto0	0,5838
ChaBugre5XChaBugre0	0,5826
PatadeVaca5XPatadeVaca0	0,5597
ErvadeBicho5XErvadeBicho0	0,5514
Doradinha5XDoradinha0	0,5235
Alecrim5XAlecrim0	0,5054

Assim, sendo pode-se afirmar, o problema precisa ser refinado, por exemplo, por analogia do Id, utilizando a especialização em sub-bibliotecas, que, também funcionam na análise de *clusters*, além do estudo e implementação das bandas de confiança. Sabe-se claramente que a implementação entao nesta sub-biblioteca de *melhorias, aperfeiçoamentos no trato dos picos, seu mapeamento de ambas as amostras (teste e referência)*, já representa um método fuzzy outro, novo, mais refinado para distinguir estes sinais que se especializam e foram separados (sugerindo confusão) como é o caso, das plantas que são variedades de uma mesma espécie.

O sistema em desenvolvimento esta sendo totalmente supervisionado.

Por outro lado, precisa ser garantido que o sistema desconhece, quando tem que desconhecer, dentro do restante, abaixo de 0,5, estão distantes e/ou zerados, os casos de amostras diferentes, com possível existência de algum erro, sendo apurado. Existem situações em que se tem conhecimento (da farmacognosia) da possibilidade de similaridade de espectros, ainda que de se trate de diferentes plantas. Inclusive neste sentido o sistema detecta tal similaridade.

Ainda, precisa ser considerado que o comportamento do modelo pode variar conforme o tamanho da base, claro, daí, decorre a separabilidade, modularidade. Isto leva a dedução que, além dos ajustes já referidos e no sentido da investigação para o aprimoramento do sistema, foi executado o segundo grupo de testes só com as variedades do Jaborandi.

E, o mesmo poderia ter sido procedido com o grupo das drogas vegetais comumente denominadas boldo.

Também, outra análise pode partir dos resultados, onde pode-se perceber que a saída através do mínimo não foi eficiente, uma vez que é muito sensível às variações isoladas, o que pode comprometer a resposta, pois se for testado “n” picos e apenas um dos “n” picos resultar em grau de pertinência baixo, o resultado será dado por este grau de pertinência baixo, podendo este não ser condizente com a realidade. Isto deve ser considerado com cautela, pois sabe-se que a análise de picos precisa ser aperfeiçoada nesta presente versão Beta. Ainda, com isto foram agrupadas todas as médias dos testes entre as plantas testes obtidas da reprodutibilidade 5 e com a respectiva planta referência de reprodutibilidade zero. Os resultados estão sendo mostrados em ordem nas tabelas 5.1, e 5.2.

Em seguida foi observada a maior média dos testes entre as plantas testes de reprodutibilidade 5 e de alguma referência diferente da referência da planta teste. Os resultados podem ser visualizados nas tabelas 5.3 e 5.4. Com base nas observações anteriores, foi constatado que a maior média ocorrida entre plantas diferentes foi de 0,3993, que foi o teste “Avenca5 X Losna0”. Então, para esta base de referência, se for dado uma margem de erro de, por exemplo, 10% pode-se validar testes para valores de média acima de 0,4392, ou seja, se um teste nesta base de referência resultar em uma média superior a 0,4392 então esta planta teste pode ser considerada como sendo a planta da referência a qual o teste resultou nesta média. Porém, se este valor for inferior, a este valor não se pode

afirmar que esta planta teste não é a planta da referência. Partindo destes valores os resultados que podem ser considerados válidos são apresentados nas tabelas 5.5, 5.6.

- O segundo grupo foi escalado um teste onde os Jaborandis estão separados, como um cluster em um sub-diretório a exemplo do procedimento usado no Id. Este grupo escolhido foi testado de acordo com os modos listados também, no capítulo anterior.

De um total de 159 testes, ocorreram 36 acertos considerando um número fuzzy maior que 0.5., aproximadamente 23% de acertos. Ocorre que a implementação fuzzy era a mesma! No entanto, já se tem melhorias na tendência dos reconhecimentos. Tende a reafirmar o caso que a especialização da base a exemplo dos *clusters* do *software* Id, apresenta oportunidades de melhores resultados.

Tabela 5.2: Valores médios agrupados de testes entre as plantas testes obtidas da reprodutibilidade 5 e da respectiva planta referência de reprodutibilidade zero.

PlantaTeste versus Planta Referência	Número <i>Fuzzy</i>
Oregano5XOregano0	0,4934
Maracuja5XMaracuja0	0,4881
Hortela5XHortela0	0,4635
Cidreira5XCidreira0	0,4588
FolhadeLaranja5XFolhadeLaranja0	0,4582
Funcho5XFuncho0	0,4496
Guinkgo5XGuinkgo0	0,4422
Camomila5XCamomila0	0,4241
AssaPeixe5XAssaPeixe0	0,4152
Alfazema5XAlfazema0	0,3725
Panaceia5XPanaceia0	0,3722
UnhadeGato5XUnhadeGato0	0,3515
ChaBranco5XChaBranco0	0,3484
Amoreira5XAmoreira0	0,2979
PinhaoRoxo5XPinhaoRoxo0	0,2955
Avenca5XAvenca0	0,2852
Arnica5XArnica0	0,2462
Salvia5XSalvia0	0,2412
DenteLeao5XDenteLeao0	0,2263
Cavalinha5XCavalinha0	0,2233
Losna5XLosna0	0,2156
PauTenente5XPauTenente0	0,2142
CipoCabeludo5XCipoCabeludo0	0,1988
Baleeira5XBaleeira0	0,1905
Porongaba5XPorongaba0	0,1895
Alfafa5XAlfafa0	0,1892
Stevia5XStevia0	0,1721
SeteSangria5XSeteSangria0	0,1709
Pulmonaria5XPulmonaria0	0,1357
Tilia5XTilia0	0,1285
Melissa5XMelissa0	0,1134
Jambolao5XJambolao0	0,0877
Tanchagem5XTanchagem0	0,0715
CentellaAsiatica5XCentellaAsiatica0	0
Carqueja5XCarqueja0	0
ErvaSaoJoao5XErvaSaoJoao0	0
Carobinha5XCarobinha0	0
JabmicroST5XJabmicroST0	0
Hibisco5XHibisco0	0

Tabela 5.3: Resultado dos testes entre maiores números *fuzzy* das plantas de reprodutibilidade 5 e qualquer planta de referência diferente da planta teste.

PlantaTeste versus Planta Referência	Número <i>Fuzzy</i>
Avenca5XLosna0	0,3993
CipoCabeludo5XErvadeBicho0	0,3724
Guinkgo5XCidreira0	0,3633
PinhaoRoxo5XSalvia0	0,3474
Pulmonaria5XPanaceia0	0,3273
Tanchagem5XAvenca0	0,3239
Cidreira5XMaracuja0	0,3133
CentellaAsiatica5XAmoreira0	0,2919
Oregano5XVerbena0	0,2697
SeteSangria5XLosna0	0,2496
AssaPeixe5XAlcachofra0	0,2481
Carqueja5XGuine0	0,2368
PatadeVaca5XTanchagem0	0,235
CardoSanto5XArnica0	0,23
UnhadeGato5XTanchagem0	0,2293
Porongaba5XAlfafa0	0,2241
ErvaSaoJoao5XJabmicroST0	0,2232
ErvadeBicho5XAmoreira0	0,2231
PauTenente5XMentruz0	0,2185
DenteLeao5XCarqueja0	0,2105
Baleeira5XPorongaba0	0,1998
Funcho5XJambolao0	0,1968
Alfafa5XErvadeBicho0	0,1962
Tilia5XHortela0	0,195
Doradinha5XPulmonaria0	0,1942
Carobinha5XPinhaoRoxo0	0,1934
Camomila5XLosna0	0,193
ErvaStaMaria5XCavalinha0	0,1928
Salvia5XAssaPeixe0	0,1927
Guine5XVerbena0	0,191
Maracuja5XGuine0	0,1869
Alfazema5XAvenca0	0,1859
Panaceia5XAlecrim0	0,1843
Verbena5XGuinkgo0	0,1841
ChapeudeCouro5XCardoSanto0	0,1837
Alcachofra5XChaBugre0	0,1798
Mentruz5XGuinkgo0	0,1739
Hortela5XAlecrim0	0,1705
BoldoChile5XAvenca0	0,1654
Arnica5XFuncho0	0,1626
Melissa5XGuinkgo0	0,1624
JabpennaNutri5XBarbatimao0	0,1614
Cavalinha5XPinhaoRoxo0	0,1611
Graviola5XVerbena0	0,1587

Tabela 5.4: Resultado dos testes entre maiores números *fuzzy* das plantas de reprodutibilidade 5 e qualquer planta de referência diferente da planta teste.

PlantaTeste versus Planta Referência	Número <i>Fuzzy</i>
Stevia5XAlfafa0	0,158
ChaBugre5XGuine0	0,1566
Amoreira5XPatadeVaca0	0,1542
Jambolao5XGuassatonga0	0,1534
Poejo5XHortela0	0,1476
Abacateiro5XAmoreira0	0,1462
FolhadeLaranja5XFuncho0	0,1461
Losna5XCarqueja0	0,141
JabmicroST5XJambolao0	0,1354
Guaco5XPinhaoRoxo0	0,1353
ChaBranco5XAbacateiro0	0,1186
Menta5XSene0	0,1161
CatingaMulata5XMentruz0	0,1049
Alecrim5XHortela0	0,1023
Malva5XErvaoSaoJoao0	0,0826
Guassatonga5XFolhadeLaranja0	0,0779
Espinheira5XErvadoce0	0,0772
Barbatimao5XJabpennaNutri0	0,059
Ervadoce5XEspinheira0	0,0119
Sene5XSeteSangria0	0
Hibisco5XMelissa0	0

Tabela 5.5: Conforme valores das tabelas 5.1, 5.2, 5.3, 5.4 tem-se os resultados considerados válidos:

PlantaTeste versus Planta Referência	Número <i>Fuzzy</i>
Ervadoce5XErvadoce0	0,9695
Sene5XSene0	0,933
Espinheira5XEspinheira0	0,9223
Menta5XMenta0	0,8839
Barbatimao5XBarbatimao0	0,882
Malva5XMalva0	0,8351
Guassatonga5XGuassatonga0	0,8102
Guaco5XGuaco0	0,7543
Verbena5XVerbena0	0,7459

Tabela 5.6: Conforme valores das tabelas 5.1, 5.2, 5.3, 5.4 tem-se os resultados considerados válidos:

PlantaTeste versus Planta Referência	Número <i>Fuzzy</i>
CatingaMulata5XCatingaMulata0	0,7457
Abacateiro5XAbacateiro0	0,699
JabpennaNutri5XJabpennaNutri0	0,6961
Graviola5XGraviola0	0,6864
BoldoChile5XBoldoChile0	0,666
Guine5XGuine0	0,6623
ChapeudeCouro5XChapeudeCouro0	0,6453
Mentruz5XMentruz0	0,6362
ErvaStaMaria5XErvaStaMaria0	0,6025
Poejo5XPoejo0	0,6004
Alcachofra5XAlcachofra0	0,5882
CardoSanto5XCardoSanto0	0,5838
ChaBugre5XChaBugre0	0,5826
PatadeVaca5XPatadeVaca0	0,5597
ErvadeBicho5XErvadeBicho0	0,5514
Doradinha5XDoradinha0	0,5235
Alecrim5XAlecrim0	0,5054
Oregano5XOregano0	0,4934
Maracuja5XMaracuja0	0,4881
Hortela5XHortela0	0,4635
Cidreira5XCidreira0	0,4588
FolhadeLaranja5XFolhadeLaranja0	0,4582
Funcho5XFuncho0	0,4496
Guinkgo5XGuinkgo0	0,4422

No caso do **modelo das diferenças**, este apresenta um comportamento diferente, para se conseguir a variabilidade no reconhecimento, precisam ser relaxados determinados parâmetros que são análogos ao Id, como é o caso do limiar e também concorre aí, a banda de confiança.

- Com o modelo das diferenças foram procedidos testes com a base de referência de uma massa de testes de 22579, o sistema acerta completamente, *sem erro*, para uma DEM (Distância Euclideana Média) de até 0.0064799, variância de 4.87E-06 e uma correlação (medida de similaridade) maior que 0.99998.

Pode-se concluir que o sistema das diferenças é “estrito”, mas funciona. Também precisam ser feitos ajustes para que se calcule uma margem de erro aceitável, por exemplo. Daí, também deve ser considerada a análise de picos, segundo Ebel, pois o conjunto de informações advindas deste método é “completo”.

Para ilustrar segue parte do código de separação das duas massas de testes aplicado no modelo Bafflert, implementado usando SED e AWK. Ainda para ilustrar o comportamento do Id é apresentada uma comparação rápida do *Pilocarpus microphyllus* utilizado como marcador e a identificação procedida com sucesso do *Jaborandi* proveniente da Faculdade de Ciências Farmêuticas (FCF) da USP, conforme figura ??.

Outra ilustração de resultados esta na tabela 5.7, estes resultados foram obtidos para os testes procedidos entre a planta teste, *Jaborandi* proveniente da FCF USP e diferentes reproduções do marcador (diversas referências do mesmo marcador), com base no método das diferenças, as correlações, também, são apresentadas.

Tabela 5.7: Método das diferenças e resultados, como as correlações para planta test-explanta de referência com reprodutibilidade.

PlantaTeste versus Planta Referência ¹	DEM	Variância	Correlação
JaborandiFolRedUsp3XJaborandiFolRedSTech0	0.041229	5.52E-05	0.99796
JaborandiFolRedUsp3XJaborandiFolRedSTech1	0.040002	5.34E-05	0.99802
JaborandiFolRedUsp3XJaborandiFolRedSTech2	0.038839	5.75E-05	0.9979
JaborandiFolRedUsp3XJaborandiFolRedSTech3	0.038899	6.05E-05	0.99781
JaborandiFolRedUsp3XJaborandiFolRedSTech4	0.040082	5.69E-05	0.99792
JaborandiFolRedUsp3XJaborandiFolRedSTech5	0.039707	5.68E-05	0.99793

Os valores das avaliações quadrática e triangular dos testes desta tabela resultam em zero.

```
% o arquivo geral_66 contem a massa de testes gerada de 200 paginas.

% o arquivo jaborandis contem a massa de testes gerada de 100 paginas.

% a seguinte linha de codigo awk separa os registros de numero fuzzy 0.5

% da tabela de 66 plantas e gera a saida classificada geral66_sorted

cat geral_66.txt |awk 'BEGIN{ FS = "\n" ;RS = "" }

{for(i=1;i<=NF;i+ +) {str=substr($i,1,5);

    if (str == "Media") {print $i, $1} }}' | sort geral66_sorted.txt

% a seguinte linha de codigo separa os registros de

% numero fuzzy 0.5 da tabela de jaborandis

% e gera a saida classificada jaborandis_sorted

cat jaborandis.txt |awk 'BEGIN{ FS = "\n" ;RS = "" }

{for(i=1;i<=NF;i++) {str=substr($i,1,5);

    if (str == "Media") {print $i, $1} }}' | sort > jaborandis_sorted.txt

% A seguir os arquivos classificados sao Identificados ou nao

cat geral66_sorted.txt | awk '{ if ( $1 == "Media")

if ( $3 >= 0.5 ) print $0 "Identificado"}' > Identgeral66.txt

cat geral66_sorted.txt | awk '{ if ( $1 == "Media")
```

```
if ( $3 < 0.5 ) print $0 "NoIdentificado"}' > NoIdentgeral66.txt

cat jaborandis_sorted.txt | awk '{ if ( $1 == "Media") if ( $3 >= 0.5 )

print $0 "Identificado"}' > IdentJaborandi.txt

cat jaborandis_sorted.txt | awk '{ if ( $1 == "Media") if ( $3 < 0.5 )

print $0 "NoIdentificado"}' > NoIdentJaborandi.txt

% estatisticas

Total_testes='cat geral_66.txt | grep Teste | wc -l'

echo $Total_testes

Total_hit='cat Identgeral66_sorted.txt | grep "Identificado"| wc -l'

echo $Total_hit

Total_nohit='cat NoIdentgeral66.txt | grep "NoIdentificado"| wc -l'

echo $Total_nohit

Total_testesJ='cat Jaborandis_sorted.txt | grep Teste | wc -l'

echo $Total_testes

Total_hitJ='cat IdentJaborandi.txt | grep "Identificado"| wc -l'

echo $Total_hitJ

Total_nohitJ='cat NoIdentJaborandi.txt | grep "NoIdentificado"| wc -l'
```

```
echo $Total_nohitJ
```

Figura 5.1: Agrupamento de média dos testes.

6 Conclusões

O presente trabalho obteve êxito na sua finalidade que foi a identificação de plantas medicinais ou drogas vegetais usando espectroscopia na região do infravermelho próximo. Foram identificadas as plantas de referência contra a própria base de referência, a começar. O que foi provado, indubitavelmente, especialmente no caso da implementação *fuzzy*, o adequado funcionamento e resultados obtidos, também.

O que significa dizer que os métodos estudados fazem o restrito reconhecimento de um espectro que é igual a ele mesmo. Este foi apenas o primeiro sucesso, porque considerando que os espectros gerados pelo espectrômetro num dado instante, da mesma amostra, já diferem, cada reconhecimento executado sobre diferentes reproduções da mesma amostra versus uma dada reprodução (desta amostra) na base de referência também representa um reconhecimento e tem diversos detalhes para serem observados.

Foi constatado que existem muitas possibilidades de alteração da configuração de teste, mas, antes ainda, mais ainda, a necessidade da aprofundada análise dos resultados

Sem dúvida o sucesso primeiro com árduo trabalho foi obtido com uso dos equipamentos e software da Bruker, o IFS no LSCP-Engenharia Química da USP e posteriormente com o MPA, indo além da utilização do Opus, com o uso do Ident e da análise de *clusters*. É fantástica a observação do potencial de inteligência “virtual” distribuído pela Bruker nos seus equipamentos, no seu software.

Acredita-se que não somente a Bruker, como seus usuários, por exemplo o LSCP, tem interesse no desenvolvimento de aplicações chamadas “laboriosas”, mas que podem representar significativos avanços para a farmacognosia.

Logo, esta oportunidade, com o objetivo principal alcançado e com diversas perspectivas de continuidade do trabalho existentes, sem dúvida demonstram que já se tem mais um trabalho seguindo métodos científicos teóricos e práticos que demonstram a

possibilidade, ferramental e método, para de se obter e testar bases de referência de drogas vegetais concretamente.

Ilustra-se o caso com um exemplo de “quick compare” do OPUS que do mesmo modo foi testado nas duas abordagens *fuzzy*

E, justamente, este estudo precisa ser feito, comparar as correlações do Id com cada um dos dois métodos *fuzzy*.

Para ilustrar, entre outras possibilidades, observar o comportamento do sistema fuzzy para amostras de diferentes granularidades. Sabe-se que o Id resolve a identificação corretamente, mesmo sendo as amostras de diferentes granularidades, acrescentamos mais este exemplo. No exemplo o Id esta identificando corretamente a amostra com a respectiva granularidade sem se preocupar que sua base de referência é composta de espécimens de diferentes granularidades. Era esperado, pois então esta comprovado. E isto se repete para todas as amostras de diferentes granularidades da base de referência, o que esclarece que a análise usando a distância estatística (distância Mahalanobis) resolve a separação com exatidão como era esperado.

Ainda voltando a questão das distâncias espectrais, pode-se analisar o comportamento destas no que se refere ao esperado na base de referência, isto é, medir o que pode ser aceito e criar tabelas de decisão, ou regras para um motor de inferência “resolver” aceitar ou não determinada variabilidade de uma espécie, a princípio nas etapas de aprendizado com supervisão, além da parametrização inicial para o teste, e evoluindo para formas onde o sistema consegue identificar uma planta com base no seu “conhecimento” adquirido, no sentido de funcionar como uma ferramenta de auxílio aos especialistas, pessoas.

A visão dos componentes principais, tridimensional, é uma ferramenta didática do Id, usada para apresentar o problema da separabilidade e a necessidade da especialização, constatada em ambas as implementações (Análise de Fatores e Sistemas *fuzzy*).. Por outro lado uma vez resolvida a problemática da separabilidade, esta ferramenta novamente demonstra claramente os resultados esperados, que culminam com o devido reconhecimento das amostras.

6.1 Trabalhos Futuros

Existem vários enfoques de oportunidades para a continuidade do trabalho. Esta autora fica muito gratificada pelo fato de que com certeza o trabalho continuará.

Uma questão relevante independente de todas as “frentes” de continuidade do trabalho é a necessidade de cooperar tanto na rede mundial ou a antiga RNP, com grupos que tenham um trabalho avançado neste sentido e pode-se identificar que no sentido de equipamentos sem dúvida a empresa Bruker Optics é muito importante de se ter uma parceria Universidade Empresa, do mesmo modo, pode-se falar de pesquisadores localizados na África do Sul, com um trabalho muito avançado impulsionados pelo PAT.

Sem dúvida a expectativa muito grande vem na pesquisa do uso do EIVP para análise química quantitativa, também, como é o exemplo da quantificação da Pilocarpina no Jaborandi. Os métodos que esta autora tem conhecimento sem auxílio de equipamentos avançados parecem ser muito dependentes do especialista farmacognosista, o que dificulta e mantém a certa problemática levantada por Jorge(JORGE, 2000).

Outrossim, na linha da extensão deste trabalho vão pelo menos objetivos que culminem com uma base de referência também esperada por taxonomistas para auxiliar seu trabalho. Este desenvolvimento futuro a partir dos resultados já desenvolvidos tem sido anseio manifestado por parte da representação do herbário ESA.

E assim seguem as perspectivas de novos projetos, para citar com base na experiência das pesquisas envolvendo a automação do controle de qualidade de proteína, fibras em alimentos para animais, na Espanha, usando EIVP, onde um sistema completo (hardware e software) foi desenvolvido num período de aproximadamente 5 anos.

Observa-se que podem existir pelo menos duas “frentes”(pesquisa na universidade versus pesquisa em empresas) para a continuidade do trabalho, que na verdade se reúnem, porque como poderíamos compilar o kernel do sistema se nos faltasse um compilador, que já vem embutido nele como pensaram os pesquisadores do Bell Labs, quando criaram o sistema operacional Unix e a linguagem C, nos anos 70.

Portanto, neste sentido, a continuidade do trabalho precisa passar pela parceria entre empresas e Universidade. A Universidade também precisa ser parceira fazendo um trabalho de pesquisa e ensino voltado para o bem estar da a começar da comunidade

Universitária, pois assim o faz, fará para fora, de seus domínios geográficos, de modo análogo a questão da parceria com as empresas. Neste sentido, também, no estado de São Paulo, a empresa SourceTech Química Ltda tem sido uma parceira durante o desenvolvimento desta pesquisa.

APÊNDICE A

A.1 Introdução

A seguir são apresentados aspectos teóricos que permeiam a interdisciplinaridade deste trabalho. Neste momento nosso objetivo é esclarecer inicialmente conceitos básicos estudados da espectroscopia IV e IVP, com aplicação tecnológica, baseada nos equipamentos, programas e documentos da Bruker Optics do Brasil, também são introduzidos conceitos da lógica fuzzy. A expectativa é que se possa conseguir esclarecer o tema, embora com a ciência de que a opção aqui foi por mostrar aspectos tecnológicos históricos e atuais. Por exemplo, foi investigado os motivos que levaram esta pesquisa para a utilização do IVP, Um dos motivos sondados é a profundidade de penetração na amostra (este assunto será discutido na subseção Métodos para Análise de Plantas Medicinais, todavia esta pesquisa ainda carece de ensaios práticos sobre esta questão.

A.2 Espectroscopia IV

O que é a espectroscopia IV?

A espectroscopia na região do infravermelho é baseada na absorção da luz infravermelho pela substância a ser medida. Esta absorção excita vibrações¹ e rotações das moléculas, as quais tem frequências que são as mesmas dentro da faixa infravermelho do espectro eletromagnético.

A matéria exposta a uma radiação eletromagnética como a luz infravermelho pode fazer com que a radiação seja absorvida, transmitida, refletida, espalhada, ou sofrer

¹Alguns tipos de vibrações são o estiramento simétrico, assimétrico e deformação (ou flexão). A luz excita os átomos nas moléculas para vibrar. Assim a intensidade da luz irradiada é enfraquecida. A espectroscopia do infravermelho é também chamada espectroscopia vibracional. O espectro vibracional é único como uma assinatura digital (*fingerprint*).

fotoluminescência. Fotoluminescência designa os efeitos que incluem fluorescência, fosforescência e espalhamento Raman.

O modelo do oscilador harmônico² é usado para descrever a absorção IV. Considerando os átomos como partículas com uma dada massa, então as vibrações de uma molécula diatômica (como HCl) podem ser descritas por FigA.1:

Figura A.1: Modelo mecânico da molécula diatômica vibrando.

A molécula consiste das massas m_1 e m_2 , ligadas por uma mola. Em equilíbrio, a distância das duas massas é r_0 . Quando a molécula é estirada por uma quantia $\Delta r = x_1 + x_2$, então a força restauradora (da posição de equilíbrio) F é produzida. Quando a mola é liberada, o sistema vibrará em torno da posição de equilíbrio ($x = 0$). Conforme a Lei de Hooke, para deflexão (ou deformação) pequena, a força F é proporcional a deflexão:

$$F = -k \cdot \Delta r \quad (\text{A.1})$$

Dado que a força age na direção oposta a deflexão (deslocamento), a constante de proporcionalidade (ou força constante) k , também chamada constante de mola k , tem sinal negativo. Na molécula esta força é a medida da força de ligação entre os átomos.

Para um oscilador harmônico é possível calcular a frequência vibracional ν da molécula diatômica como³:

²Sistema que apresenta movimento harmônico de oscilação, isto é, por exemplo uma mola que oscila, indo e vindo, em torno de uma posição de equilíbrio, sendo seu movimento descrito por uma função harmônica temporal, solução não trivial da equação de Laplace, com primeira e segunda derivada contínuas.

³Este “arranjo” particular dos conceitos de espectroscopia IV, advém majoritariamente da cortesia da empresa Bruker Optics do Brasil, atribui-se, inclusive as figuras, ao material de cursos desta empresa, além dos manuais, ver (BRUKER, 2006a), (BRUKER, 2006b), (BRUKER, 2008a),

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (\text{A.2})$$

$$\mu = \frac{m_1 \cdot m_2}{m_1 + m_2} \quad (\text{A.3})$$

Sendo μ a massa reduzida. Com base na equação tem-se: (1) Quanto maior a força constante k , isto é, a força da ligação (força adesiva), maior será a frequência vibracional (em números de onda), conforme figura A.2; (2) Quanto maior a massa atômica de vibração menor é a frequência vibracional (em números de onda), conforme figura A.3. Os números de onda de espectroscopia IV, são traçados da direita para esquerda, i.é, maior número de onda esta na esquerda.

Figura A.2: Três picos de absorção para constantes de forças diferentes.

Figura A.3: Três picos de absorção para massas atômicas diferentes.

Para o modelo do oscilador harmônico, figura A.4, o potencial de energia é simétrico (considerando o potencial de energia simétrico, as deduções são simplificadas) e de acordo com os princípios de mecânica quântica, as vibrações moleculares podem ocorrer de modo discreto, igualmente espaçado, nos níveis vibracionais, r é a distância ou separação internuclear, sendo a energia de vibração :

$$E_v = \left(v + \frac{1}{2}\right)h \cdot \nu \quad (\text{A.4})$$

Onde $\nu = 0, 1, 2, 3, \dots$, E = energia, h = constante (da lei da radiação) de Plank e ν = frequência linear ou número quântico vibracional.⁴ Caso $\nu = 0$ é definido como

(BRUKER, 2008b), bem como documentos *on-line* que acompanham seu espectrômetro MPA FT-NIR e o *software* Opus.

⁴Dependente da massa do átomo e da força da ligação. A radiação eletromagnética é definida pelo comprimento de onda λ ou frequência linear ν . Dado que as ondas eletromagnéticas viajam na

sendo ao nível vibracional básico, a molécula vibra:

$$E_v = \frac{1}{2}h.v \quad (\text{A.5})$$

Quando a **absorção** ocorre, a molécula adquire uma quantia de energia bem definida⁵ $E = h.v$, vinda da radiação e move a molécula para o próximo nível de vibração ($\Delta v = +1$).

Caso a molécula se movimente para o próximo nível inferior ($\Delta v = -1$), uma certa quantidade de energia é emitida na forma de radiação . Isso é chamado **emissão**.

Para o oscilador harmônico as únicas transições permitidas pela mecânica quântica são superiores ou inferiores ao próximo nível vibracional ($\Delta v = \pm 1$).

Um modelo mais acurado de uma molécula é o oscilador não harmônico, figura A.5, ou anarmônico, O potencial de energia é calculado pela equação de Morse e é assimétrico. Os níveis de energia possuem espaçamento desiguais:

$$E_v = (v + \frac{1}{2})h.v - (v + \frac{1}{2})^2 X_{GL}h.v \quad (\text{A.6})$$

X_{GL} é a constante anarmônica. O modelo do oscilador não harmônico permite dois importantes efeitos: (1) A medida que dois átomos aproximam-se entre si, a repulsão cresce rapidamente; (2) Caso uma energia vibracional suficientemente grande alcance a molécula, esta irá se dissociar (desmembrar-se). Esta é a chamada energia de dissociação.

No caso de oscilador anarmônico, as transições vibracionais já não irão apenas obedecer a regra de seleção $\Delta v = \pm 1$. Esse tipo transição vibracional é chamado vibração fundamental. Transições com $\Delta v = \pm 1, \pm 2, \pm 3, \dots$ também, são possíveis e são chamadas sobretons.

Quais os tipos de moléculas que absorvem luz infravermelho? A luz infravermelho só pode ser absorvida por uma molécula se o momento dipolo do grupo de átomos específico altera-se durante a vibração . Quanto maior for a alteração no momento

velocidade da luz, a frequência correspondente a um dado comprimento de onda pode ser calculada por $\nu = c/\lambda$. Para um comprimento de onda, em micras, o número de onda (número de comprimentos de onda por unidade de distância) ν em cm^{-1} é dado por $\nu = 1000/\lambda$.

⁵A frequência da radiação eletromagnética é proporcional a sua energia

Figura A.4: Curva do potencial de energia para um oscilador harmônico.

Figura A.5: Curva do potencial de energia para um oscilador não harmônico.

dipolo, mais forte será a correspondente banda de absorção IV. Moléculas diatômicas homonuclear como H_2 ou O_2 são chamadas IV inativas.⁶⁷

O IV Médio, conforme figura A.6 é a faixa espectral dos números de onda 4.000 a 400 cm^{-1} . Nesta faixa são tipicamente excitadas as vibrações fundamentais. Na faixa do IV Próximo coberta pelos números de onda de 15.000 a 4.000 cm^{-1} (666-2500 nm) vibrações combinadas (das transições da vibração fundamental) e sobretons são excitadas. O IV Distante, cuja faixa espectral contempla os números de onda entre 400 e 5 cm^{-1} . Esta faixa cobre as frequências das vibrações de grandes moléculas, bem como vibrações fundamentais que incluem átomos pesados, compostos inorgânicos ou organometálicos.

Como tecnologias IVP tem-se monocromática e transformada de Fourier (TF); Da tecnologia monocromática derivam os filtros, grade com matriz de diodos e a AOTF (Acousto-Optic Tunable Filters), conforme figuras A.7, A.8, A.9, respectivamente.

Na primeira geração todos os espectrômetros eram dispersivos. Inicialmente os elementos dispersivos eram prismas, depois eram grades de difração. No meio dos anos 60 a espectroscopia IR surgiram os espectrômetros que utilizam a transformada de Fourier (TF).

Esta segunda geração de espectrômetros com o interferômetro Michelson integrado,

⁶Espalhamento Raman ocorre se as alterações são polarizadas. Vibrações IV inativas são Raman ativas se ocorrem mudanças de polarizabilidade.

⁷Sylvia Ceyer, Principles of Chemical Science, Advanced Track (Massachusetts Institute of Technology: MIT Open Courseware), <http://ocw.mit.edu> (Accessed January, 23, 2009). Licence: Creative Commons Attribution-NonCommercial-ShareAlike 3.0

Figura A.6: Separação em três faixas espectrais do Infravermelho.

Figura A.7: Primeiras tecnologias IV monocromáticas.

Figura A.8: Evolução das tecnologias monocromáticas.

possui inúmeras vantagens em relação aos espectrômetros dispersivos. O princípio de trabalho do espectrômetro IV inicia com a luz IV emitida de uma fonte (por exemplo Glowbar (SIC/Carborundum-Silicon carbide Rod-form)) é direcionada sobre o interferômetro, o qual modula a luz. Após o interferômetro a luz passa através do compartimento (e também da amostra) e é então focalizada no detector.

O interferômetro de Michelson é a principal parte de um espectrômetro IV. A luz colimada da fonte IV incide no separador de feixe, que idealmente transmite 50% da luz e reflete a parte restante. Tendo viajado a uma distância L , a luz refletida atinge um espelho fixo $M1$, onde é refletida e atinge os refletores novamente após

Figura A.9: AOTF e técnica de modulação usando FT

um trajeto total de $2L$. A parte transmitida do feixe é direcionada para o espelho móvel M2. Como este espelho se move da frente para trás em torno de L , com uma distância Δx , o comprimento do percurso total é de $2(L + \Delta x)$. A luz que volta dos dois espelhos é recombinada no separador de feixe com os dois feixes tendo uma diferença no comprimento do caminho de $2\Delta x$. Os feixes são espacialmente coerentes e interferem um com o outro quando combinados.

O interferômetro separa e recompina os dois feixes com uma relativa diferença de fase, que depende do deslocamento do espelho ou retardo óptico. Os dois feixes sobre interferência construtiva produzem um sinal máximo no detetor, se o atraso óptico é uma integral múltipla de comprimento de onda λ , isto é se:

$$\Delta x = n \cdot \lambda$$

Sendo $n=0,1,2,\dots$. Quando a interferência é destrutiva um sinal mínimo é detectado, o que ocorre se $2\Delta x$ é um múltiplo de $\frac{\lambda}{2}$. O relacionamento funcional completo entre $I(x)$ e x é tal que a função coseno:

$$I(x) = S(\nu) \cdot \cos(2\pi \cdot \nu \cdot x)$$

Da qual são usados os números de onda $\nu = \frac{1}{\lambda}$, muito comum em espectroscopia IV-TF. $S(\nu)$ é a intensidade de uma linha espectral monocromática, muito útil, em número de onda ν . A intensidade é mostrada como uma função da frequência e é chamada espectro, podendo ser obtida pela TF do sinal que é uma função do retardo óptico.

O padrão de interferência cosenoidal proveniente de uma fonte monocromática permite um controle muito preciso do espelho móvel. Espectrômetros IV-TF usam o padrão de interferência de luz monocromática emitida por laser HeNe para monitorar a posição do espelho. O interferograma é digitalizado exatamente no cruzamento zero do interferograma com laser.

Considerando espectrômetros que são equipados com uma fonte de luz policromática (diversos comprimentos de onda) a interferência mencionada ocorre em cada onda. Os padrões de interferência produzidos por cada comprimento de onda são somados para obter o interferograma resultante. Com a diferença de caminho em zero do espelho móvel ($\Delta x = 0$), em ambos caminhos, todos os comprimentos de onda tem uma diferença de fase igual a zero e portanto sofrem uma interferência construtiva. A intensidade tem, portanto, um valor máximo. Como os retardos ópticos de interferência aumentam, cada onda sofre interferência construtiva e destrutiva em diferentes posições do espelho.

Os espectrômetros são equipados com uma fonte de luz banda larga, a qual produz comprimentos de onda de número infinito, contínuo. O interferograma é a soma contínua, isto é, a integral de todos os padrões de interferência produzidos por cada comprimento de onda. Isto resulta na curva de intensidade como uma função do retardo óptico. No caminho de diferença zero do interferograma ($\Delta x = 0$) todas os comprimentos de onda experimentam a interferência construtiva e somam-se para um sinal máximo. Como o aumento do retardo óptico comprimentos de onda diferentes sofrem interferência construtiva e destrutiva em pontos diferentes, e a intensidade, conseqüentemente, muda com o atraso. Para fonte banda larga, no entanto, todos os padrões de interferência nunca estarão simultaneamente em fase exceto no ponto de diferença de caminho zero e o sinal máximo ocorre somente neste ponto. O sinal de máximo é referido como o *centerburst*. Como vantagens da espectroscopia IV-TF tem-se: (1) O intervalo de amostragem do interferômetro, δx , é a distância entre os cruzamentos zero

do interferogram com laser HeNe e é consequentemente precisamente determinado pelo comprimento de onda do laser. O espaçamento pontual no espectro resultante δ_ν é inversamente proporcional a δx . Espectrometria IV-TF tem intrinsecamente uma escala de números de onda altamente precisos (tipicamente alguns centésimos de número de onda). Esta vantagem é conhecida como vantagem do CONNE; (2) A vantagem Jaquinot provém do fato de que as aberturas circulares tem uma área maior do que as fendas usadas em grades de difração, permitindo assim um maior rendimento de radiação; (3) Em espectrômetros de grade de difração o espectro $S(\nu)$ é medido diretamente pelo registro da intensidade, em sucessivas, estreitas faixas de comprimento de onda. No espectrômetro IV-TF todos os comprimentos de onda provenientes da fonte IV incidem simultaneamente sobre o detetor. Isso leva a multiplexação ou vantagem de Fellgett.

A combinação das vantagens Jaquinot e Fellgett significa que a relação sinal-ruído de um espectrômetro TF pode ser 10 vezes superior do que a de um espectrômetro dispersivo.

Entre as vantagens da espectroscopia IVP podem ser citadas:

- 1 podem ser usadas amostras puras;
- 2 podem ser medidas através de materiais como o vidro e embalagens;
- 3 uso de fibra ótica, facilmente adaptada;
- 4 relativamente rápido e barato;
- 5 alta reprodutibilidade.

A aquisição de dados resulta de um interferograma, $I(x)$, que é convertido em um espectro por meio da operação matemática chamada TF. Em geral a TF é aplicável a sinais contínuos. Como o sinal (interferograma) é digitalizado, consistindo de pontos N , discretos e equidistantes, a versão discreta da TF deve ser usada:

$$S(k.\Delta\nu) = \sum I(n.\Delta x). \exp(i2\pi k.\frac{n}{N})$$

As variáveis contínuas x e ν são substituídas com $n.\Delta x$ e $k.\Delta\nu$ representando os n discretos pontos do interferograma e k discretos pontos do espectro. O fato de que se

tem uma função discreta, descontínua e que só é calculado para uma gama limitada de n (isto é, o interferôgrama medido tem um comprimento finito) conduz a efeitos como *picket-fence* e *leakage*

Princípios da espectroscopia na faixa do infravermelho próximo:

(1) excitação de vibrações com sobretons de bandas e combinação de bandas. Requer a comparação com dados conhecidos - modelos de quantificação e qualificação. São avaliados grupos funcionais como OH, CH e NH.

O espectro na faixa do infravermelho próximo contém informações sobre a forma física do material e sua estrutura química. Tamanho de partícula e densidade diferentes são vistos como deslocamentos(ou compensações) de base. Este espectro amplo contempla a sobreposição de bandas de absorção de vários componentes e/ou grupos funcionais.

O interferômetro pode ser polarizado ou interferômetro com espelhos Michelson ou CubeCorner.

A espectroscopia vibracional engloba o infravermelho próximo, Raman e infravermelho médio. Estas técnicas complementares tem sido usadas comumente para o estudo de amostras em estado sólido. Elas oferecem a análise qualitativa e quantitativa e podem ser empregadas em linha (*in-line*). E, tem sido muito empregadas na indústria farmacêutica. Quase todos os compostos químicos numa amostra podem ser visualizados em minutos. Aplicações tem crescido em vários campos como a qualidade de alimentos, para análise de tecidos biológicos. PAT é um termo genérico para monitoração de parâmetros de qualidade e desempenho em sistemas baseados em produção. São dois os objetivos:

- requer monitoração do produto ao longo de toda a linha para garantir a qualidade;
- encoraja o desenvolvimento de métodos de e sistemas novos para a monitoração de processo, avançando o entendimento do processo em si;

A espectroscopia vibracional engloba importantes aspectos do processamento digital de imagens, como o tempo de aquisição, calibração e resolução (espectral e espacial).

A espectroscopia vibracional usa as faixas espectrais particulares do espectro eletromagnético ($13.000\text{--}10\text{ cm}^{-1}$ ou $0,76\text{--}1000\text{ cm}^{-1}$, conforme a unidade usada) bandas

químicas nas quais as moléculas vibram. Esta região espectral engloba as três divisões: infravermelho distante (IVD: 400–10 cm^{-1} ou 26–10000 m), infravermelho m'dio (IVM: 4000–400 cm^{-1} ou 2,6–26 m) e infravermelho próximo (IVP: 13.000–4000 cm^{-1} ou 0,76–2,6m), assim chamado em relação a região do visível.

Duas unidades são usadas na espectroscopia vibracional: cm^{-1} (número de onda) ou nm. A escolha das unidades depende do tipo de espectrômetro (dispersivo ⁸ vs. transformada de Fourier(TF)) ou também para evitar números grandes no IVP. nm é a unidade mais frequentemente usada:

$$[cm^{-1}] = \frac{1}{[nm] \times 10^7}$$

A.3 IVP

Na espectroscopia para cada comprimento de onda na região do infravermelho próximo, a absorbância da amostra é gravada em IVP. O espectro IVP surge dos sobretons moleculares em combinação com as vibrações. Combinação de sobretons e vibrações são os dados gravados na faixa IVP. Picos são largos e se sobrepõem complicando a identificação clara de químicos a partir de seus picos característicos. Análise estatística do espectro é frequentemente requerida para identificação e quantificação da molécula. Entretanto, uma vez que o sinal da absorbância é fraco, devido pequena probabilidade de ocorrência da combinação de sobretons e vibrações, as amostras não requerem preparação e a análise é não destrutiva. As amostras podem ser analisadas através do vidro, porque este é transparente para o IVP devido a carência de X-H bandas significativas e também, através do plástico dependendo da janela espectral. O desenvolvimento de recursos computacionais, combinado com dispositivos amigáveis (detectores que não necessitam ser congelados) faz da espectroscopia a escolha para aplicações em-linha.

A profundidade de penetração é uma resposta para a pergunta - porque a escolha da região do IVP para obter-se a assinatura espectral de plantas? Sendo a espectrometria de reflexão difusa usada para o propósito de controle de qualidade, é essencial saber o volume da amostra investigada, o que é a equivalente a profundidade

⁸detalhes do dispersivo

de penetração da radiação ou o tamanho efetivo da amostra m_{eff} . Na publicação de Berntsson et al. (BERNTSSON, 1999), o tamanho efetivo da amostra de pós farmacêuticos foi investigado e apresentado por aproximação de três fluxos, usando o chamado método da equação de transferência radiativa (ERT), e um método empírico, chamado método de espessura de camada variável (VLT). Nesta publicação o tamanho efetivo da amostra m_{eff} é definido como a massa, por área da amostra, na qual a sua reflectância difusa tem encontrado 98% de reflectância difusa, da espessura da amostra opticamente correspondente. Para obter m_{eff} usando a aproximação de três fluxos do ERT, os coeficientes de espalhamento e a absorção do pó investigado são obtidos das medições da reflectância difusa e transmitância de amostras opticamente finas, conforme Burger (BURGUER, 1997), (BURGUER, 1998). Então, em cada comprimento de onda, a reflectância difusa é calculada em função dos coeficientes de espalhamento e absorção, para um incremento gradual na espessura da amostra e comparada com a reflectância difusa de uma amostra de espessura infinita ($R\sigma$). Para que o limite seja superior a 98% seja atingido, a espessura da amostra corresponde ao tamanho efetivo da amostra m_{eff} . Este método é comparado com um procedimento totalmente independente, o método VLT, onde os espectros de reflexão difusa são recolhidos em várias espessuras do pó controlado. Supondo que o suporte(apoio) é completamente transparente ou completamente opaco, a reflectância difusa das camadas de pó cresce com o aumento da espessura da amostra até que a reflectância de uma amostra opticamente espessa ($R\sigma$) é atingida.

Para cada medida de comprimento de onda, uma função exponencial é montada para os dados do experimento (De $\log(1/R)$ versus espessura da amostra, onde R é a reflectância medida).

Usando o limite de 98%, o tamanho efetivo da amostra m_{eff} pode ser obtido a partir do ajuste exponencial

Por exemplo considerando o tamanho efetivo da amostra, um pó de celulose microcristalina (MCC, intervalo do tamanho da partícula: 65-300 μm) na região IVP.

Assumindo a densidade do pó de 0,30g cm^3 , a efetiva massa por área é transformada em uma profundidade de penetração eficaz em milímetros.

A dependência do comprimento de onda do tamanho da amostra efetiva, que é inversamente correlacionada ao coeficiente de absorção, e a boa correspondência entre

os dois métodos pode ser vista claramente.

O limite superior para o método VLT foi 1400nm para a amostra MCC, porque o ajuste da curva torna-se instável, se a massa efetiva da amostra por área é inferior à menor massa do pó por área utilizados nas medições.

No entanto, ambos os métodos são adequados para determinar o volume da amostra sondada de uma medição de reflexão difusa.

Outra abordagem para a determinação da profundidade de penetração em termos do número de partículas questionado foi investigada por Olinger e Griffiths.

APÊNDICE B

B.1 Modelo e Lógica Fuzzy

Zadeh (ZADEH, 1992) diz ser importante reconhecer-se que o termo lógica fuzzy é usado em dois sentidos diferentes sendo um mais amplo que o outro. No sentido restrito a lógica fuzzy é um sistema lógico cujo objetivo é oferecer um modelo para os modos de raciocínio humanos, os quais são mais aproximados do que exatos. No sentido amplo a lógica fuzzy é a teoria dos conjuntos fuzzy, sendo esta a teoria das classes com limites nebulosos (*unsharp*). É a área de administrar as incertezas. O raciocínio lógico fuzzy exato (*crisp*)¹ é visto como um caso limitante do raciocínio aproximado. Na lógica fuzzy tudo é problema de grau. Três são os conceitos chaves das aplicações de lógica fuzzy:

- variável linguística - seus valores são palavras ou sentenças em linguagem natural ou sintética: jovem, muito jovem, não muito jovem, velho, ...
- forma canônica - expressa uma proposição como uma restrição de elasticidade (*fuzzy*) sobre uma variável $p \rightarrow X$ é A, onde X é a variável restritora explícita ou implícita em p; A é a restrição elástica sobre X, o qual está implícito ou explícito em p. Ex. Maria é jovem $\rightarrow$ Idade (Maria) é jovem.
- raciocínio interpolativo é a interpolação de uma relação *fuzzy* a qual é parcialmente definida por uma coleção de regras *fuzzy if-then*;

O significado de uma variável linguística, V é definido pela sua distribuição de possibilidades $\Phi_V(u) = \text{Poss} \{X=u \mid X \text{ e } V\}$, $\text{Poss} \{\text{Paulo é 45} \mid \text{Paulo é meia-idade}\}=0.8$.

¹Pode ser entendido como binário ou advindo da lógica clássica, raciocínio exato.

A inferência usa o esquema:

$p \leftarrow$ premissa

$q? \leftarrow$ consulta (*query*)

$r \leftarrow$ resposta

Exemplo: Maria tem 25

Pat é três anos mais velha que Maria

Qual a idade de Pat?

=====

Pat tem 28

B.1.1 Modelo Fuzzy

A lógica fuzzy ou nebulosa é melhor compreendida no contexto do conjunto de seus membros. Quer-se colocar o dia de hoje numa série de dias chuvosos. Um conjunto binário (*crisp*) é formado de declarações que são verdadeiras ou falsas, do dia de hoje, logo a sua inclusão no conjunto de dias chuvosos é baseada na certeza.

Ao lidar com a lógica nebulosa, deve-se incluir o conjunto do dia de hoje, em dias chuvosos ordenados através de um par, como o (hoje, 0,8). O primeiro membro de tal par ordenado é um candidato para a inclusão no conjunto e, o segundo membro é um valor entre 0 e 1, inclusive, chamado o grau de adesão(ou de pertinência) do conjunto.

A inclusão do grau de adesão (ou de pertinência), no conjunto torna conveniente para desenvolvedores avançar com uma teoria de conjuntos baseada em lógica nebulosa, tal como se desenvolveu a teoria de conjuntos regulares.

Conjuntos nebulosos são conjuntos em que os membros são apresentados como pares ordenados que incluem informação sobre seu grau de adesão. Um conjunto tradicional de k elementos, é um caso especial de um conjunto nebuloso, onde cada um desses k elementos tem 1 para o grau de adesão(de pertinência), e qualquer outro elemento do conjunto universal tem um ou o grau de pertinência 0, motivo para o qual não será

acomodado à lista.

Operações com Conjuntos Fuzzy Operações comuns são a união, onde os todos os elementos estão em um conjunto ou em outro e a intersecção, no qual os elementos estão em ambos conjuntos.

Com a união quer-se saber o grau de pertinência que deverá ter um elemento no novo conjunto fuzzy, união de dois conjuntos fuzzy.

Se a, b, c, d são tais que os seus graus de pertinência a um conjunto fuzzy são 0,9, 0,4, 0,5 e 0, respectivamente, então o conjunto fuzzy A é dado pelo vetor de ajuste (0,9, 0,4, 0,5, 0). Os componentes deste vetor são chamados valores de ajuste a, b, c, e d.

União de Conjuntos Fuzzy Considere uma união de dois conjuntos tradicionais e um elemento que pertence a apenas um desses conjuntos. Caso estes conjuntos sejam tratados como lógica fuzzy, este elemento tem um grau de pertinência de 1 em um caso e 0 no outro, uma vez que pertence a um conjunto e não a outro. Este elemento está sendo colocado na união. O critério usado na presente ação tem a ver com os respectivos graus de adesão. Dados os dois graus de adesão, ou seja, 0 e 1, toma-se o valor mais elevado dos dois, a saber:

1. Tem-se o grau de pertinência de um elemento, quando listado na união de dois conjuntos fuzzy, na união é listado o valor máximo dos seus graus de pertinência dos dois conjuntos.

Sendo a, b, c, d com os respectivos graus de pertinência em conjuntos fuzzy A, B,

$$A = (0,9,0,4,0,5,0) \text{ e } B = (0,7,0,6,0,3,0,8), \text{ logo } A[uni]B = (0,9,0,6,0,5,0,8).$$

Intersecção e Complemento de Dois Fuzzy Sets Analogamente, o grau de pertinência de um elemento na intersecção de dois conjuntos fuzzy é o mínimo, ou o menor valor do seu grau de pertinência individualmente nos dois conjuntos formando a intersecção.

Por exemplo, se tem hoje 0,8 de grau de pertinência no conjunto de dias chuvosos e 0,5 de grau de pertinência no conjunto de dias de trabalho completo, então hoje pertence ao conjunto de dias chuvosos no qual o trabalho é concluído com grau de

0,5, o menor entre 0,5 e 0,8. Neste exemplo, a intersecção dos conjuntos fuzzy A e B, é obtida usando de cada componente, o menor dos valores encontrados entre os componente de A e B. Assim, em

$$A[\text{int}]B = (0,7,0,4,0,3,0).$$

O complemento de um conjunto fuzzy, é obtido da seguinte forma . No caso dos conjuntos fuzzy, se o grau de adesão é de 0,8 para um membro, então o membro não esta definido dentro do conjunto para um grau de $1,0 - 0,8 = 0,2$. Então, pode-se definir o grau de pertinência, dentro do complemento fuzzy definido para o complemento de conjunto fuzzy para o complemento em relação a 1. Voltando ao cenário de ter um grau de 0,8 no conjunto de dias chuvosos, em seguida, hoje se tem pertinência de grau 0,2 no conjunto de dias sem chuva.

Complemento denotado A por $A' = (0,1, 0,6, 0,5, 1)$ e $B' = (0,3, 0,4, 0,7, 0,2)$. Note que $A'[\text{uni}]B' = (0,3, 0,6, 0,7, 1)$, é também o complemento de A $[\text{int}]$ B. $A[\text{uni}]A' = (0,9,0,6,0,5,1)$ e $A[\text{int}]A' = (0,1,0,4,0,5,0)$. A e A' serão iguais, no sentido de que os seus vetores próprios são os mesmos, se cada componente do vetor de ajuste é igual a 0,5.

APÊNDICE C

C.1 Exsicatas e conjunto de amostras

Figura C.1: Caixa de Especimens

Figura C.2: Drogas Vegetais

Figura C.3: Pilocarpus microphyllus SourceTech

Figura C.4: P. microphyllus SourceTech

Figura C.5: Poejo exsicata

Figura C.6: Guaco exsicata

Figura C.7: Pata de vaca exsicata

Figura C.8: Pilocarpus pennatifolius
Praça Sé Figura C.9: Pilocarpus pennatifolius
Praça Sé Figura C.10: Pilocarpus jaborandi Praça
da Sé

Figura C.11: Drogas Vegetais

Figura C.12: Drogas Vegetais

Figura C.13: Drogas Vegetais

Figura C.14: Hibiscus sabdariffa L.

Figura C.15: Hibiscus sabdariffa L.

Figura C.16: Drogas Vegetais

Figura C.17: Alecrim exsiccata

Figura C.19: Drogas Vegetais

Figura C.20: Espinheira santa exsicata

Figura C.21: Maytenus ilicifolia Reissek

Figura C.22: Maytenus ilicifolia Reissek

Figura C.23: Maytenus ilicifolia Reissek

Figura C.24: Pilocarpus microphyllus SourceTech

Figura C.25: Urucum exsicata

Figura C.26: ESALQ Universidade de São Paulo

APÊNDICE D

D.1 Detalhes das Implementações

Programa Principal (chamado Principal.m)

Principal.m

```
clear all;
```

```
num_plan_ref = input('Digite o numero de planilhas do xls das plantas Referencias: ');
num_plan_teste = input('Digite o numero de planilhas do xls das plantas Teste: ');
```

```
% Carga dos nomes das referencias
```

```
[ndata,nome_planta] = xlsread('Ordem_Referencia.xls');
```

```
% Carga dos espectros referencias
```

```
planta = [];
for j=1:num_plan_ref
    p = xlsread('Referencia.xls',j);
    planta = [planta,p];
end
```

```
% Carga do nome da planta "Teste"
```



```
[ndata,nome_planta_teste] = xlsread('Ordem_Testes.xls');

% Carga do espectro da planta "Teste"

for m = 1:num_plan_teste
    planta_teste = xlsread('Testes.xls',m);

% Carrega o 'xls' das frequencias dos picos

picos = xlsread('Picos.xls',m);
espectros = [planta,planta_teste];

% Monta uma matriz com os espectros
% (Deve se colocar a planta teste
% SEMPRE por ultimo)

numpico = length(picos);

% Encontra o numero de picos

Saida = [];
for i = 1:numpico

% Intensidade dos picos

[intensidades] = intensidadepico(espectros,picos(i));

% Fuzzificacao

[S] = fuzzificacao(intensidades);

% Saidas

Saida = [Saida;S];
```

```
end

for k=1:num_plan_ref
    Respostas = [];

    for l=1:numpico
        Respostas = [Respostas;cellstr('Pico'),num2cell(l),num2cell(Saida(l,k))];
    end

    Teste = [cellstr(nome_planta_teste(m)),cellstr('X'),cellstr(nome_planta(k))]
    Respostas

    % Defuzzificacao

    soma = sum(Saida(:,k));

    % Soma as saidas para cada referencia(somas
    % as linhas de cada coluna)

    Media = soma/numpico

    % Faz a media das saidas para cada referencia

    Ord_Saida = sort(Saida(:,k));

    % Ordena, crescentemente, as saidas para
    % cada referencia
    % (Ordena todas as linhas para cada coluna)

    Minimo = Ord_Saida(1,:)

    % Devolve o minimo das saidas para cada referencia

end end end
```

Função intensidadepico.m

```
intensidade.m
```

```
function [intensidade] = intensidadepico (planta,ind_pico);  
planta_size = size(planta,1);
```

```
% Numero de linhas
```

```
colunas = size(planta,2);
```

```
% Numero de colunas
```

```
numplantas = colunas/2;
```

```
intensidade = [];
```

```
for k = 0:2:(colunas-2)
```

```
% Para fazer para todas as plantas
```

```
% Acha os limites
```

```
j = 1;
```

```
while planta(j,1+k)> ind_pico
```

```
j = j+1;
```

```
end
```

```
if (j+15) <= planta_size
```

```
lim_inf = planta((j+15),(1+k));
```

```
else
```

```
lim_inf = planta(planta_size,(1+k));
```

```
end
```

```
if (j-15) >= 1
```

```
lim_sup = planta((j-15),(1+k));
```

```
else
```

```
lim_sup = planta(1,(1+k));  
end  
  
% Tamanho da variavel (numero de linhas)  
  
x = 1;  
for i = 1:planta_size  
    if (planta(i,(1+k))>=lim_inf)& (planta(i,(1+k))<=lim_sup)  
        pico(x) = planta(i,(2+k));  
        x = x+1;  
    end  
end  
pico_ord = sort(pico);  
pico_size = length(pico);  
intensidade = [intensidade,pico_ord(pico_size)];  
end
```

Função fuzzificacao.m

```
fuzzificacao.m  
  
function [output] = fuzzificacao (intensidades);  
numplantas = length(intensidades);  
  
%% Gerando o Fuzzy  
% Ordenar as intensidades  
  
ord = sort(intensidades(1:(numplantas-1)));  
ord_size = size(ord,2);  
  
% Range Minimo  
  
RangeMin = ord(1)-(ord(2)-ord(1));
```

```
% Range Maximo

RangeMax = ord(ord_size)+(ord(ord_size)-ord(ord_size-1));
ord2 = [intensidades,RangeMax,RangeMin];

% Ordena todas as intensidades
% das referencias, das amostras
% e os Ranges maximo e minimo

ord2 = sort(ord2);
ord2_size = size(ord2,2);

% Logica para encontrar o grau de pertinencia

saida = zeros(1,(numplantas-1));
for i = 1:ord2_size
if ord2(i) == intensidades(numplantas)

% Acha a posicao da intensidade da
% amostra

if (i == 1)|(i == ord2_size)

% Se a intensidade da amostra estiver na primeira
% ou na ultima posicao,
% ou seja e' menor ou maior, repectivamente, que os Ranges Minimo
% e Maximo a saida sera' "Zero" para todas as referencias

output = saida;
else
delta = ord2(i+1)-ord2(i-1);

% Acha a distancia entre as intensidades das
% referencias (ou Range)
```

```
% onde a intensidade da amostra esta entre

if (i == 2)|(i == (ord2_size-1))
if (i == 2)
for j = 1:(numplantas-1)
if (ord2(i+1)== intensidades(j))
saida(1,j) = (ord2(i)- ord2(i-1))/delta;
end
end
else
for j = 1:(numplantas-1)
if (ord2(i-1)== intensidades(j))
saida(1,j) = (ord2(i+1)- ord2(i))/delta;
end
end
end
else
for k = 1:(numplantas-1)
if (ord2(i-1)== intensidades(k))
for j = 1:(numplantas-1)
if (ord2(i+1)== intensidades(j))
saida(1,j) = (ord2(i)- ord2(i-1))/delta;
saida(1,k) = (ord2(i+1)- ord2(i))/delta;
end end end end end end end end
output = saida;
%final
```

BRASIL. Agência Nacional de Vigilância Sanitária. Resolução **RDC n. 48, de 16 de março de 2004**. Dispõe sobre o registro de medicamentos fitoterápicos. Diário Oficial da República Federativa do Brasil, Brasília, DF, 18 mar. 2004. Disponível em <http://e-legis.anvisa.gov.br/leisref/public/showAct.php?id=10230&word=> Acesso em: 10 abr. 2009.

Brasil. Agência Nacional de Vigilância Sanitária (ANVISA). **Instrução Normativa no. 5** LISTA DE MEDICAMENTOS FITOTERÁPICOS DE REGISTRO SIMPLIFICADO. Diário Oficial da União. Brasília, 12 de dezembro de 2008.

Agência Nacional de Vigilância Sanitária (ANVISA). **RE no. 88, de 16 de março de 2004**. Determina a publicação da LISTA DE REFERÊNCIAS BIBLIOGRÁFICAS PARA AVALIAÇÃO DE SEGURANÇA E EFICÁCIA DE FITOTERÁPICOS. Diário Oficial da União. Brasília, 18 de março de 2004.

Bakalli, R.I., Pesti, G.M., Etheridge, R.D. Comparison of a Commercial Near-Infrared Reflectance Spectroscopy and Standard Chemical Assay Procedures for Analyzing Feed Ingredients: Influence of Grinding Methods **J. Appl. Poult Res** 2000 9:204-213

M. Baranska, H. Schulz, P. Rosch, M.A. Strehle, J. Popp (2004) Identification of secondary metabolites in medicinal and spice plants by NIR FT Raman microscopic mapping. **Analyst** 129, 926-930

O. Berntsson, T. Burguer, S. Folestad, L.G.Danielsson, J. Kuhn and J. Fricke **Anal. Chem. Acta** 71:617-623 1999

Blaffert, T., Computer-Assisted Multicomponent Spectral Analysis with Fuzzy Data Sets. **Analytica Chimica Acta**, 161 (1984) 135-148 Elsevier Science Publishers B.V., Amsterdam, The Netherlands.

Bruker Optic GmbH, **OPUS Spectroscopy Software User Manual Reference** Ettlingen, Germany, 2006.

Bruker Optic GmbH, **Application and Measurement Software Reference Manual** Ettlingen, Germany, 2006.

Bruker Optic GmbH, **MPA User Manual** Ettlingen, Germany, 2008.

Bruker Optic GmbH, **Training Course - Introduction to FT-IR Spectroscopy** Ettlingen, Germany, 2008.

Bruneton, Jean, **Farmacognosia, Fitoquímica, Plantas Medicinales** 3a Ed Acribia SA, Zaragoza (Espana), 2001.

T. Burguer, J. Kuhn, R. Caps and J. Fricke **Appl. Spectroscopie** 51:1323-1329;1997

T. Burguer, P. Keym, J. Kuhn and J. Fricke **J. Near Infrared Spectrosc.** 6:33;1998

Capellari Jr., Lindolpho **Sistemática de Plantas Medicinais e Aromáticas**. Material Didático. Depto de Ciências Biológicas ESALQ-USP, 2009.

Chalmers, J.M., Griffiths, P.R., **Handbook of Vibrational Spectroscopy** New York: J. Wiley, 2002.

T. Chiueh Optimization of Fuzzy Logic Inference Architecture **IEEE Computer** 1992 Vol. 25 No. 5 May pp. 67-71.

O. Cordon and F. Herrera and A. Peregrí, Applicability of the fuzzy operators in the design of fuzzy logic controllers **Fuzzy Sets and System** 1997 Vol. 86 pp. 15-41.

P. Corti, E. Dreassi, G.G. Franchi, G. Corbini, A. Moggi, S. Gravina Application of NIR Reflectance Spectroscopy to the Identification of Drugs Derived from Plants. **Pharmaceutical Biology**. Volume 28 Issue 3 September 1990 pages 185 - 192.

Costa, Aloísio Fernandes **Farmacognosia**. Fundacao Calouste Gulbenkian, Lisboa, 1987.

D. Driankov and H. Hellendoorn and M. Reinfrank **An Introduction to Fuzzy Control** Spriger-Verlag 1993.

Ebel, S., Janosh, J., Schutz, U., Fuzzy Methods for the analysis of NIR spectra. **Making Light Work: advances in near infrared spectroscopy**; developed from the 4th International Conference on Near Infrared Spectroscopy, Aberdeen, Scotland, August 19-23, 1991 ed. by Ian Murray and Ian A. Cowe. Weinheim; NY; Basel; Cambridge: VCH, 1992.

BRASIL. Ministério da Saúde. **Farmacopéia Brasileira**. 4 ed. São Paulo: Editora Atheneu, 1988.

GONZALEZ, Rafael C., WOODS, Richard E. 2002. **Digital Image Processing**. Upple Saddle River Prentice Hall Inc New Jersey.

Jorge, Luzia Ilza Ferreira **Botânica aplicada ao controle de qualidade de alimentos e de medicamentos** São Paulo: Editora Atheneu, 2000.

Kalra, Yash P. **Handbook of Reference Methods for Plant Analysis** Boca Raton: CRC Press, 1998.

D. Park and Zhiquiang Cao, Investigations on the applicability of fuzzy inference **Fuzzy Sets and Systems** 1992 Vol. 49 pp. 151-169.

G. J. Klir and Tina A. Folger **Fuzzy Sets, Uncertainty and information**. Prentice Hall, 1988.

B. Kosko **Fuzzy Engineering**. Prentice Hall, 1997.

LORENZI, H., MATOS, F.J.A. **Plantas Medicinais do Brasil**. Nova Odessa Instituto Plantarum, 2a Ed., 2008. São Paulo, p. 544.

Maurer, Lene Near infrared spectroscopy/imaging and terahertz pulsed spectroscopy/imaging for the analysis of solid dosage forms. **Diss. Phil.-Nat. Univ. Basel**, 2008.

J. M. Mendel Fuzzy Logic Systems for Engineering: A Tutorial **Proceeding of IEEE**. 1995, Vol. 83, No. 3. March, pp. 378-406.

Brasil Ministério da Saúde , Aprova a Política Nacional de Práticas Integradas e Complementares (PNPIC) no Sistema Único de Saúde. **Portaria n. 971, de 3 de maio de 2006**.

K. Nozaki and H. Ishibuchi and H. Tanaka, A simple but powerful heuristic method for generating fuzzy rules from numerical data **Fuzzy Sets and Systems**, 1997 Vol. 86 pp. 251-270.

OLIVEIRA, Flávia C.C. Modelos de calibração multivariada associados a espectroscopia vibracional para análise de misturas diesel-óleos vegetais. **Tese de Mestrado**.

PPGQ - IQ - UNB, 2006.

Oliveira, Fernando de; Akisue, Gokithi; Akisue, Maria Kubota, **Farmacognosia**. São Paulo : Atheneu, 1998.

Osborne B.G., Fearn, T. **Food Chemistry**, 29 (1988) 233-238.

Otto, M, Bandemer, H, A fuzzy method for component identification and mixture evaluation in the ultraviolet spectral range **Analytica Chimica Acta**, 191(1986)193-204 Elsevier Science Publications B.V., Amsterdam., The Neederlands.

W. Pedrycz **Fuzzy control and Fuzzy Systems** John Wiley e Sons Inc 1993 Second Ed.

L. Renhou and Z. Yi, Fuzzy logic controller based on genetic algorithms **Fuzzy Sets and Systems** 1996 Vol. 83, pp. 1-10.

SALA, Oswaldo 1996. **Fundamentos da Espectroscopia Raman e no Infravermelho**. São Paulo Editora UNESP.

Simões, Cláudia Maria Oliveira; Schenkel, Eloir Paulo; Grosmann, Grace **Farmacognosia : da planta ao medicamento** 5.ed. Porto Alegre-Florianópolis : Editora da UFRGS-Ed. da UFSC, 2004.

H. Schulz (2004) Application in analysis of coffee, tea, cocoa, tobacco, spices, medicinal and aromatic plants, and related products. In: Roberts, C.; Workman, J.; Reeves, J. (Eds.): **Near-infrared Spectroscopy in Agriculture**, American Society of Agronomy - Crop Science of America - Soil Science Society of America, Madison, USA, Agronomy Monograph, No. 44, 345-376.

H. Schulz, R. Quilitzsch, H. Kruger (2003) Rapid evaluation and quantitative analysis of thyme, origano and chamomile essential oils by ATR-IR and NIR spectroscopy; **J. Mol. Struct.** 661-662, 299-306.

H. Schulz, M. Baranska, H.H. Belz, P. Rosch, M. Strehle, J. Popp (2004) Chemotaxonomic characterisation of essential oil plants by vibrational spectroscopy measurements. **Vib. Spectrosc.** 35, 81-86.

B. Schrader, H. Schulz, M. Baranska, G.N. Andreev, C. Lehner, J. Sawatzki (2004) Non-destructive Raman analyses Polyacetylenes in plants. **Spectrochim. Acta** (accepted for publication).

H. Schulz, M. Baranska, R. Baranski (2004) Potential of NIR-FT-Raman spectroscopy in natural carotenoid analysis; **Biopolymers** (accepted for publication).

H. Nakanisnd I. B. sen and M. Sugeno A review and comparison of six reasoning methods **Fuzzy Sets and Systems** 1993 Vol. 57, pp. 257-294.

M. Sugeno and G. T. Kang Structure Identification of Fuzzy Model. **Fuzzy Sets and Systems** 1988 Vol. 28 pp. 15-33.

STAFFORD, F.E., HOLT, C.W., PAULSON, G.L. 1963. Vibration-Rotation Spectrum of HCl: A physical chemistry experiment. **Journal of Chemical Education**. Vol. 40 Number 5 May 1963 pp. 245-249.

L. A. Zadeh, The Concept of a Liguistic Variabel and Its Application to Approximate Reasoning. **Inf. Science** 1975 Vol. 9 pp. 43-80.

L. Zadeh, Soft computing and fuzzy Logic. **IEEE Software** 1994 Vol. 11 num. 6 Nov. pp. 48-56.

L. A. Zadeh, Fuzzy sets. **Information and Control** 1965 Vol. 8 pp. 338-353.

L. A. Zadeh, Fuzzy Logic: Advanced Concepts e Strucutures. **IEEE Educational Activities Video Tutorial** 1992.

L. A. Zadeh, Ouline of a new approach to the analysis of complex systems and decision processes. **IEEE Trans. Syst. Man and Cybern.** 1973 Vol. 3 January pp. 28-44.

H. J. Zimmermann **Fuzzy Set Theory and Its Applications** Kluwer Academic Publishers 1996 3nd Ed.

H. Zimmermann and M. Mizumoto Compartation of fuzzy reasoning methods. **Fuzzy Sets and Systems** 1982 Vol. 8 pp. 253-283.

S. T. Welstead **Neural network and fuzzy logic applications in C/C++**. John Wiley e Sons Inc., 1994.

Wenning, M., Seiler, H., Scherer, S. **Applied and Environmental Microbiology**, Oct. 2002, p.4717-4721.

Yoon, W.L., Jee, R.D., Charvill, A., Lee, G., Moffat, A.C., **J. of Pharmaceutical and Biomedical Analysis** 34 (2004) 933-944.