**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS**

Marcelo Camponez do Brasil Cardinali

**Diagnóstico de *Huanglongbing* (HLB) em citros
utilizando técnicas fotônicas**

São Carlos

2012

Marcelo Camponez do Brasil Cardinali

Diagnóstico de *Huanglongbing* (HLB) em citros utilizando técnicas fotônicas

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para a obtenção do título de Mestre em Ciências.

Área de Concentração: Física Aplicada
Orientador: Dra. Débora Marcondes Bastos Pereira Milori

Versão Corrigida

(Versão original disponível na unidade que aloja o programa)

São Carlos

2012

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação do IFSC,
com os dados fornecidos pelo(a) autor(a)

Camponez do Brasil Cardinali, Marcelo
Diagnóstico de Huanglongbing (HLB) em citros o
utilizando técnicas fotônicas / Marcelo Camponez do
Brasil Cardinali; orientadora Débora Marcondes
Bastos Pereira Milori - versão corrigida -- São
Carlos, 2012.
117 p.

Dissertação (Mestrado - Programa de Pós-Graduação em
Física Aplicada) -- Instituto de Física de São Carlos,
Universidade de São Paulo, 2012.

1. Espectroscopia de fluorescência. 2.
Espectroscopia de infravermelho. 3. Huanglongbing.
4. Greening. 5. Regressão por mínimos quadrados
parciais. I. Marcondes Bastos Pereira Milori,
Débora, orient. II. Título.

Dedico este trabalho aos meus pais Hebe e Toni (in memoriam) e a minha namorada e futura esposa Beatriz, por sempre acreditarem no meu potencial.

AGRADECIMENTOS

À DEUS, por abençoar todos os dias de minha vida.

Aos meus pais Hebe e Toni (*in memorian*) pela construção do meu caráter e me fornecer as bases para o que sou hoje.

À minha namorada e futura esposa Beatriz, por todo apoio nas horas boas e ruins, pela paciência de me ver apenas nos finais de semana. pelo amor e carinho que, com certeza, fizeram toda diferença no desenrolar do trabalho... Te amo muito!

Às minhas queridas irmãs Lívia e Vanessa, pela amizade, companheirismo, conselhos e incentivos ao longo de todos os anos de convivência.

À minha orientadora e amiga Dra. Débora, pelos ensinamentos, por sempre transmitir ânimo, pela paciência, pelos conselhos nas horas difíceis e principalmente pela confiança depositada em mim e no meu trabalho.

Ao Dr. Paulino pela amizade, grande apoio intelectual no trabalho, companheirismo nas escritas e correções do artigo e pelos valiosos conselhos acadêmicos.

À todos os companheiros de laboratório pela amizade, apoio e por tornar o laboratório um excelente ambiente de trabalho: Aline, Camila, Cleber, Edilene, Gustavo, Marina, Renan, Rafael e Thiago.

Aos meus amigos de república Elton, Foca, Loro, Milhouse, Orlando e Splinter pela verdadeira amizade e companheirismo durante esses anos de república.

À minha família por todo incentivo acadêmico que sempre recebi, em especial ao meu Avô Moacyr Camponez do Brasil (*in memorian*). Sempre vou lembrar da nossa convivência e de sua paixão em transmitir conhecimento.

À todos os funcionários da Embrapa Instrumentação que de alguma forma contribuíram para o desenvolvimento do trabalho, seja diretamente ou por tornar a Embrapa um excelente ambiente de trabalho. Agradecimento especial ao Ednaldo pelas conversas esclarecedoras de estatística e mineração de dados.

À Citrosuco pelo fornecimento de amostras, em especial ao Helton e André, por acreditarem no potencial das pesquisas por nós desenvolvida.

Ao Instituto de Física de São Carlos da Universidade de São Paulo, por tornar possível a realização do projeto.

À todos os funcionários da Pós-Graduação, em especial a Patricia e Ricardo, por sempre ajudar os alunos com eficiência, rapidez e bom humor.

À CAPES e ao CNPq pelo financiamento da bolsa e do projeto.

“Ser feliz sem motivo é a mais autêntica forma de felicidade”
— CARLOS DRUMMOND DE ANDRADE

RESUMO

CARDINALI, M. C. B. *Diagnóstico de Huanglongbing (HLB) em citros utilizando técnicas fotônicas*. 2012. 117 p. Dissertação (Mestrado em Física Aplicada) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012

A laranja é uma das frutas mais produzidas e consumidas no mundo, sendo o Brasil o maior produtor e exportador do seu suco concentrado. Entretanto, pragas e doenças comprometem consideravelmente sua produção. Atualmente, a doença mais preocupante é o *Greening*, também conhecida mundialmente como *Huanglongbing* (HLB). A doença não possui cura, apresenta longa fase assintomática e não possui um método eficiente de controle. Além disso, não existem métodos de diagnóstico aplicáveis em larga escala. Neste trabalho são propostas as técnicas fotônicas de fluorescência induzida por laser e de infravermelho por transformada de fourier para o diagnóstico do HLB. Para a realização das medidas, foram coletadas folhas de árvores saudáveis, doentes com HLB e doentes com a clorose variegada dos citros, sendo esta incluída nos estudos para verificar a capacidade de diferenciação entre as doenças. Foram coletadas quatro classes de folhas nessas plantas: sadia, HLB-sintomática, HLB-assintomática e CVC-sintomática. As folhas foram medidas em laboratório e seus espectros foram pré-processados para indução de um classificador via regressão por mínimos quadrados parciais. Além das folhas, foram medidas amostras dos seguintes metabólitos primários e secundários para entendimento espectral: amido, glicose, sacarose, hesperidina, naringina e umbeliferona. Taxas de acerto superiores a 89% foram obtidas na classificação das folhas nas técnicas de fluorescência e infravermelho, sendo superior às taxas dos métodos de manejo empregados atualmente no campo. A fluorescência induzida por laser apresenta um grande potencial para uso em campo devido a possibilidade de miniaturização de seus componentes. Os espectros dos metabólitos secundários apresentam fortes indícios de que a alteração de suas concentrações podem contribuir na detecção de doenças pelas técnicas fotônicas.

Palavras-chave: Espectroscopia de fluorescência. Espectroscopia de infravermelho. *Huanglongbing*. *Greening*. Regressão por mínimos quadrados parciais.

ABSTRACT

CARDINALI, M. C. B. *Huanglongbing (HLB) diagnosis in citros using photonic techniques*. 2012. 117 p. Dissertação (Mestrado em Física Aplicada) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012

Sweet orange is one of the most produced and consumed fruit in the world, and Brazil is the largest producer and exporter of this fruit. However, pests and diseases significantly reduce the worldwide production. Currently, the most destructive disease in the field is called greening, also known as huanglongbing (HLB). There is no control for HLB. In addition, the disease presents a long asymptomatic phase. Furthermore, no diagnostic methods are available to use in large scale. In this study are proposed fluorescence and infrared spectroscopy for the HLB diagnosis. For the measurements were collected leaves from healthy, HLB- and citrus variegated chlorosis-infected plants, being the last one to comparison between the diseases. It were collected four classes of leaves: healthy, HLB-asymptomatic, HLB-symptomatic and CVC-symptomatic. The leaves were measured and their spectra were pre-processed for the induction of classifier via partial least squares regression. In addition, samples of plant metabolites were measured for leaves spectral interpretation: starch, glucose, sucrose, hesperidin, naringin and umbelliferone. Success rates above 89% were obtained through both photonic techniques, higher compared to the success rates of the actual management methods. The metabolites spectra have shown strong evidence that their concentrations changes could contribute to the diagnosis of the diseases by photonic techniques. Particularly, the fluorescence spectroscopy seems interesting because it has a great potential for field application due to the existence of portable photonic devices.

Keywords: Fluorescence spectroscopy. Infrared spectroscopy. *Huanglongbing*. *Greening*. Partial least squares regression.

LISTA DE FIGURAS

- Figura 2.1 - Psílídeo *Diaphorina citri*, vetor do HLB. O inseto mede de 2 a 3 mm de comprimento. Imagem disponível em: <<http://www.research.ufl.edu/publications/explore/v11n1/story1.html>>. Acesso em: 22 fev 2012. 32
- Figura 2.2 - Sintomas típicos de HLB nas folhas (a), na casca do fruto (b) e dentro do fruto (c). Figura adaptada de imagens disponíveis em: <<http://www.fundecitrus.com.br/ImageBank/FCKEditor/image/Doen%C3%A7as/greening%202009%20018.jpg>>. Acesso em: 22/02/2012; <http://fabianoasato.blogspot.com/2010/04/greening-hlb-huanglongbing_06.html>. Acesso em: 22/02/2012; Manual técnico HLB (1). 33
- Figura 2.3 - Levantamento do HLB no estado de São Paulo e no sul de Minas Gerais. Mapa adaptado disponível em: <[http://www.fundecitrus.com.br/ImageBank/FCKEditor/file/pdf/relatorio_2010\(1\).pdf](http://www.fundecitrus.com.br/ImageBank/FCKEditor/file/pdf/relatorio_2010(1).pdf)>. Acesso em: 22 fev 2012. 34
- Figura 3.1 - Diagrama de Jablonski. 45
- Figura 3.2 - Espectro eletromagnético. 46
- Figura 3.3 - Diagrama simplificado de um espectrofluorímetro. 47

Figura 3.4 - Esquema de um Mini-espectrômetro composto por: 1: Conector SMA para fibra óptica; 2: Fenda; 3: Filtro; 4: Espelho colimador; 5: Grade de difração; 6: Espelho focalizador; 7: Lente para focalização da luz nos pixels da CCD; 8: Detector CCD. Figura extraída do manual USB2000 da Ocean Optics. Disponível em: < http://www.oceanoptics.com/technical/USB2000\%20operating\%20Instructions.pdf >. Acesso em: 15/02/12.	48
Figura 3.5 - Tipos de vibrações moleculares.	50
Figura 3.6 - Esquema do interferômetro de Michelson contido em um sistema FTIR.	51
Figura 3.7 - Passagem de um interferograma para o espectro de frequências utilizando a transformada de Fourier. Note que o espectro dado corresponde à absorbância.	52
Figura 3.8 - Representação gráfica das matrizes e vetores do método PLSR	54
Figura 3.9 - Representação gráfica das matrizes e vetores do método PLSR	55
Figura 5.1 - Exemplos típicos de folhas CVC (a), sadia (b), HLB-sintomática (c) e HLB-assintomática (d). Figura extraída de Cardinali et al. (2).	60
Figura 5.2 - Esquema do sistema de LIFS portátil desenvolvido no laboratório de Ótica e Lasers da Embrapa Instrumentação.	64
Figura 5.3 - Sistema de LIFS-405 portátil.	65
Figura 5.4 - Esquema de um espectrômetro FTIR. Figura adaptada disponível em: < http://mmrc.caltech.edu/FTIR/FTIRintro.pdf >. Acesso em: 22/01/2012.	66
Figura 5.5 - Esquema de um cristal de ATR. Nos pontos de reflexão do cristal em contato com a amostra, as amostras absorvem parte da radiação modificando o feixe transmitido para o detector.	67

Figura 5.6 - Diagrama ilustrando o algoritmo para uso de um método de regressão em um problema de classificação. A figura foi baseada no diagrama feito por Frank et al. (3).	70
Figura 5.7 - Distribuição normal padronizada (DNP) em comparação com a distribuição t de Student para os graus de liberdade $\nu = 1$ e $\nu = 10$. Quanto maior for o número de graus de liberdade, mais a distribuição se aproxima da DNP.	73
Figura 6.1 - Espectro de emissão de fluorescência médio das folhas com excitação em 560 nm. As amostras HLB-sintomáticas apresentam uma relação entre os picos 685 e 735 nm diferente em relação as demais classes. . .	76
Figura 6.2 - Espectro de emissão de fluorescência médio das folhas com excitação em 400 nm.	77
Figura 6.3 - Espectros médios de emissão de fluorescência das folhas de citros com excitação Laser em 561 nm	78
Figura 6.4 - Taxa de acerto do classificador em função do número de componentes utilizados no classificador via PLSR. A maior taxa de acerto pode ser observada em torno de 70% para 4 componentes e 9 componentes. . . .	79
Figura 6.5 - RMSE do classificador em função do número de componentes utilizados no classificador via PLSR. Nota-se uma diminuição do erro até o uso de 10 componentes. O desvio padrão aumenta conforme aumenta-se o número de componentes.	79
Figura 6.6 - Espectros médios de fluorescência das folhas de citros com excitação Laser em 405 nm	83
Figura 6.7 - Taxa de acerto do classificador em função do número de componentes utilizados no classificador via PLSR. Nota-se uma estabilização da taxa de acerto próximo de 11 componentes.	84

Figura 6.8 - RMSE do classificador em função do número de componentes utilizados no classificador via PLSR. Nota-se os valores de erros mais baixos entre 13 e 20 componentes.	84
Figura 6.9 - Espectro de fluorescência 3D da Umbeliferona dissolvida em água com concentração de 10 ppb.	89
Figura 6.10 - Espectro de emissão de fluorescência da Umbeliferona dissolvido em água com concentração de 2 ppm com excitação em 405 nm	89
Figura 6.11 - Espectro 3D da Hesperidina dissolvido em água com concentração de 40 ppm	91
Figura 6.12 - Espectro 3D da Naringina dissolvida em água com concentração de 20 ppm	91
Figura 6.13 - Espectro de FTIR típico de cada uma das classes de folhas. Figura adaptada de Cardinali et al. (2).	92
Figura 6.14 - Espectros normalizados de FTIR das pastilhas de KBr de amostras padrão de amido, glicose e sacarose. Figura adaptada de Cardinali et al. (2). . .	96
Figura 6.15 - Espectro de FTIR das pastilhas de KBr dos metabólitos secundários hesperidina, naringina e umbeliferona. Figura adaptada de Cardinali et al. (2).	96
Figura A.1 - Dimensões das matrizes envolvidas no MLR.	112
Figura A.2 - Dimensões das matrizes envolvidas no PCA	113

LISTA DE TABELAS

Tabela 5.1 - Número de folhas de cada planta utilizadas para a realização das medi- das de cada uma das técnicas.	61
Tabela 6.1 - Matriz de confusão obtida na execução de 10 validações cruzadas com 10 folds nos espectros do LIFS-561. As colunas correspondem às predições de classes feitas pelo classificador e as linhas às classes nominais. . . .	80
Tabela 6.2 - Matriz de confusão obtida na execução de 10 validações cruzadas com 10 folds dos espectros do LIFS-405. As colunas correspondem às predições de classes feitas pelo classificador e as linhas às classes nominais. . . .	84
Tabela 6.3 - Histograma acumulado das probabilidades de predição corretas e incor- retas das amostras do LIFS-405.	85
Tabela 6.4 - Matriz de confusão obtida na execução de 10 validações cruzadas com 10 folds nos espectros do FTIR. As colunas correspondem às predições de classes feitas pelo classificador e as linhas às classes nominais. Tabela extraída de Cardinali et al. (2).	93
Tabela 6.5 - Histograma acumulado das probabilidades de predição corretas e incor- retas das amostras do FTIR. Tabela adaptada de Cardinali et al. (2). . . .	94

LISTA DE ABREVIATURAS

<i>ATR</i>	<i>Attenuated total reflectance (Reflectância total atenuada)</i>
<i>CVC</i>	<i>Clorose Variegada dos Citros</i>
<i>FTIR</i>	<i>Fourier transform Infrared (Infravermelho por transformada de Fourier)</i>
<i>HLB</i>	<i>Huanglongbing</i>
<i>IR</i>	<i>Infrared (Infravermelho)</i>
<i>LIFS</i>	<i>Laser Induced Fluorescence Spectroscopy (Espectroscopia de fluorescência induzida por laser)</i>
<i>LIFS-405</i>	<i>LIFS com excitação laser em 405 nm</i>
<i>LIFS-561</i>	<i>LIFS com excitação laser em 561 nm</i>
<i>MLR</i>	<i>Multiple Linear Regression (Regressão linear múltipla)</i>
<i>PCR</i>	<i>Polymerase Chain Reaction (Reação em cadeia da polimerase)</i>
<i>PC</i>	<i>Principal Component (Componente principal)</i>
<i>PCA</i>	<i>Principal Component Analysis (Análise de componentes principais)</i>
<i>PLSR</i>	<i>Partial Least Squares Regression (Regressão por mínimos quadrados parciais)</i>
<i>PMT</i>	<i>Photomultiplier (fotomultiplicadora)</i>
<i>RMSE</i>	<i>Root Mean Squared Error (Raiz quadrada do erro quadrático médio)</i>
<i>RT-qPCR</i>	<i>Real time-quantitative PCR (Reação da polimerase em cadeia quantitativa em tempo real)</i>
<i>UV/VIS</i>	<i>Ultraviolet/Visible (Ultravioleta/Visível)</i>

SUMÁRIO

1	Introdução	27
2	Revisão bibliográfica	31
2.1	Huanglongbing (HLB)	31
2.2	Clorose variegada dos citros	35
2.3	Reações das plantas à doenças e estresses	36
2.4	Métodos de diagnóstico de estresses em citros	38
2.4.1	Reação em cadeia da polimerase	38
2.4.2	Espectroscopia de Fluorescência	39
2.4.3	Espectroscopia de Infravermelho	41
3	Fundamentos teóricos	43
3.1	Técnicas fotônicas	43
3.1.1	Espectroscopia de fluorescência	43
3.1.2	Espectroscopia de absorção no infravermelho	48
3.2	Análise multivariada	53
3.2.1	Regressão por mínimos quadrados parciais	53
4	Objetivos	57

5 Materiais e Métodos 59

5.1	Amostras	59
5.1.1	Folhas de citros	59
5.1.2	Carboidratos e metabólitos secundários	61
5.2	Técnicas fotônicas	62
5.2.1	Fluorescência convencional e 3D	62
5.2.2	LIFS	64
5.2.3	FTIR	66
5.3	Tratamento dos espectros	68
5.3.1	Fluorescência convencional e 3D	68
5.3.2	LIFS	68
5.3.3	FTIR	69
5.4	Método de classificação	69

6 Resultados 75

6.1	Fluorescência	75
6.1.1	Varredura de excitação	75
6.1.2	LIFS-561	77
6.1.3	LIFS-405	82
6.1.4	Metabólitos secundários	87
6.2	FTIR	92
6.2.1	Metabólitos	95

7	Conclusões	99
	REFERÊNCIAS	105
	Apêndice A – Regressão Linear Múltipla e Análise de Componentes Principais	111
	Apêndice B – Algoritmo PLS1	115
B.1	Algoritmo PLS1-calibração	115
B.2	Algoritmo de predição	117

1 Introdução

A produção de citros é uma das atividades econômicas mais importantes do setor agropecuário. Aproximadamente 122 milhões de toneladas de frutas cítricas, incluindo laranja, limão, toranja e tangerina são produzidas, correspondendo a quase 17 bilhões de dólares nas vendas de frutas frescas e de suco concentrado no mundo (4). O Brasil é o maior produtor e exportador de laranja do mundo, contribuindo com mais de 18 milhões de toneladas anuais e sendo quase 80% desta produção feita no estado de São Paulo (5). Além disso, 50% da produção mundial de suco concentrado é nacional e 98% desta é exportada (6). O setor ainda gera 230 mil empregos diretos em mais de 300 municípios (6).

Todo esse potencial econômico, entretanto, vem sendo ameaçado nos últimos anos devido à presença de pragas e doenças cada vez mais destrutivas. Atualmente, a doença de campo mais preocupante é o *Huanglongbing* (HLB), mais conhecida no Brasil como *Greening*. A doença teve origem na China em 1919 e dizimou grande parte de seus pomares de citros. Ao longo dos anos a doença foi se disseminando pelo mundo, sendo hoje encontrada em mais de quarenta países (7). O Brasil e os Estados Unidos, maiores produtores mundiais de citros, tiveram seu primeiro caso de HLB em 2004 e 2005, respectivamente e, atualmente, possuem cerca de 10% dos seus pomares infectados pela doença.

O HLB é causado pela bactéria provisoriamente chamada de *Candidatus Liberibacter* spp., que provoca a obstrução do floema e impede a distribuição de seiva do floema na planta (8). Ela é transmitida através de um psílídeo conhecido como *Diaphorina citri*, um pequeno inseto de 2 a 3 mm de comprimento que se alimenta da seiva floemática da planta (9). A transmissão da doença também pode ocorrer através da enxertia de mudas contaminadas.

Apesar do HLB ter presença mundial conhecida há quase 100 anos, a doença ainda não

possui controle ou método eficaz. O manejo realizado pelos produtores consiste no controle do vetor, erradicação de árvores sintomáticas e plantio de mudas certificadas. Entretanto, essas ações não tem sido suficientes para o controle da doença. A doença apresenta uma fase assintomática longa, que pode durar de 6 meses a 2 anos (10), tornando a planta uma fonte de inóculo da doença. Além disso, as inspeções visuais apresentam baixa eficiência (8) e os sintomas iniciais do HLB podem ser parecidos com o de outras doenças, como a clorose variegada dos citros (CVC) e com deficiências nutricionais, como zinco, manganês, magnésio e cobre (1). A ineficácia do manejo é evidente, visto o aumento dos casos ao longo dos anos.

Um método laboratorial empregado atualmente no diagnóstico do HLB é a análise de PCR, sigla do inglês *Polymerase Chain Reaction*. Essa técnica é baseada na comparação entre os DNAs presentes na folha com os possíveis DNAs da bactéria do HLB presentes na folha (11, 12). Dificuldades neste diagnóstico são encontradas em árvores que ainda não manifestam sintomas, pois a baixa quantidade de bactéria pode não ser detectada nas análises. Uma outra versão mais sensível do método usualmente empregada é o PCR em tempo real, conhecido pela sigla RT-qPCR, do inglês *real time - quantitative PCR*, o qual pode detectar e quantificar a bactéria em árvores sem sintomas visuais. Entretanto, ambas as técnicas tem um custo muito elevado e exigem um meticuloso preparo prévio das amostras, sendo, portanto, inviáveis para o uso em larga escala.

O Laboratório de Ótica e Lasers da Embrapa Instrumentação tem avaliado o potencial do uso de técnicas espectroscópicas como uma alternativa de diagnóstico da HLB (13–18). A hipótese básica para esta proposta é que a partir do momento em que a planta foi infectada, ela passa por alterações em seu metabolismo e, conseqüentemente, ocorrem alterações químicas e físicas que podem ser detectadas por técnicas espectroscópicas. A vantagem em relação ao método atual de diagnóstico é a redução significativa no custo da análise, aumento da velocidade do processo e possibilidade do desenvolvimento de equipamentos portáteis para uso em campo.

Este trabalho propõe dar continuidade na busca de um método alternativo para detecção de HLB no campo. Foram empregadas técnicas de espectroscopia de fluorescência convenci-

onal, espectroscopia de fluorescência induzida por laser, do inglês *Laser-induced fluorescence spectroscopy* (LIFS) e espectroscopia de infravermelho por transformada de Fourier, do inglês *Fourier transform infrared* (FTIR). As medidas foram realizadas em folhas de citros frescas, coletadas de árvores do campo sadias e infectadas com HLB. Além disso, foram avaliadas árvores infectadas com CVC para verificar se as técnicas fotônicas são capazes de distinguir as diferentes doenças. Para o entendimento de quais compostos químicos são responsáveis pelas alterações espectrais das folhas, amostras padrão dos carboidratos amido, glicose e sacarose e dos metabólitos secundários hesperidina, naringina e umbeliferona, foram também avaliados utilizando-se as mesmas técnicas fotônicas. A fluorescência convencional foi utilizada para verificar qual o comprimento de onda de excitação é mais adequado para diagnosticar as folhas e para excitar os metabólitos secundários. As técnicas de LIFS e FTIR foram utilizadas para verificar a potencialidade no diagnóstico do HLB nas folhas de citros. Além disso, os metabólitos primários e secundários foram avaliados no FTIR em busca de uma melhor interpretação dos espectros das folhas. Foi utilizada a regressão por mínimos quadrados parciais, do inglês *partial least squares regression* (PLSR), juntamente com uma metodologia proposta por Frank et al. (3) para utilização de ferramentas de regressão para problemas de classificação.

O presente trabalho se divide em sete capítulos e dois apêndices. O capítulo 2 descreve a revisão bibliográfica referente ao tema abordado. O capítulo 3 faz uma breve introdução teórica às técnicas fotônicas e ferramentas de extração de conhecimento utilizadas. O capítulo 4 apresenta os objetivos do trabalho. O capítulo 5 descreve todos os materiais utilizados e a metodologia aplicada e o capítulo 6, os resultados obtidos. Por fim, o último capítulo traz a conclusão do trabalho. Os apêndices A e B descrevem duas outras ferramentas muito utilizadas na quimiometria e o algoritmo da ferramenta PLSR.

2 Revisão bibliográfica

2.1 Huanglongbing (HLB)

O HLB ou *Greening*, também conhecido no Brasil como *amarelão*, é uma doença bacteriana que teve sua primeira manifestação conhecida em 1919 no sul da China. Atualmente é conhecida por ocorrer em diversos países da Ásia, África, Oceania e Américas do norte e do sul (7). No Brasil, maior produtor de citros do mundo, as primeiras plantas com sintomas da doença foram descobertas em março de 2004, na região de Araraquara, estado de São Paulo, sendo a primeira detecção no continente americano (8). Em poucos anos, a doença já havia se espalhado para grande parte do estado de São Paulo e em pomares de Minas Gerais e Paraná (8). Nos Estados Unidos, o segundo maior produtor do mundo, a doença chegou no estado da Flórida em 2005 (10) e, que tornou o país mais dependente da importação do suco de laranja.

O HLB é uma doença que obstrui o floema das árvores infectadas, impedindo o transporte da seiva elaborada para os órgãos da planta (7). Ainda não existe cura ou variedade imune ou resistente à doença. Quando contaminada, plantas novas não chegam a produzir frutos e plantas adultas tornam-se improdutivas em poucos anos (1). Os sintomas da doença podem demorar de 6 meses a 2 anos para aparecerem (10), dificultando as inspeções visuais e a ação dos produtores para impedir o avanço da doença no campo. Estima-se que atualmente haja pelo menos uma árvore assintomática para cada árvore sintomática encontrada no campo (10).

O agente causador da doença é a bactéria *Candidatus Liberibacter* spp, sendo duas espécies causadoras do HLB no Brasil: *asiaticus* e *americanus*. A *Candidatus Liberibacter asiaticus* é a mais comum nos pomares brasileiros devido a sua tolerância a altas temperaturas, de até

35 °C (19). A bactéria é transmitida principalmente por um psílídeo conhecido como *Diaphorina citri*, um inseto de coloração cinza que mede de 2 a 3 mm de comprimento (ver Fig. 2.1). O psílídeo se hospeda em todas as variedades de citros e em plantas ornamentais conhecidas como falsa murta, que também são hospedeiras da bactéria (1). Os adultos se alimentam da seiva floemática da planta, geralmente em brotos novos e permanecem sobre as folhas e ramos. O inseto adquire a doença se alimentando de plantas doentes, quer elas apresentem sintomas ou não. Durante sua alimentação, ele permanece em uma posição inclinada com um ângulo de 45 °, facilitando sua identificação. Além da transmissão via vetor, a doença pode ser transmitida através da enxertia de borbulhas de citros contaminadas. As borbulhas são partes do tecido da planta responsáveis pela propagação da espécie. A enxertia é realizada a fim de unir tecidos de plantas diferentes, para melhorar a qualidade da produção e a resistência contra pragas.

Figura 2.1 – Psílídeo *Diaphorina citri*, vetor do HLB. O inseto mede de 2 a 3 mm de comprimento. Imagem disponível em: <<http://www.research.ufl.edu/publications/explore/v11n1/story1.html>>. Acesso em: 22 fev 2012.

Os sintomas típicos do HLB que aparecem nas plantas ocorrem nas folhas e nos frutos, como mostra a Fig. 2.2. As plantas infectadas sintomáticas apresentam em geral um mosqueado amarelado assimétrico nas folhas e frutos deformados (8). Em alguns casos, as folhas são mais espessas e apresentam um engrossamento nas nervuras. É comum também o aparecimento de sintomas de deficiência de zinco em alguns ramos (9). Os frutos geralmente são menores, mais ácidos, com o albedo (parte branca da casca) espesso e sementes abortadas. Alguns frutos também podem apresentar um mosqueado amarelado na casca (9).

Figura 2.2 – Sintomas típicos de HLB nas folhas (a), na casca do fruto (b) e dentro do fruto (c). Figura adaptada de imagens disponíveis em: <<http://www.fundecitrus.com.br/ImageBank/FCKEditor/image/Doen%C3%A7as/greening%202009%20018.jpg>>. Acesso em: 22/02/2012; <http://fabianoasato.blogspot.com/2010/04/greening-hlb-huanglongbing_06.html>. Acesso em: 22/02/2012; Manual técnico HLB (1).

O controle da doença no campo é realizado através de inspeções visuais frequentes para erradicação de plantas sintomáticas e controle do vetor através de pulverizações constantes. Além disso, é recomendado a aquisição de mudas e borbujas de citros certificadas para novos plantios (8). A instrução normativa nº 53, em vigor desde outubro de 2008, regulamenta por lei as obrigações dos produtores em eliminar as plantas sintomáticas da fazenda, cumprindo no mínimo uma inspeção visual a cada três meses. A Fundecitrus foi responsável pelas fiscalizações nas fazendas até janeiro de 2010 (20). Após esta data, a Fundecitrus passou apenas a orientar os citricultores no manejo devido a pouca eficácia das fiscalizações.

A Fig. 2.3 mostra um levantamento da doença no estado de São Paulo e sul de Minas Gerais, realizado pela Fundecitrus em 2010. A incidência de HLB dobrou em dois anos no estado de São Paulo e Minas Gerais. O relatório apontou que o HLB atingiu cerca de 53,3% dos talhões dos pomares e 3,8% das plantas (20). O primeiro percentual indica que pelo menos uma planta se encontra infectada em um determinado talhão. Nota-se que a região central do estado é mais

afetada pela doença, com aproximadamente 73,5% de talhões infectados. Apesar desse fato, a incidência de plantas infectadas nessa região é de 6%.

Figura 2.3 – Levantamento do HLB no estado de São Paulo e no sul de Minas Gerais. Mapa adaptado disponível em: <[http://www.fundecitrus.com.br/ImageBank/FCKEditor/file/pdf/relatorio_2010\(1\).pdf](http://www.fundecitrus.com.br/ImageBank/FCKEditor/file/pdf/relatorio_2010(1).pdf)>. Acesso em: 22 fev 2012.

A baixa eficiência da inspeção visual para encontrar plantas sintomáticas, por volta de 27% (8) e a falta de cumprimento da instrução normativa n° 53 por parte de alguns produtores agravam ainda mais a disseminação da doença no campo. Atualmente, é muito comum na região central do estado de São Paulo encontrar pequenas e médias fazendas produtoras de citros praticamente tomadas pelo HLB. Alguns produtores deixam suas fazendas nesse estágio até que a colheita não valha mais a pena para trocar de cultura. Belasque et al (8) mostrou que as medidas de manejo só são efetivas quando todas as fazendas de uma região as realizam.

Ainda não há um método de detecção preciso do HLB aplicável no campo. Atualmente, a técnica padrão para detecção da doença em laboratório é o RT-qPCR. Esta técnica é mais sensível que o PCR comum e ainda faz a quantificação da bactéria presente na amostra anali-

sada. Como discutido anteriormente, o RT-qPCR visa encontrar o DNA da bactéria nas folhas de citros de árvores infectadas. Entretanto, a distribuição da bactéria nos galhos não é uniforme e se encontra geralmente nos galhos afetados pelos sintomas (11). Além disso, a técnica é cara, necessita de um preparo cuidadoso das amostras para evitar contaminações e é inviável para aplicação em larga escala. Esse e os demais fatores associados ao controle efetuado pelos citricultores demonstram que há urgência no desenvolvimento de um método de detecção de HLB.

2.2 Clorose variegada dos citros

A Clorose variegada dos citros (CVC), também conhecida como amarelinho, é uma doença causada pela bactéria *Xylella fastidiosa* que coloniza o xilema e interrompe o transporte de água e nutrientes da raiz para a copa. A obstrução provoca a diminuição do fruto, tornando-os inviáveis para venda e consumo. A doença afeta todas as variedades de laranja doce igualmente e pode ser transmitida por até 12 espécies de cigarrinhas, mudas e borbulhas contaminadas (21).

A doença foi observada pela primeira vez no Brasil em 1987 na região do triângulo mineiro e norte do estado de São Paulo. Em 2005 já estava presente em todas as regiões produtoras de citros no estado de São Paulo (22). A distribuição da doença no estado é não-uniforme, se concentrando mais na região central. Estima-se que o gasto no controle da doença no estado de São Paulo seja de US\$ 120 milhões por ano (22).

Os sintomas de CVC aparecem nas folhas, ramos e frutos. Nas folhas aparece um amarelado assimétrico com lesões espalhadas na superfície da face adaxial (superior) (23). Nos sintomas iniciais, a doença pode ser facilmente confundida com HLB. Em folhas mais velhas, as lesões se estendem até a superfície abaxial (inferior) das folhas. Os frutos não amadurecem, permanecendo pequenos e duros na época da colheita (24). Em estágios mais avançados, os galhos mais altos da planta secam e apresentam queda de folhas (21).

O manejo realizado no campo consiste no plantio de mudas certificadas, controle dos vetores e arranque dos galhos afetados em árvores mais velhas (acima de 6 anos) e erradicação das árvores mais novas (abaixo de 5 anos) (22). Da mesma forma que para o HLB, as técnicas de PCR são as únicas ferramentas eficazes no diagnóstico da doença utilizado atualmente em laboratório. Pelos mesmos motivos descritos para o HLB, o PCR não é aplicado em larga escala.

2.3 Reações das plantas à doenças e estresses

A presença de patógenos e deficiências nutricionais nas plantas podem acarretar em alterações químicas e físicas nas plantas. Estas podem ser oriundas de uma reação específica das plantas ou pela interrupção de um determinado processo fisiológico. As substâncias químicas envolvidas nesses processos são chamados de metabólitos, que por definição, são produtos do metabolismo de um determinado organismo vivo. Nas plantas, os metabólitos primários estão envolvidos no crescimento e manutenção da planta, como por exemplo: lignina, celulose, proteínas, lipídios, carboidratos e os pigmentos. As plantas ainda produzem metabólitos secundários, muitas vezes relacionados ao mecanismo de defesa da planta. Como exemplo de metabólitos secundários pode-se citar: alcalóides, flavonóides, terpenóides, compostos fenólicos, aldeídos, álcoois, cetonas e as fitoalexinas.

Dentre os grupos de metabólitos primários presentes nas plantas, a clorofila e os carotenóides se sobressaem pela importância biológica, pois participam do processo fotossintético da planta. As clorofilas, que conferem a coloração verde às folhas, são presentes nas angiospermas como clorofilas *a* e *b*. Os carotenóides, em sua maioria, são pigmentos amarelados ou alaranjados, e suas cores não aparecem visivelmente nas folhas devido à maior presença das clorofilas. Estudos realizados anteriormente em diferentes variedades de laranjas, com folhas de árvores saudáveis e infectadas com as doenças Morte súbita dos citros, Declínio e HLB, sugerem que as concentrações das clorofilas se alteram conforme as doenças progridem na planta (13). Além disso, estresses abióticos também podem afetar a concentração da cloro-

fila (25, 26). Buschmann (25) explica que a excitação da clorofila com luz visível promove a emissão de duas bandas de fluorescência no vermelho e infravermelho próximo, sendo a intensidade delas relacionadas à concentração de clorofila nas folhas. A razão entre os máximos dessas bandas é, em geral, utilizada para determinar a concentração de clorofila nas folhas, mas também pode ser utilizado para detectar estresses bióticos e abióticos diversos (25).

Outros metabólitos primários de grande importância nas plantas são os carboidratos (e.g. amido, sacarose, glicose, maltose, frutose entre outros). O amido é produzido para ser utilizado como reserva energética. Em diferentes estudos, foi observado que a concentração do amido pode variar de acordo com estresses sofridos pelas plantas. Achor et al. (27) usou a técnica de Microscopia Eletrônica de Transmissão para relacionar os sintomas apresentados pelo HLB nas folhas com o aumento do amido e obstrução do floema. O aumento do amido é causado pela obstrução do floema, que impede o seu transporte, levando ao seu acúmulo nas faces adaxial até a abaxial. Han et al. (26) verificou também o acúmulo de amido nas folhas de plantas de citros com deficiência de Boro.

Fan et al. (28) encontrou alterações nas concentrações de diferentes carboidratos em folhas saudáveis, HLB-sintomáticas e HLB-assintomáticas. Foi observado que enquanto o amido e a sacarose aumentaram suas concentrações em folhas HLB-sintomáticas e HLB-assintomáticas quando comparadas com folhas saudáveis, a maltose decresceu progressivamente de folhas saudáveis para folhas HLB-sintomáticas. Houve também um aumento significativo na concentração de frutose em folhas HLB-assintomáticas comparadas com folhas saudáveis e HLB-sintomáticas.

Alguns trabalhos tem evidenciado alterações nas concentrações de metabólitos secundários devido a presença de patógenos. Tais substâncias são mais conhecidas como fitoalexinas, importantes devido à característica de defesa contra patógenos (29). Tais metabólitos secundários devem aumentar em plantas doentes. Ribeiro et al. (30) observou um aumento na concentração do flavonóide hesperidina em células mesófilas de folhas de árvores infectadas com CVC. Esse fato sugere que a hesperidina pode funcionar como uma barreira para a colônia de *Xylella fastidiosa* ou como precursora de outros elementos que poderiam combater a bactéria.

Afek et al. (31) verificou o acúmulo de escorparona (da família das cumarinas) nos galhos de citros após a inoculação da bactéria *Phytophthora citrophthora*, causadora do mal de fitoftora. As plantas foram tratadas previamente com fosetyl-Al. Essa substância atuou indiretamente na planta ativando o acúmulo da escorparona, que funciona como mecanismo de defesa para a planta.

Nota-se portanto uma série de trabalhos que evidenciam alterações químicas em metabólitos primários e secundários em plantas doentes. Como visto, é possível explorar essas alterações para detecção de doenças.

2.4 Métodos de diagnóstico de estresses em citros

Como descrito anteriormente, ainda não existem métodos aplicáveis no campo para detecção de HLB e CVC. Atualmente, a identificação de plantas doentes é realizada através de inspeções visuais frequentes no pomar. Essa atividade é pouco eficaz, pois podem levar a escapes de plantas doentes sintomáticas e a diagnósticos equivocados devido a confusões com outras doenças. Nesta seção, além da demonstração dos trabalhos realizados com a técnica padrão PCR na detecção de HLB e CVC, foram revisados os trabalhos realizados utilizando as espectroscopias de fluorescência e de infravermelho na detecção de estresses em plantas.

2.4.1 Reação em cadeia da polimerase

A Reação em cadeia da polimerase, da sigla em inglês PCR, é uma técnica de análise biológica muito utilizada na bioquímica e no diagnóstico de doenças. Especificamente na fitopatologia, seu uso é denominado padrão na detecção de doenças como HLB e CVC. A técnica se baseia na replicação de um determinado fragmento de DNA. No caso do diagnóstico de doenças, a amplificação do material genético ocorre apenas na presença do patógeno. Na detecção de

HLB e de CVC, muitos trabalhos foram realizados demonstrando a capacidade da técnica e suas limitações (e.g. (11, 12, 32)).

Os trabalhos realizados até o momento demonstraram a eficiência de alguns tipos de PCR, sendo que uns são mais ou menos sensíveis à detecção da bactéria. Um trabalho importante foi o do mapeamento da bactéria do HLB nos galhos de uma árvore sintomática realizado por Teixeira et al. (11). Foram utilizados três diferentes versões de PCR (diferentes sensibilidades) para o mapeamento da bactéria. Foi possível a identificação da bactéria apenas nos galhos afetados. Em galhos não sintomáticos não foram encontradas bactérias nas amostras. Mostrou-se também que para detecção em folhas assintomáticas, o uso de PCR quantitativo em tempo real foi mais adequado, pois com ela é possível detectar a bactéria em baixíssimas concentrações (aproximadamente 10 bactérias por grama de amostra).

Apesar das altas taxas de detecção nos galhos sintomáticos, é possível ver que o PCR apresenta problemas em detectar a doença em uma planta devido à distribuição não uniforme da bactéria, mesmo nas versões mais sensíveis. Outros problemas relativos à técnica são a necessidade do preparo prévio das amostras, alto custo por análise (em torno de 30 dólares por árvore) e grande cuidado na preparação das amostras para evitar contaminações. Esses fatores tornam a técnica inviável para uso em larga escala.

2.4.2 Espectroscopia de Fluorescência

A técnica de fluorescência encontra ampla aplicação no estudo de compostos orgânicos heterogêneos, como solos e tecidos vivos animais e vegetais, nos quais a presença de fluoróforos (moléculas emissoras de luz quando excitadas) é sempre muito provável. Existem inúmeras aplicações da técnica em diversas áreas, sendo as mais comuns em química analítica e físico-química (e.g. (33)). No caso do diagnóstico de estresses e doenças em citros, a técnica é promissora pelo fato de poder detectar as alterações químicas e físicas provocadas na planta.

Em relação a metodologias de diagnóstico de estresses bióticos e abióticos em plantas, Lichtenthaler et al. (34) mostrou que é possível detectar estresses em plantas através da reflectância e da fluorescência. De acordo com o estudo de fluorescência, as folhas quando excitadas no ultravioleta apresentam uma banda larga de emissão no azul e verde e as emissões características da clorofila na região do vermelho e infravermelho. As emissões da região azul e verde correspondem em grande parte aos ácidos ferúlicos e a uma menor parte a outros compostos tais como flavonóides, cumarinas, quercetinas, entre outros (35). Através das razões entre os picos da região azul-verde e entre os máximos das bandas da clorofila foi possível diferenciar folhas de plantas com deficiências diversas de plantas saudáveis (34). Esses diagnósticos indicam que tanto as clorofilas quanto os compostos emissores na região azul-verde alteram suas emissões devido as suas alterações de concentrações, de acordo com a evolução de um determinado estresse.

Kancheva et al. (36) verificou que através da fluorescência foliar é possível detectar alterações na concentração de clorofila. Foi observado que o estresse induzido leva a diminuição dos compostos fotossintéticos das plantas, que afetam a emissão desses compostos na região do vermelho e vermelho distante (entre 600 e 800 nm). Foram evidenciadas as diferenças espectrais em folhas com diferentes concentrações de clorofila através da razão entre os picos de emissão.

Pereira et al. (17) utilizou a espectroscopia de imagens de fluorescência induzida por laser para estudar a evolução do HLB em mudas inoculadas e comparar com mudas saudáveis. Foi possível diagnosticar a presença da doença já no primeiro mês após a inoculação. Nesse mesmo estudo de mudas, Venâncio (14) demonstrou resultados promissores utilizando LIFS, conseguindo o diagnóstico precoce de aproximadamente 60% das mudas a partir do 3º mês após a inoculação. O sistema utilizado foi desenvolvido no laboratório de ótica e lasers da Embrapa Instrumentação e se encontra na patente internacional *WO2010069017-A1*. O laser de excitação utilizado foi de 561 nm (amarelo), embora na literatura alguns trabalhos dão evidências que excitações mais energéticas provém uma emissão na região do azul-verde importante para o diagnóstico de HLB e estresses (17, 34, 35). Vale ressaltar que em ambos trabalhos citados nesse parágrafo não houve testes com outras doenças para comparação.

Os trabalhos já realizados evidenciam uma grande potencialidade do LIFS para detecção de doenças e estresses em citros. Entretanto, é importante avaliar a excitação das folhas de citros em outros comprimentos de onda e comparar com outras doenças. Nesse trabalho a fluorescência foi avaliada para o diagnóstico de folhas de HLB, CVC e sadias, varrendo em diferentes comprimentos de onda no visível. Além disso, um estudo com alguns metabólitos secundários foi realizado para interpretação dos espectros das folhas.

2.4.3 Espectroscopia de Infravermelho

A espectroscopia de infravermelho tem encontrado muitas aplicações, sobretudo na química, com destaque para a físico-química e a química analítica. No contexto da citricultura, a técnica FTIR já foi utilizada no diagnóstico de HLB, de outras doenças e de deficiências diversas (37–39). Além disso, foi utilizada também no entendimento das alterações químicas que as doenças podem provocar nas plantas.

Hawkins et al. (37) empregou o FTIR para comparar diferentes doenças como tristeza, declínio dos citros, cancro cítrico e algumas deficiências nutricionais. As folhas de plantas com as deficiências nutricionais e com cancro cítrico não puderam ser distinguidas de HLB. Vale ressaltar que as folhas utilizadas apresentavam sintomas característicos em cada doença ou deficiência. Em outras análises, Hawkins et al. (38) comparou plantas saudáveis com plantas infectadas com HLB em diferentes estágios da doença, incluindo estágios assintomáticos. Com o uso do FTIR, foi possível atingir a taxa de acerto de 95% no geral. Entretanto, o número de árvores assintomáticas utilizado foi de 6 num universo de 124 árvores HLB-sintomáticas. Além disso, o diagnóstico correto ocorreu em apenas 4 das árvores assintomáticas. Em ambos os trabalhos, as amostras de folhas foram secas em um microondas e trituradas para a realização das medidas, o que torna o processo mais demorado.

Usando Espectroscopia de infravermelho médio, Sankaran et al. (39) distinguiu folhas HLB-sintomáticas de saudáveis e outras com deficiências nutricionais, com taxas de acerto

superiores a 95%. Além disso, foi observado também o acúmulo de amidos nas folhas infectadas com HLB através de testes químicos. Nesse estudo, as folhas coletadas foram previamente preparadas para as análises. As folhas analisadas foram misturadas a nitrogênio líquido e trituradas, sendo colocadas em seguida para secagem durante 2 dias.

Nota-se, portanto, que embora a espectroscopia de infravermelho já tenha sido aplicada no diagnóstico de doenças cítricas, ainda faltam muitos estudos. As plantas utilizadas nesses estudos foram provenientes de regiões de clima temperado. Além disso, houve preparo prévio de amostras, dificultando o uso em larga escala. No presente estudo, as folhas utilizadas são provenientes de plantas do clima tropical e suas medidas foram realizadas de maneira direta sem preparo prévio, i.e., nas folhas frescas coletadas da árvore. Além disso, este estudo contemplou a comparação entre as doenças HLB e CVC utilizando FTIR, sendo inédito até o momento a comparação entre essas doenças.

3 Fundamentos teóricos

3.1 Técnicas fotônicas

A fotônica é o ramo da física que se refere à pesquisa, desenvolvimento e aplicação da ótica e lasers. As aplicações envolvem a faixa do espectro eletromagnético que se estende do ultra-violeta até o infravermelho distante. Nesta seção será apresentada uma breve introdução teórica das espectroscopias de fluorescência e de infravermelho.

3.1.1 Espectroscopia de fluorescência

A interação da radiação com a matéria pode promover diferentes fenômenos físicos, tais como a reflexão, absorção, transmissão, espalhamento, luminescência, reações fotoquímicas, entre outros. Quando a radiação é absorvida pela matéria e promove transições eletrônicas com emissão de radiação, chamamos o fenômeno de luminescência (40). O termo luminescência é bastante amplo, abrangendo qualquer emissão que não seja provocada devido à temperatura do corpo. Em compostos orgânicos, a luminescência estimulada por radiação é dividida em duas categorias: fluorescência ou fosforescência, dependendo da natureza do estado excitado.

A transição eletrônica é bastante conhecida na espectroscopia, sendo compreendida através da mecânica quântica. Dessa forma, os processos envolvidos são probabilísticos e cada fóton da radiação incidente provê um quantum de energia. Para uma dada frequência de radiação ν , os fótons da radiação possuem uma energia quantizada dada pela relação de Planck

$$E = h\nu, \quad (3.1)$$

onde h é a constante de Planck. Para que a absorção dessa radiação ocorra e o elétron do estado fundamental seja excitado para outro estado de maior energia, a frequência desta deve corresponder à separação de energia entre os dois estados eletrônicos, i.e.,

$$\Delta E = E_m - E_n = h\nu, \quad (3.2)$$

onde E_n e E_m são as energias dos estados eletrônicos inicial e final, respectivamente. A transição do elétron para a camada mais energética é tipicamente da ordem de 10^{-15} s, tempo muito curto para movimentos moleculares como diz o princípio de Franck-Condon (40).

Após a absorção da radiação, muitos processos podem ocorrer. O diagrama de Jablonski da figura 3.1 ilustra os possíveis processos de maneira simplificada, não levando em conta outros fenômenos tais como o *quenching* (redução da intensidade da fluorescência por um determinado processo) e interações entre solventes. No diagrama, os estados eletrônicos singletos estão representados pela letra S e o estado eletrônico tripleto pela letra T . onde seus sub-índices são relacionados ao estado fundamental (0) ou estados excitados (1,2,...). Cada linha do estado eletrônico correspondem aos níveis de energia vibracionais, representados pelos números 0, 1 e 2. O fluoróforo é excitado para um estado vibracional de energia mais elevada e de mesma multiplicidade de spin, como o exemplo do diagrama, para estados singletos S_1 ou S_2 . Logo após a excitação, parte da energia do fóton incidente é dissipada por vibrações moleculares. Esse processo, chamado de relaxação vibracional ou conversão interna, ocorre tipicamente em 10^{-12} s. Como o tempo de vida da fluorescência e da fosforescência são tipicamente maiores, a emissão de luz ocorre sempre em níveis de energia menores que o inicialmente excitado.

O retorno do elétron para o seu estado de origem ocorre com a emissão de radiação. A transição entre os estados do tipo singleto (e.g. S_1 e S_0) com emissão de radiação é chamada de fluorescência. Essa transição é permitida quanticamente por não ocorrer inversão de spin. O tempo de vida da fluorescência é de aproximadamente 10^{-8} s. Devido a esse curto intervalo

Figura 3.1 – Diagrama de Jablonski.

de tempo comparado com a escala de tempo visual, a fluorescência ocorre tipicamente apenas durante a excitação com radiação.

As moléculas no estado S_1 podem sofrer uma inversão de spin, ocorrendo a transição para um estado tripleto, como para o primeiro estado tripleto T_1 . Esse processo é chamado de cruzamento intersistema. A transição de um estado tripleto para um singlete é proibida devido às regras de seleção quânticas, pois ocorre inversão de spin. Sendo assim, a probabilidade da transição é muitas ordens de grandeza inferior a fluorescência. A emissão de luz de um estado tripleto para um singlete, como no caso dos estados T_1 para S_0 , é chamada de fosforescência. O tempo de vida da fosforescência é longo, tipicamente da ordem de segundos a milissegundos. Geralmente a emissão ocorre em grandes comprimentos de onda (i.e., com energias menores) comparando-se com a fluorescência (40).

A fluorescência convencional é observada para excitações entre o ultravioleta e o visível (ver Fig. 3.2). Nos espectrofluorímetros utilizados neste trabalho, tem-se a disponibilidade de excitar as amostras com comprimentos de onda na faixa entre 250 nm e 700 nm. Acima de 700 nm (abaixo de $4,3 \cdot 10^{14}$ Hz de frequência) de excitação, a energia das radiações vão, em geral, apenas promover transições não radioativas, tais como as relaxações vibracionais e rotacionais.

Em compostos orgânicos, a fluorescência ocorre tipicamente em estruturas rígidas, i.e, cadeias fechadas e com ligações duplas. A falta de rigidez, como a de moléculas com cadeias

Figura 3.2 – Espectro eletromagnético.

lineares e longas, aumenta a taxa de relaxação vibracional. A rigidez de compostos aromáticos garante que a energia da radiação incidente não seja perdida completamente em relaxações vibracionais (40). As cadeias cíclicas de ligações duplas diminuem o número de modos normais de vibração e de rotação. Nesse aspecto, uma estrutura suficientemente rígida minimiza perdas de energia por processos não radioativos, aumentando a probabilidade da emissão de radiação. Dessa forma, a estrutura da molécula juntamente com suas ligações químicas são fatores importantes na emissão da fluorescência.

A espectroscopia de fluorescência pode ser realizada utilizando equipamentos conhecidos como fluorímetros e espectrofluorímetros. O esquema básico desse equipamento segue ilustrado na Fig. 3.3. Nesses equipamentos, é comum o uso de lâmpadas para excitação no ultravioleta e visível, o que torna possível a varredura de excitação. O espectrofluorímetro faz uso de monocromadores (sistema contendo basicamente fenda e grade de difração) para ser possível excitar a amostra com radiação de determinada faixa de comprimento de onda. A fluorescência emitida pela amostra passa por um filtro para retirar a luz de excitação e por um monocromador de emissão. Em seguida, a luz segue até atingir o detector, geralmente um PMT, sigla do inglês *photomultiplier* (fotomultiplicador). O PMT converte a radiação incidente em corrente elétrica, que é amplificada e analisada para posterior interpretação espectral da luz. É comum a presença de um PMT para coleta de um sinal de referência para correções do espectro coletado. Os sinais são enviados para um computador para a geração do espectro da amostra. Vale ressaltar que o

computador faz o controle dos monocromadores de acordo com o modo de aquisição. Para o modo de emissão, o monocromador de excitação é fixo para uma determinada frequência e o de emissão é variável, de modo a varrer a faixa de frequência escolhida. No modo de excitação, os monocromadores apresentam o comportamento inverso.

Figura 3.3 – Diagrama simplificado de um espectrofluorímetro.

É comum atualmente a montagem de sistemas de bancada, onde usualmente são utilizadas fontes de luz laser, sigla do inglês *light amplification by stimulated emission of radiation*. O uso do laser é devido a sua alta intensidade, monocromaticidade, coerência e alta direcionalidade. Os lasers mais utilizados nas aplicações tecnológicas são os de diodo, devido ao menor custo, baixa manutenção e tamanho reduzido em comparação às demais tecnologias.

Outro componente muito utilizado em sistemas de bancada são os mini-espectrômetros, tal como apresentado na Fig. 3.4. Os mini-espectrômetros apresentam-se em um formato compacto, contendo os componentes essenciais para a aquisição de um espectro de uma determinada radiação. Os principais componentes consistem de uma fenda de abertura para entrada do sinal de luz, uma grade de difração para dispersão e um detector do tipo *linear silicon CDD array*, onde CCD é a sigla do inglês *Charge-coupled device* - dispositivo de carga acoplada. Os mini-espectrômetros garantem medidas rápidas e simultâneas devido às características do

detector.

Figura 3.4 – Esquema de um Mini-espectrômetro composto por: 1: Conector SMA para fibra óptica; 2: Fenda; 3: Filtro; 4: Espelho colimador; 5: Grade de difração; 6: Espelho focalizador; 7: Lente para focalização da luz nos pixels da CCD; 8: Detector CCD. Figura extraída do manual USB2000 da Ocean Optics. Disponível em: <<http://www.oceanoptics.com/technical/USB2000%20operating%20Instructions.pdf>>. Acesso em: 15/02/12.

O detector tipo CCD capta a luz de todos os pontos do espectro simultaneamente. O CCD é composto por uma rede linear de elementos fotossensíveis, sendo cada um deles chamados de pixels. A luz de entrada interage com cada elemento fotossensível liberando elétrons através do efeito fotoelétrico. A quantidade de elétrons liberada é proporcional à intensidade da luz incidente. Deste modo, o sinal analógico de saída do CCD é convertido em digital para obtenção do espectro de luz de entrada. Nos espectrômetros, cada um dos pixels é responsável por uma frequência de luz, sendo por esse motivo o uso de uma grade de difração.

3.1.2 Espectroscopia de absorção no infravermelho

A espectroscopia de IR, sigla do inglês *infrared*, é um tipo de espectroscopia de absorção que utiliza a porção do IR da radiação eletromagnética. A técnica explora o fato de que as ligações químicas das moléculas presentes em uma determinada amostra absorvem frequências específicas que são características de sua estrutura. As ligações químicas das substâncias possuem frequências de vibração específicas que correspondem aos níveis de energia vibracionais

da molécula. A radiação incidente é absorvida por uma dessas moléculas no caso em que esta altera seu momento dipolar durante as vibrações, devido as regras de seleção quânticas (41). O resultado obtido é um espectro rico em informações das moléculas contidas na amostra.

A radiação do IR envolve a região do espectro eletromagnético entre a radiação visível e de microondas (ver Fig. 3.2). A região mais utilizada na prática para a identificação de compostos químicos nas amostras é compreendida entre 4000 a 400 cm^{-1} (2500 a 25000 nm) (41). A unidade cm^{-1} é chamada de número de onda e é uma grandeza física inversamente proporcional ao comprimento de onda, sendo definida por

$$\tilde{\nu} = 1/\lambda \quad (3.3)$$

O uso da unidade número de onda decorre do fato dela ser proporcional a energia e a frequência da radiação. Dessa forma, números de onda mais baixos estão associados a energias mais baixas, enquanto que os números de onda maiores são mais energéticos (41).

O número de modos vibracionais moleculares é descrito pelos graus de liberdade internos, dados através das relações $3n - 6$ e $3n - 5$, válidas para as vibrações de moléculas lineares e não lineares, respectivamente. Nessas relações, n é o número de átomos na molécula e as subtrações numéricas correspondem aos graus de liberdade de translação e rotação. Dentre as vibrações, existem dois tipos fundamentais que podem ocorrer nas moléculas: estiramento e deformação. A Fig. 3.5 mostra um exemplo de cada um dos tipos de vibração fundamental. A vibração de estiramento ocorre ao longo do eixo de ligação, aumentando ou diminuindo a distância interatômica. A vibração de deformação leva a alteração dos ângulos no eixo de ligação. Como dito anteriormente, apenas as vibrações que provocam mudanças no momento dipolar são registradas no espectro de IR, pois o dipolo oscilante gera um campo elétrico que interage com a componente elétrica da radiação incidente (42). Sendo assim, apenas as vibrações assimétricas são ativas no IR, pois seus momentos dipolares parciais não se anulam em qualquer que seja a distância em relação à posição de equilíbrio. As vibrações assimétricas são os modos vibracionais nos quais enquanto um átomo se afasta o outro átomo se aproxima do átomo central e vice

versa.

Figura 3.5 – Tipos de vibrações moleculares.

Os espectros de IR são representados normalmente em termos de transmitância ou absorbância, com relação ao número de onda da radiação incidente. Os espectros de IR permitem a identificação de compostos se comparado com os de uma amostra conhecida (42). Em amostras complexas, a identificação se torna muito complicada devido a possíveis sobreposições de bandas de absorção. Entretanto, a técnica pode ser utilizada apenas para diferenciação entre dois tipos de amostras diferentes, sem o interesse da identificação dos grupos funcionais. Folhas de citros apresentam espectros de difícil interpretação, mas que podem ser utilizados para caracterização de doenças como HLB (e.g. (37–39)).

Existem dois tipos de sistemas para realização da espectroscopia de infravermelho: os espectrômetros dispersivos e os não dispersivos (FTIR). Os espectrômetros dispersivos apresentam em geral feixe de radiação duplo, onde um deles passa pela amostra e outro pela referência. O funcionamento depende do monocromador para enviar luz monocromática para incidência na amostra e captação no detector. Esse sistema apresenta uma sensibilidade limitada, depende de uma grande precisão no alinhamento de todo o sistema ótico e é lento nas medidas. O espectrômetro mais recente não dispersivo, conhecido como FTIR, faz uso de um interferograma para a composição do espectro, tornando o sistema mais rápido e sem necessidade do uso de monocromadores. Esse tipo de sistema mede todas as frequências simultaneamente. A Fig. 3.6 apresenta o esquema típico de um interferômetro de Michelson, contido em um sistema FTIR.

A base do funcionamento do interferômetro consiste na geração de um feixe de luz modulado, i.e, com intensidade variável. A radiação policromática de IR penetra no interferômetro e

Figura 3.6 – Esquema do interferômetro de Michelson contido em um sistema FTIR.

é separada em um *beamsplitter* (divisor de feixes). O divisor possui um determinado material e espessura de modo que a reflexão e transmissão apresentem a mesma intensidade. Um dos feixes é transmitido e atinge um espelho fixo. O outro feixe é refletido e atinge um espelho móvel, que é responsável por introduzir a diferença de caminho ótico. Os dois feixes são refletidos e se recombinam no divisor de feixes, onde ocorrem as interferências. Para uma dada diferença de caminho ótico, a interferência é construtiva para algumas frequências e destrutiva para outras. Como o caminho ótico se altera constantemente, é criada uma onda modulada, i.e, um feixe com intensidade flutuante. O feixe modulado passa pela amostra e segue até o detector, sendo que ele contém toda a informação espectral dentro da faixa de radiação utilizada. O registro desse feixe é mais conhecido como interferograma, que nada mais é do que a relação entre a intensidade de energia em função da diferença de caminho ótico. A expressão matemática da intensidade do interferograma F é dada por

$$F(x) = \int_0^{+\infty} I(\tilde{\nu}) \cdot \cos(2\pi\tilde{\nu}x) d\tilde{\nu}. \quad (3.4)$$

Nessa expressão, F é uma função da distância x do espelho móvel a uma determinada

referência fixa. O interferograma depende basicamente da intensidade de uma determinada frequência e do fator de defasagem das vibrações luminosas. Como dito anteriormente, o interferograma contém toda informação na faixa de luz aplicada. Entretanto, para uma melhor visualização e entendimento, o interferograma é convertido para um espectro aplicando a transformada de Fourier:

$$I(\tilde{\nu}) = \int_{-\infty}^{+\infty} F(x) \cdot \cos(2\pi\tilde{\nu}x) dx \quad (3.5)$$

onde $I(\nu)$ é a transformada de Fourier de $F(x)$ e corresponde a intensidade de luz que atinge o detector em função do número de onda. A Fig. 3.7 mostra a transformação de um interferograma para o espectro de frequências.

Figura 3.7 – Passagem de um interferograma para o espectro de frequências utilizando a transformada de Fourier. Note que o espectro dado corresponde à absorbância.

Vemos na equação 3.5 que o espectro é obtido com a integração da posição variando de menos a mais infinito. Entretanto, dado que a posição x do espelho varia de posições finitas, a transformada de Fourier é calculada de maneira aproximada, utilizando a transformada de Fourier discreta ou a transformação rápida de Fourier. Este último é mais comum devido a maior rapidez nos cálculos.

O sistema FTIR utilizado nesse trabalho terá o seu esquema de medidas e seus parâmetros melhor descritos na seção de *materiais e métodos*.

3.2 Análise multivariada

A análise multivariada é caracterizada pelo uso de métodos para a análise simultânea de muitos dados na extração de conhecimento (43, 44). Nesta seção será apresentada uma breve introdução à regressão por mínimos quadrados parciais, ferramenta utilizada neste trabalho.

3.2.1 Regressão por mínimos quadrados parciais

O método de regressão por mínimos quadrados parciais, do inglês *partial least squares regression* (PLSR), é um método de análise multivariada que se caracteriza pela transformação linear de um conjunto de dados. A técnica é relativamente recente, tendo sido criada na década de 70 por S. Wold (45). O PLSR é comumente utilizado na quimiometria (46) para determinação da concentração de compostos químicos nas amostras através de espectros de emissão, absorção, reflectância ou transmitância. O PLSR se tornou uma ferramenta promissora nesses casos por possibilitar o tratamento de grandes quantidades de dados com ruído e altamente correlacionados (46).

A metodologia do PLSR generaliza e combina elementos da regressão linear múltipla, do inglês *multiple linear regression* (MLR), e da análise de componentes principais, do inglês *principal component analysis* (PCA). Maiores detalhes sobre os métodos MLR e PCA podem ser encontrados no apêndice A. O método PLSR encontra uma transformação linear entre as variáveis de dados de entrada (e.g. pontos espectrais, como nesse trabalho) que determina a melhor correlação com as variáveis de resposta (e.g. concentração dos compostos de referência).

O primeiro passo do PLSR consiste em decompor a matriz de dados de entrada $\mathbf{X}$, contendo n observações nas linhas e m variáveis nas colunas, em um produto entre um conjunto de fatores ortogonais $\mathbf{T}$ e um conjunto de coeficientes $\mathbf{P}$, acrescido de uma matriz de resíduos $\mathbf{E}$

$$\mathbf{X} = \sum_{h=1}^a \mathbf{t}_h \mathbf{p}_h^T + \mathbf{E} = \mathbf{TP}^T + \mathbf{E}. \quad (3.6)$$

A matriz de dados de entrada agora é representada pela soma dos produtos entre os vetores de *scores* $\mathbf{t}$ e de *loadings* $\mathbf{p}$, nomes dados por analogia ao PCA. Os *scores* são as projeções dos dados nas novas componentes e os *loadings*, os coeficientes da combinação linear das variáveis antigas. A soma é realizada de 1 até a , sendo a o número de variáveis latentes ou componentes. Com isso, a dimensionalidade dos dados é dada pelo número de variáveis latentes necessárias para modelar os *scores*. As matrizes $\mathbf{T}$ e $\mathbf{P}$ correspondem à junção dos vetores $\mathbf{t}$ e $\mathbf{p}$, respectivamente, e a matriz de resíduos $\mathbf{E}$, à parte não modelada de $\mathbf{X}$. A representação gráfica das matrizes com suas dimensões pode ser visualizada na Fig. 3.8. Essa nova descrição da matriz garante uma representação com um número reduzido de variáveis sem perdas significativas de informação.

Figura 3.8 – Representação gráfica das matrizes e vetores do método PLSR

A primeira equação descrita acima corresponde exatamente à do método PCA, que tem por objetivo projetar os dados em novas componentes que fornecem a maior variância dos dados. Entretanto, o método PLSR estabelece e encontra as componentes que apresentam a maior correlação com as respostas. A grande vantagem do método é a decomposição do vetor de respostas $\mathbf{y}$, seguindo o mesmo princípio da matriz de dados de entrada:

$$\mathbf{y} = \mathbf{Tc}^T + \mathbf{f}, \quad (3.7)$$

Para facilitar a visualização das dimensões das variáveis da equação, a Fig. 3.9 mostra a representação gráfica das matrizes e vetores na decomposição de $\mathbf{y}$. No caso do PLSR, o vetor

y é descrito em função dos mesmos *scores* da matriz de dados X . O vetor de *loadings* de y , mais conhecido como coeficientes de regressão, são representados pelo vetor c e o vetor de resíduos por f . A decomposição da matriz de dados e do vetor de respostas é realizado de forma conjunta para garantir que as componentes calculadas apresentem alta correlação com o vetor de respostas y . O método PLSR gera um modelo de regressão tal que os resíduos sejam os menores possíveis. O algoritmo do método PLSR está descrito no apêndice B.

$$\begin{array}{c} 1 \\ \boxed{y} \\ n \end{array} = \begin{array}{c} a \\ \boxed{T} \\ n \end{array} \begin{array}{c} 1 \\ \boxed{c^T} \\ a \end{array} + \begin{array}{c} 1 \\ \boxed{f} \\ n \end{array}$$

Figura 3.9 – Representação gráfica das matrizes e vetores do método PLSR

A vantagem dessa técnica reside na redução de variáveis, usando a correlação das variáveis de entrada para melhor estimar as variáveis de resposta. Neste trabalho, o método PLSR foi utilizado para identificar as folhas em quatro diferentes classes: sadia, HLB-sintomática, HLB-assintomática ou CVC. Como o PLSR é um método de regressão, é necessário uma adaptação do método para problemas de classificação, de modo a transformar as classes em números. Essa metodologia foi proposta por Frank et al. (3). O funcionamento do método e do classificador induzido via regressão será melhor descrito na seção de *materiais e métodos*.

4 Objetivos

Este trabalho teve como objetivo geral avaliar o diagnóstico de HLB através das técnicas fotônicas LIFS e FTIR e verificar a hipótese de que a doença induz alterações físicas e químicas na planta. A caracterização ótica de carboidratos e metabólitos secundários presentes em folhas de citros foi realizada para o entendimento dos espectros foliares. Os objetivos específicos deste trabalho foram:

1. Avaliar o melhor comprimento de onda de excitação para emissão de fluorescência nas folhas de citros, de modo que possam evidenciar melhores diferenças entre folhas de plantas saudáveis, infectadas com HLB (folhas sintomáticas e assintomáticas) e CVC.
2. Avaliar o potencial de LIFS na discriminação de folhas saudáveis, HLB-sintomáticas, HLB-assintomáticas e CVC, coletadas de árvores de campo.
3. Relacionar os espectros de LIFS dos metabólitos secundários com os espectros de folhas de citros.
4. Avaliar o potencial de FTIR na discriminação de folhas saudáveis, HLB-sintomáticas, HLB-assintomáticas e CVC, coletadas de árvores de campo.
5. Relacionar os espectros de FTIR das folhas de citros com os espectros de absorção de carboidratos e de metabólitos secundários.

5 Materiais e Métodos

Nesse capítulo serão descritas as amostras e apresentadas as técnicas e ferramentas utilizadas neste trabalho. Na seção 5.1 serão descritas detalhadamente a forma de coleta das folhas, quais árvores foram utilizadas, seu preparo para a realização dos experimentos e o seu acondicionamento durante as medidas. Além disso, será descrita a preparação das amostras de metabólitos de plantas para análises com as técnicas espectroscópicas para comparação com os espectros de folhas. Na seção *Técnicas fotônicas e Tratamento dos espectros* serão apresentados os parâmetros, formas de medida e pré-processamento dos espectros. A seção *Metodologia de Classificação* apresenta todo modelo de classificação empregado e seus parâmetros.

5.1 Amostras

5.1.1 Folhas de citros

Folhas de citros foram coletadas de árvores cultivadas em campo, enviadas pelo Grupo Fisher e oriundas da Fazenda Citrícola, situada na cidade Matão/SP. Laranjeiras adultas da variedade Valência enxertadas em citrumelo Swingle foram utilizadas no experimento. Três conjuntos de árvores foram avaliados no campo para extração de folhas: árvores saudáveis, árvores infectadas com HLB e com CVC. As plantas saudáveis utilizadas se encontravam livres de qualquer deficiência nutricional e em talhões de baixa infestação de HLB e CVC.

Oito plantas de cada caso, i.e., sadia, infectada com HLB e CVC foram utilizadas no ex-

perimento. A coleta das folhas foi realizada por uma equipe de inspetores treinados em um único dia. Folhas HLB-sintomáticas e HLB-assintomáticas foram coletadas de árvores com sintomas de HLB. As folhas sintomáticas foram retiradas de galhos sintomáticos e as folhas assintomáticas de diferentes galhos ao redor da árvore sem sintomas e em regiões externas da copa à meia altura. Todas as folhas de plantas com CVC foram coletadas de galhos sintomáticos e apresentando sintomas. As folhas saudáveis foram coletadas em diferentes galhos ao redor da árvore e na parte externa da copa à meia altura. Neste trabalho, cada uma das condições das folhas coletadas: sadia, CVC, HLB-assintomática e HLB-sintomática, serão chamadas de classe. Um total de 40 folhas foram coletadas de cada classe. A Fig. 5.1 exibe uma folha típica de cada classe utilizada no experimento.

Figura 5.1 – Exemplos típicos de folhas CVC (a), sadia (b), HLB-sintomática (c) e HLB-assintomática (d). Figura extraída de Cardinali et al. (2).

Todas as folhas utilizadas foram limpas com algodão umedecido em água destilada e mantidas em geladeira a aproximadamente 4° C. Todos os experimentos foram realizados em até quatro dias após a coleta das folhas para evitar a degradação da amostra. A tabela 5.1 exibe o número de folhas utilizadas em cada uma das técnicas para atingir os objetivos propostos. Para verificação do melhor comprimento de onda de excitação, foram utilizadas 15 folhas de cada classe. Para a obtenção dos espectros nos dois sistemas portáteis de LIFS e no FTIR, foram utilizadas as mesmas 40 folhas de cada classe coletadas.

Tabela 5.1 – Número de folhas de cada planta utilizadas para a realização das medidas de cada uma das técnicas.

Tipos de folhas	Fluorescência convencional	LIFS exc. em 405 e em 561 nm	FTIR
Sadia	15	40	40
HLB-assintomática	15	40	40
HLB-sintomática	15	40	40
CVC	15	40	40

5.1.2 Carboidratos e metabólitos secundários

Amostras padrão dos carboidratos amido, glicose e sacarose e amostras comerciais de metabólitos secundários de citros naringina, hesperidina, da classe flavonóides, e, umbeliferona, pertencente à classe das cumarinas, foram preparadas em pastilhas de brometo de potássio (KBr) para medidas no FTIR. Em relação ao grau de pureza dos metabólitos secundários utilizados, a hesperidina e naringina era acima de 90% e a umbeliferona acima de 99%. As pastilhas de KBr foram feitas com uma quantidade de 100 mg de KBr em pó seco para cada 1 mg de cada uma das amostras. As misturas foram trituradas em um almofariz e pistilo e em seguida inseridas em um pastilhador da Perkin-Elmer. Para cada uma delas, o pastilhador foi submetido a uma pressão de 5 toneladas durante 5 minutos para a formação da pastilha. Uma pastilha de KBr puro também foi feita com o mesmo procedimento.

Para medidas na fluorescência convencional e 3D, os flavonóides e a cumarina foram também preparadas em soluções aquosas. As soluções foram preparadas em diferentes concentrações, de acordo com a máxima diluição de cada composto na água e com a emissão abaixo do limite do detector. Diferentes massas de cada elemento foram trituradas em um almofariz e pistilo e pesadas em uma balança de alta precisão, sendo posteriormente diluídas em água deionizada. Para os flavonóides, 4 mg e 2 mg de hesperidina e naringina, respectivamente, foram diluídas em balões volumétricos de 100 mL, para a obtenção de 40 ppm e 20 ppm, respectivamente. Para a cumarina umbeliferona, 1 mg de massa foi diluída em um balão volumétrico de 500 mL. Nesse caso foi necessário o uso de um ultra-som para uma melhor diluição. Em seguida, através de

uma pipeta de 1 mL, foi coletado uma alíquota de 0,25 mL e diluído em um balão volumétrico de 50 mL para a obtenção de uma solução de 10 ppb. As soluções prontas foram mantidas isoladas de radiação eletromagnética para evitar qualquer tipo de degradação do material.

5.2 Técnicas fotônicas

5.2.1 Fluorescência convencional e 3D

Medidas de folhas de citros foram realizadas no espectrofluorímetro RF-5301 da Shimadzu, cujo esquema simplificado foi exibido na figura 3.3. O equipamento consiste basicamente de uma lâmpada de excitação de Xenônio de 150 W, monocromadores de excitação e emissão, fotomultiplicadores (PMT) para detecção da emissão da amostra e para referência, que mede a intensidade e a fase da luz incidente. O equipamento ainda permite a inserção de filtros no módulo ótico (após a amostra) para eliminação da luz de excitação e suas ordens de difração superiores. Para o controle dos parâmetros e coleta dos espectros foi utilizado o software RF-5301PC.

Os parâmetros do sistema ajustáveis são: modo de aquisição, comprimento de onda de excitação ou emissão, faixa de comprimentos de onda, abertura da fenda do monocromador de excitação e do monocromador de emissão, intervalo de leitura, velocidade de escaneamento, sensibilidade do fotomultiplicador e tempo de resposta. O modo de aquisição permite a escolha do tipo de operação do equipamento, i.e., aquisição do espectro de emissão (excitação da luz fixa e varredura da emissão da amostra) ou do espectro de excitação (emissão em um comprimento de onda fixo e varredura da excitação). O intervalo de leitura corresponde à faixa espectral que o espectro será coletado. A sensibilidade do fotomultiplicador e o tempo de resposta correspondem a parâmetros de intensidade e tempo da detecção da luz, respectivamente.

A obtenção dos espectros das folhas foi realizada com os seguintes parâmetros: modo de aquisição em *emissão*, comprimento de onda de excitação variável de *400 a 600 nm*, intervalo de leitura de *400 a 900 nm*, abertura das fendas do monocromador de excitação e de emissão de *5 e 10 nm*, respectivamente, velocidade de escaneamento *média* (aproximadamente de *500 nm/min*), intervalo de leitura de *0.2 nm*, *alta sensibilidade* e tempo de resposta *automático*. Foi utilizado um filtro da *Melles Griot* com transmitância de 75% acima de 470 nm e 90% acima de 500 nm, de modo a remover os picos de excitação abaixo de 450 nm e suas segundas ordens de difração.

As medidas dos metabólitos secundários em solução aquosa foram realizadas em um espectrofluorímetro de luminescência LS50B da Perkin-Elmer, equipamento com montagem similar ao da figura 3.3. Os parâmetros ajustáveis desse sistema são um pouco diferentes do sistema da Shimadzu. A sensibilidade e o tempo de resposta do fotomultiplicador é definida através da escolha do modo de luminescência (no nosso experimento, fluorescência). Outros recursos diferentes fornecem a possibilidade de realizar varredura em 3D (varredura emissão–excitação) de forma automática e escolher filtros e polarizadores já embutidos no equipamento. O software de aquisição dos espectros foi o FL Winlab versão 3.0.

Para cada metabólito foi realizada uma varredura de excitação e emissão 3D, fenda de excitação e emissão com abertura de *10 nm*, velocidade de escaneamento em *500 nm/min*. O intervalo de leitura da excitação foi de *200 a 600 nm* e o de emissão foi de *300 a 900 nm*. Um filtro de emissão foi utilizado com transmitância de 95% acima de 290 nm, a fim de eliminar os picos de excitação inferiores a 290 nm e suas respectivas ordens de difração. Foram realizadas medidas espectrais da mesma água deionizada utilizada como solvente para garantir que a mesma se encontrava livre de impurezas que pudessem afetar os espectros.

5.2.2 LIFS

Para as medidas foram utilizados dois sistemas de LIFS portáteis desenvolvidos no Laboratório de Ótica e Lasers da Embrapa Instrumentação. O esquema de ambos os sistemas se encontra descrito na Fig. 5.2. Ambos os sistemas desenvolvidos possuem basicamente o mesmo aparato, com diferença nos lasers de excitação e nos espectrômetros. Os sistemas possuem um laser de diodo Coherent, sendo um modelo CUBE com excitação em 405 nm e o outro modelo COMPASS com excitação em 561 nm e um mini-espectrômetro USB2000–Ocean Optics de alta sensibilidade com intervalo de 194 a 894 nm no sistema de excitação em 405 nm e de 500 a 1200 nm no de excitação em 561 nm. Um esquema do mini-espectrômetro foi exibido na Fig. 3.4. Os demais constituintes consistem de um conjunto de sete fibras ópticas fixas em uma ponteira de aço inoxidável – seis fibras de iluminação em volta de uma fibra de leitura da Ocean Optics, um filtro ótico ajustável para barrar as respectivas reflectâncias do laser e um notebook com software desenvolvido para aquisição e tratamento de dados. Por simplicidade, vamos nos referir a cada um dos sistemas como LIFS-405 e LIFS-561, em referência aos diferentes lasers de excitação. A Fig. 5.3 exhibe o sistema portátil LIFS-405 desenvolvido.

Figura 5.2 – Esquema do sistema de LIFS portátil desenvolvido no laboratório de Ótica e Lasers da Embrapa Instrumentação.

Figura 5.3 – Sistema de LIFS-405 portátil.

Os parâmetros do espectrômetro passíveis de ajustes são: *tempo de integração*, *número de médias* e *boxcar*. O *tempo de integração* consiste no tempo em que o espectrômetro fica captando fótons emitidos pela amostra antes da próxima coleta. O *número de médias* consiste na coleta de vários espectros para a obtenção de um espectro médio para diminuição de ruído. O *boxcar* é um parâmetro de suavização, no qual é realizada uma média da vizinhança de cada ponto do espectro. O valor atribuído ao *boxcar* define o número de vizinhos, de cada lado do ponto a ser levado em consideração na suavização.

As medidas foram realizadas nas folhas frescas coletadas, na parte inferior da face abaxial, evitando-se a nervura central. Para o sistema LIFS-405, utilizaram-se os seguintes parâmetros de aquisição: tempo de integração de 60 ms, 20 médias espectrais e boxcar 2. Para o sistema LIFS-561, os parâmetros foram: tempo de integração de 2 ms, 20 médias espectrais e boxcar 2. Para cada medida, a coleta do espectro foi realizada após um intervalo de tempo de 2,5 s, tempo suficiente para a intensidade ter atingido seu valor máximo. A sala de medidas foi mantida a temperaturas abaixo de 23°.

5.2.3 FTIR

Foi utilizado nos experimentos o equipamento da Perkin-Elmer Spectrum 1000, cujo esquema básico é ilustrado na figura 5.4. O sistema consiste de uma fonte de IR, interferômetro, laser, detector feito de cristal de LiTaO_3 (Tantalato de lítio) e um software de aquisição Spectrum versão 5.3. O laser do equipamento tem por função monitorar a posição do espelho móvel durante a leitura dos espectros e guiar a posição de colocação de amostra para o usuário.

Figura 5.4 – Esquema de um espectrômetro FTIR. Figura adaptada disponível em: <<http://mmrc.caltech.edu/FTIR/FTIRintro.pdf>>. Acesso em: 22/01/2012.

Os parâmetros ajustáveis do sistema consistem no intervalo e resolução espectral, intervalo entre pontos e número de médias. Em todas as medidas foram feitas 32 médias no intervalo espectral de 4000 a 700 cm^{-1} , com resolução espectral de 4 cm^{-1} e intervalo entre pontos de 1 cm^{-1} . A sala foi mantida em temperaturas entre 20° e 22° C e com umidade relativa do ar entre 35% a 40%. Espectros de fundo foram coletados antes e durante as medidas das amostras para remover as características atmosféricas e do equipamento.

Para as medidas das folhas de citros frescas coletadas, foi utilizado um cristal de ATR, sigla

do inglês *attenuated total reflection* (reflexão total atenuada), como o representado na Fig. 5.5. O ATR é feito de Seleneto de Zinco (ZnSe) e possui alto índice de refração. Sua função é promover múltiplas reflexões do feixe de IR dentro do cristal que se encontra em contato com a amostra. Nos pontos de reflexão, parte da onda penetra na amostra, formando uma onda evanescente. A radiação penetra alguns microns na amostra (0,5 a $5\mu\text{m}$) e absorve parte da radiação IR, modificando o feixe que é transmitido para o detector. O uso do ATR garante medidas rápidas sem necessidade de preparo de amostras.

Figura 5.5 – Esquema de um cristal de ATR. Nos pontos de reflexão do cristal em contato com a amostra, as amostras absorvem parte da radiação modificando o feixe transmitido para o detector.

As medidas foram realizadas nas folhas frescas coletadas, na face adaxial evitando-se a nervura central. A região de medida da folha foi pressionada sobre o cristal de ATR para medir o espectro de superfície. Entre cada medida de folha, o cristal de ATR foi limpo com algodão umedecido em acetona. A cada cinco folhas medidas, foi coletado um espectro de fundo, i.e, sem a presença da folha para eliminar as influências atmosféricas e do equipamento. As medidas foram feitas fora de ordem para garantir a diferenciação entre elas nos classificadores.

Para as medidas com pastilhas de KBr preparadas com carboidratos e metabólitos secundários, foi utilizado um porta pastilhas para utilizar o método da transmitância. A radiação infravermelha é incidida sobre a pastilha, onde parte dela é absorvida e outra parte é transmitida. A radiação transmitida segue para a leitura no detector, sendo possível assim saber a parcela de radiação que foi absorvida pela amostra. Para tanto, entre cada medida das pastilhas, foi realizada a coleta de um espectro de fundo com a pastilha de KBr puro para remoção da absorção do KBr e das características atmosféricas e do equipamento.

5.3 Tratamento dos espectros

5.3.1 Fluorescência convencional e 3D

Para tratamento dos espectros de emissão das folhas, primeiramente foi realizada a escolha da região espectral de análise. A região de cada espectro foi selecionada de modo a retirar o pico de reflectância da lâmpada e regiões sem emissão. Em seguida, cada espectro teve sua linha base corrigida, subtraindo a menor intensidade no intervalo. Por fim, os espectros foram normalizados pela área para avaliação do perfil espectral e das relações entre os picos.

Nos espectros de fluorescência 3D, a linha base foi corrigida subtraindo o menor valor de intensidade do espectro. O espectro foi avaliado nas regiões de emissão entre 300 e 900 nm e de excitação entre 200 e 500 nm. A partir disso, foi possível a construção de um gráfico 3D: comprimento de onda de emissão x comprimento de onda de excitação x intensidade. Todos os tratamentos foram realizados no software livre *Scilab 5.3*.

5.3.2 LIFS

Todos os espectros de LIFS foram pré-tratados corrigindo a linha base. No sistema LIFS-561, a linha base foi corrigida subtraindo a média de intensidades no intervalo compreendido entre 1050 a 1070 nm e o intervalo espectral de análise foi de 600 a 1140 nm. No sistema LIFS-405, a linha base foi corrigida subtraindo a média de intensidades entre 425 e 430 nm e o intervalo espectral de análise foi entre 360 e 892 nm. Os espectros foram finalmente normalizados pela área para enfatizar as diferenças locais entre os picos e as bandas (perfil espectral) e diminuir a influência da intensidade do espectro. Por fim, os valores de intensidades relativas serviram de entrada para o treinamento e validação dos classificadores, totalizando 1593 e 1678

pontos para os sistemas LIFS-405 e LIFS-561, respectivamente. Todos os tratamentos foram realizados no software livre *Scilab 5.3* e depois implementados no software do equipamento de LIFS.

5.3.3 FTIR

O mesmo procedimento foi realizado para os espectros das folhas e das pastilhas de KBr. O primeiro tratamento realizado nos espectros foi a correção da linha base usando uma função quadrática. Esse tratamento foi realizado no software *Spectrum 5.1* do equipamento da Perkin Elmer. Em seguida, os espectros foram exportados para o software livre *Scilab 5.3* para os demais ajustes. A região de análise foi de 1530 a 700 cm^{-1} , removendo a região de bandas vibracionais da água e do gás carbônico. Nessa região, a linha base foi ajustada para o valor mínimo de absorção em zero e os espectros foram normalizados pela área.

5.4 Método de classificação

Os espectros das folhas de citros adquiridos e pré-tratados nas técnicas LIFS e FTIR, foram exportados para o software livre WEKA (47) para indução do classificador via regressão por mínimos quadrados parciais (PLSR). Embora esse método seja geralmente aplicado em quimiometria (46) para determinar a concentração de compostos químicos, o PLSR pode ser aplicado em problemas de classificação transformando as classes em números, metodologia proposta por Frank et al. (3). O esquema simplificado do método é ilustrado na Fig. 5.6.

O classificador induzido via regressão associa as classes nominais, i.e. sadia, CVC, HLB-sintomática e HLB-assintomática, à números, usando o processo de binarização. Para iniciar o processo, é necessário um conjunto de treino para calibração do modelo de regressão e um conjunto de teste para validação do modelo gerado. Um conjunto de dados de treino igual ao

Figura 5.6 – Diagrama ilustrando o algoritmo para uso de um método de regressão em um problema de classificação. A figura foi baseada no diagrama feito por Frank et al. (3).

original é criado e para uma determinada classe é atribuído o valor 1 e 0 para as demais. O conjunto de treino é usado para o ajuste do método de regressão. Com o modelo ajustado, um segundo conjunto, chamado de teste, valida o modelo criado. Para cada elemento do conjunto de teste, o modelo ajustado retorna valores entre 0 e 1. Quanto maior o valor, maior é a similaridade entre o espectro de teste e a classe de referência. O mesmo procedimento é repetido para todas as classes usando o mesmo conjunto de treino e teste. Assim, um valor entre 0 e 1 é obtido para cada amostra de teste. A amostra testada é classificada de acordo com o maior valor

entre 0 e 1 retornado pelos métodos de regressão. Para cada amostra testada, é possível avaliar a probabilidade de predição para cada uma das classes, que é proporcional ao valor retornado pelos modelos de regressão. Essa probabilidade pode ser usada como um indicador da qualidade da classificação realizada.

Para avaliação do conjunto de dados no classificador PLSR, a separação em conjunto de treino e teste foi realizada através do método da validação cruzada. Esse método é amplamente utilizado para avaliar a capacidade de generalização do modelo de predição. A validação cruzada realiza a separação do conjunto de dados em n folds ou grupos, utilizando $n - 1$ deles para o treino do classificador e o grupo restante para teste. Em seguida, o grupo deixado de lado é reintegrado ao grupo e outro é retirado. Esse procedimento é repetido até que todos os grupos tenham sido treinados e testados. Para uma maior generalização do modelo, é possível repetir o procedimento de validação cruzada aleatorizando as amostras que compõe cada grupo. Cada procedimento da validação cruzada leva a diferentes acertos e erros, possibilitando uma estimativa melhor do seu real funcionamento. Assim, realizando k execuções da validação cruzada temos no final do processo kn resultados.

Para início da classificação, cada atributo do espectro (intensidade) foi centrado na média para facilitar os cálculos. A classificação do conjunto de dados de espectros foi realizada com 10 execuções de uma validação cruzada randômica com 10-folds. O mesmo procedimento ainda foi repetido variando o número de componentes do PLSR, mantendo fixo os mesmos subconjuntos de cada execução de validação cruzada para comparações de acurácia. A escolha do número de componentes é feita com aquele que apresenta a maior taxa de acerto e a menor raiz do erro quadrático médio, do inglês root mean squared error (RMSE). A taxa de acerto é obtida através da razão entre o número de amostras corretas e o número total de amostras. O RMSE é calculado através da soma dos quadrados das diferenças entre os valores esperados y_i e os valores preditos $\hat{y}_i$ de cada instância i , como demonstra a equação 5.1:

$$RMSE = \sqrt{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots + (y_n - \hat{y}_n)^2} \quad (5.1)$$

onde n é o número total de predições realizadas. O RMSE é a ferramenta mais comumente utilizada nas medidas de erro das análises de performance de um classificador, devido a sua simplicidade nos cálculos e de apresentar a mesma dimensão dos valores preditos.

O interesse aqui é encontrar o número de componentes do PLSR que oferece a melhor taxa de acerto média e menor RMSE comparado à outros esquemas. Nem sempre essa tarefa é simples, pois o que temos em geral é uma estimativa da média e desvio padrão de cada esquema através das amostras e não da população como um todo. Os valores de média amostral obtidos em cada esquema podem apresentar uma proximidade de tal forma que apenas um teste de significância pode descobrir se um esquema de classificação é melhor que outro ou se são equivalentes. Esse trabalho pode ser realizado pelo método estatístico conhecido como teste t de Student, ou simplesmente teste t (43, 48).

O teste t é uma ferramenta estatística comumente utilizada nas análises de significância, onde o interesse é rejeitar uma hipótese nula e aceitar uma hipótese alternativa. A hipótese nula é referente ao que se quer comprovar que não é verdadeiro e a hipótese alternativa é o que se quer aceitar como verdadeiro dentro de um intervalo de confiança. No caso aqui presente, a hipótese nula seria propor que não há diferença significativa entre os resultados obtidos entre dois métodos. A hipótese alternativa é a de que as diferenças existem e que usar um método produz resultados diferentes (melhor ou pior), dentro do intervalo de confiança escolhido. Para o uso dessa ferramenta, os objetos de teste devem seguir a distribuição t de Student, o que ocorre naturalmente quando são coletadas amostras de uma população que segue a distribuição normal. A diferença entre essas distribuições é que a distribuição t de Student é mais alargada nas bordas, aumentando a probabilidade de apresentar valores mais extremos. A distribuição t de Student depende do número de graus de liberdade das amostras dado por $\nu = n - 1$, onde n é o número de amostras. Quanto maior for n , mais a distribuição se aproxima da distribuição normal (ver Fig. 5.7).

A teoria e as inúmeras aplicações do teste t de Student são amplas e fogem do foco deste trabalho. Sendo assim, neste trabalho será ilustrado a aplicação em dados pareados, onde o

Figura 5.7 – Distribuição normal padronizada (DNP) em comparação com a distribuição t de Student para os graus de liberdade $\nu = 1$ e $\nu = 10$. Quanto maior for o número de graus de liberdade, mais a distribuição se aproxima da DNP.

método é conhecido com teste t pareado. Para um entendimento mais básico e exemplos de outras aplicações segue a referência (48). Vamos considerar que $x_1, x_2, \dots, x_n$ e $y_1, y_2, \dots, y_n$ são as amostras obtidas por dois diferentes esquemas de classificação sob as sucessivas validações cruzadas. Essas amostras podem ser tanto a taxa de acerto, RMSE ou outro parâmetro desejado. Sendo a média das amostras denotadas por $\bar{x}$ e $\bar{y}$, queremos saber se x é significativamente diferente de y . Como sabemos que x e y seguem a distribuição t de Student, nós vamos utilizar a diferença d_i entre cada uma das amostras, onde $d_i = x_i - y_i$. Além disso, isso pode ser feito porque as amostras são pareadas, i.e., as partições realizadas nas validações cruzadas são as mesmas para ambos esquemas. No teste t pareado, é possível calcular o grau de desvio da média amostral $\bar{d}$ de tamanho n da média da população μ_d através do valor de t ,

$$t = \frac{\bar{d} - \mu_d}{\sqrt{s_d^2/n}} \quad (5.2)$$

onde s_d é o desvio amostral dado por

$$s_d^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2.$$

A média amostral da diferença é dada por $\bar{d} = \bar{x} - \bar{y}$ e a média da população μ_d é estabelecida pela hipótese nula. Nesse caso, a hipótese nula é a de que não há diferenças entre x e y , e portanto, $\mu_d = 0$. Dessa forma, o teste t pareado assume a forma:

$$t = \frac{\bar{d}}{\sqrt{s_d^2/n}} \quad (5.3)$$

A partir do valor de t calculado na amostra, é possível comparar com os valores de t tabelados da distribuição t de Student para dizer com um nível de confiança que a hipótese nula pode ser rejeitada. Para a leitura da tabela é necessário o número de graus de liberdade e o nível de confiança, que é geralmente 95% ou 99%. Se $|t|$ calculado for maior que o valor de t tabelado, a hipótese nula pode ser rejeitada com o nível de confiança escolhido. Se $|t|$ for menor que o valor tabelado, não se pode rejeitar a hipótese nula. Dado o vasto uso dessa ferramenta em análise de classificadores, o teste t já se encontra implementado no software WEKA para uso direto nos resultados de classificação. Para maiores informações de sua aplicação na avaliação de classificadores, segue a referência do livro texto de Witten et al. (43).

Após a escolha do número de componentes, as matrizes de confusão foram extraídas a partir dos seus resultados de classificação. Além disso, foi possível montar um histograma acumulado com as probabilidades de predição.

6 Resultados

6.1 Fluorescência

6.1.1 Varredura de excitação

As figuras 6.1 e 6.2 exibem os perfis espectrais médios de emissão de cada uma das classes de folhas, i.e., Sadia, HLB-sintomática, HLB-assintomática e CVC, para excitações em 560 nm e 400 nm, respectivamente.

Na excitação com comprimento de onda em 560 nm (Fig. 6.1), é possível observar duas bandas de emissão no vermelho–infravermelho com máximos em 685 e 735 nm. Segundo Buschmann (25), essas bandas são associadas em grande parte à emissão da clorofila *a* e em menor parte a estrutura da folha, atividade fotossintética e propriedades ópticas da folha. Em relação às diferenças espectrais entre as classes, nota-se que a HLB-sintomática possui o espectro mais diferenciado. É observado um deslocamento lateral nas duas bandas da clorofila e uma relação entre os seus máximos diferente em relação às outras classes. As demais apresentam espectros médios muito semelhantes entre si.

Na excitação das folhas com comprimento de onda em 400 nm (Fig. 6.2), observamos com mais detalhes as bandas na mesma região das mostradas no espectro da Fig. 6.1 e uma banda na região entre 450 e 650 nm (azul–verde) associada em grande parte aos ácidos ferúlicos e em menor parte a outros compostos, tais como cumarinas, flavonóides, entre outros (34). Visualmente, as diferenças espectrais entre as classes são mais acentuadas na excitação com

Figura 6.1 – Espectro de emissão de fluorescência médio das folhas com excitação em 560 nm. As amostras HLB-sintomáticas apresentam uma relação entre os picos 685 e 735 nm diferente em relação as demais classes.

400 nm do que com a excitação em 560 nm. Através da relação entre os máximos em 685 e 735 nm é possível distinguir HLB-sintomática e CVC das demais. Além disso, a região azul-verde ainda pode contribuir bastante, pois a relação entre as intensidades relativas em 550 e 685 nm diferenciam largamente HLB-sintomática de CVC-sintomática.

As excitações intermediárias realizadas entre 560 nm e 400 nm apresentaram uma transição entre os espectros observados. Foi possível observar a emissão azul-verde completa para excitações abaixo de 420 nm. Excitações na região do ultra-violeta não foram realizadas devido às lesões causadas nas folhas devido ao tempo de exposição com radiação ionizante.

Qualitativamente, é possível dizer que a excitação mais adequada na faixa do visível é a realizada entre 400 e 420 nm, pois evidencia emissões em praticamente todas as regiões do espectro, desde o visível até início do infravermelho e garante diferenças visuais mais acentuadas. Pereira et al. (17) mostrou que através das imagens de fluorescência com excitação em

Figura 6.2 – Espectro de emissão de fluorescência médio das folhas com excitação em 400 nm.

473 nm, a emissão na cor verde foi a que mais contribuiu para a discriminação entre as folhas doentes e saudáveis. A excitação com luz com comprimento de onda menor que esse privilegia uma emissão mais intensa no verde e ainda permite a emissão no azul. Para confirmação das afirmações tomadas aqui, foi possível utilizar dois sistemas portáteis de LIFS com excitações em 561 e 405 nm, onde foi feito um tratamento espectral para indução de classificadores e extração de resultados quantitativos.

6.1.2 LIFS-561

A figura 6.3 mostra os espectros típicos médios das folhas de citros medidas no sistema LIFS-561. Notamos que os espectros são ligeiramente diferentes dos espectros coletados no fluorímetro 6.1. As diferenças ocorrem devido a diferentes fatores, tais como o detector de CCD, a grade de difração e excitação via laser.

Figura 6.3 – Espectros médios de emissão de fluorescência das folhas de citros com excitação Laser em 561 nm

A fluorescência observada na excitação em 561 nm apresenta o perfil relacionado a concentração de clorofila *a*, da estrutura da folha, da eficiência de transmissão de energia, entre outros. Na Fig. 6.3 é possível observar que a razão entre os picos em 685 nm e 735 nm é ligeiramente diferente em todas as classes. Os espectros da HLB-sintomática e CVC-sintomática apresentam relação de picos invertida relativamente à das sadias. A HLB-assintomática tem um comportamento intermediário de transição. Essas diferenças demonstram que os estresses causados pelas doenças levam a alterações na concentração das clorofilas e possivelmente nas demais características associadas a eficiência da emissão, como já verificado em trabalhos anteriores (25, 34).

Para extrair a capacidade de discriminação das classes, os espectros foram pré-tratados e serviram de entrada para a indução do classificador via PLSR. Foram realizadas várias execuções de validação cruzada para a determinação do número de componentes que produz melhores resultados. As Figs. 6.4 e 6.5 exibem as taxas de acerto médias e o RMSE em função do número de componentes, respectivamente.

Figura 6.4 – Taxa de acerto do classificador em função do número de componentes utilizados no classificador via PLSR. A maior taxa de acerto pode ser observada em torno de 70% para 4 componentes e 9 componentes.

Figura 6.5 – RMSE do classificador em função do número de componentes utilizados no classificador via PLSR. Nota-se uma diminuição do erro até o uso de 10 componentes. O desvio padrão aumenta conforme aumenta-se o número de componentes.

Na Fig. 6.4 vemos que o uso de 4 e 9 componentes apresentam as maiores taxas de acerto. Embora o uso de 9 componentes apresente uma taxa de acerto média pouco maior do que para 4 componentes, estatisticamente ambos são equivalentes. Essa constatação pode ser observada na Fig. 6.4 ou de maneira melhor através do teste t de Student. Como descrito em materiais e métodos, nesse caso foi aplicado o teste t pareado, tomando como hipótese nula que cada

taxa de acerto comparada são equivalentes. Como era de se esperar, verificou-se que ambos são estatisticamente equivalentes com nível de confiança de 95%.

A Fig. 6.5 exibe a curva do RMSE em função do número de componentes. Verificou-se que o RMSE para 9 componentes apresentou um valor médio menor do que para 4 componentes, embora apresente um maior desvio padrão. Como era de se esperar, o desvio padrão do RMSE aumenta conforme aumenta o número de componentes. Esse fato é devido ao aumento do ruído e da correlação entre os dados. Utilizando o teste t pareado, foi possível verificar que usar 9 componentes confere um menor RMSE com nível de confiança de 95%. Entretanto, visualmente vemos que os erros não devem ser muito menores estatisticamente. O teste t apenas mostrou que o uso de 4 e 9 componentes conferem resultados diferentes.

A tabela 6.1 exibe a matriz de confusão com os resultados da classificação para 9 componentes. As colunas da matriz representam as classes preditas pelo classificador e as linhas às respectivas classes das folhas. Cada valor da tabela corresponde a porção de uma classe nominal que foi assinalada pelo classificador induzido para uma certa classe. A soma de cada linha é de 100%. Os elementos em destaque representam os acertos do classificador, sendo a taxa de acerto global de 69% e desvio padrão de 10%.

Tabela 6.1 – Matriz de confusão obtida na execução de 10 validações cruzadas com 10 folds nos espectros do LIFS-561. As colunas correspondem às predições de classes feitas pelo classificador e as linhas às classes nominais.

Classe nominal	CVC	Sadia	HLB-assint.	HLB-sint.
CVC	57	12,75	17,75	12,5
Sadia	8	76,5	15,5	0
HLB-assint.	9,5	31	59,5	0
HLB-sint.	14,5	0	2,75	82,75

A maior taxa de acerto foi de aproximadamente 83% e ocorreu para HLB-sintomática. Esse fato já era esperado devido ao espectro bastante diferenciado em relação aos demais (ver Fig. 6.3). Além da relação entre os picos de emissão ser visualmente diferente, a intensidade relativa da reflectância na superfície da folha, visualizada através do primeiro harmônico do laser (em 1122 nm) deve contribuir na diferenciação. A maior taxa de confusão ocorreu com a classe

CVC. As confusões entre HLB-sintomática e HLB-assintomática não devem ser consideradas um erro, já que ambas são as mesmas doenças em estágios diferentes. A segunda maior taxa de acerto foi referente a classe sadia, que apresentou taxa de acerto de aproximadamente 77%.

A alta taxa de acerto da HLB-sintomática deve estar relacionada à alteração na concentração da clorofila nas folhas. Zaghi (13) extraiu pigmentos de folhas sadias e com HLB e seus resultados sugeriram que a doença provocou uma diminuição na concentração, já que a intensidade da emissão da clorofila foi mais baixa. Nos espectros obtidos aqui foi observado um comportamento que evidencia a redução na concentração da clorofila nas folhas HLB-sintomáticas. O aumento do pico em 690 nm relativo ao pico em 735 nm sugere a redução da concentração da clorofila nas folhas. Buschmann (25) explica que a fluorescência induzida em folhas *in natura* provocam a reabsorção da luz emitida em 680 nm devido às diversas camadas de clorofila, o que aumenta a intensidade da emissão de fluorescência em 735 nm. A reabsorção é maior quanto maior for a concentração da clorofila nas folhas. Dessa forma, se a concentração da clorofila é reduzida, a reabsorção será menor, provocando um aumento no pico em 680 nm e uma redução no pico em 735 nm. Esse fato torna o espectro de HLB-sintomática diferente dos demais.

A menor taxa de acerto foi referente às folhas CVC-sintomáticas, com apenas 57%. Muito próximo disso, HLB-assintomática apresentou uma taxa de acerto um pouco maior, de aproximadamente 60%. Um agravante é a alta confusão de HLB-assintomática com as classes sadia e CVC, sem nenhuma ocorrência com HLB-sintomática. A confusão com a classe sadia era esperada, já que os sintomas da doença podem não ser uniformes nos galhos da planta, o que acarreta na presença de folhas aparentemente saudáveis nas plantas doentes. Entretanto, a alta confusão com CVC mostra a dificuldade de encontrar um padrão espectral que defina as classes. Isso pode ser atribuído às bandas excitadas, que não são suficientes para diagnosticar as doenças de forma precisa.

Uma possível explicação para a pouca eficiência global é a presença única das bandas de emissão da clorofila, que embora contribuam para a diferenciação de HLB-sintomática das demais, apresenta pouca informação para a diferenciação das demais classes. Além disso, muitas

doenças e estresses acarretam em alterações na concentração de clorofila, tornando difícil a determinação de uma doença em específico.

Em relação à repetibilidade desses resultados em outros conjuntos, os resultados em sua grande maioria variaram de 65 a 75%. Houve casos de resultados superiores a 80% obtidos por Venâncio (14), onde não foram levadas em consideração folhas de outras doenças como no estudo aqui apresentado. Entretanto, é importante enfatizar, como discutido na revisão bibliográfica, que apenas a emissão da clorofila pode não ser suficiente para a determinação do HLB, pois outras doenças e deficiências nutricionais levam à variações na concentração da clorofila. Assim sendo, o uso dessa metodologia poderia ser útil apenas em fazendas altamente controladas, em que a incidência de outras doenças e de deficiências nutricionais seja desprezível.

6.1.3 LIFS-405

A Fig. 6.6 exibe os espectros típicos médios das folhas de citros medidas no sistema LIFS-405. Os espectros são claramente diferentes dos coletados no fluorímetro (ver Fig. 6.2). Como dito anteriormente, as diferenças são esperadas devido aos componentes dos equipamentos serem diferentes.

Os espectros das folhas com excitação em 405 nm apresentam as emissões da clorofila visivelmente diferentes das encontradas na excitação em 561 nm. A emissão observada nas folhas com a excitação em 561 nm (ver Fig. 6.3) é característico da clorofila quando ocorre o fenômeno da reabsorção da fluorescência emitida em 680 nm. Na excitação com comprimentos de onda menores (violeta-azul), o fenômeno da reabsorção não é expressivo devido à radiação penetrar pouco nas folhas. Isso ocorre porque nessa faixa de comprimento de onda a clorofila possui maior eficiência de absorção. Sendo assim, a luz é absorvida e emitida nas clorofilas mais próximas à superfície foliar (25). Dessa forma, pouca fluorescência será reabsorvida pelas clorofilas. No estudo realizado por (34), foi possível constatar que a emissão natural da clorofila com excitação em 430 nm apresenta um perfil de emissão muito próximo ao observado na

Figura 6.6 – Espectros médios de fluorescência das folhas de citros com excitação Laser em 405 nm

Fig. 6.6, confirmando a baixa reabsorção da luz nessa faixa de excitação.

Os espectros das folhas coletados no sistema LIFS-405 foram pré-tratados e serviram de entrada para a indução do classificador via PLSR. Novamente foram realizadas várias execuções de validação cruzada para determinar o número mais adequado de componentes. As Figs. 6.7 e 6.8 exibem as taxas de acerto médias e o RMSE em função do número de componentes, respectivamente.

Na Fig. 6.7 vemos que as taxas de acerto são maiores no intervalo de 13 a 20 componentes. Utilizando o teste t para a escolha do número de componentes, observamos que o uso de 13 componentes apresenta resultados equivalentes aos demais nessa faixa com confiança estatística de 95%. A escolha é também baseada no menor desvio padrão e no menor número de componentes possível. Pode-se confirmar na Fig. 6.8 que nesse número de componentes o RMSE é um dos menores e o teste t garante com confiança estatística de 95%. Com isso, segue abaixo a matriz de confusão extraída do classificador via PLSR para 13 componentes.

Figura 6.7 – Taxa de acerto do classificador em função do número de componentes utilizados no classificador via PLSR. Nota-se uma estabilização da taxa de acerto próximo de 11 componentes.

Figura 6.8 – RMSE do classificador em função do número de componentes utilizados no classificador via PLSR. Nota-se os valores de erros mais baixos entre 13 e 20 componentes.

Tabela 6.2 – Matriz de confusão obtida na execução de 10 validações cruzadas com 10 folds dos espectros do LIFS-405. As colunas correspondem às predições de classes feitas pelo classificador e as linhas às classes nominais.

Classe nominal	CVC	Sadia	HLB-assint.	HLB-sint.
CVC	82,75	0	12,25	5
Sadia	0	96,5	3,5	0
HLB-assint.	7,25	5,5	86,75	0,5
HLB-sint.	5	0	1,75	93,25

A matriz de confusão da Tabela 6.2 apresenta taxa de acerto médio de 90%, com desvio padrão de 7%. A classe sadia apresentou a maior taxa de acerto e seus erros ocorreram apenas com HLB-assintomática. Importante ressaltar também que não houve confusões entre HLB-sintomática e CVC com sadias, reforçando a robustez da classe.

A classe HLB-sintomática apresentou a segunda maior taxa de acerto, apresentando uma confusão de 5% com CVC. A confusão dela com HLB-assintomática não deve ser considerada um erro, pois a doença é a mesma em estágios diferentes. Dessa forma podemos considerar que o classificador apresentou um índice de acerto de 95% para folhas HLB-sintomáticas e 88% para as folhas HLB-assintomáticas. A ausência de confusão da HLB-sintomática com sadias é importante, pois minimiza os erros referentes a atestar uma planta doente como saudável, deixando mais uma fonte de bactéria no campo.

Além das taxas de acerto que a matriz de confusão apresenta, é possível verificar a qualidade das predições realizadas pelo classificador. A tabela 6.3 mostra um histograma acumulado das probabilidades de predição obtidas. As colunas 1 e 3 dizem respeito ao intervalo de probabilidade de predição e as colunas 2 e 4 correspondem a frequência de amostras que o classificador atribuiu no respectivo intervalo de probabilidades, sendo a coluna 2 relativo as corretas e a 4 as incorretas. Essa tabela é bastante útil na verificação da qualidade das predições efetuadas pelo classificador.

Tabela 6.3 – Histograma acumulado das probabilidades de predição corretas e incorretas das amostras do LIFS-405.

Intervalo de probabilidades (%)	Frequência das corretas (%)	Intervalo de probabilidades (%)	Frequência das erradas (%)
30-100	100,0	0-30	0,0
40-100	98,8	0-40	9,3
50-100	92,4	0-50	35,4
60-100	77,8	0-60	79,5
70-100	56,9	0-70	94,4
80-100	27,7	0-80	99,4
90-100	9,7	0-90	100,0

Podemos observar nessa tabela as informações que são omitidas pela matriz de confusão. O classificador faz a predição com uma determinada probabilidade, que serve como critério de

decisão em amostras de teste. No nosso caso, a tabela serve como uma estimativa de como o classificador se comportaria com amostras de testes, já que aqui foi realizado o procedimento de validação cruzada. Nota-se que, embora as taxas de acerto apresentadas na matriz de confusão sejam elevadas, nem sempre a predição correta tem uma probabilidade elevada. Além disso, nem sempre as amostras preditas incorretamente apresentam probabilidades baixas. Vemos na tabela 6.3 que aproximadamente 92% das amostras preditas corretamente apresentaram probabilidades de acerto inferiores a 50% e que 35% das amostras incorretas apresentaram probabilidades de acerto abaixo de 50%.

Uma informação relevante que pode ser extraída da tabela 6.3 é o intervalo de probabilidades no qual é garantido que praticamente todas as amostras sejam corretas. Aproximadamente 28% das amostras corretamente classificadas tiveram probabilidades de predição acima de 80% e 99,4% das amostras incorretamente classificadas tiveram probabilidades de predição inferiores a 80%. Nessa condição, a classificação de uma amostra com probabilidade acima de 80% indica que a predição é correta com uma taxa de erro de apenas 0,6%. Embora a taxa de acerto seja elevada nesse caso, 72% das amostras corretamente classificadas tiveram probabilidades de predição inferiores a esse valor. Portanto, vemos que é possível estabelecer um critério de classificação confiável e suspeita.

Um critério menos rígido do que o anterior seria verificar a margem na qual engloba grande parte das corretamente classificadas, com uma margem de erro relativamente aceitável. Nota-se que aproximadamente 78% das amostras classificadas corretamente tiveram probabilidades de predição acima de 60% e quase 80% das incorretas, abaixo de 60%. Esse critério é um pouco mais abrangente, sendo que agora a classificação de uma amostra com probabilidade acima de 60% indica uma predição correta com uma margem de erro de aproximadamente 20%. Uma predição com probabilidade abaixo de 60% indicaria, portanto, uma classificação suspeita. Apesar do aparente alto nível de erro, a inspeção visual no campo apresenta eficiência de 47% em encontrar árvores doentes sintomáticas (8).

Dessa forma, a tabela 6.3 exhibe o comportamento esperado que o classificador deverá apre-

sentar quando aplicado em um conjunto de teste. Essa informação deve ser útil para a tomada de decisão por parte do produtor de citros, pois remete ao estabelecimento de critérios tais como os exibidos acima. No nosso caso, conhecemos as classes das folhas previamente, mas na prática o produtor não vai possuir essa informação, sendo importante o estabelecimento de critérios para uma melhor tomada de decisão, seja para a erradicação das árvores ou para deixar as árvores suspeitas em quarentena.

Em relação a repetibilidade dos resultados obtidos, em outros conjuntos de calibração sem a presença de folhas de CVC foram obtidos resultados globais entre 75 e 90%. A variação dos resultados é observada devido a distribuição não uniforme dos sintomas do HLB, de tal forma que parte das folhas rotuladas como assintomáticas podem, na verdade, apresentar características químicas de folhas saudáveis. Dessa forma, a aplicação de classificadores supervisionados são inadequados e apresentam maiores erros nas classificações entre folhas assintomáticas e saudáveis. O estudo mais adequado nesses casos seria o uso de métodos de clusterização, i.e., uso de classificadores não supervisionados para tentar descobrir quais folhas assintomáticas que de fato são doentes. Esses estudos serão realizados em pesquisas posteriores.

6.1.4 Metabólitos secundários

Comparando os sistemas LIFS utilizados, vemos que a diferença nos resultados do LIFS-405 em relação ao sistema LIFS-561 é provavelmente devido à contribuição de informação da emissão azul-verde do espectro. Essa emissão, aliada a emissão da clorofila, caracteriza de maneira mais eficiente as classes de folhas usadas. A emissão azul-verde está fortemente associada à presença de metabólitos secundários, sendo grande parte deles responsáveis pela defesa da planta. Na banda é possível notar alguns ombros que podem indicar a presença de diferentes substâncias químicas emitindo fluorescência. Na literatura, alguns pesquisadores inferem que essa emissão das folhas provém dos ácidos ferúlicos em grande parte e de outros compostos, tais como os flavonóides, cumarinas, quercetina, entre outros (25, 35). A maioria desses com-

postos possui a característica de atuarem contra patógenos, tais como fungos e bactérias. Além disso, podem atuar como precursores de outros compostos que contribuam no mecanismo de defesa da planta.

Um estudo foi realizado, com a colaboração do laboratório de produtos naturais da Universidade Federal de São Carlos, para verificar a capacidade de emissão de fluorescência dos seguintes metabólitos secundários: umbeliferona, hesperidina e naringina. Esses compostos são muito comuns em citros e podem atuar de alguma maneira no mecanismo de defesa da planta. No caso desses metabólitos serem fluoróforos, poderia haver alguma contribuição espectral que evidencie diferenças entre as folhas sadias e doentes.

A figura 6.9 exibe a varredura 3D, i.e. , de excitação e emissão, realizada para a solução aquosa da cumarina umbeliferona. As linhas pretas representam as curvas de nível e as cores a intensidade de emissão. No gráfico, os picos ou faixas diagonais com intensidades mais elevadas constituem as excitações da lâmpada e as segundas ordens de difração, que não foram cortadas pelo filtro utilizado (transmitância de 95% acima de 290 nm). A banda presente entre 400 e 500 nm é a emissão característica da umbeliferona. A banda observada entre 850 e 900 é uma falsa emissão, correspondendo a primeira ordem de difração dessa emissão. Sua intensidade mais elevada é devido à mudança de detector, que apresenta uma sensibilidade mais elevada.

Em primeira análise, notou-se que a umbeliferona é um composto altamente fluorescente, mesmo em concentrações baixas como a utilizada para obter o espectro mostrado na Fig. 6.10. Nessa concentração (10 ppb), a intensidade máxima de emissão ocorreu para a excitação em 320 nm aproximadamente. Nas demais regiões, o composto apresentou uma fluorescência muito baixa, não sendo detectada. Em uma solução aquosa com concentração de Umbeliferona mais elevada (2 ppm), foi possível observar a emissão de fluorescência para excitação em 405 nm, como mostra a Fig. 6.10. Novamente a emissão no infravermelho é a primeira ordem de difração da emissão verdadeira na região azul-verde. Essa concentração não foi utilizada para a varredura 3D devido a elevada intensidade de emissão observada para excitações no

Figura 6.9 – Espectro de fluorescência 3D da Umbeliferona dissolvida em água com concentração de 10 ppb.

Figura 6.10 – Espectro de emissão de fluorescência da Umbeliferona dissolvido em água com concentração de 2 ppm com excitação em 405 nm

ultravioleta, que saturava o equipamento.

A emissão da umbeliferona na região azul-verde aparentemente coincide com as emissões

observadas nos espectros das folhas (ver Fig. 6.6). Mesmo em baixas concentrações, o composto é altamente fluorescente. É possível que este composto esteja contribuindo para as emissões das folhas na mesma região. A planta pode estar reagindo contra a presença da doença, produzindo esse e outros metabólitos secundários.

O estudo foi realizado também com os flavonóides hesperidina e naringina. Ambos os compostos apresentaram emissão de fluorescência baixa para as concentrações utilizadas. A Fig. 6.11 exibe a fluorescência 3D da hesperidina para uma concentração de 40 ppm. Além das faixas diagonais referentes à excitação e primeira ordem de difração, é possível observar o início da segunda ordem de difração próximo à 900 nm e algumas faixas diagonais mais fracas de espalhamento da luz ao lado das excitações. Em relação ao composto analisado, notamos que a intensidade da emissão entre 300 e 530 nm foi bastante baixa e presentes apenas para excitações entre 200 e 230 nm. A emissão entre 750 e 900 nm equivale a primeira ordem de difração da emissão anterior. Vale ressaltar que não foi possível fazer uma concentração maior que 40 ppm devido a não dissolução do composto em água acima deste valor. A emissão do composto apresenta fluorescência muito baixa e é difícil concluir sobre sua contribuição no espectro das folhas. Entretanto, foi possível descobrir a faixa de emissão de fluorescência do composto, que é muito próxima da região de emissão azul-verde observada nas folhas.

A figura 6.12 exibe a fluorescência 3D do flavonóide naringina, para uma concentração de 20 ppm. Devido à baixa intensidade, novamente é observado as faixas de espalhamento ao lado das faixas de excitação e segunda ordem de difração. A emissão, de baixa intensidade, é presente no intervalo de 350 a 530 nm, sendo entre 820 a 900 nm a primeira ordem de difração. Nada poderíamos afirmar à respeito de sua contribuição nos espectros das folhas, mas foi possível verificar que sua emissão ocorre próxima da região azul-verde observada nos espectros das folhas. Entretanto, vale ressaltar que a naringina não é encontrada em folhas saudáveis de árvores valência (49). De qualquer forma, sua produção poderia ocorrer no caso da presença de doenças, podendo assim, interferir nos espectros das folhas.

A escolha da água como solvente dos metabólitos secundários foi devido a sua incapacidade

Figura 6.11 – Espectro 3D da Hesperidina dissolvido em água com concentração de 40 ppm

Figura 6.12 – Espectro 3D da Naringina dissolvida em água com concentração de 20 ppm

de emissão de fluorescência. Determinados solventes podem emitir fluorescência e esconder a emissão do fluoróforo em questão. Além disso, a água é um meio na qual é mais próximo

daquele presente nas folhas. É sabido que o solvente altera as propriedades da emissão de fluorescência, provocando deslocamentos e alargamentos ou compressão das bandas espectrais.

6.2 FTIR

A figura 6.13 mostra os espectros típicos de uma folha Sadia, HLB-assintomática, HLB-sintomática e CVC, normalizado pela área abaixo da curva. Para cada classe, o espectro foi um dos que apresentaram probabilidade de predição máxima dada pelo classificador, cujos resultados serão apresentados mais adiante. A região espectral utilizada foi escolhida de tal forma a eliminar o pico de água (região entre 1900 a 1700 cm^{-1}) e que levaram aos melhores resultados do classificador via PLSR induzido.

Figura 6.13 – Espectro de FTIR típico de cada uma das classes de folhas. Figura adaptada de Cardinali et al. (2).

Na Fig. 6.13, muitas diferenças entre os espectros das folhas podem ser observados, espe-

cialmente entre 1500 e 900 cm^{-1} . O pico presente em 1437 cm^{-1} é bem definido apenas para o espectro da folha saudável. A banda em 1250 cm^{-1} , relativa ao espectro da HLB-sintomática, é intenso, deformado e deslocado em comparação a mesma banda das demais classes. A razão entre as intensidades relativas dos picos também é útil para diferenciar os tipos de folhas, como mostra as razões entre os picos em 1170 e 1105 cm^{-1} , 1170 e 1035 cm^{-1} , 1419 e 1377 cm^{-1} , 913 e 877 cm^{-1} e 832 e 720 cm^{-1} . Além dessas, outras diferenças não visíveis na Fig. 6.13 podem ajudar na distinção entre as classes das folhas. Pelo fato da busca por padrões espectrais ser uma tarefa difícil, especialmente quando as diferenças não são evidentes, o uso de classificadores é indispensável para definir cada tipo de classe.

Os espectros FTIR foram pré-processados e serviram de entrada para o classificador induzido via PLSR. Um estudo do número de componentes foi também realizado para garantir as maiores taxas de acerto. A tabela 6.4 exibe a matriz de confusão das 10 execuções da validação cruzada com 10 *folds*, obtidas pelo classificador induzido com 20 componentes. O traço da matriz indica a taxa de acerto global média do classificador, que foi de aproximadamente 94%. O desvio padrão, que pôde ser calculado através de todas as execuções, foi de 6%.

Tabela 6.4 – Matriz de confusão obtida na execução de 10 validações cruzadas com 10 folds nos espectros do FTIR. As colunas correspondem às predições de classes feitas pelo classificador e as linhas às classes nominais. Tabela extraída de Cardinali et al. (2).

Classe nominal	CVC	Sadia	HLB-assint.	HLB-sint.
CVC	95,00	2,50	2,50	0
Sadia	3,25	89,25	1,50	6,00
HLB-assint.	3,25	3,00	93,75	0
HLB-sint.	0	2,50	0,25	97,25

De acordo a matriz de confusão, HLB-sintomática é a classe com a maior taxa de acerto, seguido da CVC e HLB-assintomática. A pior classificação obtida foi para a sadia, cuja taxa de acerto foi de 89.25%. Curiosamente, a confusão entre sadia e HLB-assintomática, que são muito similares visualmente (ver Fig. 5.1), foi uma das menores. No caso das confusões entre HLB-sintomática e HLB-assintomática não podem ser consideradas um erro visto que é a mesma doença em diferentes estágios. No presente caso, a taxa de confusão entre HLB-assintomática e HLB-sintomática foi muito baixo. Embora a confusão entre as diferentes classes é em geral

baixa (abaixo de 6%), pode ser concluído que há de fato um desbalanço químico entre as classes, que foi evidenciada pela espectroscopia de infravermelho. Essas diferenças puderam ser descobertas pelo classificador via PLSR.

Além das análises da matriz de confusão, foi possível estudar a probabilidade de predição de cada folha como foi feito no sistema LIFS-405. A tabela 6.5 exibe o histograma acumulado das probabilidades de predição efetuadas pelo classificador em função da frequência de amostras. Numa análise rápida, vemos que aproximadamente 27% das amostras corretas apresentaram probabilidade de predição superior a 80% e que 99% das amostras incorretas ficaram abaixo dessa probabilidade de predição.

Tabela 6.5 – Histograma acumulado das probabilidades de predição corretas e incorretas das amostras do FTIR. Tabela adaptada de Cardinali et al. (2).

Intervalo de probabilidades (%)	Frequência das corretas (%)	Intervalo de probabilidades (%)	Frequência das erradas (%)
30-100	100,0	0-30	0,0
40-100	98,9	0-40	8,1
50-100	90,0	0-50	48,5
60-100	73,7	0-60	80,8
70-100	53,6	0-70	93,9
80-100	26,9	0-80	99,0
90-100	8,9	0-90	100,0

Tomando o mesmo critério apresentado no sistema LIFS-405, vemos que 73,7% das predições corretas apresentaram probabilidades superiores a 60% e 80,8% das predições incorretas apresentaram probabilidades de predição abaixo de 60%. Dessa forma, esse critério aplicado em um conjunto teste levaria a classificações corretas com erro da ordem de aproximadamente 19%. Esse erro foi extraído tomando a frequência de amostras incorretas acima de 60%. Abaixo dos 60% teríamos uma condição de predição suspeita. Reiterando, esse tipo de análise permite uma melhor tomada de decisão por parte dos produtores.

Conforme discutido nos sistemas LIFS, a repetibilidade e confiabilidade nos resultados obtidos no sistema FTIR são bastante estabelecidos. Resultados obtidos no laboratório em um experimento com mudas saudáveis e inoculadas com HLB mostraram que a detecção da doença pelo FTIR foi possível a partir do 3º mês após a inoculação (dados não publicados). Esses

resultados corroboraram em grande parte com as outras técnicas utilizadas, tais como LIFS e imagens de fluorescência, discutidas anteriormente na revisão bibliográfica (14, 17). Além disso, conforme descrito na revisão bibliográfica, os estudos realizados por Sankaran e Hawkins et al. (37–39) confirmam a alta sensibilidade apresentada pela espectroscopia infravermelho na detecção de HLB.

A robustez do classificador induzido via PLSR pode ser associada as diferenças espectrais encontradas entre as classes e a eficiência do modelo PLSR em encontrar tais diferenças. Para entendimento dos sinais espectrais e dos excelentes resultados do classificador obtidos no FTIR, um estudo com metabólitos primários e secundários foi realizado na subseção que segue.

6.2.1 Metabólitos

Voltando aos espectros de FTIR das folhas de citros da Fig. 6.13, o intervalo espectral de 1175 a 900⁻¹ é geralmente associado a absorção de amido (37–39), cuja concentração é esperada ser diferente entre as folhas sadias e de árvores afetadas com HLB (27, 28, 39, 50). Entretanto, não é possível afirmar que esse intervalo seja apenas caracterizado pela absorção de carboidratos, já que outras substâncias químicas tais como metabólitos secundários, podem estar absorvendo luz nessa faixa. Além disso, as concentrações dos metabólitos secundários devem ser diferentes em folhas de árvores sadias e doentes. A Fig. 6.14 mostra o espectro normalizado de amostras padrões de amido, glicose e sacarose. A Fig. 6.15 mostra os espectros dos flavonóides hesperidina e naringina e a cumarina umbeliferona produzida por citros.

Os espectros obtidos mostrados nas Fig. 6.14 e 6.15 são similares aos encontrados no banco de dados de compostos orgânicos da *National Institute of Advanced Science and Technology* (51). Em relação aos flavonóides, a hesperidina é um dos metabólitos secundários mais abundantes nas folhas de citros da variedade Valência (por volta de 5 mg/g) (49). No caso da naringina, não é de se esperar sua contribuição espectral no espectro das folhas sadias, pois sua presença na mesma variedade de citros não é verificada (49). É possível observar na Fig. 6.15

Figura 6.14 – Espectros normalizados de FTIR das pastilhas de KBr de amostras padrão de amido, glicose e sacarose. Figura adaptada de Cardinali et al. (2).

Figura 6.15 – Espectro de FTIR das pastilhas de KBr dos metabólitos secundários hesperidina, naringina e umbeliferona. Figura adaptada de Cardinali et al. (2).

que a hesperidina apresenta maior absorção nas bandas entre 1100 e 950 cm^{-1} , dentro da região de maior influência dos carboidratos (ver Fig. 6.14). Mudanças na concentração desse composto poderiam acarretar mudanças espectrais nas folhas nessa região, junto com os demais carboidratos. Mesmo a naringina poderia interferir nessa região no caso de sua produção ser provocada pela infecção de HLB ou CVC. Embora existam outros flavonóides abundantes nas folhas de citros, tais como rutin, diosmin, nobiletin e isorhoifolin, que não foram considerados nessas análises, os espectros dos metabólitos apresentam fortes evidências de que eles são cruciais para o entendimento dos picos e bandas fora da região de principal influência dos carboidratos, i.e., nos intervalos espectrais de 1530 cm^{-1} a 1175 cm^{-1} e 900 cm^{-1} a 700 cm^{-1} (ver Fig. 6.13). As bandas em 1515 e 1450 cm^{-1} na Fig. 6.13 podem ser associados aos modos vibracionais da hesperidina e umbeliferona (ver Fig. 6.15). Dessa forma, estudos ainda devem ser conduzidos a fim de descobrir quais compostos químicos estão provocando as diferenças espectrais observadas no infravermelho, principalmente nas regiões de baixa absorção dos carboidratos.

7 Conclusões

Os experimentos realizados nesse trabalho permitiram comprovar a hipótese levantada. Através dos espectros obtidos com as folhas saudáveis e de árvores infectadas com HLB, foi observado que a doença promove alterações químicas e físicas na planta e nas folhas, permitindo o uso das técnicas fotônicas utilizadas aqui para diagnóstico da doença.

A primeira parte do experimento revelou qualitativamente uma melhor condição espectral para excitações com comprimentos de onda compreendidos entre 400 e 420 nm. A excitação das folhas nesse intervalo provê a emissão completa na região do visível e início do infravermelho. A emissão na região azul-verde é importante devido a sua associação com a presença de ácidos ferúlicos e outros metabólicos secundários nas folhas, que apresentam relação direta com o mecanismo de defesa da planta. A variação da concentração desses metabólitos altera o padrão de emissão de fluorescência dessa região. A emissão da clorofila na região do vermelho e infravermelho também apresenta variação devido ao desbalanço de sua concentração nas folhas doentes. Entretanto diferentes estresses podem causar sua variação nas folhas. A união dessas duas regiões de emissão de fluorescência garante um resultado promissor no diagnóstico de doenças em citros.

Os resultados extraídos dos sistemas LIFS-561 e LIFS-405 ilustraram de forma mais concreta as justificativas descritas no parágrafo anterior. A indução do classificador via PLSR nos espectros de LIFS-405 resultou em uma taxa de acerto média de 90% contra 69% do sistema LIFS-561. Claramente o sistema LIFS-561 é prejudicado devido aos espectros das folhas apresentarem principalmente as emissões de fluorescência da clorofila. Mesmo assim, ambos os sistemas apresentaram resultados muito superiores se comparado com a inspeção visual no campo, com o adicional de diagnosticar folhas assintomáticas. Se a análise for realizada com

folhas sintomáticas, como é feito nas inspeções visuais, as taxas de acerto ficam em torno de 85% para o LIFS-561 e 95% para o LIFS-405. No LIFS-405 foi possível diagnosticar o HLB em folhas sintomáticas e assintomáticas com taxas superiores a 87%, com confusões entre sadias inferiores a 5,5%. Com o histograma acumulado de probabilidades de predição do classificador no sistema LIFS-405, foi possível estimar o comportamento do classificador quando aplicado em amostras de teste. Essa tabela permite o estabelecimento de critérios por parte dos produtores em relação à confiabilidade da predição do classificador.

Embora as taxas de acerto globais do sistema LIFS-561 foram mais baixas em relação ao sistema LIFS-405, o seu estudo contribuiu para um melhor entendimento dos espectros. Um fato interessante é o de que a reabsorção da clorofila, fenômeno físico que é mais visível na excitação em 561 nm, evidencia de forma mais clara que o HLB leva a uma maior diminuição de clorofila nas folhas em relação ao CVC. Isso porque quando a concentração da clorofila se altera, pode ocorrer a inversão da relação entre os picos em 680 e 730 nm, como ocorre para o CVC e principalmente para o HLB. No sistema LIFS-405 esse fenômeno é observado não com a inversão de picos, mas com o aumento do pico em 680 em relação ao pico em 730 nm.

Em relação aos metabólitos secundários medidos no fluorímetro, apenas a umbeliferona apresentou fluorescência observável com excitação em 400 nm. Apesar da emissão de fluorescência apresentar intensidade baixa na hesperidina e naringina, a região de emissão dos três compostos analisados é praticamente na mesma região, próximo à região da emissão azul-verde observada nas folhas. Com as análises realizadas aqui, não podemos afirmar que a naringina ou a hesperidina não contribuem na diferenciação espectral das folhas doentes. Porém, observamos que a presença desses compostos nas folhas poderiam acarretar na emissão de fluorescência na região azul-verde.

Os resultados da espectroscopia FTIR mostraram que é possível diagnosticar HLB e CVC em folhas de citros. A taxa de acerto total do FTIR foi de aproximadamente 94%. É importante enfatizar que o método diferencia folhas saudáveis de folhas HLB-sintomáticas com taxa de acerto acima de 95%. As altas taxas de acerto podem estar associadas às variações nas

concentrações de carboidratos e de metabólitos secundários, tais como a hesperidina e umbeliferona. Enquanto a hesperidina é um dos metabólitos secundários de maior concentração em folhas de citros do cultivar Valência, a umbeliferona é da família das cumarinas, compostos associados na defesa da planta contra patógenos, podendo ser até precursora de outras cumarinas que tem o mesmo propósito.

Um fato importante observado nesse trabalho foi a corroboração dos resultados obtidos pelas técnicas LIFS-405 e FTIR. Ambas as técnicas apresentaram taxas de acerto elevadas, mostrando principalmente que a grande maioria das folhas HLB-assintomáticas utilizadas apresentaram sinais da doença. Entretanto, é sabido que a distribuição dos sintomas na planta não é uniforme e, portanto, a eficiência do classificador nesses casos pode ser comprometida. Em outros conjuntos de calibração, foram observadas taxas de acerto inferiores, claramente devido a esse fator. Estudos mais cuidadosos nesses casos ainda deverão ser realizados para verificar a distribuição dos sintomas em árvores infectadas com HLB, inclusive em níveis de severidade diferentes.

Os estudos realizados neste trabalho evidenciaram novas metodologias que são muito promissoras no diagnóstico de doenças em citros. Ambas as técnicas LIFS-405 e FTIR são de baixo custo e não exigem o preparo de amostras, o que tornam suas análises ainda mais rápidas. Vale ressaltar que as taxas de acerto aqui apresentadas são referentes às folhas de cada classe. Os resultados de diagnóstico por árvore devem aumentar ainda mais a eficiência do método. Outro atrativo é que as taxas de acerto de ambas as técnicas são muito superiores as das inspeções visuais realizadas atualmente no campo.

Em relação ao possível uso das técnicas em campo, o LIFS-405 é o mais apropriado devido à capacidade de miniaturização de seus componentes ou até mesmo pela substituição do laser por led. Além disso, o sistema já se apresenta em um formato portátil com um software capaz de fazer o diagnóstico em tempo real. O FTIR foi realizado em um equipamento comercial fabricado pela Perkin Elmer, onde seu uso deve ser realizado em um ambiente com temperatura e umidade controladas. A técnica poderia ser utilizada na fazenda em uma sala com temperatura

e umidade controladas. As amostras de folhas seriam coletadas no campo e levadas para o local para efetuar o diagnóstico. Essa técnica poderia ser aliada ao uso do LIFS-405 no campo, de modo a melhorar ainda mais as taxas de acerto.

Muitas pesquisas devem ser realizadas a fim de disponibilizar um sistema portátil de diagnóstico de HLB e de outras doenças em citros. Abaixo segue uma lista com as principais atividades futuras:

- 1) Realizar testes em campo com a técnica LIFS-405 para validar os resultados aqui obtidos.
- 2) Realizar estudos da interferência de fatores externos na aplicação das técnicas no diagnóstico de HLB, tais como sazonalidade, temperatura, região, solo, entre outros.
- 3) Verificar a possibilidade de generalizar o classificador para identificação de outras doenças e deficiências nutricionais em plantas. Essa etapa é importante para entender se inclusive podem ocorrer confusões entre HLB e outras doenças ou deficiências nutricionais.
- 4) Desenvolver um classificador por planta. Os resultados obtidos até agora foram para as folhas, entretanto, é importante fazer um estudo do número de folhas necessárias para garantir um diagnóstico preciso e de baixo índice de erro por planta. Com os resultados apresentados aqui, os classificadores por planta com certeza apresentarão resultados superiores.
- 5) Verificar a distribuição da doença nas plantas infectadas com HLB. Esse estudo é importante para entender a distribuição dos sintomas, inclusive em árvores em estágios iniciais e avançados da doença.
- 6) Uso de outras técnicas para melhor entendimento das alterações físicas e químicas provocadas pelo HLB na planta. Entendendo quais compostos estão se alterando nas plantas, é possível desenvolver um método que melhor avalie essas alterações específicas.
- 7) Em conjunto com os estudos propostos no item anterior, medir metabólitos secundários na fluorescência e infravermelho para continuar na busca pelo entendimento dos espectros foliares. Importante medir a absorção da clorofila no infravermelho para verificar se há influência da mesma nos espectros das folhas.

8) Estudo de outras técnicas que possam ser aplicadas no diagnóstico de HLB, tais como transmitância, fluorescência resolvida no tempo, imagens de fluorescência, imagens de infravermelho, imagens multi/hiper-espectrais, entre outras. Outros sistemas podem oferecer medidas em campo mais rápidas e eficazes.

9) Uso de outras ferramentas estatísticas de classificação e reconhecimento de padrões, tais como support vector machines (Maquinas de vetores de suporte), redes neurais artificiais, árvores de decisão, métodos bayesianos, entre outros. É importante testar diferentes ferramentas para extrair os melhores resultados possíveis da maneira mais simples possível.

REFERÊNCIAS

- 1 FUNDECITRUS. *Manual técnico - Greening (HLB)*. 2009. Disponível em: <<http://www.fundecitrus.com.br/ImageBank/PageFlip/pageflip.aspx?idPage=143>>. Acesso em: 27 Dez 2011.

- 2 CARDINALI, M. C. do B. et al. Infrared spectroscopy: a potential tool in huanglongbing and citrus variegated chlorosis diagnosis. *Talanta*, v. 91, n. 0, p. 1–6, 2012.

- 3 FRANK, E. et al. Using model trees for classification. *Machine Learning*, v. 32, n. 1, p. 63–76, 1998.

- 4 FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO). *Produção total de citrus em 2009*. 2009. Disponível em: <<http://www.fao.org>>. Acesso em: 05 set 2011.

- 5 INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Levantamento sistemático da produção agrícola: pesquisa mensal de previsão e acompanhamento das safras agrícolas no ano civil. 2011. Disponível em: <http://www.ibge.gov.br/home/estatistica/indicadores/agropecuaria/lspa/lspax_201101.pdf>. Acesso em: 08 nov 2011.

- 6 NEVES, M. F. et al. *O retrato da citricultura brasileira*. 2010. Disponível em: <http://www.citrusbr.com.br/download/biblioteca/Apresentacao_Marcos_Fava_evento_valor.pdf>. Acesso em: 04 set 2011.

- 7 BOVÉ, J. M. Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. *Journal of Plant Pathology*, v. 88, n. 1, p. 7–37, 2006.

- 8 BELASQUE JUNIOR, J. et al. Scientific basis for the eradication of symptomatic and asymptomatic plants in Huanglongbing (HLB, Greening) aimed at effective disease control. *Tropical Plant Pathology*, v. 34, n. 3, p. 137–145, 2009.

- 9 GOTTWALD, T. R.; GRAÇA, J. V.; BASSANEZI, R. B. *Citrus Huanglongbing: the pathogen and its impact*. 2007. Disponível em: <<http://ddr.nal.usda.gov/bitstream/10113/12085/1/IND44030779.pdf>> Acesso em: 09 out 2011.

- 10 SPANN, T. M. et al. IFAS guidance for Huanglongbing (Greening) management. *Citrus Industry*, v. 91, n. 4, p. 6–13, 2010.
- 11 TEIXEIRA, D. C. et al. Distribution and quantification of *Candidatus Liberibacter americanus*, agent of huanglongbing disease of citrus in São Paulo State, Brasil, in leaves of an affected sweet orange tree as determined by PCR. *Molecular and Cellular Probes*, v. 22, n. 3, p. 139–150, 2008.
- 12 LI, W.; HARTUNG, J. S.; LEVY, L. Quantitative real-time PCR for detection and identification of *Candidatus Liberibacter* species associated with citrus Huanglongbing. *Journal of Microbiological Methods*, v. 66, n. 1, p. 104–115, 2006.
- 13 ZAGHI, A. F. *Diagnóstico de doenças de citros utilizando espectroscopia de fluorescência*. 2008. 97p. Dissertação (Mestrado) – Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, 2008.
- 14 VENÂNCIO, A. L. *Avaliação da precocidade do diagnóstico do Greening por técnicas de fluorescência*. 2010. 111p. Dissertação (Mestrado) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2010.
- 15 PEREIRA, F. M. V. et al. Investigation of the stages of citrus greening disease using micro synchrotron radiation X-ray fluorescence in association with chemometric tools. *Journal of Analytical Atomic Spectrometry*, v. 25, n. 3, p. 351–355, 2010.
- 16 PEREIRA, F. M. V. et al. Fluorescence images combined to statistic test for fingerprinting of citrus plants after bacterial infection. *Analytical Methods*, v. 3, n. 3, p. 552–556, 2011.
- 17 PEREIRA, F. M. V. et al. Laser-induced fluorescence imaging method to monitor citrus greening disease. *Computers and Electronics in Agriculture*, v. 79, n. 12, p. 90–93, 2011.
- 18 PEREIRA, F. M. V. et al. Evaluation of the effects of *Candidatus Liberibacter asiaticus* on inoculated citrus plants using laser-induced breakdown spectroscopy (LIBS) and chemometrics tools. *Talanta*, v. 83, n. 2, p. 351–356, 2010.
- 19 LOPES, S. A. et al. Liberibacters associated with citrus Huanglongbing in brazil: *Candidatus Liberibacter asiaticus* is heat tolerant, *Ca. L. americanus* is heat sensitive. *Plant Disease*, v. 93, n. 3, p. 257–262, 2009.
- 20 FUNDECITRUS. *Relatório Anual*. 2011. Disponível em: <<http://www.fundecitrus.com.br/Pagina/cancrocitrico,315>> Acesso em: 05 set 2011.
- 21 FUNDECITRUS. *Manual técnico - Clorose variegada dos citros (CVC)*. 2009. Disponível em: <<http://www.fundecitrus.com.br/ImageBank/PageFlip/pageflip.aspx?idPage=137>>. Acesso em: 29 dez 2011.

- 22 BOVÉ, J. M.; AYRES, A. J. Etiology of three recent diseases of citrus in são paulo state: Sudden death, variegated chlorosis and huanglongbing. *IUBMB Life*, v. 59, n. 4, p. 346–354, 2007.
- 23 SLUYS, M. A. V. et al. Comparative analyses of the complete genome sequences of pierce's disease and citrus variegated chlorosis strains of xylella fastidiosa. *Journal of Bacteriology*, v. 185, n. 3, p. 1018–1026, 2003.
- 24 SOUZA, A. et al. Analysis of gene expression in two growth states of xylella fastidiosa and its relationship with pathogenicity. *Molecular Plant-Microbe Interactions*, v. 16, n. 10, p. 867–875, 2003.
- 25 BUSCHMANN, C. Variability and application of the chlorophyll fluorescence emission ratio red/far-red of leaves. *Photosynthesis Research*, v. 92, n. 2, p. 261–271, 2007.
- 26 HAN, S. et al. Boron deficiency decreases growth and photosynthesis, and increases starch and hexoses in leaves of citrus. *Journal of Plant Physiology*, v. 165, n. 13, p. 1331–1341, 2008.
- 27 ACHOR, D. S. et al. Sequence of anatomical symptom observations in citrus affected with Huanglongbing disease. *Plant Pathology Journal*, v. 9, n. 2, p. 56–64, 2010.
- 28 FAN, J. et al. Changes in carbohydrate metabolism in citrus sinensis infected with "Candidatus Liberibacter asiaticus". *Plant Pathology*, v. 59, n. 6, p. 1037–1043, 2010.
- 29 HAMMERSCHMIDT, R. phytoalexins: What have we learned after 60 years? *Annual Review of Phytopathology*, v. 37, p. 285–306, 1999. DOI: 10.1146/annurev.phyto.37.1.285.
- 30 RIBEIRO, A. B. et al. Chemical characterization of citrus sinensis grafted on c. limonia and the effect of some isolated compounds on the growth of xylella fastidiosa. *Journal of Agricultural and Food Chemistry*, v. 56, n. 17, p. 7815–7822, 2008.
- 31 AFEK, U.; SZTEJNBERG, A. Effects of Fosetyl-Al and Phosphorous Acid on Scorparone, a Phytoalexin Associated with Resistance of Citrus to *Phytophthora Citrophthora*. *Phytopathology*, v. 79, n. 7, p. 736–739, 1989.
- 32 PERONI, L. A. et al. Assessment of the diagnostic potential of immunocapture-PCR and immuno-PCR for Citrus Variegated Chlorosis. *Journal of Microbiological Methods*, v. 75, n. 2, p. 302–307, 2008.
- 33 MILORI, D. M. B. P.; NETO, L. M.; BAYER, C. *Fluorescência induzida por laser para análise da matéria orgânica de solos intactos*. São Carlos: EMBRAPA, 2002. 18p.
- 34 LICHTENTHALER, H. K. et al. Plant stress detection by reflectance and fluorescence. *Annals New York Academy of Sciences*, v. 851, p. 271–285, 1998. DOI: 10.1111/j.1749-6632.1998.tb09002.x.

- 35 LICHTENTHALER, H. K.; SCHWEIGER, J. Cell wall bound ferulic acid, the major substance of the blue-green fluorescence emission of plants. *Journal of Plant Physiology*, v. 152, n. 2-3, p. 272–282, 1998.
- 36 KANCHEVA, R. H.; BORISOVA, D. S.; ILIEV, I. T. Chlorophyll fluorescence plant as a stress indicator. *Recent Developments in Remote Sensing from Space*, v. 5, p. 301–306, 2008.
- 37 HAWKINS, S. A. et al. Comparison of FTIR spectra between Huanglongbing (Citrus Greening) and other citrus maladies. *Journal of Agricultural and Food Chemistry*, v. 58, n. 10, p. 6007–6010, 2010.
- 38 HAWKINS, S. A. et al. Detection of citrus huanglongbing by Fourier transform infrared-attenuated total reflection spectroscopy. *Applied Spectroscopy*, v. 64, n. 1, p. 100–103, 2010.
- 39 SANKARAN, S.; EHSANI, R.; ETXEBERRIA, E. Mid-infrared spectroscopy for detection of Huanglongbing (Greening) in citrus leaves. *Talanta*, v. 83, n. 2, p. 574–581, 2010.
- 40 LAKOWICZ, J. R. *Principles of fluorescence spectroscopy*. 3rd ed. New York: Springer, 2006.
- 41 SILVERSTEIN, R. M.; WEBSTER, F. X.; KIEMLE, D. J. *Spectrometric identification of organic compounds*. 7th ed. Hoboken: John Wiley & Sons, Inc., 2005.
- 42 LARKIN, P. J. *Infrared and raman spectroscopy: principles and spectral interpretation*. New York: Elsevier, 2011. ISBN: 978-0-12-386984-5.
- 43 WITTEN, I.; FRANK, E.; HALL, M. *Data mining: practical machine learning tools and techniques*. 3rd ed. Burlington: Morgan Kaufmann, 2011.
- 44 MANLY, B. F. J. *Multivariate statistical methods: a primer*. 3rd ed. Florida: Chapman & Hall/CRC, 2004.
- 45 WOLD, S. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics*, v. 20, n. 4, p. 397–405, 1978.
- 46 WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, v. 58, n. 2, p. 109–130, 2001.
- 47 HALL, M. A. et al. The WEKA data mining software: an update. *SIGKDD Explorations*, v. 11, n. 1, p. 10–18, 2009.
- 48 CHRISTIAN, G. *Analytical chemistry*. 5th ed. Australia: John Wiley and Sons, 1994.

- 49 KAWAII, S. et al. Quantitative study of flavonoids in leaves of citrus plants. *Journal of Agricultural and Food Chemistry*, v. 48, n. 9, p. 3865–3871, 2000.
- 50 ETXEBERRIA, E. et al. Anatomical distribution of abnormally high levels of starch in HLB-affected valencia orange trees. *Physiological and Molecular Plant Pathology*, v. 74, n. 1, p. 76–83, 2009.
- 51 NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST). *Spectral database for organic compounds*. 2011. Disponível em: <http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/direct_frame_top.cgi>. Acesso em: 30 oct 2011.

APÊNDICE A – Regressão Linear Múltipla e Análise de Componentes Principais

Nesse apêndice será dada uma breve apresentação da Regressão Linear Múltipla (MLR) e da Análise de Componentes Principais (PCA) para um melhor entendimento da metodologia do PLSR e o porque do PLSR nas análises dos dados.

O MLR é um método utilizado para relacionar linearmente as variáveis de dados com as variáveis de resposta. Essa relação é representada através de

$$y = b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_mx_m + e, \quad (\text{A.1})$$

onde x_i são as variáveis de dados ou variáveis independentes, b_i os pesos ou parâmetros a serem estimados, m o número total de variáveis, y a variável de resposta ou variável dependente e e o erro na regressão. Para o caso de mais amostras, é possível generalizar esse método para uma matriz de dados, onde ficamos com

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}. \quad (\text{A.2})$$

sendo a representação gráfica das matrizes exibida na Fig.A.1,

onde n representa o número de amostras. O uso dessa ferramenta é muito útil quando os dados são multi-dimensionais e pouco correlacionados entre si. Entretanto, quando os dados são muito correlacionados e apresentam ruídos, o método apresenta dificuldades na regressão. Nesses casos é recomendado o uso de outras ferramentas para reduzir a dimensionalidade e

Figura A.1 – Dimensões das matrizes envolvidas no MLR.

ruído dos dados. Uma ferramenta muito utilizada é a análise de componentes principais (PCA). O PCA realiza uma transformação nos dados de forma a projetá-los em novas componentes, que são descorrelacionadas entre si, i.e., ortogonais. Por se tratar de um método de redução de variáveis, a metodologia busca pelas componentes que apontam para a direção de maior variância dos dados, eliminando as demais componentes encontradas. Tais componentes são usualmente chamadas de componentes principais, do inglês *principal components* (PC).

A metodologia do PCA se baseia na decomposição da matriz de dados de posto r em r matrizes de posto 1.:

$$\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 + \dots + \mathbf{M}_r \quad (\text{A.3})$$

O posto ou *rank* indica o número de linhas ou colunas linearmente independentes na matriz. Dessa forma, como cada nova matriz apresenta posto 1, elas são ortogonais entre si e representam uma componente. Como nesse caso estamos lidando com espectros, são obtidos muitos dados e não é necessário o cálculo de todas essas matrizes. Nesses casos, são calculadas apenas as primeiras componentes principais. As novas matrizes, decompostas em um produto de vetores $\mathbf{t}$ e $\mathbf{p}$, chamados de *scores* e *loadings*, respectivamente, podem ser calculadas em um processo iterativo

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \mathbf{t}_3 \mathbf{p}_3^T + \dots + \mathbf{E} \quad (\text{A.4})$$

Os *loadings* descrevem as relações entre as variáveis dos dados, i.e., cada um de seus va-

lores são os coeficientes da combinação linear de cada uma das variáveis dos dados. Esses coeficientes indicam o grau de importância de cada variável na respectiva componente. Os *scores* são as projeções dos dados na respectiva componente. A matriz **E** são os resíduos, i.e., as matrizes não calculadas nesse processo iterativo. Podemos juntar os *scores* e *loadings* nas matrizes **T** e **P**, respectivamente:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (\text{A.5})$$

com representação gráfica dada pela Fig. A.2.

Figura A.2 – Dimensões das matrizes envolvidas no PCA

considerando n amostras com m variáveis e o cálculo de a componentes principais. O algoritmo mais usual para cálculo das PCs em conjuntos de dados com um número de variáveis muito grande é o NIPALS (Nonlinear Iterative Partial Least Squares). O algoritmo calcula em um processo iterativo cada um dos *scores* e *loadings* de cada PC. As primeiras PCs calculadas já se apresentam em ordem decrescente de variância, sendo o critério de parada a escolha do número de PCs ou quando a variância dos dados já esteja bastante concentrada nas novas PCs, i.e., quando $|\mathbf{E}|$ é pequeno.

Os métodos MLR e PCA podem ser utilizados conjuntamente, realizando primeiramente o PCA sobre os dados para redução da dimensionalidade e em seguida o MLR sobre os *scores* para a regressão. Entretanto, como visto nos cálculos acima, a PCA não leva em consideração o vetor de respostas. Portanto, nem sempre as PCs calculadas se correlacionam com as respostas esperadas. Nesses casos, o PLS é mais adequado considerar a variância dos dados e as respostas no cálculo das componentes. Esse processo torna a redução de dimensionalidade e a regressão mais eficiente. Por esse motivo, foi utilizado o PLS, explicado na seção *fundamentos teóricos* e

com o algoritmo expresso no *apêndice*.

APÊNDICE B – Algoritmo PLS1

Nesse apêndice será descrito o algoritmo do PLSR (PLS1) para encontrar os dados de regressão. Como dito anteriormente, o objetivo do PLSR é estabelecer relações entre a matriz de dados $\mathbf{X}$ e o vetor de respostas $\mathbf{y}$ de modo a criar um modelo preditivo.

B.1 Algoritmo PLS1-calibração

A primeira etapa antes do início do algoritmo PLS1 é a centralização de $\mathbf{X}$ e $\mathbf{y}$. A centralização dos dados em torno da média é realizada subtraindo a média dos dados. No caso da matriz $\mathbf{X}$, o procedimento é realizado subtraindo a média de cada intensidade de emissão.

Sendo i a variável de iteração do algoritmo, que no caso desse algoritmo corresponde aos cálculos de uma determinada componente. O algoritmo PLS1 inicia fazendo $i = 1$, $\mathbf{X}_i = \mathbf{X}$ e $\mathbf{y}_i = \mathbf{y}$.

Passo 1: $\mathbf{w}_i = \mathbf{X}_i^T \mathbf{y}_i / \|\mathbf{X}_i^T \mathbf{y}_i\|$

Passo 2: $\mathbf{t}_i = \mathbf{X}_i \mathbf{w}_i$

Passo 3: $c_i = \mathbf{t}_i^T \mathbf{y}_i / \mathbf{t}_i^T \mathbf{t}_i$

Passo 4: $\mathbf{p}_i = \mathbf{X}_i^T \mathbf{t}_i / \mathbf{t}_i^T \mathbf{t}_i$

Passo 5: $\mathbf{X}_{i+1} = \mathbf{X}_i - \mathbf{t}_i \mathbf{p}_i^T$ e $\mathbf{y}_{i+1} = \mathbf{y}_i - \mathbf{t}_i c_i^T$

Passo 6: Parar se $i = a$; Senão $i = i + 1$ e retorna ao passo 1.

No fim do algoritmo, as matrizes $\mathbf{T}$, $\mathbf{P}$, $\mathbf{W}$ e o vetor coluna $\mathbf{c}$ podem ser formados a partir de cada um dos seus respectivos valores $\mathbf{t}_i$, $\mathbf{p}_i$, $\mathbf{w}_i$ e c_i . Através da calibração dos valores realizada, é possível estimar os valores de $\mathbf{X}$ e $\mathbf{y}$ fazendo:

$$\hat{\mathbf{X}} = \mathbf{T}\mathbf{P}^T \quad (\text{B.1})$$

e

$$\hat{\mathbf{y}} = \mathbf{T}\mathbf{c}^T. \quad (\text{B.2})$$

Com essas matrizes de estimação, é possível verificar se o modelo gerado consegue representar bem os dados utilizados para calibração. Cada um dos passos do algoritmo seguem explicados abaixo:

Passo 1: No modelo PLSR, procuramos por uma direção no espaço de $\mathbf{X}$ que dê a máxima covariância entre $\mathbf{X}$ e $\mathbf{y}$, i.e., aquela na qual grandes variações em $\mathbf{X}$ levam a grandes variações em $\mathbf{y}$. Essa direção é encontrada através do produto entre $\mathbf{X}$ e $\mathbf{y}$, sendo descrita pelo vetor unitário $\mathbf{w}_i$, usualmente é chamado de vetor de pesos.

Passo 2: O vetor de scores $\mathbf{t}_i$ é formado através da combinação linear da coluna de $\mathbf{X}$ com os pesos $\mathbf{w}_i$

Passo 3: Cálculo do coeficiente de regressão c_i através da regressão linear de $\mathbf{y}$ em $\mathbf{t}_i$.

Passo 4: Vetor de *loadings* $\mathbf{p}_i$, obtido através de uma regressão linear simples das colunas de $\mathbf{X}$ em $\mathbf{t}_i$.

Passo 5: A matriz $\mathbf{X}_{i+1}$ corresponde aos resíduos após a projeção de $\mathbf{X}_i$ em $\mathbf{t}_i$, assim como $\mathbf{y}_{i+1}$ são os resíduos após a projeção de $\mathbf{y}_i$ em $\mathbf{t}_i$. Esse passo é importante para garantir que os vetores $\mathbf{t}_i$ sejam ortogonais.

B.2 Algoritmo de predição

Para o uso do modelo gerado na fase de calibração em amostras testes, são utilizadas as matrizes $\mathbf{W}$, $\mathbf{P}$ e $\mathbf{c}$ extraídas. Considerando apenas uma amostra de teste não centralizada, chamada de $\mathbf{z}$, devemos centralizá-la de acordo com a matriz de dados $\mathbf{X}$ antes de sua centralização, utilizando $\bar{\mathbf{X}}$. Dessa forma, a amostra teste centralizada, que chamaremos de $\mathbf{x}$, é obtida através de $\mathbf{x} = \mathbf{z} - \bar{\mathbf{X}}$. Os passos da predição de y dessa amostra são os seguintes:

Passo 1: $t_i = \mathbf{x}_i \mathbf{w}_i$

Passo 2: $\mathbf{x}_{i+1} = \mathbf{x}_i - t_i \mathbf{p}_i^T$

Passo 3: Parar se $i = a$; Senão $i = i + 1$ e retorna ao passo 1.

No fim dos três passos, o vetor $\mathbf{t}$ pode ser formado a partir de cada t calculado. Sendo assim, o cálculo da resposta y para essa amostra teste é realizado no passo 4:

Passo 4: $y = \bar{y} + \mathbf{t} \mathbf{c}^T$

É importante notar que como o modelo foi gerado com os dados centralizados, a predição de y é, portanto, centralizada. Por esse motivo, foi somada a média do vetor de respostas das amostras de calibração $\bar{y}$ em y para que a resposta não seja centralizada. Para a predição de mais amostras, basta repetir as 4 etapas acima para cada amostra.