



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE QUÍMICA

DEPARTAMENTO DE QUÍMICA ANALÍTICA

**DETERMINAÇÃO DE PARÂMETROS DE QUALIDADE EM
MADEIRA UTILIZANDO ESPECTROSCOPIA NO
INFRAVERMELHO PRÓXIMO E CALIBRAÇÃO
MULTIVARIADA**

Dissertação de Mestrado

Laura Barbosa Lima Nascimbem de Godoy

Orientador: Prof. Dr. Ronei Jesus Poppi

Campinas, 15 de julho de 2010

**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE
QUÍMICA DA UNICAMP**

G548d Godoy, Laura Barbosa Lima Nascimbem de.
Determinação de parâmetros de qualidade em
madeira utilizando espectroscopia no infravermelho
próximo e calibração multivariada / Laura Barbosa Lima
Nascimbem de Godoy. -- Campinas, SP: [s.n], 2010.

Orientador: Ronei Jesus Poppi.

Dissertação - Universidade Estadual de Campinas,
Instituto de Química.

1. NIR. 2. Calibração multivariada. 3. Madeira.
4. Teor de umidade. I. Poppi, Ronei Jesus.
II. Universidade Estadual de Campinas. Instituto de
Química. III. Título.

Título em inglês: Determination of wood quality parameters using near infrared spectroscopy and multivariate calibration

Palavras-chaves em inglês: NIR, Multivariate calibration, Wood, Moisture content

Área de concentração: Química Analítica

Titulação: Mestre em Química na área de Química Analítica

Banca examinadora: Prof. Dr. Ronei Jesus Poppi (orientador), Prof. Dr. Edson Irineu Müller (DQ-UFSM), Prof. Dr. Jarbas José Rodrigues Rohwedder (IQ-UNICAMP)

Data de defesa: 15/07/2010

AGRADECIMENTOS

- Ao meu orientador Prof. Dr. Ronei Jesus Poppi pela oportunidade de realizar este trabalho, pela orientação, conhecimentos transmitidos e amizade.
- Aos amigos do grupo LAQQA: Danilo Althmann, Patrícia Valderrama, Paulo H. Março, Marcelo G. Trevisan, Guilherme P. Sabin, Werickson Rocha, Márcio Pedroso, Luiz A.F. de Godoy Junior, Mônica López, Renato Carneiro pelo convívio e amizade.
- À Fibria Celulose SA, por permitir a obtenção dos dados experimentais na empresa, estrutura física e pelo financiamento do projeto.
- A Bibiana Ribeiro Rubini e Alessandra Foresti Caldeira pela orientação e acompanhamento do trabalho.
- Aos colegas da Fibria – ES: Bruno, Lousival, Fátima, Eliene, Gil, Jair, FEM, Renato, Helder, Marcelo, Roberto, Jucemir, Jocemar, Mila, Flávia, Jaqueline, Deusa, por toda colaboração, incentivo, apoio durante a realização do trabalho e a receptividade e amizade.
- Ao Gilmar e a todos os pesquisadores da Fibria pela oportunidade de convívio e pelo aprendizado.
- Aos professores Maria Izabel Maretti Silveira Bueno, Jarbas José Rodrigues Rohwedder e Edson Irineu Müller pelas contribuições dadas ao trabalho no exame de qualificação e na defesa da dissertação.
- À UNICAMP pelo fornecimento de toda estrutura física e tecnológica para a realização deste trabalho.
- A todos que contribuíram direta ou indiretamente para a execução este trabalho.
- Aos meus amigos (as): Marina, Marla, Renan, Rebeca, Camila, Nathalie, Mônica, Márcio, Humberto, Ana pelo companheirismo e amizade durante todo esse tempo.
- A minha família: em especial ao meu pai Luiz Nascimbem e minha avó Fanny Nascimbem e meu marido Luiz Antonio F. de Godoy Junior por todo esforço realizado comigo, apoio, compreensão, força e por estarem ao meu lado sempre, nos momentos mais alegres e mais difíceis também. Sem vocês eu não teria conseguido chegar até aqui.

CURRICULUM VITAE

Dados Pessoais

Nome – Laura Barbosa Lima Nascimbem de Godoy

Nascimento – 05/04/1984

Naturalidade – Campinas –SP

E-mail – lauranascimbem@uol.com.br

Formação

Graduação – Bacharelado em Química – 2003-2008

Graduação – Bacharelado em Química Tecnológica – 2003-2009

UNICAMP – Universidade Estadual de Campinas – Campinas-SP

Experiência Profissional

- Fibria Celulose SA – 08/2008-07/2010

Função – Pesquisa e desenvolvimento de metodologia utilizando espectroscopia no infravermelho próximo para controle de qualidade de cavacos de madeira

- Contech Produtos Biodegradáveis Ltda – 01/2007-04/2008

Função – Estagiária de Pesquisa e Desenvolvimento

Publicações

- Nascimbem, L.B.L.R.; Almeida, M.R.; Alves, R.S.; Stephani, R.; Oliveira, L. F. C.; Poppi, R. J. “Determination of amylose content in starch using Raman Spectroscopy and Multivariate Calibration analysis”. *Anal. Bioanal. Chem.*, **2010** – *in press*.

- Nascimbem, L.B.L.R.; Barreto, R.L.; Pavam, C.H.; Rezende, L.F. “Controle Químico de *Pitch* e *Stickies*: Abordando suas Causas, Efeitos Adversos sobre os Processos de Fabricação de Papel & Celulose e Experiências de Remediação Bem Sucedidas”; *Pasta e Papel*, dezembro, **2007**;

- Nascimbem, L. B. L. R.; Barreto, R.L.; Correia, C. R. D. “Synthesis of the 3-(3,4,5-trimethoxyphenyl)-pyrrolidine: a new conformationally constrained mescaline analogue”. *Synth. Commun.* **2007**, *12*, 2011.

- Patente depositada no INPI: “Processo de preparação dos compostos fluorescentes anidridos malêicos arilados, acrilatos substituídos arilados e maleimidas ariladas.” Sob o protocolo de número 018060068565, de 30.06.06.

Trabalhos em Eventos Internacionais

- Nascimbem, L.B.L.R.; Rubini, B.R; Caldeira, A.F.; Poppi, R.J. “Determination of quality parameters in woodchips with different moisture contents using NIR spectroscopy” – VII Colloquium Chemiometricum Mediterraneum, Granada – Spain – 21 a 24 de junho de 2010.

Trabalhos em Eventos Nacionais

- 33ª Reunião Anual da Sociedade Brasileira de Química, realizada de 28 a 31/05/2010 na cidade de Águas de Lindóia –SP com painel sob título: “Classificação de amostras de madeira conforme o teor de umidade utilizando Espectroscopia no Infravermelho Próximo e PLS-DA”.

- 15º ENQA – Encontro Nacional de Química Analítica realizado em Salvador – Bahia nos dias 18 à 21/10/2009 com painel sob título: “Determinação de densidade diretamente em cavacos de madeira com Espectroscopia no Infravermelho Próximo e Calibração Multivariada”

- 1º Seminário Tecnológico CONTECH, Valinhos-SP, 23 a 25 de maio, 2007;

- 29ª Reunião Anual da Sociedade Brasileira de Química, realizada nos dias 22 a 24/05/2006 na cidade de Águas de Lindóia-SP com painel sob título “Síntese de novos análogos da mescalina conformacionalmente restringidos”.

RESUMO

DETERMINAÇÃO DE PARÂMETROS DE QUALIDADE EM MADEIRA UTILIZANDO ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E CALIBRAÇÃO MULTIVARIADA

Autora: Laura B. L. Nascimbem de Godoy

Orientador: Ronei Jesus Poppi

Neste trabalho, foi utilizada a Espectroscopia no Infravermelho Próximo (NIR) na região de $4.000 - 10.000 \text{ cm}^{-1}$ e modelos de calibração multivariada para a previsão de densidade básica, lignina total e extrativos em madeira. Para isso, foi utilizado o algoritmo dos mínimos quadrados parciais por intervalos (iPLS) obtendo espectros NIR diretamente sobre os cavacos de madeira secos. Um total de 107 e 53 amostras compõem os conjuntos de calibração e validação, respectivamente, sendo a divisão destas amostras realizada pelo algoritmo de Kennard – Stone. A calibração foi otimizada pela eliminação das amostras anômalas (*outliers*), com base nas amostras com *leverage* extremo, resíduos não modelados nos dados espectrais e os resíduos com base na repetibilidade espectral. Também foi realizado um estudo visando à classificação e a determinação destes parâmetros em amostras com diferentes teores de umidade. Neste estudo, o método dos mínimos quadrados parciais para análise discriminante (PLS-DA) foi utilizado para a classificação das amostras de acordo com o teor de umidade apresentado e as máquinas de vetores de suporte por mínimos quadrados (LS-SVM) foi utilizada para determinação de cada um dos parâmetros em cada classe com teor de umidade diferente. Os resultados obtidos indicam que os modelos desenvolvidos para os cavacos secos podem ser utilizados na indústria papelreira como alternativa aos métodos convencionais e os modelos desenvolvidos no estudo para as amostras úmidas podem ser utilizados para a classificação de uma amostra desconhecida e, posteriormente, a previsão de cada parâmetro em cada um dos teores de umidade.

ABSTRACT

DETERMINATION OF WOOD QUALITY PARAMETERS USING NEAR INFRARED SPECTROSCOPY AND MULTIVARIATE CALIBRATION

Author: Laura B. L. Nascimbem de Godoy
Adviser: Ronei Jesus Poppi

In this work, it was used the near infrared spectroscopy (NIRS) in region of $4.000 - 10.000 \text{ cm}^{-1}$ and multivariate calibration models to predict basic density, total lignin and extractives in woodchips. The spectra were obtained directly on the dry woodchips and interval partial least squares (iPLS) was used to develop the calibration model. A total of 107 and 53 samples constitute the calibration and validation sets, respectively, split by Kennard – Stone algorithm. The calibration set was optimized by outliers elimination based on data with extreme leverage, unmodelled residuals in spectral data and unmodelled residuals in the dependent variable. It was also performed a study aiming the classification and determination of this parameters in samples with different moisture contents. In this study, the discriminant analysis with partial least squares method (PLS-DA) was used to classify samples according to the moisture content and the least squares supporting vector machines (LS-SVM) was used to determine density, total lignine and extractives in each class with different moisture content. The results indicate that the models developed for the dry woodchips can be used in the pulp and paper industry as an alternative to the conventional methods and the models developed in the study with samples presented different moisture content can be used to classify an unknown sample and to predict each parameter in each moisture content class.

LISTA DE ABREVIATURAS

- AOTF** – Filtro Óptico Acústico Sintonizável (do inglês, Acousto-Optic Tunable Filter)
- ASTM** – American Society for Testing and Materials
- CV** – Validação Cruzada (do inglês, Cross Validation)
- DB** – Densidade Básica
- E** – Extrativos
- FAR** – Infravermelho Distante (do inglês, Far Infrared)
- iPLS** – Mínimos Quadrados Parciais por Intervalos (do inglês, Interval Partial Least Square)
- LS-SVM** – Máquinas de Vetores de Suporte por Mínimos Quadrados (do inglês, Least Square Supporting Vector Machines)
- MID** – Infravermelho Médio (do inglês, Middle Infrared)
- MLR** – Regressão Linear Múltipla (do inglês, Multiple Linear Regression)
- MSC** – Correção do Espalhamento Multiplicativo (do inglês, Multiple Scattering Correction)
- NIPALS** – Nonlinear Iterative Partial Least Squares
- NIR** – Infravermelho Próximo (do inglês Near Infrared)
- PCs** – Componentes Principais (do inglês, Principal Components)
- PCA** – Análise de Componentes Principais (do inglês, Principal Components Analysis)
- PCR** – Regressão por Componentes Principais (do inglês, Principal Component Regression)
- PDS** – Método para Selecionar Amostras (do inglês, Piecewise Direct Standardisation)
- PLS** - Mínimos Quadrados Parciais (do inglês, Partial Least Squares)
- PLS-DA** – Análise Discriminante com método dos Mínimos Quadrados Parciais (do inglês, Partial Least Squares with Discriminant Analysis)
- RBF** – Função de Base Radial (do inglês, Radial-Base Function)
- RMSEC** – Raiz Quadrada do Erro Médio Quadrático da Calibração (do inglês, Root Mean Square Error of Calibration)

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

RMSEP - Raiz Quadrada do Erro Médio Quadrático de Previsão (do inglês, Root Mean Square Error of Prediction)

SVD – Decomposição por Valor Singular (do inglês, Singular Value Decomposition)

UV – Ultra Violeta

UVE-PLS – Mínimos Quadrados Parciais com Eliminação de Variáveis não Informativas (do inglês, Elimination of Uninformative Variables in Partial Least Square)

VIS – Visível

VL – Variável Latente

v - estiramento

δ – deformação

LISTA DE TABELAS

Tabela 1. Composição dos materiais ligninocelulósicos das madeiras	4
Tabela 2. Regiões espectrais do infravermelho.	17
Tabela 3. Valores previstos pelo modelo.	66
Tabela 4. Atribuição das bandas no NIR.	69
Tabela 5. Valores previstos pelo modelo.	72
Tabela 6. Atribuição das bandas no NIR.	75
Tabela 7. Valores previstos pelo modelo.	77
Tabela 8. Atribuição das bandas no NIR.	80
Tabela 9. Características dos modelos construídos para cada classe com o LS-SVM.	93
Tabela 10. Características dos modelos construídos para cada classe com o LS-SVM.	96
Tabela 11. Características dos modelos construídos para cada classe com o LS-SVM.	99

LISTA DE FIGURAS

Figura 1. Aspectos macroscópicos da madeira em uma seção transversal do tronco.	3
Figura 2. Representação da estrutura da molécula de celulose.	4
Figura 3. Estrutura da lignina	5
Figura 4. Diagrama de energia potencial para (A) oscilador harmônico, (B) oscilador anarmônico.	21
Figura 5. Construção da matriz X para calibração multivariada.	32
Figura 6. Representação esquemática da decomposição por PCA.	33
Figura 7. Decomposição em variáveis latentes das matrizes X e Y para modelos PLS.	37
Figura 8. Mudança do espaço dos dados através da função Kernel (ϕ).	41
Figura 9. Espectros originais das 160 amostras de cavaco.	63
Figura 10. Primeira derivada dos espectros originais para as 160 amostras.	64
Figura 11. iPLS com 7 intervalos, onde as barras indicam o erro de cada intervalo, a linha tracejada o erro obtido com o espectro inteiro e em itálico o número de VLs para o modelo em determinado intervalo e a linha contínua é o espectro médio obtido para o modelo	65
Figura 12. Valores de densidade de referência (análise química convencional) <i>versus</i> previstos pelo modelo para as amostras de calibração (●) e validação (▲).	66
Figura 13. Gráfico com barras de erro relativo para cada amostra de validação.	67
Figura 14. a) Gráfico do <i>leverage</i> das amostras de calibração. b) Gráfico dos resíduos espectrais das amostras de calibração.	68
Figura 15. a) Gráfico do <i>leverage</i> das amostras de validação. b) Gráfico dos resíduos espectrais das amostras de validação.	69
Figura 16. Espectros das 160 amostras após pré-processamento com MSC.	70

Figura 17. iPLS com 7 intervalos, onde as barras indicam o erro de cada intervalo, a linha tracejada o erro obtido com o espectro inteiro e em itálico o número de VLs para o modelo em determinado intervalo e a linha contínua é o espectro médio obtido para o modelo	71
Figura 18. Valores de lignina de referência (análise química convencional) <i>versus</i> previstos pelo modelo para as amostras de calibração (●) e validação (▲).	72
Figura 19. Gráfico com barras de erro relativo para cada amostra de validação.	73
Figura 20. a) Gráfico do <i>leverage</i> das amostras de calibração. b) Gráfico dos resíduos espectrais das amostras de calibração.	74
Figura 21. a) Gráfico do <i>leverage</i> das amostras de validação. b) Gráfico dos resíduos espectrais das amostras de validação.	74
Figura 22. Espectros das amostras pré-processadas com o algoritmo MSC.	76
Figura 23. iPLS com 7 intervalos, onde as barras indicam o erro de cada intervalo, a linha tracejada o erro obtido com o espectro inteiro e em itálico o número de VLs para o modelo em determinado intervalo e a linha contínua é o espectro médio obtido para o modelo	76
Figura 24. Valores de extrativos em etanol/tolueno de referência (análise química convencional) <i>versus</i> previstos pelo modelo para as amostras de calibração (●) e validação (▲).	77
Figura 25. Gráfico com barras de erro relativo para cada amostra de validação.	78
Figura 26. a) Gráfico do <i>leverage</i> das amostras de calibração. b) Gráfico dos resíduos espectrais das amostras de calibração.	79
Figura 27. a) Gráfico do <i>leverage</i> das amostras de validação. b) Gráfico dos resíduos espectrais das amostras de validação.	80
Figura 28. Espectros originais das 92 amostras com diferentes teores de umidade.	83

Figura 29. Espectros dos cavacos com vários teores de umidade diferente, mostrando a influência da água.	84
Figura 30. Espectros das 92 amostras após utilização do MSC como pré-processamento.	85
Figura 31. Gráfico dos escores da primeira componente principal <i>versus</i> escores da segunda componente principal. Amostras secas (◆), aspensão (▼), vaporização (✦) e imersão (■).	85
Figura 32. Gráfico dos pesos do PCA na PC1 (esquerda) e na PC2 (na direita).	86
Figura 33. Classificação das amostras da classe 1 (aspensão ▼) e também de amostras de validação (●).	87
Figura 34. Classificação das amostras da classe 2 (vaporização ✦) e também de amostras de validação (●).	88
Figura 35. Classificação das amostras da classe 3 (imersão ■) e também de amostras de validação (●).	88
Figura 36. Classificação das amostras da classe 4 (secas naturalmente ao ar ◆) e também de amostras de validação (●).	89
Figura 37. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de aspensão. b) Gráfico dos valores de densidade de referência <i>versus</i> previstos pelo modelo de LS-SVM para a classe de aspensão com as amostras de calibração (●) e validação (▲).	91
Figura 38. a) Superfície mostrando os valores dos parâmetros γ e σ^2 para amostras de vaporização. b) Gráfico dos valores de densidade de referência <i>versus</i> previstos pelo modelo de LS-SVM para a classe de vaporização para amostras de calibração (●) e validação (▲).	92

Figura 39. a) Superfície mostrando os valores para os parâmetros γ e σ^2 em amostras de imersão. b) Gráfico dos valores de densidade de referência *versus* previstos pelo modelo de LS-SVM para a classe de imersão para amostras de calibração (●) e validação (▲). 92

Figura 40. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras secas naturalmente ao ar. b) Gráfico dos valores de densidade de referência *versus* previstos pelo modelo de LS-SVM para a classe de amostras secas naturalmente ao ar para amostras de calibração (●) e validação (▲). 93

Figura 41. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de aspersão. b) Gráfico dos valores de lignina de referência *versus* previstos pelo modelo de LS-SVM para a classe de aspersão para amostras de calibração (●) e validação (▲). 94

Figura 42. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de vaporização. b) Gráfico dos valores de lignina de referência *versus* previstos pelo modelo de LS-SVM para a classe de vaporização para as amostras de calibração (●) e validação (▲). 95

Figura 43. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de imersão. b) Gráfico dos valores de lignina de referência *versus* previstos pelo modelo de LS-SVM para a classe de imersão para as amostras de calibração (●) e validação (▲). 95

Figura 44. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras secas naturalmente ao ar. b) Gráfico dos valores de lignina de referência *versus* previstos pelo modelo de LS-SVM para a classe de amostras secas naturalmente ao ar para as amostras de calibração (●) e validação (▲). 96

Figura 45. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de aspersão. b) Gráfico dos valores de extrativos de referência *versus* previstos pelo modelo de LS-SVM para a classe de aspersão com as amostras de calibração (●) e validação (▲). 97

Figura 46. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de vaporização. b) Gráfico dos valores de extrativos de referência *versus* previstos pelo modelo de LS-SVM para a classe de vaporização para as amostras de calibração (●) e validação (▲). 98

Figura 47. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de imersão. b) Gráfico dos valores de extrativos de referência *versus* previstos pelo modelo de LS-SVM para a classe imersão com as amostras de calibração (●) e validação (▲). 98

Figura 48. a) Superfície mostrando os valores obtidos para os parâmetros γ e σ^2 para amostras secas naturalmente ao ar. b) Gráfico dos valores de extrativos de referência *versus* previstos pelo modelo de LS-SVM para a classe de amostras secas naturalmente ao ar para as amostras de calibração (●) e validação (▲). 99

ÍNDICE

CAPÍTULO 1 – Madeira e a indústria papelreira	1
1.1 Madeira	3
1.2 Controle de Qualidade	7
1.3 Tendências tecnológicas do setor de papel e celulose	10
CAPÍTULO 2 – Espectroscopia no Infravermelho	13
2.1 Histórico	15
2.2 Aplicações da Espectroscopia no Infravermelho	16
2.3 Fundamentos da Espectroscopia no Infravermelho	19
2.4 Aplicações da Espectroscopia no Infravermelho no setor de papel e celulose	24
CAPÍTULO 3 – Métodos de análise multivariada	27
3.1 Quimiometria	29
3.2 Calibração	30
3.3 Pré-processamentos	32
3.4 Análise de Componentes Principais (PCA)	33
3.5 Mínimos Quadrados Parciais (PLS)	36
3.6 Mínimos Quadrados Parciais por intervalos (iPLS)	37
3.7 Análise Discriminante com método de mínimos quadrados parciais (PLS-DA)	39
3.8 Máquinas de Vetores de Suporte por Mínimos Quadrados (LS-SVM)	39
3.9 Algoritmo de Kennard-Stone	43
3.10 Detecção de amostras anômalas – <i>Outliers</i>	44
3.10.1 Amostras anômalas na calibração	45
3.10.2 Amostras anômalas na validação	47
CAPÍTULO 4 – Objetivos	49
CAPÍTULO 5 – Parte Experimental	53
5.1 Amostras	55
5.2 Espectros	55
5.3 Construção dos modelos quimiométricos	55
5.3 Densidade básica	56
5.4 Extrativos em etanol/tolueno	56
5.5 Lignina total	57
5.6 Amostras úmidas	58
CAPÍTULO 6 – Resultados e Discussões – Amostras secas	61
6.1 Densidade básica	64
6.2 Lignina total	70
6.3 Extrativos em etanol/tolueno	75

CAPÍTULO 7 – Resultados e Discussões – Amostras úmidas	81
7.1 Análise de Componentes Principais (PCA)	84
7.2 Análise Discriminantes com método dos mínimos quadrados parciais (PLS-DA)	87
7.3 Modelos de LS-SVM para cada classe	90
7.3.1 Densidade básica	90
7.3.2 Lignina total	94
7.3.3 Extrativos em etanol/tolueno	97
CAPÍTULO 8 – Conclusões	101
CAPÍTULO 9 - Referências	105

CAPÍTULO 1
Madeira e a
Indústria papelreira

1. Madeira e a indústria papelreira

1.1 Madeira

A madeira é um polímero natural complexo, constituída por algumas substâncias macromoleculares orgânicas como a celulose, a hemicelulose e a lignina. Além disso, a madeira possui em sua composição outros constituintes menores, porém não menos importantes como os extrativos ¹.

Desde os tempos remotos, a madeira vem sendo utilizada como material de construção e como matéria-prima para produção de carvão, alcatrão e piche. Também é a principal matéria-prima para produção de papel e celulose, que é renovável e, portanto, é um bem de valor indiscutível para toda humanidade ².

A estrutura geral da madeira possui as seguintes partes, que são distinguidas na seção transversal do tronco: casca, floema, câmbio, cerne, alburno e medula. A casca tem a função de proteger os tecidos vivos da árvore contra o ressecamento, o ataque de microrganismos e insetos e às condições mecânicas e climáticas ³.

O floema é o principal tecido condutor da seiva que carrega os nutrientes das plantas vasculares. O câmbio é o tecido responsável pela formação do xilema e do floema. O xilema é constituído pelo cerne (células inativas) e alburno (células funcionais) e é o principal tecido condutor de água. Por fim, a medula é um tecido primário, localizado na região central do caule e responsável pelo armazenamento de substâncias nutritivas para a planta ³. A Figura 1 apresenta uma seção transversal do tronco e suas partes principais.

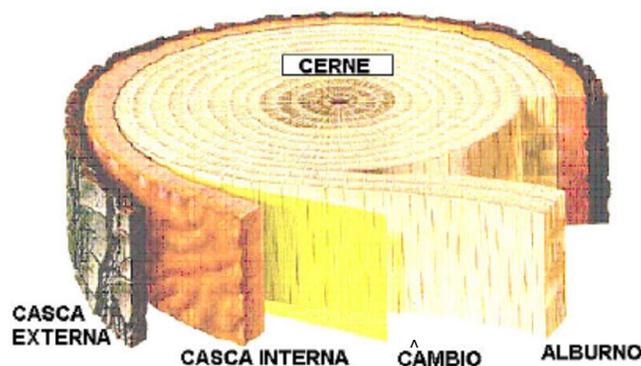


Figura 1. Aspectos macroscópicos da madeira em uma seção transversal do tronco.

Os materiais ligninocelulósicos da madeira são compostos basicamente por celulose, hemiceluloses, lignina e constituintes menores, como apresentado na Tabela 1. Os constituintes menores incluem compostos orgânicos de diversas funções e também compostos inorgânicos. Os constituintes menores que são solúveis em solventes orgânicos, em água ou em ambos são denominados extrativos³.

Tabela 1. Composição dos materiais ligninocelulósicos das madeiras.

Componente	Presença (%)
Celulose	~ 50
Hemiceluloses	~ 20
Lignina	15 a 35
Constituintes menores	Até ~ 10

A celulose é um polissacarídeo (Figura 2) formado por unidades do monossacarídeo β - D - glicose, que se ligam entre si através dos carbonos 1 e 4, originando um polímero linear.

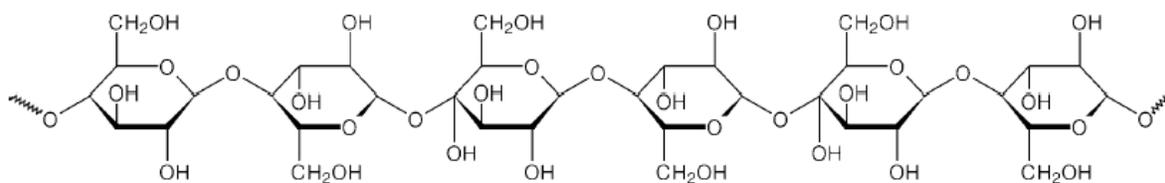


Figura 2. Representação da estrutura da molécula de celulose.

A formação da celulose ocorre através de sucessivas reações de desidratação entre as hidroxilas do carbono 1 da β - D - glicose e o carbono 4 de outras unidades deste mesmo monômero. As moléculas de celulose tendem a formar ligações de hidrogênio intramoleculares e intermoleculares. As ligações intramoleculares são responsáveis pela rigidez das cadeias e as intermoleculares pela formação da fibra vegetal³.

As hemiceluloses recebem esta designação por se tratarem de uma mistura de polímeros polissacarídeos de baixa massa molecular, os quais estão intimamente associados com a celulose nos tecidos das plantas. As hemiceluloses apresentam em sua composição açúcares condensados em proporções variadas e sempre estão presentes pelo menos dois tipos de açúcares ³.

As hemiceluloses são responsáveis por diversas propriedades importantes das pastas celulósicas. O fato de absorverem água facilmente contribui para a mobilidade interna e o aumento da flexibilidade das fibras, para a redução do tempo e da energia requeridos no refino das pastas e no aumento da área específica ou de ligação das fibras ³.

A lignina é uma substância química que confere rigidez à parede da célula e nas partes da madeira age como um agente permanente de ligação entre as células, gerando uma estrutura resistente ao impacto, compressão e dobra. Ela atua no transporte interno de água, nutrientes e metabólitos. Tecidos lignificados resistem ao ataque por microrganismos, impedindo a penetração de enzimas destruidoras da parede celular ³.

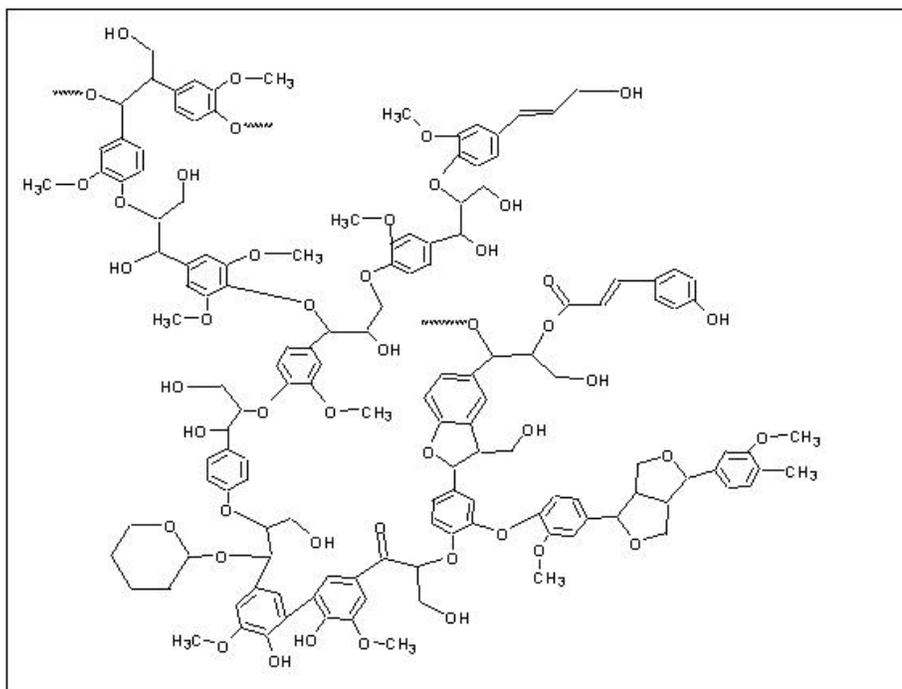


Figura 3. Estrutura da lignina.

Os constituintes menores dividem-se basicamente em duas classes: a dos extrativos, por serem extraíveis com água ou solventes orgânicos neutros e a segunda classe dos compostos que não são extraíveis nos solventes mencionados anteriormente. Esses constituintes presentes em menores quantidades são responsáveis por determinadas características das plantas como cor, cheiro, resistência natural ao apodrecimento e sabor. Sua composição e quantidade relativa dependem de diversos fatores como a espécie, idade da planta e região de ocorrência ³.

Os extrativos podem ser classificados em vários grupos de acordo com suas características estruturais. Os principais constituintes são os terpenos, lignanas, estilbenos, flavonóides, entre outros. Além disso, outros compostos orgânicos podem estar presentes nos extrativos, como gorduras, ceras, ácidos graxos, alcoóis, esteróides e hidrocarbonetos de elevada massa molecular ³.

Os não-extrativos são compostos inorgânicos e estão presentes em apenas pequena quantidade. Estes são principalmente sulfatos, fosfatos, oxalatos, carbonatos e silicatos de cálcio, potássio e magnésio. Apesar de estarem presentes em pequena quantidade, estão combinados com substâncias orgânicas e assim, exercem funções fisiológicas importantes para o metabolismo da planta ³.

No Brasil, a madeira utilizada para a produção da pasta celulósica provém principalmente das diversas espécies de eucalipto (*Eucalyptus* spp) e pinus (*Pinus* spp), sendo que as espécies de eucalipto são as mais utilizadas ³.

O eucalipto é proveniente da Austrália e possui espécies que podem suportar temperatura em torno de 35 °C, enquanto outras suportam até dezoito graus negativos. É uma das maiores árvores do mundo e possui uma taxa de crescimento muito elevada, o que a torna altamente renovável, podendo ser utilizada para fins comerciais, como para a produção de papel e celulose, construção civil, setor moveleiro, entre outros ⁴.

No Brasil, a idade ideal de corte do eucalipto para a produção de celulose, energia e carvão ocorre entre 6 e 7 anos após o seu plantio. Este rápido ciclo de crescimento do eucalipto faz com que o Brasil tenha os maiores índices de produtividade do mundo. O eucalipto pode ser cultivado em terrenos com baixa

fertilidade natural e não exige muitos nutrientes e defensivos agrícolas comparado às outras culturas ⁴.

Atualmente na indústria de papel e celulose brasileira, diferentes métodos de melhoramento genético são empregados para a obtenção de gerações avançadas de algumas espécies de eucalipto. Essas técnicas envolvem a seleção de plantas com melhor desempenho e seu inter cruzamento ao longo de sucessivas gerações. Desta forma, um conjunto de clones é utilizado para fins comerciais e estes são frequentemente substituídos, garantindo diversidade genética às florestas plantadas e reduzindo a probabilidade de ataque de pragas e doenças ⁴.

1.2 Controle de Qualidade

A qualidade da madeira como matéria-prima industrial pode ser expressa através de uma série de características intrínsecas de cada madeira. Isto definirá seu potencial para produção de pasta celulósica para a indústria papelreira.

A densidade básica é um dos parâmetros mais utilizados na indústria papelreira para a determinação da qualidade da madeira. Este parâmetro possui elevada importância, pois está altamente relacionado com a determinação de propriedades físicas e mecânicas que caracterizam as diferentes espécies de madeira, diferentes árvores de uma dada espécie e diferentes regiões de uma mesma árvore ⁵.

Na indústria de papel e celulose, a densidade básica influencia diversos processos. Como a madeira é usualmente comprada em volume, no processamento é desejável o conhecimento do seu peso seco para o controle do processo. Além disso, é importante para a determinação do rendimento do processo tanto no rendimento por digestor individual, quanto no rendimento por unidade de volume ⁵.

Outros fatores são importantes, como a existência de uniformidade dentro da mesma madeira para que o produto final esteja padronizado e principalmente o fato de a velocidade de impregnação do licor de cozimento e conseqüente ritmo

de deslignificação serem influenciados pela densidade básica, uma vez que espera-se que dentro de uma mesma espécie as madeiras menos densas sejam mais facilmente deslignificadas ⁵.

A densidade básica pode ser definida como a relação entre o peso seco da amostra analisada e seu respectivo volume saturado. Em uma mesma árvore, a densidade básica varia no sentido longitudinal (base-topo) e também no sentido transversal (medula-casca). No sentido longitudinal, os diferentes gêneros ou espécies de árvores apresentam diversos modelos de variação (lineares, quadráticos, logarítmicos, etc). Além disso, observa-se que a densidade básica aumenta com a idade da árvore até certo ponto, e depois se estabiliza ⁵.

O cálculo da densidade básica através do método de imersão ^{5,6} de cavacos de madeira em água é realizado utilizando a fórmula:

$$DB = \frac{(M_4 - M_1)}{(M_3 - M_2)} \cdot 1000 \quad (1)$$

onde M_4 é a massa seca do cesto (volume aproximado de 3000 cm³) mais a massa da amostra em gramas, DB é a densidade básica dos cavacos em Kg/m³, M_1 é a massa seca do cesto em gramas, M_2 é a massa hidrostática do cesto em gramas e M_3 é a massa hidrostática (úmida) do cesto mais a massa da amostra em gramas.

A lignina apresenta características intrínsecas, por ser uma substância estável, hidrofóbica, fundamentalmente aromática e de alto peso molecular, integrando-se às necessidades das plantas pelo sistema de vascularização. A lignina, além de ter adaptado as plantas à vida terrestre no processo evolutivo, também impermeabilizou a parede celular, permitindo o transporte de água e soluções pelo sistema vascular ⁷.

O cálculo da porcentagem de lignina total é realizado através da obtenção das concentrações da lignina solúvel e insolúvel. A concentração da lignina solúvel (em água e H₂SO₄) é realizada de acordo com a fórmula ⁸:

$$C1 = \frac{(3,595 \cdot A206) - A246}{339,5} \quad (2)$$

onde C1 é a concentração de lignina solúvel, A206 é a absorvância a 206 nm, A246 é a absorvância a 246 nm e os outros dois valores são constantes de correção.

Assim, a porcentagem de lignina solúvel é:

$$\% \text{ lignina solúvel} = \frac{(C1.3,833)}{\text{peso seco}} \times 100 \quad (3)$$

onde 3,833 é o fator de diluição utilizado.

A lignina insolúvel é determinada através da fórmula:

$$\% \text{ lignina insolúvel} = \frac{A}{W} \times 100 \quad (4)$$

onde A é o peso de lignina em gramas, W é o peso seco (massa de amostra utilizada vezes a consistência) da amostra em teste em gramas.

Com os valores das porcentagens de lignina solúvel e de lignina insolúvel é possível a determinação da lignina total, como a seguir:

$$\text{Lignina Total} = \text{lignina solúvel} + \text{lignina insolúvel}. \quad (5)$$

O teor de extrativos é um dos mais importantes indicadores de conformidade da madeira para diversos usos industriais. Durante a formação do cerne, uma ampla variedade de substâncias extrativas, incluindo taninos, corantes, óleos, gomas, resinas e sais de ácidos orgânicos acumula-se nas células e nas paredes celulares, resultando, às vezes, em uma coloração mais escura da madeira, além de um aumento da densidade e da durabilidade.

Os extrativos aumentam com a idade e com a posição do disco ao longo do tronco no sentido base-topo. Essas mesmas observações podem ser feitas para a

lignina, pois seu teor também aumenta no sentido longitudinal do tronco e com o aumento da idade da árvore.

O cálculo do teor de extrativos presentes em uma amostra é realizado de acordo com a fórmula ⁹:

$$E\% = \frac{(M_2 - M_1) - M_3}{M.S} \times 100 \quad (6)$$

onde E é a porcentagem de extrativos em etanol/tolueno, M é a massa em gramas da amostra utilizada, M₁ é a massa do copo de extração ou frasco coletor em gramas, M₂ é a massa do frasco coletor mais a massa do material extraído e seco em gramas, S é o teor seco da amostra em porcentagem e M₃ é a massa da prova em branco.

1.3 Tendências tecnológicas do setor de papel e celulose

A demanda por celulose de fibra curta de eucalipto apresenta perspectivas positivas para os fabricantes do hemisfério sul. Isto se deve a alguns fatores como o fechamento de fábricas menores e de custo mais elevado na América do Norte e na Europa, a substituição do uso de fibra longa por fibra curta pelos fabricantes de papel e ao grande crescimento do mercado chinês, que está montando um parque industrial papelero com grande capacidade, mas que não dispõe da celulose necessária para atendê-lo. Portanto, a tendência observada para a indústria de celulose de mercado no Brasil é de um novo ciclo de expansão para atender à crescente demanda de celulose de mercado ¹⁰.

Os diversos segmentos do setor de papel apresentam características e tendências distintas. O aumento nas demandas por papéis sanitários, papéis especiais e papelão ondulado é basicamente dependente do crescimento da economia brasileira, e sua estrutura produtiva comporta um grande número de empresas médias e pequenas. Já os segmentos de papel de imprimir e escrever,

kraftliner e cartão são abastecidos por grandes empresas, que encontram nas exportações um método de impulsionar as vendas, incentivando investimentos em expansão de capacidade ¹⁰.

A indústria de papel apresenta uma tendência mundial de consolidação e aumento do grau de concentração, com a formação e o crescimento de empresas globais. As grandes empresas internacionais têm elevadas escalas de produção, amplos canais de distribuição de seus produtos e baixos custos de capital. Entretanto, essas empresas não são favorecidas pelo baixo custo de matéria-prima e suas plantas menos competitivas estão diante da possibilidade de fechamento ou paralisação temporária de suas atividades. Verifica-se, portanto, uma tendência das grandes empresas internacionais em investir na implantação de unidades industriais em mercados com potencial de crescimento ¹⁰.

Um dos fatores mais importantes da indústria de papel está relacionado à reciclagem. Por se tratar de uma matéria-prima natural, a busca pela melhoria na qualidade e no processo de produção de papel reciclado são desafios futuros para a indústria de papel e celulose ¹¹.

Outra tendência relacionada ao processo e ao meio ambiente, é a redução dos odores de enxofre provenientes das mercaptanas que são liberadas durante o processo de remoção da lignina (cozimento das madeiras). Além disso, estudos visando a substituição de compostos organoclorados, formados pela reação do cloro ou hipoclorito com a lignina durante a etapa de branqueamento da celulose têm sido realizados na busca de reagentes alternativos para o branqueamento, como o ozônio e o peróxido de hidrogênio ¹¹.

Por fim, uma tendência relevante em quase todos os setores industriais atualmente é a automação na produção. No setor de papel e celulose, essa automação já é elevada, sobretudo na produção, porém o controle de qualidade ainda necessita de muita mão-de-obra e possui um elevado custo além de consumir um grande tempo para a realização das análises em laboratório. Desta forma, uma automação no controle de qualidade da matéria-prima destinada à produção, geraria uma economia e melhores previsões sobre quantidades de reagentes a serem utilizados e qualidade do produto final para esse setor.

CAPÍTULO 2
Espectroscopia no Infravermelho

2. Espectroscopia no Infravermelho

2.1 Histórico

A descoberta da Espectroscopia no Infravermelho é atribuída ao músico e astrônomo alemão Frederick William Herschel ¹². Este cientista dedicava seu tempo à composição e também à astronomia, sendo que sua grande contribuição no campo astronômico foi a descoberta do planeta Urano em 1781. Porém, a grande contribuição para ciência foi durante a execução de um experimento no qual Herschel estudava as propriedades físicas da radiação eletromagnética. Neste experimento, Herschel queria observar a contribuição de cada cor proveniente da luz branca dispersa por um prisma no aumento da temperatura das substâncias expostas à ela. Herschel continuou observando um aumento na temperatura de um termômetro abaixo do vermelho, onde não havia mais luz visível, e pode notar que esta era a região que apresentava maior temperatura. Desta forma, em 1800, a primeira parte do espectro eletromagnético não visível era relatada ¹³.

Mais tarde, essa radiação passou a ser chamada de infravermelha, sendo esta uma grande contribuição, pois foi demonstrado que existem outras formas de luz que não podem ser observadas pelo olho humano. A partir de seu experimento, outros estudos foram realizados e culminaram na descoberta de faixas espectrais como o infravermelho próximo, médio e distante, além da região do ultravioleta ¹².

Apesar da radiação no infravermelho próximo (NIR - do inglês, *Near Infrared*) ter sido descoberta anteriormente, a radiação no infravermelho médio foi a região que obteve uma maior aceitação. Isto ocorreu após um experimento realizado em 1900 por Coblentz, no qual ele obteve o espectro de substâncias puras e observou sua utilidade na detecção de grupos orgânicos funcionais. Isto ocorreu pois nesta região é possível observar bandas de absorção como N-H, C-H, O-H e C-C específicos de alguns grupos orgânicos ¹³.

Por um grande período de tempo, a espectroscopia na região do NIR não foi utilizada, pois não se observava nenhuma informação atrativa nesta região do espectro, já que esta era ocupada por bandas de absorção fracas, largas e

sobrepostas. Contudo, a década de 80 ficou marcada como um grande *boom* da técnica ¹³.

A partir da década de setenta, novos estudos espectroscópicos surgiram, principalmente devido ao desenvolvimento de espectrofotômetros com Transformada de Fourier, da informática, da possibilidade de interface entre instrumentos e computadores. Com tantas inovações, surgiu um grande interesse por parte de grupos de pesquisa, governos e indústrias no desenvolvimento de novas metodologias de análise, visando o aumento da produtividade e redução de gastos ¹².

Os primeiros trabalhos que despertaram o interesse no estudo da espectroscopia no infravermelho próximo para análises industriais foram realizados por Karl Norris, em um grupo de pesquisa no Departamento de Agricultura dos Estados Unidos ¹³. Seus estudos iniciais visavam à procura de novos métodos para a determinação de umidade em produtos agrícolas. Outra grande contribuição de Norris foi a utilização da reflectância difusa para realização de medidas não destrutivas na região do NIR, possibilitando o trabalho diretamente sobre as amostras, sem a necessidade de utilização de nenhum pré-tratamento ¹³.

Com isso, a espectroscopia NIR pode atingir um rápido desenvolvimento com as melhoras instrumentais e computacionais e também devido ao tratamento dos dados utilizando a Quimiometria, que forneceu as ferramentas necessárias para a obtenção da informação e seu uso racional ¹³.

2.2 Aplicações da Espectroscopia no Infravermelho

A radiação infravermelha pertence a uma região espectral correspondente aos números de onda no intervalo de 12800 a 10 cm⁻¹. O espectro infravermelho é subdividido em infravermelho próximo (NIR- do inglês, *Near Infrared*), médio (MID – do inglês, *Middle Infrared*) e distante (FAR – do inglês, *Far Infrared*). A Tabela 2 apresenta os limites aproximados para cada região ¹⁴.

Tabela 2. Regiões espectrais do infravermelho.

Região	Intervalo de número de onda (ν) – (cm^{-1})	Região em comprimento de onda (λ) – (nm)
Próximo (NIR)	12800 a 4000	780 a 2500
Médio (MIR)	4000 a 200	2500 a 5000
Distante (FIR)	200 a 10	5000-10000

As primeiras aplicações na região do infravermelho foram destinadas para o monitoramento da qualidade e no controle de processos de indústrias petroquímicas durante a segunda guerra mundial para análise de combustíveis, lubrificantes e polímeros ¹⁵.

Somente nos anos setenta a espectroscopia no infravermelho próximo foi utilizada como uma técnica analítica, sendo que suas aplicações se voltavam basicamente aos produtos agrícolas e alimentícios. Desta forma, a área agrícola foi a primeira a utilizar intensamente a espectroscopia NIR. Atualmente este setor já utiliza o NIR para agricultura de precisão, utilizando os instrumentos no campo, coletando dados em tempos reais e relacionando-o com a composição do solo ¹⁵.

As áreas farmacêutica ¹⁶ e petroquímica ¹⁷ também encontraram muitas aplicações para a espectroscopia NIR e outros campos como NIR para imagem ¹⁶, e tomografia óptica clínica *in-vivo* são áreas em grande crescimento.

Inicialmente, as medidas eram realizadas somente em fotômetros e espectrofotômetros dispersivos baseados em filtros e redes de difração, respectivamente. Estes instrumentos eram parecidos com os utilizados na espectroscopia na região do ultravioleta/visível (UV/VIS) e muitas vezes eram equipamentos que compreendiam as três regiões (UV/VIS/NIR) ¹⁴.

Atualmente, os equipamentos modernos para análises quantitativas são em sua maioria interferométricos com Transformada de Fourier, já que estão sendo desenvolvidos especificamente para análises na região do NIR e também apresentam acessórios para análises de amostras sólidas, líquidas e gasosas ¹².

A espectroscopia no infravermelho próximo possui características como o fornecimento de resultados mais rápidos, é um método não destrutivo e não invasivo, não gera subprodutos tóxicos, adequado para utilização *on-line* e requer mínima preparação das amostras. As maiores desvantagens da técnica são provavelmente a baixa sensibilidade a constituintes em menores concentrações e também a grande dependência de bons e aceitáveis métodos de referência ^{12,14}.

A região do infravermelho médio (MIR) é a que possui a maior parte das aplicações e começou a ser utilizada no final dos anos 50 para análise qualitativa de compostos orgânicos, já que possui muitas informações que auxiliam na caracterização funcional dos compostos orgânicos ¹⁴. Desta forma, a utilização do MIR é muito direcionada à identificação de compostos.

Na região do MIR, até o início dos anos 80, a maioria dos instrumentos era do tipo dispersivo baseado em redes de difração. Com o surgimento dos instrumentos interferométricos, a maior parte então passou a ser baseada na Transformada de Fourier. Isto gerou um aumento no número de aplicações no MIR, tanto na área qualitativa quanto na quantitativa ¹⁴, sendo boa parte das aplicações baseadas na análise quantitativa.

Apesar dos avanços instrumentais, as aplicações do MIR ainda consistem na identificação de compostos orgânicos, pois nesta região ocorrem essencialmente transições fundamentais, além de uma faixa conhecida como impressão digital. Nesta faixa, a estrutura da molécula pode ser conhecida através de mudanças nos picos de absorção do espectro com alterações na estrutura e composição de uma determinada molécula. Isto facilita a identificação dos compostos orgânicos, pois seu espectro na região do MIR pode ser comparado com os presentes em bancos de dados ¹⁴.

A região do infravermelho distante (FIR) é pouco utilizada devido às limitações instrumentais, uma vez que existem poucas fontes para este tipo de radiação e também é necessária a utilização de filtros de interferência para evitar radiações de ordem superiores no detector ¹⁴.

Esta região só se tornou acessível para pesquisas e aplicações com o desenvolvimento de instrumentos com Transformada de Fourier. Desde então, o

FIR é utilizado para estudos de compostos inorgânicos, em organometálicos onde as vibrações dos ligantes podem ser observadas e também para moléculas que apresentam apenas átomos leves, desde que estes possuam mais de dois átomos diferentes do hidrogênio e com modos de deformação angular em sua estrutura ¹⁴.

2.3 Fundamentos da Espectroscopia no Infravermelho

Para absorver radiação infravermelha a molécula precisa sofrer uma variação no momento de dipolo como consequência do movimento vibracional ou rotacional. Isto ocorre pois esta radiação não possui energia suficiente para promover transições eletrônicas e a absorção desta radiação é restrita a moléculas com pequenas diferenças de energia entre vários estados vibracionais e rotacionais ^{14,15}.

Somente com a variação no momento de dipolo o campo elétrico alternado da radiação pode interagir com a molécula e gerar variações na amplitude de um de seus movimentos ¹⁴.

As posições relativas dos átomos em uma molécula variam continuamente como consequência de inúmeros tipos de vibrações e rotações em torno das ligações da molécula. As vibrações podem ser de dois tipos: estiramentos e deformações angulares ¹⁴.

Uma vibração de estiramento envolve uma variação contínua na distância interatômica ao longo do eixo da ligação entre dois átomos. As vibrações de deformação angular são caracterizadas pela variação do ângulo entre duas ligações e podem ser de quatro tipos: tesoura (do inglês *scissoring*), balanço (do inglês *rocking*), sacudida (do inglês *wagging*) e torção (do inglês *twisting*) ¹⁴.

Na temperatura ambiente, a maioria das moléculas está em seu nível energético vibracional fundamental. Os átomos ou grupos de átomos que participam das ligações químicas são deslocados uns em relação aos outros em uma frequência que é definida pela força da ligação e pela massa dos átomos individualmente ligados ou pelos seus grupos. A amplitude destas vibrações é da

ordem de alguns nanômetros e irá aumentar conforme a energia que é transferida para a molécula ¹⁴.

Esta energia pode ser transferida de um fóton com determinado comprimento de onda (λ), nos quais a energia (E_p) pode ser dada por ¹³:

$$E_p = h\nu = \frac{hc}{\lambda} \quad (7)$$

onde h é a constante de Planck e c é a velocidade da luz.

Considerando que as vibrações ocorrem de maneira isolada em uma molécula, um modelo mecânico conhecido como oscilador harmônico pode ser utilizado para representar a interação entre a radiação e a matéria na região do NIR ^{12,13,14}.

Este modelo descreve uma molécula diatômica como duas massas esféricas ligadas por uma mola com uma constante de força (k) e as frequências fundamentais de quaisquer dois átomos ligados é calculada assumindo que a energia segue o comportamento do oscilador harmônico que obedece a Lei de Hook. Segundo a Lei de Hook, a energia (E) formada pelo sistema massa-mola é dada por ¹⁴:

$$E = \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} \quad (8)$$

onde μ é a massa reduzida:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (9)$$

A vibração molecular pode ser descrita supondo um oscilador harmônico no qual uma energia potencial (V) é dada em função do deslocamento proporcionado pela mola (x) entre os átomos ¹³:

$$V = \frac{1}{2}kx^2 \quad (10)$$

A Figura 4A mostra o comportamento da energia potencial em função do deslocamento dos átomos da posição de equilíbrio (ponto de mínimo). Pode-se observar que a energia potencial é máxima quando a mola está esticada ou comprimida e decresce para zero na posição de equilíbrio ¹³.

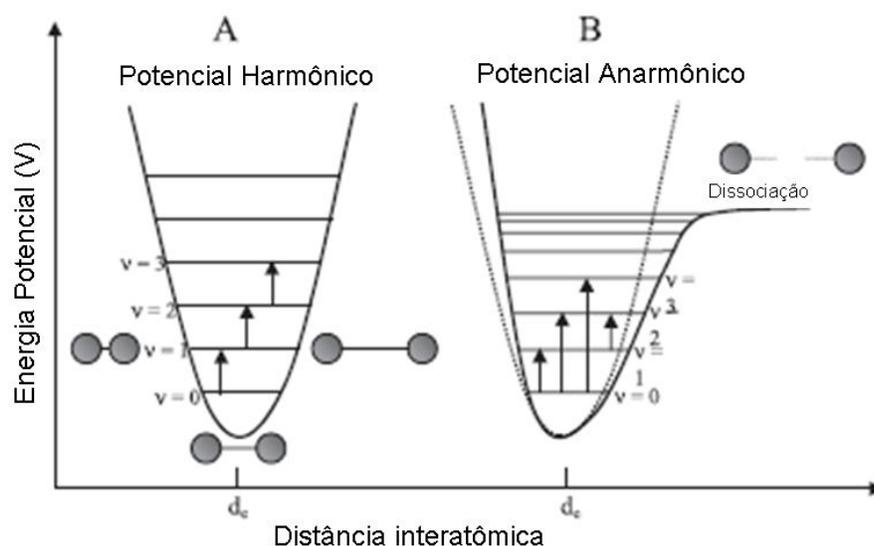


Figura 4. Diagrama de energia potencial para (A) oscilador harmônico, (B) oscilador anarmônico.

O oscilador harmônico representa uma boa aproximação para o cálculo de freqüências fundamentais de moléculas diatômicas simples. Contudo, o modelo fornece apenas uma média, ou a freqüência central de transições de estados vibracionais e rotacionais de ligações diatômicas. Em moléculas poliatômicas, os elétrons sofrem influência de grupos vizinhos, influenciando o estiramento, o comprimento e o ângulo de ligação e, portanto, a freqüência das vibrações das ligações químicas ¹².

As diferenças nas freqüências das vibrações ocorrem devido às interações e proporciona o surgimento de um espectro específico para cada substância e

valores como a constante de força de ligação (k) sofrem grande variação, possibilitando a interpretação dos espectros pelas diferenças de energia ¹².

Desta forma, o modelo do oscilador harmônico não pode mais ser utilizado, pois com sistemas moleculares, não se pode assumir o mesmo perfil de energia contínua utilizado para o sistema massas-mola do oscilador harmônico. O sistema molecular pode ter apenas alguns níveis de energia discretos (E_v) definidos pela mecânica quântica pela equação ¹³:

$$E_v = (v + \frac{1}{2})h\nu \quad (11)$$

onde v é o número quântico vibracional, que pode assumir valores positivos e inteiros incluindo o zero, E_v é a energia associada ao respectivo nível quântico e ν é a frequência vibracional fundamental.

A energia envolvida na transição do nível 1 para o nível 2 ou do nível 2 para o 3 deveria ser idêntica à da transição de 0 para 1, e a teoria quântica indica que as únicas transições que poderiam ocorrer seriam aquelas em que o número quântico vibracional sofre alteração de uma unidade. Segundo a teoria quântica, portanto, somente transições fundamentais poderiam existir sendo esta restrição denominada regra de seleção ($\Delta v = \pm 1$). Uma vez que os níveis vibracionais são igualmente espaçados, apenas um pico de absorção deveria ser observado para uma certa vibração molecular ^{12,14}.

A descrição da vibração molecular, considerando os tratamentos clássico e mecânico-quântico do oscilador harmônico é imperfeita quando se levam em consideração aproximações qualitativas. Por exemplo, a medida que dois átomos se aproximam, a repulsão coulombiana entre os dois núcleos produz uma força que age na mesma direção da força de restauração da ligação, assim, espera-se que a energia potencial cresça mais rapidamente do que é previsto pelo modelo do oscilador harmônico. Por outro lado, quando a distância entre os átomos aumenta, um decréscimo na força de restauração e, portanto, da energia

potencial, ocorre quando a distância interatômica se aproxima daquela em que ocorre a dissociação dos átomos ¹⁴.

O comportamento anarmônico conduz a desvios de duas espécies. O primeiro desvio é que os níveis de energia não são igualmente espaçados, como pode ser observado na figura 4B. Com o aumento da energia ocorre uma diminuição da diferença entre os níveis de energia. O outro desvio é que em números quânticos altos, ΔE se torna menor e a regra de seleção não é seguida rigorosamente. Como resultado, harmônicos que ocorrem em frequências com aproximadamente duas a três vezes a de uma transição fundamental são observadas, isto é, transições de $\Delta \pm 2$ ou ± 3 . Essas transições são conhecidas como sobretons (do inglês *overtone*) ^{12,14}.

A maioria dos sinais do tipo sobreton ocorre na região do infravermelho próximo e com intensidade de transição cerca de 10 a 1000 vezes menor que as observadas para transição fundamental. Este fato é inconsistente com a teoria quântica, uma vez que esta prevê que tais transições seriam proibidas ¹⁴.

Além destes desvios, pode-se citar também a presença de bandas de combinação, que ocorre quando um fóton excita dois modos vibracionais simultaneamente, ou seja, um quantum de energia é absorvido por duas ligações ao invés de uma. Tal ocorrência é comum principalmente na faixa de 5500 a 4000 cm^{-1} , e sua frequência é aproximadamente a soma ou a diferença das duas frequências fundamentais ^{14,18}.

As equações de onda da mecânica quântica permitem a obtenção de curvas de energia potencial mais corretas para as vibrações moleculares, porém a grande complexidade matemática destas equações impede sua aplicação quantitativa. Qualitativamente, as curvas possuem uma forma anarmônica, que se diferencia do comportamento harmônico em alguns aspectos, dependendo da natureza da ligação e dos átomos envolvidos. Porém, as curvas harmônicas e anarmônicas são semelhantes para energias potenciais pequenas e com isso o sucesso dos métodos de aproximação descritos é grande ^{14,19}.

Na região do infravermelho próximo, os sinais observados são essencialmente devido aos sobretons e bandas de combinação de estiramentos e

deformações angulares de transições fundamentais de ligações X-H, em que X representa átomos de oxigênio, nitrogênio, carbono de grupos aromáticos e também grupos funcionais importantes como grupos carbonila, C-O, C-N, C-C, entre outros que sofrem estiramento ¹⁸.

2.4 Aplicações da Espectroscopia no Infravermelho no setor de papel e celulose

A utilização da espectroscopia NIR no setor de papel e celulose está presente em uma série de trabalhos. Como exemplo, Schultz e Burns ²⁰ pesquisaram a composição da madeira, para a análise de lignina, hemicelulose e celulose em madeiras do tipo *pinus*.

Também foi relatado por Easty et al ²¹ a utilização do NIR para avaliação das propriedades químicas da polpa e do papel, através da quantificação das misturas de madeiras do tipo *hardwood* ou *softwood*, e a estimativa da quantidade de lignina no início dos anos 90.

Nos últimos anos, vários trabalhos envolvendo o setor de papel e celulose, a espectroscopia NIR e a quimiometria, especialmente utilizando calibração multivariada podem ser encontrados. Em um trabalho pode-se encontrar um estudo de caso da produção de celulose ²², visando à previsão do número Kappa. Também foi realizado um estudo para a determinação de umidade na madeira em condições instáveis de temperatura, utilizando NIR ²³, variando-se a temperatura de -20°C a $+25^{\circ}\text{C}$. Em outra pesquisa, feita por Poke et al ²⁴, foi realizado um estudo para previsão dos conteúdos de lignina e extrativos presentes em uma espécie de eucalipto, utilizando NIR.

Existe uma série de trabalhos no Brasil envolvendo empresas do setor de papel e celulose desenvolvendo pesquisas utilizando o NIR para a previsão de alguns parâmetros da madeira. Magalhães et al ²⁵, utilizou o NIR para a predição da densidade básica da madeira em *Pinus taeda*, e os autores relatam que os resultados foram satisfatórios se for considerada a variabilidade natural das amostras e também o pequeno tempo despendido para a realização das análises.

Outro estudo também envolveu a aplicação do NIR *on-line* ²⁶ no qual o espectro foi coletado com um equipamento de NIR instalado na correia por onde passam os cavacos. Cavacos são pequenos pedaços de madeira com dimensões de aproximadamente 2 x 2 cm. Os autores concluíram que é viável a utilização do equipamento na linha de produção, pois com este foi possível a utilização de modelos de calibração multivariada para a previsão de densidade básica e também do teor seco de amostras de cavacos.

Além dos trabalhos citados, muitos outros trabalhos envolvendo NIR podem ser encontrados na literatura e para uma pesquisa mais detalhada, pode-se citar uma revisão de Tsuchikawa ²⁷, onde se encontram as referências de diversos trabalhos utilizando NIR em madeira e papel.

CAPÍTULO 3
Métodos de análise multivariada

3. Métodos de análise multivariada

3.1 Quimiometria

A quimiometria pode ser definida como a aplicação de métodos matemáticos e estatísticos, assim como de lógica, em química visando extrair uma maior quantidade de informação química com a análise dos dados obtidos e também para planejar ou selecionar experimentos de forma otimizada ²⁸.

Os métodos utilizados na quimiometria foram desenvolvidos a princípio em outras disciplinas e que foram posteriormente aplicadas a pesquisas voltadas para o tratamento de dados químicos, gerando uma nova área dentro da química analítica ²⁸.

Essa nova área dentro da química analítica apareceu nos anos 60 com o surgimento de métodos instrumentais computadorizados para análise química, gerando uma grande quantidade de dados. Esses dados anteriormente eram poucos, uma vez que eram obtidos de forma lenta e dispendiosa ²⁸.

Desde o surgimento da quimiometria, muitos métodos foram desenvolvidos, possibilitando o processamento e a interpretação de dados complexos. Desta forma, a quimiometria tem como grande contribuição tornar métodos matemáticos complexos práticos para utilização em laboratórios industriais e para análises de rotina ²⁸.

As análises na região do infravermelho próximo são exemplos do sucesso da quimiometria, pois com a utilização de modelos de calibração multivariada, permitem a realização de determinações quantitativas. Com o crescimento das pesquisas na quimiometria, novas ferramentas foram desenvolvidas visando o tratamento dos dados. Estas ferramentas encontraram diversas aplicações, dependendo do objetivo dos estudos, como, a otimização de processos, a classificação de dados, as determinações quantitativas, entre outros. Com isso, a quimiometria foi dividida em algumas frentes de aplicação:

- Processamento de sinais analíticos
- Planejamento e otimização de experimentos
- Modelagem

- Reconhecimento de padrões e classificação de dados
- Calibração Multivariada
- Métodos de inteligência artificial
- Processamento de imagens ²⁸

A principal linha de pesquisa e aplicação da quimiometria dentro da química analítica é a de calibração multivariada, com a construção de modelos de regressão a partir de dados de primeira ordem, que são dados que podem ser representados através de um vetor para cada amostra ²⁹.

3.2 Calibração

A calibração é uma série de operações que visam à obtenção, sob condições específicas, de um algoritmo matemático que produza as propriedades de interesse a partir das medidas instrumentais.

A função matemática do modelo de calibração relaciona dois grupos de variáveis, uma delas denominada variável dependente (Y) e outra denominada variável independente (X):

$$Y = f(X) = Xb \quad (12)$$

Esta etapa representa a calibração e o conjunto dos dados empregados para esta finalidade é chamado de conjunto de calibração. Os parâmetros do modelo são chamados de coeficientes de regressão (b), determinados matematicamente ³⁰.

Após a calibração, deve-se proceder com uma etapa de validação, no qual as variáveis independentes obtidas para um conjunto de amostras são utilizadas juntamente com o coeficiente de regressão para o cálculo dos valores previstos para a variável dependente. Assim, no conjunto de validação são utilizadas amostras cujas variáveis dependentes sejam conhecidas, para que se estabeleça uma comparação entre os valores previstos pelo modelo e os valores conhecidos

previamente através de uma metodologia padrão, o que permitirá a avaliação do desempenho do modelo de calibração proposto ³⁰.

Existem diversos métodos para a construção de modelos de calibração, sendo que a função que ajusta as variáveis independentes e dependentes pode ser linear ou não, dependendo da complexidade do sistema em estudo.

Entre os métodos de calibração, os mais conhecidos ainda são os métodos de calibração univariada ou também conhecidos de calibração de ordem zero. Estes métodos recebem esses nomes por apresentarem apenas uma medida instrumental para cada amostra de calibração, ou seja, para cada amostra têm-se apenas um escalar.

Os métodos de calibração univariada estão presentes em uma infinidade de trabalhos ^{31,32,33,34} na literatura, e possui aplicação e validação relativamente fáceis. No entanto, a aplicação da calibração univariada não é geral, uma vez que, quando a amostra apresenta interferentes, e a medida é realizada diretamente na metodologia instrumental, isso pode provocar desvios na determinação da propriedade de interesse, o que pode tornar este método inviável.

Na calibração multivariada, mais de uma resposta instrumental é relacionada com a propriedade de interesse e nestes métodos, a análise pode ser realizada mesmo na presença de interferentes, desde que estes estejam presentes nas amostras utilizadas para a construção do modelo de calibração. Neste tipo de calibração, também podem ser feitas determinações simultâneas e análises mesmo sem resolução. Desta forma, os métodos de calibração multivariada atuam como uma alternativa quando os métodos univariados não encontram aplicação ³⁰.

A resposta instrumental é representada na forma de uma matriz na calibração multivariada, enquanto a propriedade de interesse, determinada por uma metodologia padrão é representada por um vetor. A Figura 5 mostra a construção de uma matriz de dados a partir de um vetor de respostas instrumental representado por um espectro.

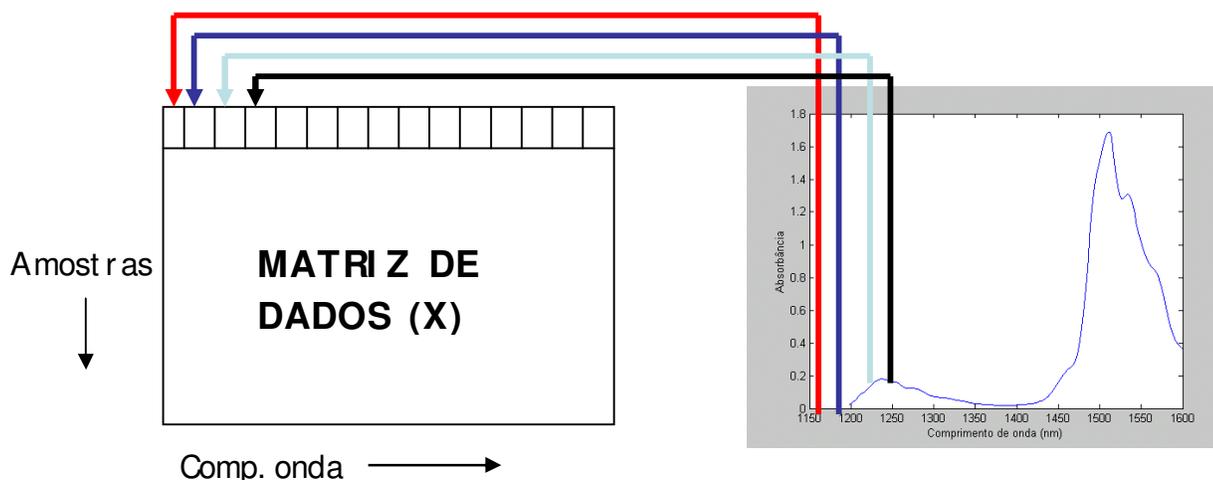


Figura 5. Construção da matriz X para calibração multivariada.

Vários métodos de regressão vêm sendo utilizados em química analítica para a construção de modelos de calibração multivariada. Os métodos mais utilizados são: regressão linear múltipla (MLR), a regressão por componentes principais (PCR) e a regressão por mínimos quadrados parciais (PLS), que são métodos para ajuste linear entre as variáveis. Muitos métodos de calibração multivariada empregados em espectroscopia são lineares, uma vez que são fáceis de interpretar e elaborar. Porém, existem também métodos não lineares, como as máquinas de vetores de suporte por mínimos quadrados (LS-SVM).

3.3 Pré- processamentos

Neste trabalho, foram utilizados três diferentes tipos de pré-processamento dos espectros, a Correção do Espalhamento Multiplicativo (MSC - do inglês, *Multiplicative Scatter Correction*), a primeira derivada (algoritmo de Savitzky-Golay) e a centragem na média.

O algoritmo de correção do espalhamento multiplicativo (MSC) foi utilizado para diminuir os efeitos de espalhamento de luz, que é muito comum em espectroscopia por reflectância difusa ³⁰.

O algoritmo de Savitzky-Golay é utilizado para derivar os espectros obtidos com no NIR. As derivadas são muito úteis para a eliminação do sinal do branco e também para melhorar a resolução visual dos picos, podendo, entretanto aumentar os ruídos ²⁸.

A centragem na média é utilizada para a eliminação de um constante deslocamento, então os dados são transladados até a origem da coordenada. Cada variável é centrada na média através da subtração da média da coluna ²⁸.

3.4 Análise de Componentes Principais (PCA) ³⁵

A análise de componentes principais (PCA) é um tipo de análise exploratória que visa extrair o máximo de informação de uma tabela de dados, convertendo-a em gráficos informatizados que mostram a relação entre amostras (linhas da matriz) e as variáveis (colunas da matriz), transformando dados complexos em informações relevantes. O novo conjunto de variáveis (componentes principais) é a combinação linear das variáveis originais e os novos eixos são ortogonais entre si. Assim, o PCA tenta agrupar as variáveis altamente correlacionadas numa nova variável que recebe o nome de componente principal.

Como representado na Figura 6, a Análise de Componentes Principais (PCA) tem como objetivo a decomposição da matriz **X** em uma matriz de escores (**T_{nxa}**) e uma matriz de pesos (**P_{mxa}**), onde *a* é o número de componentes principais.

$$\begin{array}{ccccccc}
 & m & & 1 & & m & & 1 & & m & & 1 & & m \\
 & \boxed{\mathbf{X}} & = & \boxed{\mathbf{T}_1} & \boxed{\mathbf{P}_1^T} & + & \boxed{\mathbf{T}_2} & \boxed{\mathbf{P}_2^T} & + \dots + & \boxed{\mathbf{T}_A} & \boxed{\mathbf{P}_A^T} \\
 n & & & n & 1 & & n & 1 & & n & 1 & & n \\
 & & & & & & & & & & & & &
 \end{array}$$

Figura 6. Representação esquemática da decomposição por PCA.

A matriz \mathbf{X} pode ser escrita como o produto de dois vetores, o vetor de escores t_h e o vetor de pesos p_h .

$$\mathbf{X} = t_1\mathbf{p}_1 + t_2\mathbf{p}_2 + \dots + t_a\mathbf{p}_a, \text{ para "a" componentes principais.} \quad (13)$$

Ou então:

$$\mathbf{X} = \mathbf{TP}^T \quad (14)$$

onde \mathbf{X} é o conjunto original dos dados, \mathbf{T} são escores que são as coordenadas das amostras em um novo sistema de eixos e \mathbf{P} são os pesos que contém a informação de peso de cada variável original na formação dos novos eixos.

Como usualmente a grande fração da variância é descrita nos primeiros componentes principais, é possível a visualização dos dados pelo gráfico de escores de um componente contra o outro. Os cálculos com PCA em química analítica utilizam principalmente dois algoritmos: o NIPALS (do inglês, *Nonlinear Iterative Partial Least Squares*)³⁶ e o SVD (do inglês, *Singular Value Decomposition*). No presente trabalho o PCA utilizou o algoritmo SVD.

O número de componentes principais obtidos é igual ao número de vetores de dados utilizados, sendo que nem todas as PCs possuem informações relevantes. Isso ocorre pois normalmente as últimas PCs modelam ruído inerente aos dados. Com isso, a eliminação de PCs faz com que haja um aumento na relação sinal/ruído²⁸.

O método mais utilizado para a determinação do número correto de PCs a serem utilizadas é o método da Validação Cruzada (CV-do inglês, *Cross Validation*), que é baseado na construção de um modelo com uma parte de um conjunto de dados seguido pela previsão do restante do conjunto.

A validação cruzada pode ser do tipo "*leave one out*" (deixe um de fora), onde uma amostra é deixada de fora no processo de construção do modelo e em seguida esta amostra é prevista pelo modelo construído. A validação cruzada também pode ser feita em blocos, onde determinados grupos de amostras são deixados de fora na construção do modelo e em seguida previstos por este modelo. Nos dois casos, o processo se repete até que todas as amostras tenham sido deixadas de fora e previstas pelo modelo construído. Com isso, a raiz

quadrada da soma dos quadrados dos erros da validação cruzada (RMSECV- do inglês, *Root Mean Square Error of Cross Validation*) é calculada^{28,29}:

$$\text{RMSECV} = \frac{\sum_{i=1}^k \text{RMSEPi}}{k} \quad (15)$$

em que k é o número de blocos e o RMSEP (do inglês, *Root Mean Square Error of Prediction*) é calculado como:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (16)$$

em que \hat{y}_i e y_i são os valores previstos e de referência para a propriedade de interesse, respectivamente e n é o número de amostras de calibração.

Idealmente, o número de PCs deveria ser igual ao número de espécies químicas presentes na amostra. Isto permite a determinação de apenas algumas espécies de interesse em um meio complexo. Essa propriedade recebe o nome de vantagem de primeira ordem, uma vez que os interferentes presentes nas amostras também podem ser modelados, desde que estes estejam presentes no desenvolvimento do modelo²⁹.

Desta forma, os principais objetivos do PCA são o de encontrar relações entre objetos e classificá-los de acordo com suas similaridades, tornando possível a detecção de amostras anômalas e também é uma técnica importante para a redução dos dados, o que é de grande utilidade quando muitas informações necessitam ser manipuladas.

3.5 Mínimos Quadrados Parciais (PLS) ³⁷

A regressão por mínimos quadrados parciais (PLS – do inglês, *Partial Least Squares*) é muito utilizada para construção de modelos de calibração multivariada. Este método não requer um conhecimento exato de todos os componentes presentes nas amostras, podendo realizar a previsão de amostras mesmo na presença de interferentes, desde que estes também estejam presentes na etapa de construção do modelo.

O modelo PLS é obtido através de um processo iterativo, otimizando ao mesmo tempo a projeção das amostras sobre os *loadings* para a determinação dos *scores* e o ajuste por uma função linear dos *scores* da matriz **X** aos *scores* da matriz **Y** de modo a minimizar os desvios. A otimização simultânea causa pequenas distorções nas direções dos *loadings*, fazendo com que eles percam a ortogonalidade, levando a pequenas redundâncias que otimizam a relação linear entre os *scores*. As distorções da ortogonalidade entre os componentes principais no PLS fazem com que estes não sejam mais componentes principais (que são ortogonais) e sim variáveis latentes ²⁸.

Considerando um caso geral, onde se tem mais de uma espécie de interesse, **Y** é uma matriz de dimensão (n x z), onde z é o número de colunas de **Y** e tem-se a decomposição das matrizes **X** e **Y** em suas respectivas matrizes de escores e pesos (*loadings*) ²⁸:

$$\mathbf{X} = \mathbf{T} \mathbf{P}^T + \mathbf{E}_x = \sum \mathbf{t}_A \mathbf{p}_A^T + \mathbf{E}_x \quad (17)$$

$$\mathbf{Y} = \mathbf{U} \mathbf{Q}^T + \mathbf{E}_y = \sum \mathbf{u}_A \mathbf{q}_A^T + \mathbf{E}_y \quad (18)$$

em que **X** é a matriz de respostas instrumentais, **Y** é a matriz de respostas da propriedade de interesse obtida por metodologia padrão, **T** e **U** são os *scores* de **X** e **Y**, respectivamente, **P** e **Q** são os *loadings* de **X** e **Y** respectivamente, **E_x** e **E_y**

correspondem à matriz de resíduos. Uma relação linear é, então, estabelecida entre os *scores* de \mathbf{X} e os *scores* de \mathbf{Y} ²⁸:

$$\hat{\mathbf{u}}_A = b_A \hat{\mathbf{t}}_A \quad (19)$$

em que b_A é o vetor de coeficientes de regressão obtido na etapa de calibração.

A Figura 7 ilustra a decomposição das matrizes \mathbf{X} e \mathbf{Y} no produto das matrizes de *scores* e *loadings*. Esta decomposição pode ser feita utilizando vários algoritmos (como exemplo NIPALS), que fazem a decomposição através de passos diferentes, porém obtendo resultados praticamente iguais.

$$\begin{aligned} \mathbf{X} &= \mathbf{t}_1 \mathbf{p}_1^T + \dots + \mathbf{t}_A \mathbf{p}_A^T + \mathbf{E}_x \\ \mathbf{Y} &= \mathbf{u}_1 \mathbf{q}_1^T + \dots + \mathbf{u}_A \mathbf{q}_A^T + \mathbf{E}_y \end{aligned}$$

Figura 7. Decomposição em variáveis latentes das matrizes \mathbf{X} e \mathbf{Y} para modelos PLS.

3.6 Mínimos Quadrados Parciais por Intervalos (iPLS)

A seleção de variáveis envolve a escolha de determinadas regiões do espectro (um único ou um conjunto de comprimentos de onda) independentes e mais restritas, fazendo com que o modelo de calibração minimize os erros de previsão. Tal afirmação pode ser feita com base em evidências experimentais e teóricas, nas quais uma melhora na eficiência é observada, sendo os modelos construídos mais robustos, precisos e de simples interpretação ³⁸.

Na espectroscopia no infravermelho próximo, a seleção de variáveis pode ser aplicada visando à eliminação de comprimentos de onda que induzem a ruídos, informações irrelevantes e não-linearidades ³⁹. Para isto, existem vários métodos para a escolha da região espectral, como o algoritmo genético ⁴⁰, o método de mínimos quadrados parciais com eliminação de variáveis não-informativas (UVE-PLS – do inglês, *Elimination of Uninformative Variables in Partial Least Square*) ⁴¹, entre outros.

O método dos Mínimos Quadrados Parciais por intervalos (iPLS – do inglês, *Intervals Partial Least Squares*) ^{39,42} foi o método de seleção de variáveis utilizado no presente trabalho. Esta escolha ocorreu por se tratar de um método mais simples comparado aos exemplificados anteriormente e também por selecionar faixas espectrais e não apenas variáveis isoladas. A seleção de intervalos espectrais é especialmente importante para métodos baseados na determinação de fatores, como é o caso do PLS.

O iPLS é um algoritmo que divide o conjunto de dados em um certo número de intervalos (escolhido pelo usuário) equidistantes e calcula-se um modelo de Mínimos Quadrados Parciais (PLS - *Partial Least Squares*) para cada intervalo, sendo escolhido o modelo com menor erro em relação ao modelo utilizando todo o espectro. Este método visa encontrar uma região espectral que apresente as informações mais relevantes, produza melhores resultados e pode ser muito útil para interpretação dos dados ⁴².

Desta forma, o iPLS fornece a informação do erro de validação cruzada em função de cada intervalo no qual o espectro foi sub-dividido. Com os valores de RMSECV para cada intervalo do espectro em relação ao RMSECV obtido para o modelo global, com todas as variáveis, avalia-se a previsão de cada intervalo. A melhor seleção de variáveis será representada pelo intervalo que apresentar o menor valor de RMSECV comparado ao valor obtido para o modelo global ⁴².

A construção de modelos de PLS para cada intervalo do espectro normalmente necessita de diferentes números de variáveis latentes, comparado ao necessário para a construção do modelo global para alcançar uma variância relevante referente às variáveis dependentes. Tal fato pode ser atribuído devido

ao número de substâncias que absorvem/interferem, ao ruído e à largura dos intervalos.

3.7 Análise Discriminante com método de mínimos quadrados parciais (PLS-DA) ⁴³

O PLS-DA (do inglês, *Discriminant Analysis – Partial Least Squares*) é um método multivariado utilizado para classificação de amostras onde é necessária a redução de variáveis e a variabilidade dentro do grupo é maior que a variabilidade entre grupos. O bloco **Y** em um modelo de PLS-DA indica a classe ao qual uma amostra pertence. Quando se têm duas classes a serem discriminadas, utiliza-se o PLS1. Neste caso existe uma variável dependente *y* que pode assumir os valores 0 ou 1. No caso de três ou mais classes, utiliza-se o PLS2 e a variável dependente é uma matriz com o número de colunas igual ao de classes e assumindo valores 0 ou 1 para indicar se a amostra pertence ou não a essa classe.

Os valores previstos pelo modelo do PLS-DA deveriam ser idealmente 0 ou 1, mas na prática esses valores apenas se aproximam destes valores. É calculado um valor limite, entre os valores previstos, sendo que valores acima deste limite indicam que a amostra pertence a classe modelada e valores abaixo deste limite indicam que a amostra não pertence à classe modelada. O modelo permite ainda o cálculo da probabilidade da amostra pertencer à classe que está sendo modelada.

3.8 Máquinas de Vetores de Suporte por Mínimos Quadrados (LS-SVM) ^{44,45}

O algoritmo máquinas de vetores de suporte (SVM – do inglês, *Support Vector Machines*) utilizado para regressão é uma generalização do algoritmo Generalized Portrait, que começou a ser desenvolvido nos anos 60 por Vapnik e colaboradores. Entretanto, a forma atual do LS-SVM foi desenvolvida na década de noventa, com maior ênfase nas aplicações práticas.

O SVM possui algumas vantagens, como uma alta capacidade de generalização, robustez em grandes dimensões, teoria bem estabelecida na Matemática e Estatística e convexidade da função objetivo. Porém, encontrar um modelo final no SVM pode ser computacionalmente muito difícil, pois requer a solução de um conjunto de equações não-lineares. Para simplificar esta aproximação, Suykens e colaboradores propuseram o uso de máquinas de vetores de suporte por mínimos quadrados (LS-SVM).

A máquina de vetor de suporte por mínimos quadrados (LS-SVM) é um método de calibração multivariada não-linear, que é capaz de tratar diferenças espectrais causadas pelas variações das condições físico-químicas presentes nas medidas. O LS-SVM é capaz de lidar com problemas lineares e não-lineares de calibração multivariada de uma maneira relativamente rápida.

Em princípio, no LS-SVM é feita uma estimativa linear em um novo espaço característico utilizando uma função kernel ($y = w^T \Phi(x) + b$) entre os regressores (x) e as variáveis dependentes (y). Cada kernel é associado com um parâmetro específico para funções não-lineares. Assim, ao invés de calcular um mapeamento específico para cada dimensão dos dados, uma função Kernel apropriada é selecionada e seu parâmetro específico é otimizado.

A função Kernel transforma o espaço de entrada em um espaço de alta dimensão, onde a solução pode ser apresentada como um problema linear, como mostra a figura 8.

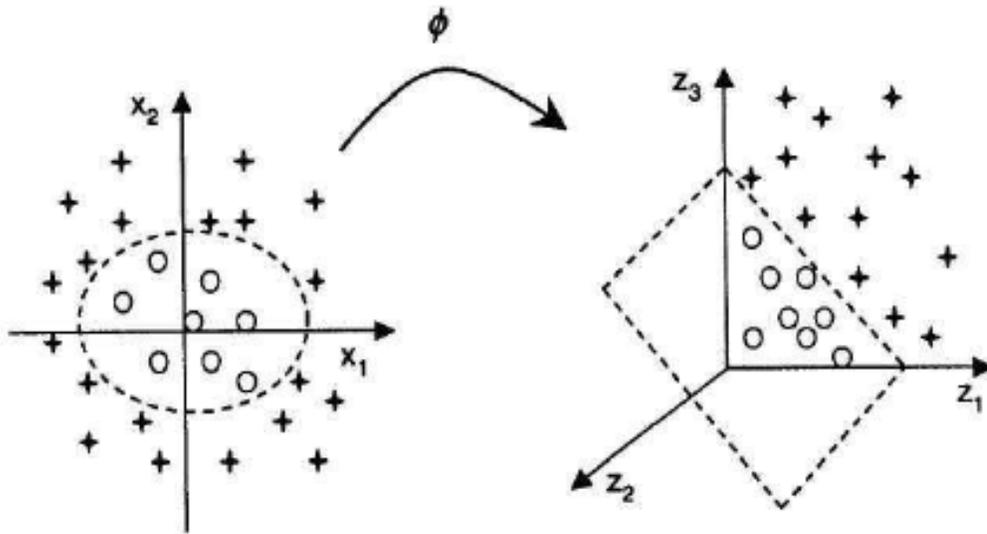


Figura 8. Mudança do espaço dos dados através da função Kernel (ϕ).

É necessário minimizar uma função custo (\mathbf{C}), contendo uma penalização (γ) no erro de regressão, como mostrado a seguir:

$$C = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad (20)$$

Tal que:

$$y_i = \mathbf{w}^T \Phi(x_i) + b + e_i \quad (21)$$

para todo $i = 1, 2, \dots, N$, onde Φ representa a função kernel.

A primeira parte desta função custo é um decaimento utilizado para penalizar o quadrado dos pesos (\mathbf{w}), visando à obtenção dos menores valores possíveis. O segundo termo representa os erros de regressão (e) para todos os conjuntos de treinamentos. O peso relativo entre as duas partes da equação é dado por γ , que deve ser otimizado pelo usuário. Com o aumento de γ , a ênfase maior é sobre a redução dos quadrados dos erros.

O LS-SVM resolve este problema de otimização restrita com o método dos multiplicadores de Lagrange. Neste método, os pesos (\mathbf{w}) podem ser escritos

como combinações lineares dos multiplicadores de Lagrange com os dados do conjunto de calibração (x_i) correspondentes.

$$L(w,b,e,\alpha) = \frac{1}{2} \|\mathbf{w}\|^2 + \gamma \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i \{ \mathbf{w}^T x_i + b + e_i - y_i \} \quad (22)$$

em que

$$y_i = \begin{bmatrix} y \\ y \\ M \\ y_N \end{bmatrix}, e_i = \begin{bmatrix} e_1 \\ e_2 \\ M \\ e_N \end{bmatrix} e \alpha_i = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ M \\ \alpha_N \end{bmatrix} \quad (23)$$

A função custo definida anteriormente, se encontra nos dois primeiros termos desta função de Lagrange (L), mas esta expressão tem multiplicadores de restrição, os chamados multiplicadores de Lagrange (α_i), cada um correspondendo a um dado de treinamento. Para obter a solução final, as primeiras derivadas parciais são acertadas para zero, e o resultado é colocado na equação de regressão original ($y = wx + b$), obtendo:

$$y = \sum_{i=1}^N \alpha_i \varphi(x_i)^T \varphi(x) + b = \sum_{i=1}^N \alpha_i \langle \varphi(x_i)^T, \varphi(x) \rangle + b \quad (24)$$

Em aplicações envolvendo regressão não-linear, é suficiente a substituição do produto interno $\langle x_i^T, x \rangle$ por uma função de kernel não-linear $k(x_i^T, x)$. A função kernel mais utilizada é a função de base radial (RBF):

$$\phi = \exp \left(\frac{-\|x_i - x\|^2}{2\sigma^2} \right) \quad (25)$$

O grau de não linearidade é ajustado alterando-se o parâmetro de variância (σ^2). Quando σ^2 aumenta, o modelo é forçado a encontrar uma solução linear.

3.9 Algoritmo de Kennard-Stone

A separação de um conjunto em amostras de calibração e validação deve ser realizada de maneira que as amostras de validação sejam bem representadas pelas amostras de calibração. Existem vários métodos para realizar essa divisão, como PDS (do inglês, *Piecewise Direct Standardisation*)⁴⁶, o algoritmo de projeções sucessivas (SPA – do inglês, *Successive Projections Algorithm*)⁴⁷ e o algoritmo de Kennard – Stone⁴⁸, entre outros.

Neste trabalho, a separação das amostras em conjuntos de calibração e validação foi realizada utilizando o algoritmo de Kennard-Stone, uma vez que este algoritmo seleciona as amostras baseadas em suas distâncias. Além disso, este algoritmo é de fácil aplicação, além de ser uma alternativa viável para o cumprimento do objetivo deste trabalho.

O algoritmo de Kennard-Stone é iniciado selecionando-se dois espectros: o mais próximo do ponto central (ou mais distante) e o segundo é o que possui a maior distância euclidiana em relação ao primeiro. Estas duas amostras são então removidas do conjunto de calibração e alocadas em um subconjunto. Para cada espectro não selecionado do conjunto de calibração, as distâncias euclidianas entre o espectro selecionado do subconjunto e os espectros restantes são computadas. A amostra do conjunto de calibração com maior valor da distância euclidiana é selecionada para o subconjunto. O procedimento segue até a obtenção do número desejado de amostras para o conjunto de calibração⁴⁸.

Normalmente, este algoritmo é utilizado para realizar a seleção de amostras que irão compor o conjunto de calibração, uma vez que este atua selecionando as amostras com maior variabilidade do conjunto total.

3.10 Detecção de amostras anômalas – *Outliers*

As amostras anômalas são amostras que apresentam um comportamento diferente das demais amostras do conjunto de dados. Esse comportamento pode ser representado por amostras irrelevantes, erros grosseiros ou anormais comparados ao restante dos dados. *Outliers* é o termo usado para designar essas amostras anômalas que podem estar presentes nos conjuntos de calibração e validação, que serão utilizadas para construção e validação de um modelo de calibração multivariada ⁴⁹.

A presença de amostras anômalas no conjunto de calibração pode gerar modelos com baixa capacidade de previsão, ou seja, com altos valores de erro. Já a presença destas amostras no conjunto de validação pode influenciar os resultados, gerando resultados que indicam que o modelo não é adequado ou possui capacidade inferior à que poderia ser apresentada na ausência destas amostras.

Quando possível, devemos tentar entender a razão de cada amostra anômala detectada. Essas podem ser boas ou ruins. As amostras anômalas boas podem apresentar objetos com concentração analítica muito alta ou muito baixa, e se a resposta instrumental é linear, essas amostras devem possuir bastante informação e, portanto, devem ser mantidas no conjunto de calibração. As amostras anômalas ruins são aquelas que não conseguimos entender e que possuem grande influência no modelo de calibração. Estas amostras podem representar simples confusões e serem corrigidas ou então levarem mais tempo e requererem nova medida instrumental ou até o seu descarte.

Sendo assim, a identificação de anomalias presentes nas amostras é de grande importância para a otimização dos conjuntos de calibração e validação e a exclusão destas amostras permite a construção de modelos mais eficientes e precisos e com melhor capacidade de previsão ³⁰.

Existem muitas técnicas que são utilizadas para a identificação de amostras anômalas, como por exemplo: a distância de Mahalanobis ⁵⁰, o método da

incerteza ⁵¹, funções potenciais ⁵¹, RHM (do inglês – *Resampling by the Half-Means*) ⁵¹, SHV (do inglês – *Smallest Half-Volume*) ⁵¹, entre outras.

No presente trabalho, a identificação das amostras anômalas nos conjuntos de calibração e validação foi realizada de acordo com as recomendações da norma E1655-05 ASTM (*American Society for Testing and Materials*) ⁵² e da referência 30.

3.10.1 Amostras anômalas na calibração

As amostras anômalas na calibração são usualmente avaliadas com base no *leverage* extremo, nos resíduos não modelados nos dados espectrais e nos resíduos não modelados na variável dependente ^{30,52}.

O *leverage* representa o quanto uma amostra está distante da média do conjunto dos dados, ou seja, o peso relativo de uma amostra em relação às demais amostras presentes no conjunto. Utilizando dados espectrais como exemplo, o *leverage* mede a diferença no espectro de uma determinada amostra comparado com os espectros das demais amostras do conjunto de dados. Desta forma, uma amostra com *leverage* baixa está próxima ao centro do conjunto de calibração, e uma amostra com alto valor de *leverage* está distante do centro do conjunto das amostras de calibração. O *leverage* pode ser representado por ^{30,52}:

$$h_i = \mathbf{t}_{A,i}^T (\mathbf{T}_A^T \mathbf{T}_A)^{-1} \mathbf{t}_{A,i} \quad (26)$$

em que \mathbf{T} são os escores de todas as amostras de calibração, \mathbf{t}_i é o vetor de escores de uma determinada amostra e A é o número de variáveis latentes.

Amostras com alto valor de h podem ser consideradas anômalas. De acordo com a norma ASTM E1655-05 ⁵², amostras com h_i maiores que um valor limite (h_{lim}) devem ser removidas e o modelo reconstruído.

$$h_{limit} = 3 \frac{A+1}{I_c}$$

onde I_c é o número de amostras de calibração. Deve-se notar, que para modelos não centrados na média o fator 1 deve ser desconsiderado.

No entanto, é comum após a eliminação de amostras com alto *leverage*, o surgimento de novos espectros com a reconstrução do modelo, que apresentem $h_i > h_{lim}$. Quando aplicações sucessivas do teste resultam na identificação de novas amostras como sendo anômalas, um fenômeno conhecido como “*snowball*” (do inglês, efeito bola de neve) pode estar acontecendo. Caso este efeito ocorra, há um indício de algum problema com a estrutura dos dados espectrais.

Se for observada a ocorrência do efeito “*snowball*”, o teste para a identificação de *outliers* com base no *leverage* pode ser relaxado. Sendo assim, o procedimento a ser seguido envolve a construção de um modelo a partir do conjunto de calibração inicial. Em seguida, os espectros da calibração que apresentarem $h_i > h_{lim}$ devem ser removidos do conjunto de calibração inicial. Um segundo modelo utilizando o mesmo número de variáveis latentes (A) deve ser construído com o subconjunto de espectros de calibração e os espectros de calibração com $h_i > h_{lim}$ são identificados no segundo modelo. O segundo modelo pode ser utilizado como critério de parada do teste de identificação de amostras anômalas baseado no *leverage*, desde que as amostras de calibração apresentem um valor de h inferior a 0,5. Se for o caso, os *outliers* para o segundo modelo são removidos e um terceiro modelo é construído ⁵².

A identificação de anomalias baseadas nos resíduos não modelados nos dados espectrais é obtida através da comparação do desvio padrão residual total ($s(\hat{e})$) com o desvio padrão residual de uma amostra particular ($s(\hat{e}_i)$), definidos como:

$$s(e)^2 = \frac{1}{I_c J - J - A_{\max}(I_c, J)} \sum_{i=1}^{I_c} \left(\sum_{j=1}^J \left(x_{ij} - \hat{x}_{ij} \right)^2 \right) \quad (27)$$

$$s(e_i)^2 = \frac{I_c}{I_c J - J - A_{\max}(I_c, J)} \sum_{j=1}^J \left(x_{ij} - \hat{x}_{ij} \right)^2 \quad (28)$$

em que J é o número de variáveis espectrais e I_c é o número de amostras da calibração. Se uma amostra apresentar $s(\hat{e}_i) > 3s(\hat{e})$ esta deve ser removida do conjunto de calibração³⁰.

A identificação de *outliers* com relação aos resíduos não modelados na variável dependente, é feita através da comparação da raiz quadrada do erro médio da calibração (RMSEC) com o erro absoluto daquela amostra específica. Se a amostra apresentar erro absoluto ($y_i - \hat{y}_i$) maior do que (3xRMSEC), esta é definida como sendo um *outlier*³⁰. O RMSEC é determinado como:

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^{I_c} (y_i - \hat{y}_i)^2}{v}} \quad (29)$$

onde v é o número de graus de liberdade, y_i é o valor de referência e \hat{y}_i é o valor estimado.

3.10.2 Amostras anômalas na validação

Para a identificação de amostras anômalas no conjunto de validação, devem ser feitos testes baseados no *leverage* e nos resíduos espectrais como foram descritos anteriormente.

Novamente, temos que as amostras com alto *leverage*, ou seja, $h_i > h_{\text{lim}}$ devem ser removidas e uma nova validação deve ser feita na ausência destas amostras. Com relação aos resíduos espectrais, a comparação entre o desvio padrão residual total e o desvio padrão residual de uma amostra deve ser feita e caso o desvio padrão residual de uma amostra for maior que três vezes o desvio padrão residual total, esta deve ser removida.

CAPÍTULO 4

Objetivos

4. Objetivos

Este trabalho tem como objetivo a realização de um estudo empregando a Espectroscopia no Infravermelho Próximo (NIR) e a calibração multivariada no desenvolvimento de metodologias para análises diretas em amostras de cavacos de madeira para determinação de parâmetros de qualidade como densidade básica, lignina total e extrativos em etanol/tolueno.

Além disso, o trabalho visa à realização de um estudo em amostras de cavaco úmidas, visando a separação e a classificação de amostras, mesmo estas sendo desconhecidas. Outro objetivo é a previsão dos mesmos parâmetros de qualidade dos cavacos mencionados anteriormente em amostras com diferentes teores de umidade.

CAPÍTULO 5
Parte Experimental

5. Parte Experimental

5.1 Amostras

As amostras foram coletadas no período de agosto a dezembro de 2008. No período de 12 de janeiro a 13 de fevereiro de 2009 a obtenção dos espectros e as análises químicas padrões foram realizadas na Fibria – unidade Barra do Riacho – ES.

Foram utilizadas 160 amostras de diferentes localidades (Bahia, Espírito Santo e Rio Grande do Sul) e com diferentes idades das árvores (3 a 29 anos). As amostras foram coletadas na forma de cavacos de madeira já secos ao ar livre e foram levadas ao laboratório do Centro de Pesquisa e Tecnologia da Fibria – unidade Barra do Riacho – ES, onde após climatização de no mínimo 1 dia foram obtidos os espectros de Infravermelho Próximo (NIR).

5.2 Espectros

Foram obtidos espectros de reflectância difusa na região de 4000 a 10000 cm^{-1} (com resolução de 4 cm^{-1}) em espectrofotômetro Antaris II – Thermo Electron Corporation para as amostras de cavaco. O espectro para cada amostra foi obtido com a média de 10 cavacos e utilizando-se 64 varreduras para cada cavaco. Os modelos foram desenvolvidos em *Matlab* 6.5 (MathWorks Co).

5.3 Construção dos modelos quimiométricos

Os modelos foram desenvolvidos em *Matlab* 6.5 (MathWorks Co). Para os modelos construídos com o algoritmo iPLS para amostras de cavacos secas, para cada parâmetro foram testados modelos com 5 a 10 intervalos, com os pré-processamentos MSC, primeira e segunda derivadas e centragem na média. Na segunda parte do trabalho, utilizando amostras de cavacos úmidas, o PLS-DA

também foi testado com outros pré-processamentos. Já os modelos construídos com o LS-SVM, os parâmetros γ e σ^2 foram otimizados com base na superfície fornecida pelo próprio algoritmo.

5.4 Densidade básica

A medida da densidade básica foi realizada no Centro de Pesquisa e Tecnologia da Fibria de acordo com a norma técnica Tappi T258 om06 ⁶. A repetitividade esperada na determinação deste parâmetro é de 18 Kg/m³.

Foram transferidas cerca de 500 g de cavacos para uma sacola plástica, que foi perfurada e imersa em tanque de água quente (60-70 °C) até que ficasse totalmente submersa (aproximadamente 3 dias). Em seguida a amostra foi transferida para um kitassato de 4000 mL contendo 3000 mL de água quente, que foi tampado com rolha de borracha e então colocado sob pressão reduzida para a retirada de bolhas de ar (aproximadamente 12 horas).

Colocou-se um recipiente plástico sobre uma balança, no qual foi adicionado um volume de água suficiente para encobrir as amostras contidas em um cesto de arame que por sua vez foi submerso no recipiente plástico. Este cesto já possui peso seco e hidrostático registrado. Então a medida do peso hidrostático das amostras foi aferida após centrifugação dos cavacos por 5 segundos.

As amostras foram secas em estufa a $105,0 \pm 3,0$ °C por um período de no mínimo 12 horas. Em seguida, as amostras foram resfriadas em dessecador e o peso seco foi registrado.

5.5 Extrativos em etanol/tolueno

A análise da porcentagem de extrativos foi realizada na Fibria – unidade Barra do Riacho – ES, de acordo com a norma Tappi 204 cm97 ⁹. A repetitividade

na determinação da quantidade de extrativos em etanol/tolueno presente na amostra é de 0,4%.

As amostras de cavacos foram secas em estufa a $50 \pm 5^\circ\text{C}$ e em seguida moídas em moinho de facas com peneira de 1 mm de diâmetro. A amostra foi então ambientada ao laboratório. Pesou-se aproximadamente 6,0 g de amostra em cartucho de extração e este foi fechado com algodão para evitar projeções e também para fixá-los ao equipamento. Adicionou-se aproximadamente 60 mL do solvente (etanol/tolueno 1:2 v/v) nos copos de extração e estes foram fixados no suporte do equipamento (Extrator Soxtec/Soxlet). Realizou-se a extração por 1,5 h, seguida de 1,5 h de lavagem com o mesmo solvente.

Após a lavagem, removeram-se os copos e secou-se em estufa a $105 \pm 3^\circ\text{C}$ até peso constante. Os copos foram transferidos para o dessecador e estes foram pesados. O procedimento foi repetido para uma amostra do branco (cartucho de extração preenchido somente com algodão).

5.6 Lignina total

A análise da porcentagem de lignina foi realizado de acordo com a norma Tappi 222 om02⁸ na Fibria – unidade Barra do Riacho – ES. A repetitividade na determinação da lignina total é de 0,7%.

Após a realização do teste de extrativos, deve-se realizar a análise de determinação da lignina, com as amostras livres dos extrativos orgânicos. Pesou-se $1,0 \pm 0,1\text{g}$ da amostra de madeira e imediatamente depois pesou-se outra amostra para determinação da umidade natural (peso seco da amostra) que foi posteriormente utilizado no cálculo da porcentagem de lignina solúvel.

Adicionou-se em béqueres de 100 mL contendo a madeira moída, 15 mL de ácido sulfúrico 12 mol.L^{-1} em banho a $(20 \pm 1)^\circ\text{C}$, macerando com bastão de vidro. O material foi mantido nesta condição por 2 horas e com agitação freqüente para que houvesse completa impregnação. Após este tempo, o material foi transferido do béquer para um erlenmeyer de 1000 mL lavando-o com água

deionizada. Completou-se o volume com água deionizada quente (100 °C) até 575 mL.

Ferveu-se a solução por quatro horas, mantendo o volume constante através de freqüente adição de água deionizada quente (100 °C). Após este período, retirou-se o erlenmeyer do aquecimento e a mistura foi mantida em repouso para precipitação por um período de 12 horas.

Filtrou-se a solução sobrenadante sem agitação do precipitado através de um cadinho de filtro (porosidade média) previamente tarado. Retirou-se uma porção deste filtrado para a determinação da lignina solúvel em ácido (através de espectroscopia no ultravioleta com medidas de absorbância em 206 e 246 nm).

Lavou-se a lignina insolúvel restante no erlenmeyer com água quente (100 °C) até que estivesse livre do ácido (verificação com papel indicador). Secou-se o cadinho contendo a lignina insolúvel em ácido em estufa a $(105 \pm 3)^\circ\text{C}$ até peso constante. Este foi esfriado em dessecador e pesado.

5.7 Amostras úmidas

O ensaio com as amostras de cavacos úmidas envolveu a simulação de quatro teores de umidade: um com a umidade natural dos cavacos (apenas secos ao ar), outro com respingos de água (aspersão), outro com imersão em água e por fim uma com vaporização de água.

Nas amostras com simulação de aspersão, os cavacos foram colocados em um béquer de 400 mL, e receberam respingos de água aleatórios. Em seguida, aguardou-se 15 minutos para absorção de água e então os espectros foram coletados. Nesse procedimento obtinha-se um teor de umidade de aproximadamente 8 a 10 %.

Nas amostras de imersão, os cavacos foram imersos em béqueres de 600 mL contendo 400 mL de água a temperatura ambiente. A amostra permanecia imersa por um período de 15 horas, e então era retirada, colocada em papel absorvente para remoção do excesso de água e então os espectros foram

coletados. Com esse procedimento, as amostras obtinham uma umidade de aproximadamente 50 a 60 %.

As amostras que recebiam vaporização de água eram colocadas em um cesto dentro de um mini-digestor, onde recebiam vapor de água a 105°C por um período de 20 minutos. Após este período os cavacos eram retirados e logo em seguida os espectros eram coletados. A umidade obtida neste procedimento é de aproximadamente 70%.

As amostras secas naturalmente ao ar possuem apenas uma umidade natural de aproximadamente 5%.

CAPÍTULO 6
Resultados e Discussões:
Amostras secas

6. Resultados e Discussões – Amostras secas

Foram obtidos os espectros de 160 amostras diretamente sobre o cavaco inteiro (não moído) seco (seco naturalmente ao ar). Na composição do vetor das variáveis dependentes (\mathbf{y}) foram utilizados os valores de referência para os parâmetros densidade básica, extrativos em etanol/tolueno e lignina total, obtidos através de procedimento padrão, conforme descrito na parte experimental. Desta forma, foi obtida uma matriz \mathbf{X} composta pelas absorvâncias nos diferentes números de onda de cada amostra, sendo que cada linha da matriz é representada por uma amostra diferente e as colunas representam as respectivas absorvâncias para as amostras estudadas em cada número de onda.

A separação das amostras nos conjuntos de calibração e validação para a construção dos modelos, utilizando o algoritmo de Kennard – Stone, resultou em 107 amostras para o conjunto de calibração e 53 amostras para o conjunto de validação.

A Figura 9 mostra os espectros, na região do infravermelho próximo, dos cavacos de madeira para todas as amostras, na região de 4000 a 10000 cm^{-1} .

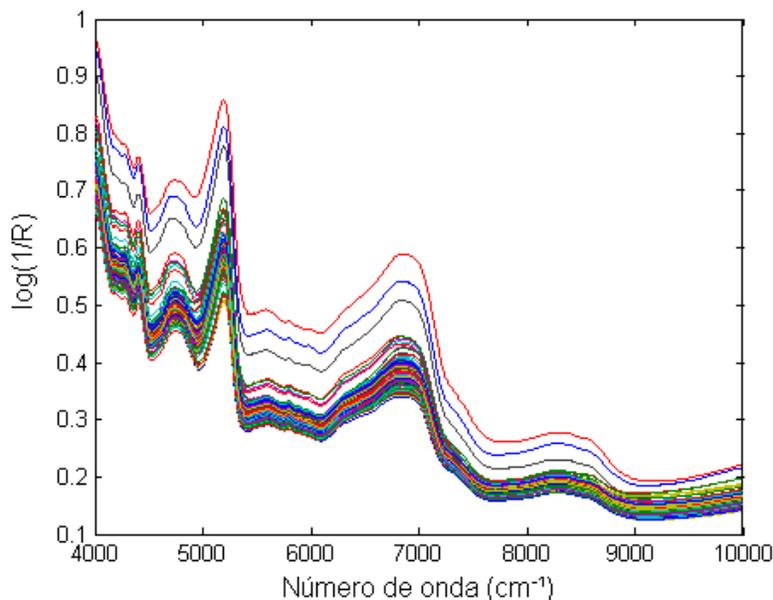


Figura 9. Espectros originais das 160 amostras de cavaco.

6.1 Densidade básica

Para o modelo final construído com o parâmetro densidade básica, foi utilizado o algoritmo iPLS com 7 intervalos, sendo que o intervalo selecionado pelo algoritmo foi o primeiro (4000-4856 cm^{-1}), com 6 variáveis latentes. Para a construção do modelo foram utilizados como pré-processamentos a primeira derivada através do algoritmo de Savitzky – Golay (Figura 10) e a centragem na média e em seguida as amostras foram separadas em conjunto de calibração (107 amostras) e de validação (53 amostras) utilizando o algoritmo de Kennard-Stone. Outros pré-processamentos e intervalos foram testados, porém este foi o que apresentou o melhor resultado.

O número de variáveis latentes (VLs) para cada modelo de calibração multivariada é determinado através dos resultados obtidos da raiz quadrada do erro quadrático médio de validação cruzada (RMSECV). Os valores de RMSECV para cada intervalo do iPLS são calculados, assim como o RMSECV do modelo global (construído utilizando-se todas as variáveis espectrais). Assim, o intervalo que possui o menor RMSECV comparado ao valor do RMSECV obtido para o modelo global é selecionado (Figura 11).

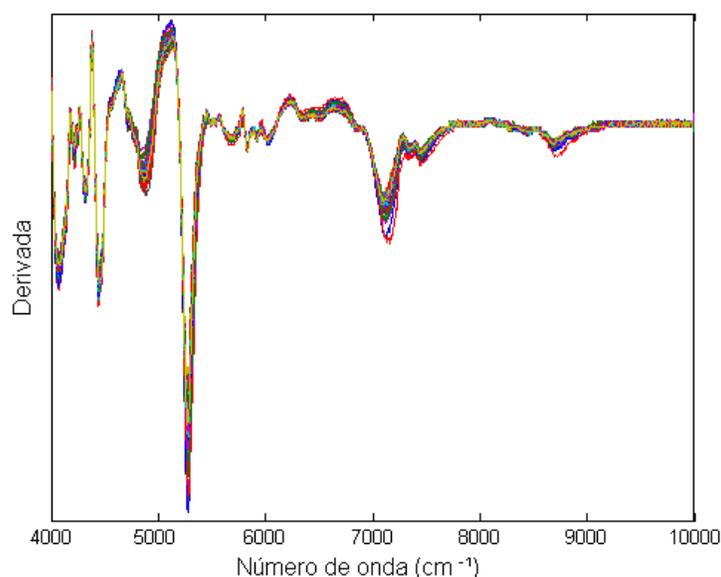


Figura 10. Primeira derivada dos espectros originais para as 160 amostras.

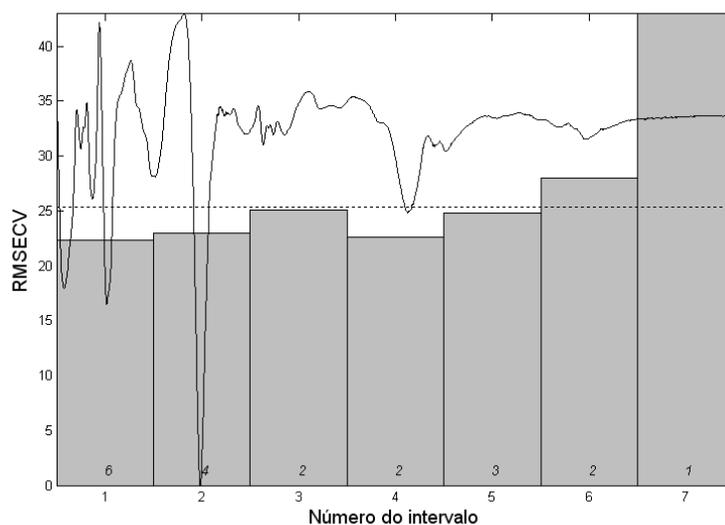


Figura 11. iPLS com 7 intervalos, onde as barras indicam o erro de cada intervalo, a linha tracejada o erro (RMSECV) obtido com o espectro inteiro, em itálico o número de VLs para o modelo em determinado intervalo e a linha contínua é o espectro médio obtido para o modelo.

Através da Figura 11 pode-se observar que os intervalos 1 e 4 possuem valores de erro (RMSECV) semelhantes, porém a escolha do primeiro intervalo e não do quarto para a construção do modelo foi baseada no fato da informação química relacionada à densidade básica estar presente no primeiro intervalo. O quarto intervalo pode ter outras informações relacionadas indiretamente com o parâmetro.

A construção do gráfico com os valores de densidade obtidos pelo método padrão e os valores previstos pelo modelo (Figura 12) indica que foi obtida uma correlação entre os valores de 0,63. O RMSEP obtido foi de apenas 22,4 kg/m³ e a média dos erros relativos para a determinação da densidade básica através do modelo construído foi de 4,4%.

O modelo construído com todo o espectro apresenta valores de RMSEP (22,2 Kg/m³), RMSECV (25,6 Kg/m³) e média de erros relativos (4,5%) próximos aos do modelo obtido com o iPLS, porém apresenta um menor valor de RMSEC (9,8 Kg/m³) o que pode ser uma indicativa de sobreajuste do modelo.

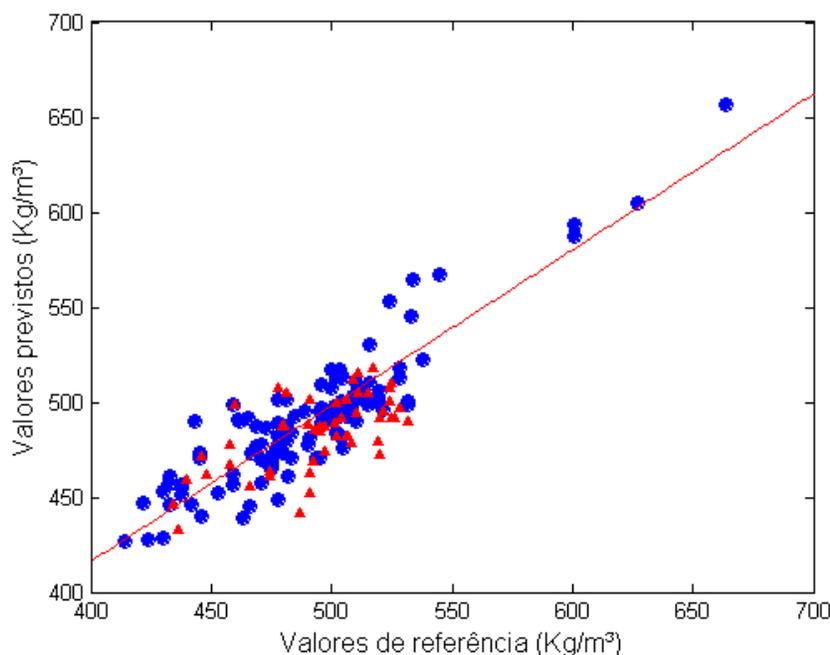


Figura 12. Valores de densidade de referência (análise química convencional) *versus* previstos pelo modelo para as amostras de calibração (●) e validação (▲).

A Tabela 3 mostra o ajuste do modelo de iPLS.

Tabela 3. Valores previstos pelo modelo.

Modelo	RMSEC (Kg/m ³)	RMSECV (Kg/m ³)	RMSEP (Kg/m ³)	Média Er (%)
iPLS	16,7	22,3	22,4	4,4
PLS	9,8	25,6	22,2	4,5

Através da comparação entre o RMSEP obtido na construção deste modelo e o valor de repetitividade do método padrão (18 Kg/m³), pode-se dizer que o modelo pode ser utilizado para a previsão da densidade básica de uma determinada amostra.

A Figura 13 mostra um gráfico com barras de erros relativos, que foram determinados segundo a equação:

$$Er = \left[\frac{y_{prev} - y_{real}}{y_{real}} \right] \times 100 \quad (30) \quad \text{Média Er} = \sqrt{\sum \frac{(Er)^2}{n}} \quad (31)$$

onde n é o número de amostras de validação.

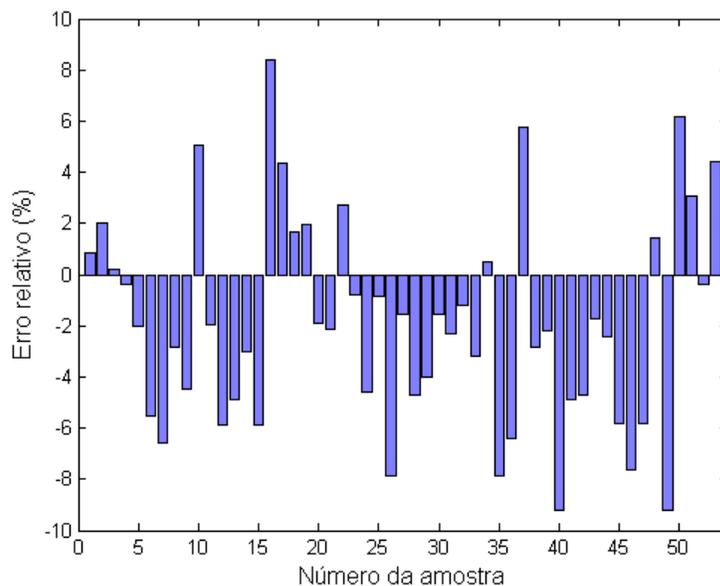


Figura 13. Gráfico com barras de erro relativo para cada amostra de validação.

O gráfico da Figura 13 indica que os erros estão relativamente distribuídos como positivos e negativos e pode ser uma indicativa de que o modelo pode ser utilizado para a previsão dos valores de densidade básica das amostras desconhecidas.

A detecção de amostras anômalas (*outliers*) foi realizada com base no *leverage* extremo, nos resíduos não modelados dos dados espectrais e nos resíduos não modelados na variável dependente. Desta forma, o *leverage* das amostras de calibração e os resíduos espectrais destas mesmas amostras estão representados na Figura 14. A ausência de uma ou mais amostras que apresentem um *leverage* alto e simultaneamente um resíduo alto indica a ausência de uma amostra anômala no conjunto de calibração, baseado nestes dois aspectos. O limite no qual acima deste a amostra apresenta *leverage* ou resíduo alto está representado por uma linha tracejada. O limite do *leverage* (h_{lim}) e o limite dos resíduos espectrais ($s(\hat{\epsilon}_i) > 3s(\hat{\epsilon})$) foram descritos anteriormente no capítulo 3.

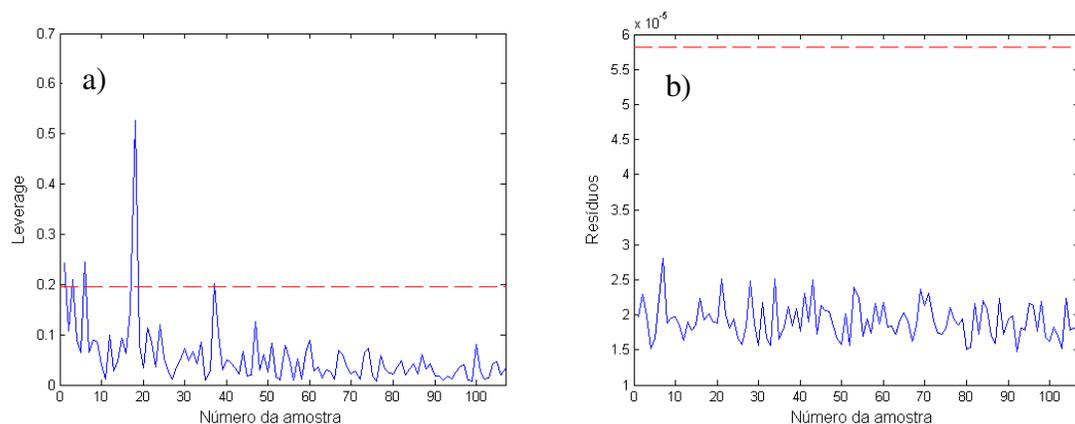


Figura 14. a) Gráfico do *leverage* das amostras de calibração. b) Gráfico dos resíduos espectrais das amostras de calibração.

Através da observação dos gráficos da Figura 14, é possível concluir que nenhuma amostra de calibração apresentou simultaneamente *leverage* extremo e resíduo alto, e, portanto pode-se concluir que nenhuma amostra de calibração deve ser considerada anômala para o parâmetro densidade básica e com base nestes dois gráficos.

Além disso, é necessário verificar a presença de amostras de calibração que possuam resíduos não modelados na variável dependente. Como neste teste de detecção de *outliers* nenhuma amostra apresentou erro absoluto superior a três vezes o RMSEC, o modelo não apresenta *outliers* na calibração³⁰.

Para as amostras de validação, o mesmo procedimento foi realizado para detecção de amostras anômalas baseadas no *leverage* extremo e nos resíduos espectrais não modelados. Os gráficos estão representados na Figura 15 e as amostras de validação não apresentaram *leverage* altos ou resíduos espectrais não modelados, uma vez que os valores para cada amostra nestes dois gráficos são inferiores ao limite estabelecido. Portanto, para o parâmetro densidade básica, não foi constatada a presença de nenhuma amostra anômala de validação.

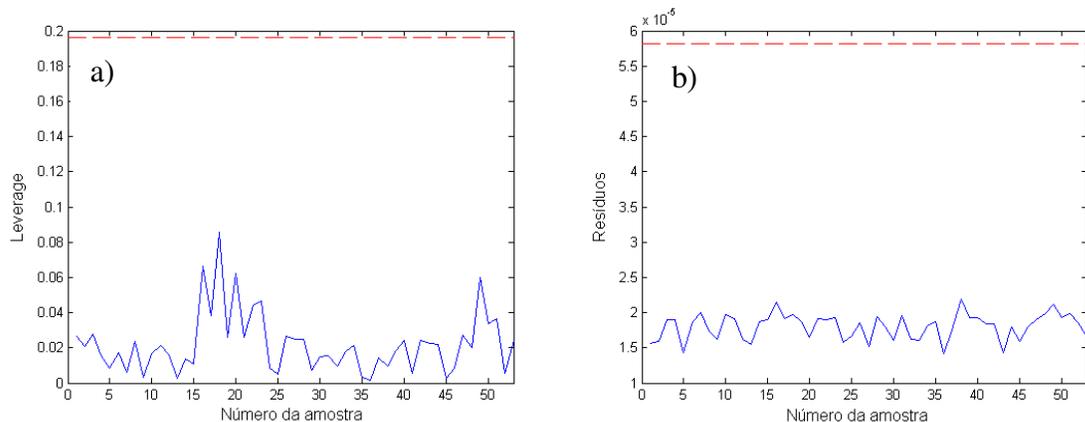


Figura 15. a) Gráfico do *leverage* das amostras de validação. b) Gráfico dos resíduos espectrais das amostras de validação.

Foram realizadas atribuições de bandas na região do infravermelho próximo, especificamente na região espectral selecionada pelo algoritmo do iPLS ($4000\text{-}4856\text{ cm}^{-1}$). Essa região espectral tem grande importância para o desenvolvimento de um modelo de calibração pois apresenta bandas características da celulose, que por ser um dos principais constituintes da madeira exerce grande influência no valor da densidade básica. A Tabela 4 relata a região (em número de onda) e as ligações responsáveis por tal absorção.

Tabela 4. Atribuição das bandas no NIR.

Número de onda (cm^{-1})	Atribuição
4019	ν C-H e combinação ν C-H
4252	δ C-H
4261	Combinação ν CH ₂ simétrico e δ CH ₂
4283	δ C-H mais combinação de δ CH ₂

6.2 Lignina Total

Para a construção do modelo com o parâmetro lignina total foi utilizado o MSC (Correção de Espalhamento Multiplicativo) como pré-processamento (Figura 16) e as amostras foram também centradas na média. Em seguida as amostras foram separadas em conjunto de calibração (107 amostras) e de validação (53 amostras) utilizando o algoritmo de Kennard-Stone.

Para a construção do modelo, utilizou-se o algoritmo do iPLS com 7 intervalos. O intervalo selecionado pelo algoritmo foi o primeiro (4000-4856 cm^{-1}), com 6 variáveis latentes (Figura 17). Novamente, o intervalo selecionado foi o que apresentou o menor RMSECV comparado aos erros de validação cruzada para os outros intervalos e também com o valor obtido para o modelo global (construído com espectro inteiro e cujo valor de RMSECV está representado pela linha tracejada do gráfico).

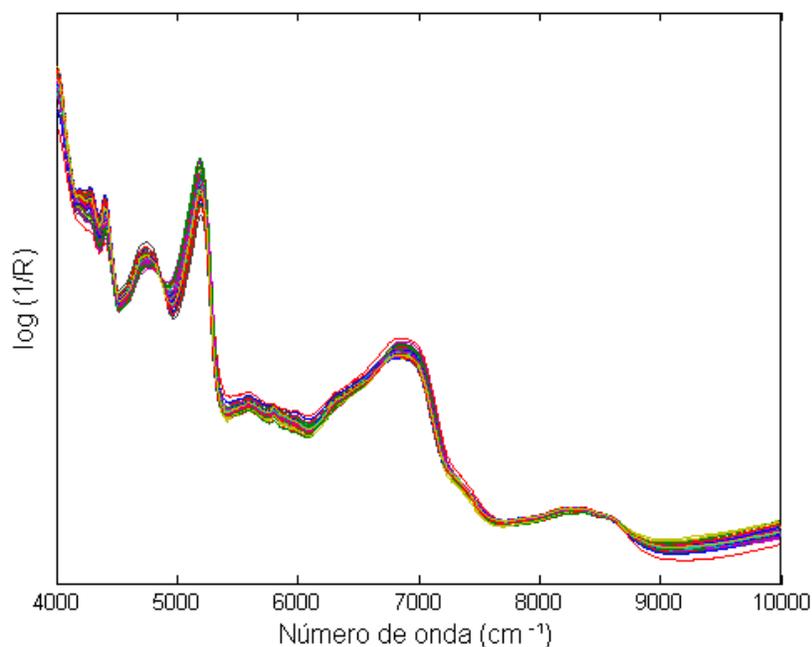


Figura 16. Espectros das 160 amostras após pré-processamento com MSC.

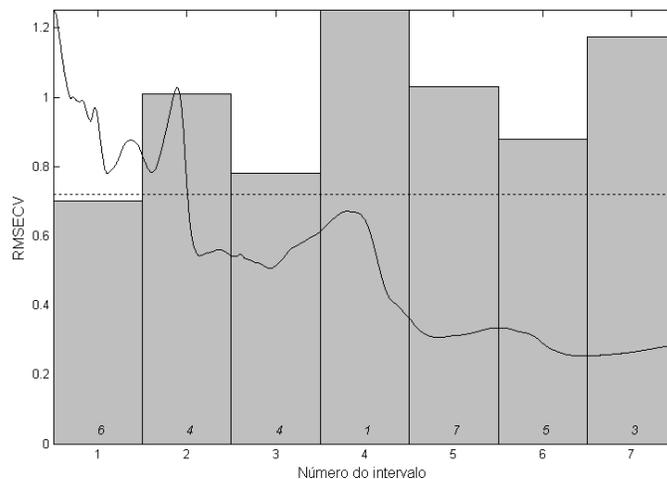


Figura 17. iPLS com 7 intervalos, onde as barras indicam o erro de cada intervalo, a linha tracejada o erro obtido com o espectro inteiro e em itálico o número de VLs para o modelo em determinado intervalo.

A construção do gráfico dos valores medidos de lignina contra os valores previstos pelo modelo (Figura 18), nos permite concluir que houve uma boa correlação entre os dados, o que também pode ser confirmado pelo baixo valor de RMSEP (0,73) fornecido pelo modelo. Além disso, foi calculada a média dos erros relativos das amostras de validação, obtendo-se o valor de 2,5%.

O modelo construído para o primeiro intervalo comparado ao modelo de PLS utilizando o espectro inteiro é similar, com valores de RMSEP bem similares, porém com média de erros relativos ligeiramente superiores (2,7%).

Com a realização de um teste F com nível de confiança de 95%, pode-se notar que não houve uma diferença significativa entre os modelos construídos com o PLS e com o iPLS. Porém a escolha do modelo de iPLS como o melhor modelo foi baseada na utilização de um algoritmo no qual uma atribuição química pudesse ser feita em uma certa região espectral característica.

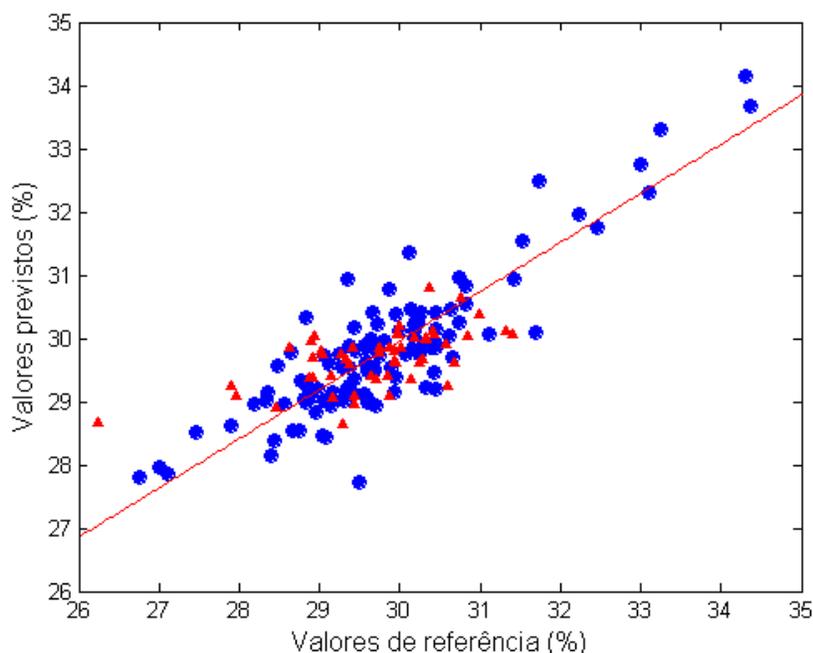


Figura 18. Valores de lignina de referência (análise química convencional) *versus* previstos pelo modelo para as amostras de calibração (●) e validação (▲).

A Tabela 5 mostra o ajuste do modelo de iPLS.

Tabela 5. Valores previstos pelo modelo.

Modelo	RMSEC (%)	RMSECV (%)	RMSEP (%)	Média Er (%)
iPLS	0,61	0,71	0,73	2,5
PLS	0,67	0,75	0,77	2,7

Comparando-se o valor de RMSEP obtido com o modelo construído para a determinação de lignina e o valor da repetibilidade do método padrão, pode-se observar que ambos são semelhantes e que, portanto, o modelo pode ser utilizado em substituição à análise química convencional para a previsão da lignina total presente em uma amostra desconhecida.

A Figura 19 mostra o gráfico de barras contendo o erro relativo da previsão das amostras de validação com o parâmetro lignina.

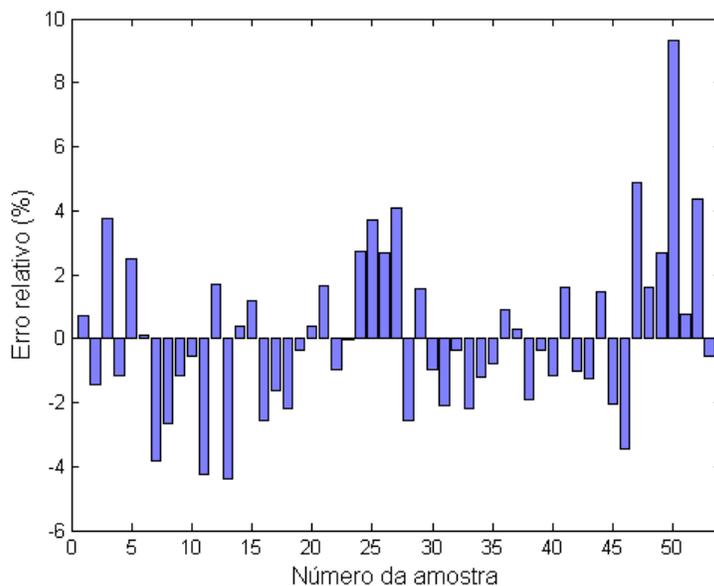


Figura 19. Gráfico com barras de erro relativo para cada amostra de validação.

O gráfico da Figura 19 indica que os erros estão aleatórios e é uma indicativa de que o modelo pode ser utilizado para a previsão dos valores de lignina total das amostras desconhecidas.

A detecção de amostras anômalas (*outliers*) foi realizada com base no *leverage* extremo, nos resíduos não modelados dos dados espectrais e nos resíduos não modelados na variável dependente. Desta forma, o *leverage* das amostras de calibração e os resíduos espectrais destas mesmas amostras estão representados na Figura 20. O limite no qual acima deste a amostra apresenta *leverage* ou resíduo alto está representado por uma linha tracejada. Para as amostras de calibração com o parâmetro lignina total, nenhuma amostra deste conjunto apresentou simultaneamente um elevado *leverage* e um elevado valor de resíduo. Sendo assim, com base nestes dois testes de detecção, não existem amostras anômalas no conjunto de calibração.

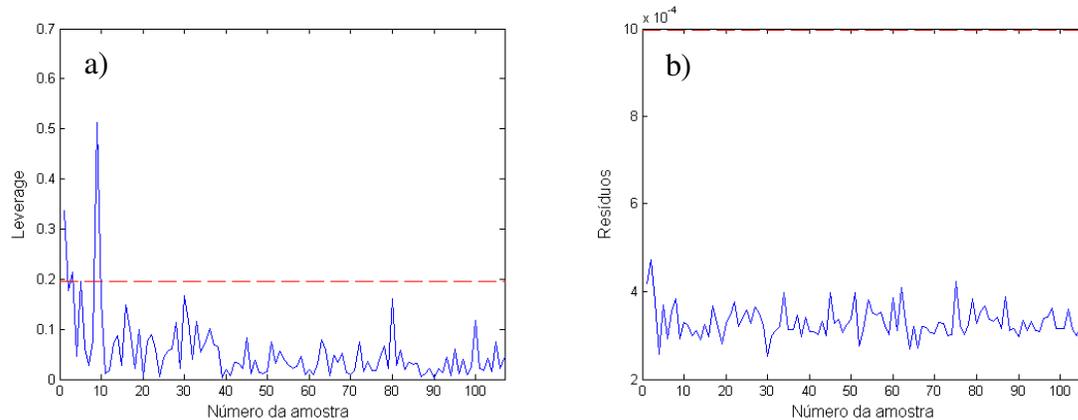


Figura 20. a) Gráfico do *leverage* das amostras de calibração. b) Gráfico dos resíduos espectrais das amostras de calibração.

Foi realizado o teste de detecção de *outliers* baseado nos resíduos não modelados na variável dependente e não havia nenhuma amostra cujo erro absoluto fosse mais de três vezes superior ao valor do RMSEC. Desta forma, não existem amostras anômalas no conjunto de calibração.

Para as amostras de validação, os testes de detecção de amostras anômalas baseados nos resíduos não modelados dos dados espectrais e nos *leverages* extremos não resultaram em nenhum *outlier*, uma vez que não existem amostras de validação com altos valores de *leverage* e de resíduos simultaneamente, como pode ser observado na Figura 21.

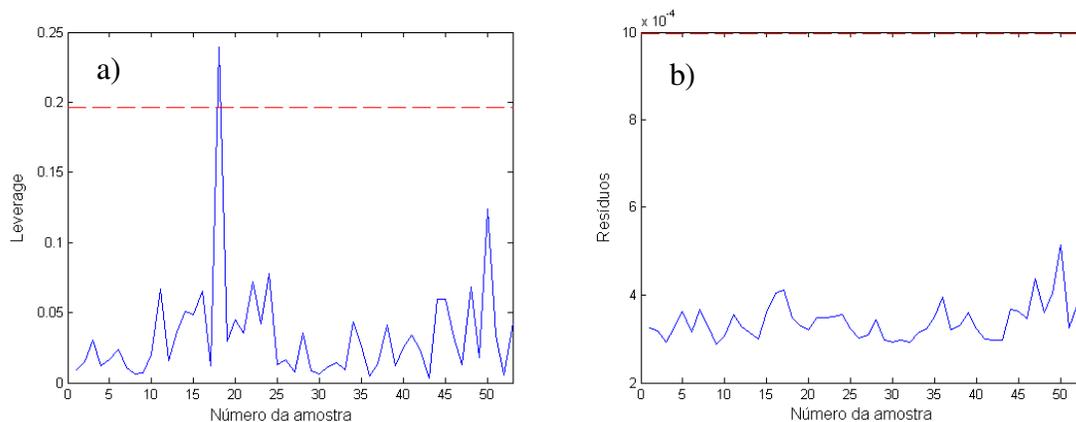


Figura 21. a) Gráfico do *leverage* das amostras de validação. b) Gráfico dos resíduos espectrais das amostras de validação.

Foram realizadas atribuições de bandas na região do infravermelho próximo, especificamente na região espectral selecionada pelo algoritmo do iPLS (4000-4856 cm^{-1}). Essa região espectral tem grande importância para o desenvolvimento de um modelo de calibração pois apresenta bandas características da lignina. A Tabela 6 relata a região (em número de onda) e as ligações responsáveis por tal absorção.

Tabela 6. Atribuição das bandas no NIR.

Número de onda (cm^{-1})	Atribuição
4545	ν C-H
4545	Combinação de C=O

6.3 Extrativos em etanol/tolueno

Para a construção do modelo com o parâmetro extrativos em etanol/tolueno, foi utilizado a Correção do Espalhamento Multiplicativo (MSC) como pré-processamento (Figura 22), e os dados também foram centrados na média. Além disso, o modelo utilizou 4 variáveis latentes.

Para a construção do modelo foi utilizado o algoritmo iPLS com 7 intervalos. Novamente o primeiro intervalo (4000-4856 cm^{-1}) foi selecionado (Figura 23) por apresentar o menor erro de validação cruzada comparado aos outros intervalos e também ao modelo global, construído utilizando o espectro inteiro. As amostras foram separadas, utilizando-se o algoritmo de Kennard-Stone em 107 amostras de calibração e 53 amostras de validação.

Os valores previstos pelo modelo para as amostras de cavaco e sua comparação com os valores de extrativos medidos através da análise química convencional (Figura 24) resultou em uma média dos erros relativos de 27,5%.

Comparando o modelo obtido para o primeiro intervalo com um modelo utilizando o espectro sem a seleção de parte das variáveis (espectro inteiro), pode-se concluir que o modelo utilizando o espectro inteiro obteve piores

resultados, tais como RMSEC (0,46%), RMSECV (0,54%), RMSEP (0,32%), além da média dos erros relativos ser superior (29,0%).

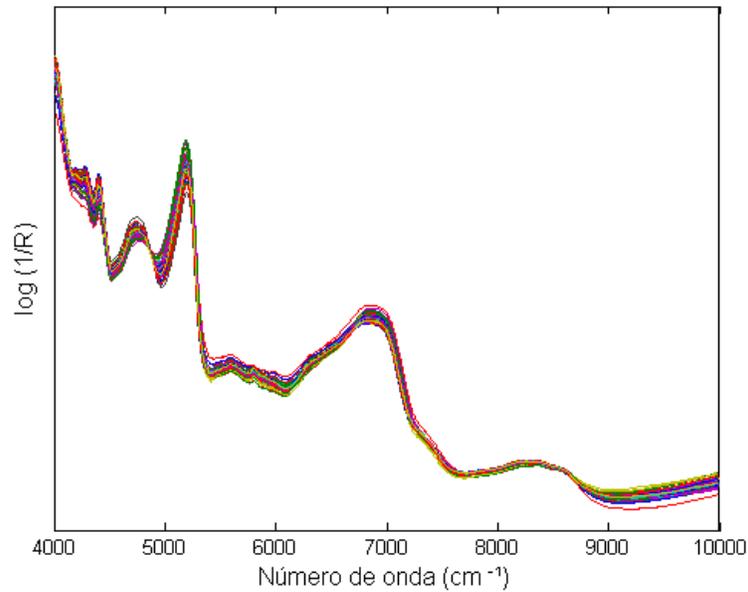


Figura 22. Espectros das amostras pré-processadas com o algoritmo MSC.

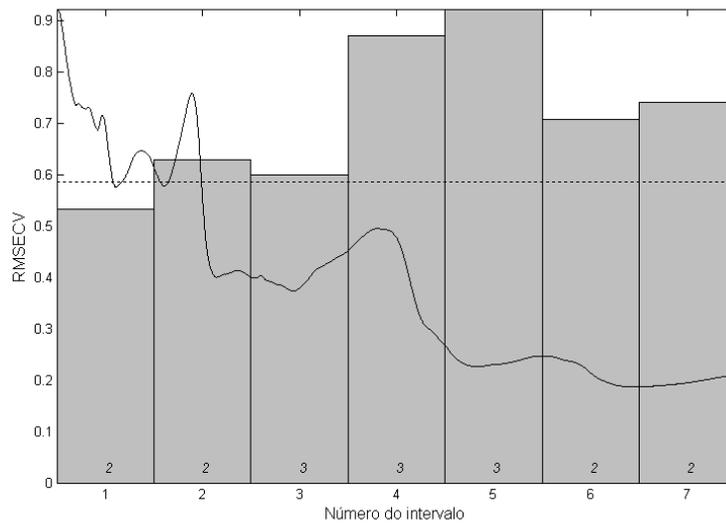


Figura 23. iPLS com 7 intervalos, onde as barras indicam o erro de cada intervalo, a linha tracejada o erro obtido com o espectro inteiro e em *ítilo* o número de VLs para o modelo em determinado intervalo.

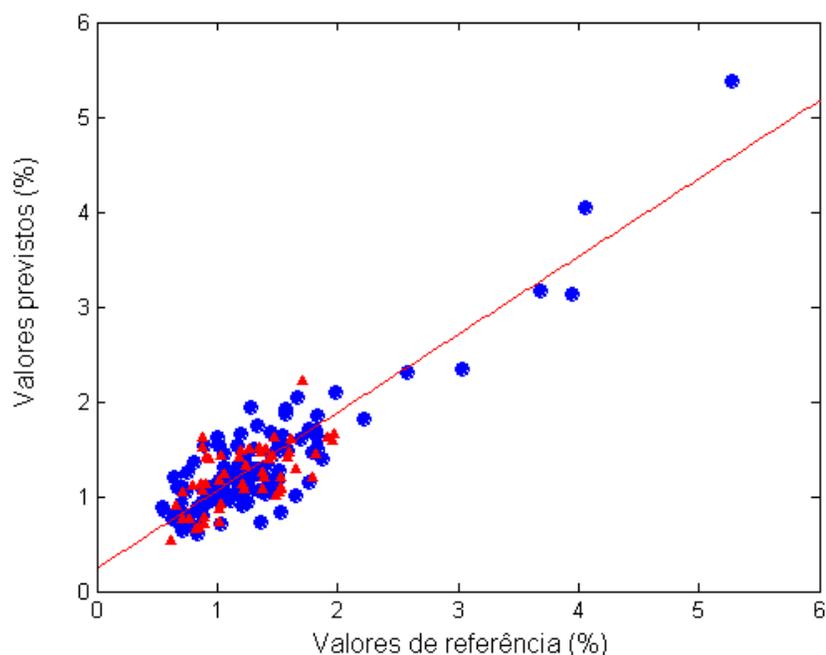


Figura 24. Valores de extrativos em etanol/tolueno de referência (análise química convencional) *versus* previstos pelo modelo para as amostras de calibração (●) e validação (▲).

A tabela 7 mostra o ajuste do modelo.

Tabela 7. Valores previstos pelo modelo.

Modelo	RMSEC (%)	RMSECV (%)	RMSEP (%)	Média Er (%)
iPLS	0,30	0,33	0,30	27,5
PLS	0,46	0,54	0,32	29,0

O valor de RMSEP obtido na construção do modelo é próximo ao valor de repetitividade do método padrão. Assim, o modelo construído pode ser utilizado na previsão do valor de extrativos em uma amostra.

Na Figura 25 pode-se observar o gráfico de barras contendo o erro relativo da previsão das amostras de validação para os extrativos em etanol/tolueno.

Os valores obtidos para o modelo construído com o PLS e utilizando todas as variáveis espectrais resultou em um modelo com maiores valores de RMSEC (0,32%) e RMSEP (0,31%) e com média de erros relativos semelhantes. Desta forma, o melhor modelo obtido foi o construído utilizando o iPLS.

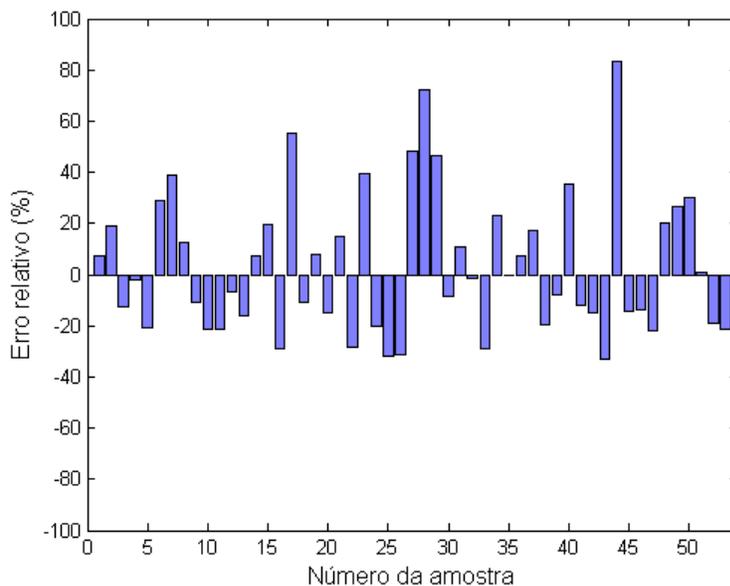


Figura 25. Gráfico com barras de erro relativo para cada amostra de validação.

O gráfico de barras indica que o modelo está adequadamente construído, uma vez que os erros estão distribuídos aleatoriamente, e é uma indicativa de que o modelo pode ser utilizado para a previsão da porcentagem de extrativos em etanol/tolueno para novas amostras. Pode-se atribuir os erros relativos elevados envolvidos na análise deste parâmetro ao fato de existirem muitas substâncias químicas com características diferentes que podem ser extraídas com a mistura etanol/tolueno e, portanto, talvez seja necessária a construção de um modelo para cada classe de substâncias químicas, sendo que estas substâncias juntas podem colaborar para o erro maior do modelo.

A detecção de amostras anômalas foi realizada com base no *leverage* extremo e nos resíduos espectrais não modelados e através da Figura 26 pode-se observar que nenhuma amostra de calibração apresentou valores acima dos limites estabelecidos simultaneamente nos dois gráficos. Assim, com base nestes dois tipos de detecção não houve nenhum *outlier* nas amostras de calibração.

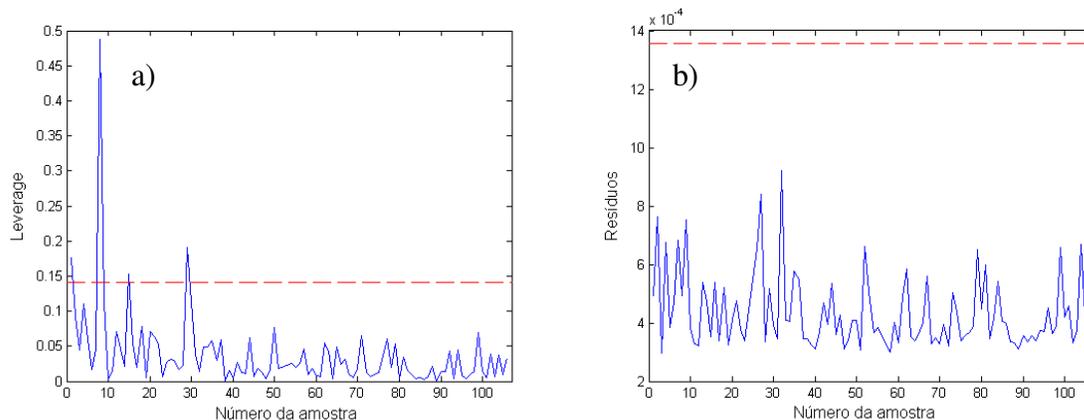


Figura 26. a) Gráfico do *leverage* das amostras de calibração. b) Gráfico dos resíduos espectrais das amostras de calibração.

Apesar dos gráficos da Figura 26 não indicarem a presença de nenhum *outlier* devido ao fato de nenhuma amostra possuir simultaneamente um *leverage* alto e um resíduo alto, foi detectada a presença de uma amostra anômala da calibração no teste dos resíduos não modelados da variável dependente. A amostra número 1 da calibração apresentou um valor de erro absoluto superior a três vezes o valor do RMSEC. O valor de RMSEC obtido para o primeiro modelo construído (com todas as amostras de calibração) foi de 0,43. Assim, $3 \times \text{RMSEC} = 1,29$. O erro absoluto (diferença entre o valor previsto pelo modelo e o valor real) da amostra 1 da calibração foi de $|2,56|$ e todas as outras amostras apresentaram erro absoluto inferior a 1,29. Desta forma, a amostra 1 foi retirada do conjunto de calibração e o modelo foi novamente construído, obtendo-se melhores valores de RMSEC e RMSEP.

A detecção de amostras anômalas no conjunto de validação foi realizada para *leverages* extremos e resíduos espectrais não modelados e os gráficos obtidos estão na Figura 27. Novamente, como não existem amostras acima dos limites estabelecidos em ambos os gráficos, pode-se afirmar que não existem *outliers* no conjunto de validação.

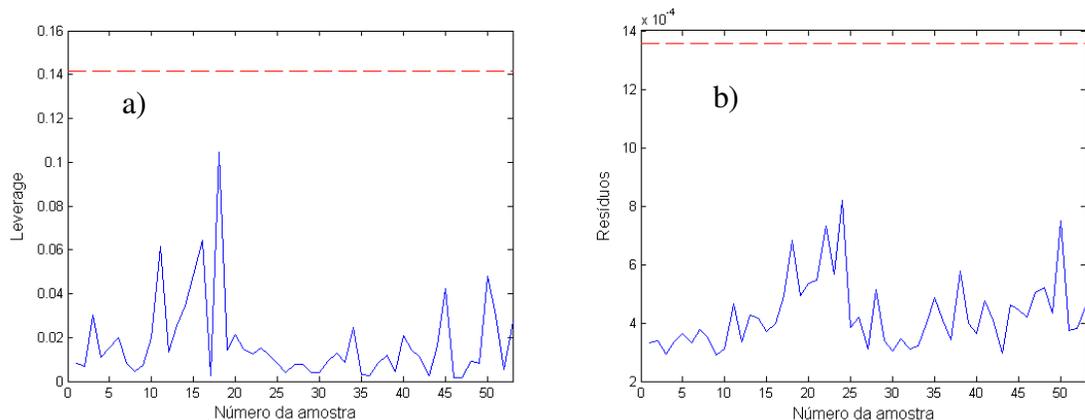


Figura 27. a) Gráfico do *leverage* das amostras de validação. b) Gráfico dos resíduos espectrais das amostras de validação.

Foram realizadas também as atribuições das bandas do infravermelho próximo na região espectral selecionada pelo algoritmo iPLS (4000-4856 cm^{-1}). Essa região espectral tem grande importância para o desenvolvimento de um modelo de calibração pois apresenta bandas características da celulose. Estas bandas relativas à celulose também estão relacionadas ao teor de extrativos, devido a presença ou não de substâncias orgânicas que podem ser extraídas por estes solventes na celulose. A Tabela 8 relata a região (em número de onda) e as ligações responsáveis por tal absorção.

Tabela 8. Atribuição das bandas no NIR.

Número de onda (cm^{-1})	Atribuição
4019	ν C-H e combinação ν C-H
4252	δ C-H
4261	Combinação ν CH ₂ simétrico e δ CH ₂
4283	δ C-H mais combinação de δ CH ₂

CAPÍTULO 7
Resultados e Discussões:
Amostras úmidas

7. Resultados e Discussões – Amostras úmidas

Foram realizados ensaios visando à determinação da influência que a água exerce sobre os espectros dos cavacos de madeira, e então, coletaram-se espectros dos cavacos com diferentes teores de umidade (Figura 28). Isso foi realizado para estudo da possibilidade da utilização do NIR para monitoramento dos parâmetros (densidade básica, lignina total e extrativos em etanol/tolueno) diretamente na linha de produção. Esses ensaios foram feitos através de uma simulação de várias intensidades de chuva que poderiam estar presentes na pilha de cavacos ao ar livre no pátio de produção.

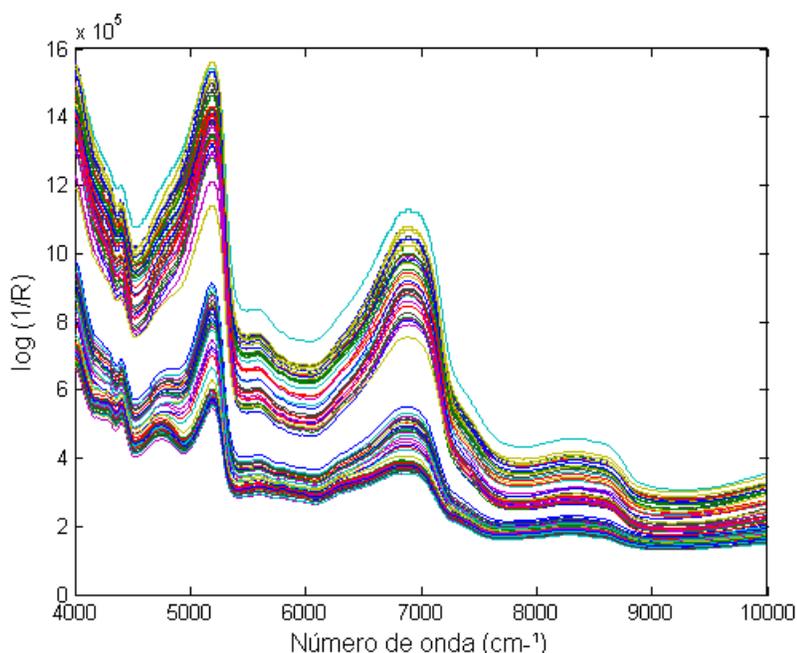


Figura 28. Espectros originais das 92 amostras com diferentes teores de umidade.

Inicialmente, estudou-se a influência da água nos espectros dos cavacos (Figura 28). Para isso, coletaram-se espectros de amostras de cavaco com umidades variadas (5 - 70 %) e observou-se um grande aumento na intensidade das bandas relacionadas à água (aproximadamente 5200 cm⁻¹ e 7000 cm⁻¹).

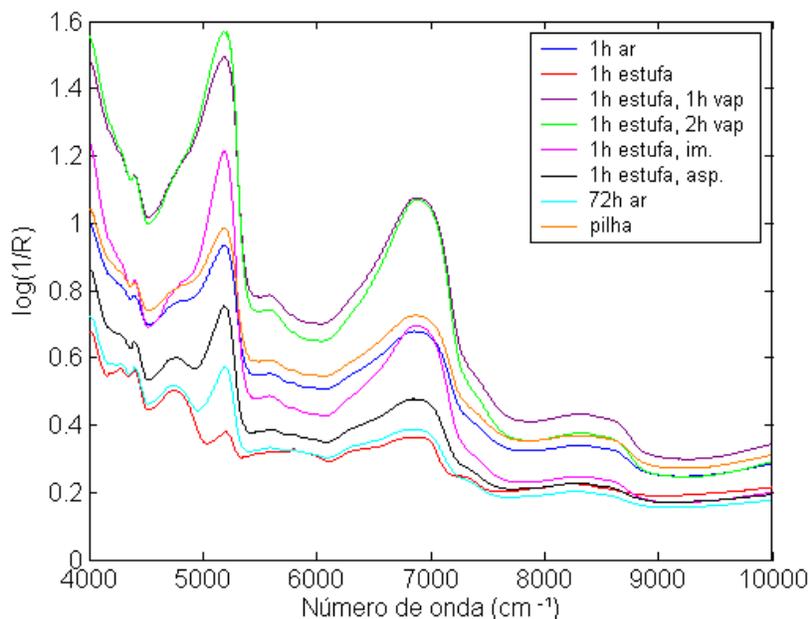


Figura 29. Espectros dos cavacos com vários teores de umidade diferente, mostrando a influência da água.

Através do gráfico da Figura 29, podemos observar que nos espectros obtidos diretamente sobre os cavacos, houve um aumento significativo na intensidade das bandas relacionadas à água (aproximadamente 5200 cm^{-1} e 7000 cm^{-1}) mesmo após a secagem dos cavacos em estufa e em seguida a realização de um dos processos que promove o aumento da umidade nas amostras. Com isso, fica comprovada a capacidade de penetração da água nos poros da madeira, promovendo novamente a inserção da água, mesmo após a secagem em estufa. Além disso, pode-se observar que o aumento na intensidade das bandas relacionada à água é proporcional a eficiência da geração de umidade, ou seja, quando maior a umidade proporcionada por um dos processos, maior a intensidade da banda no espectro dos cavacos.

7.1 Análise de Componentes Principais (PCA)

Foi realizada uma análise de componentes principais nos espectros das amostras que receberam diferentes teores de umidade. Com isso poderíamos

verificar se as amostras seriam separadas em classes de acordo com o teor de umidade que apresentavam. Utilizou-se o PCA com pré-processamentos MSC (Figura 30) e centrado na média, e foram necessários 2 componentes principais. Os resultados estão apresentados na Figura 31.

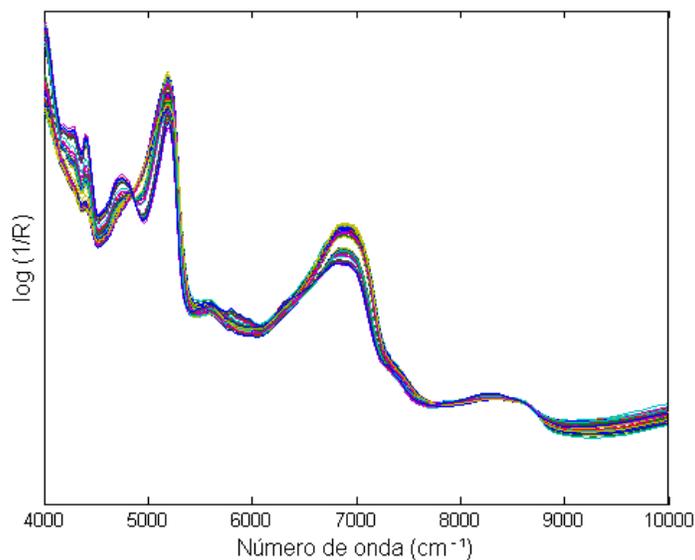


Figura 30. Espectros das 92 amostras após utilização do MSC como pré-processamento.

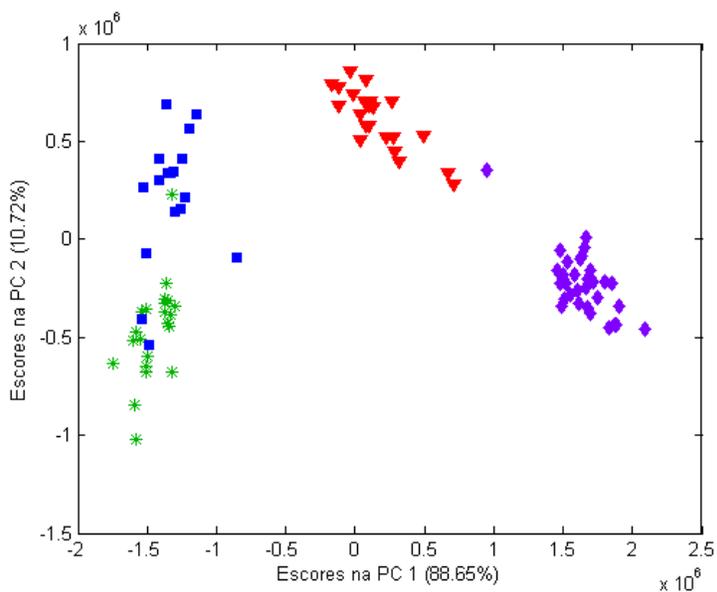


Figura 31. Gráfico dos escores da primeira componente principal *versus* escores da segunda componente principal. Amostras secas (◆), aspersão (▼), vaporização (✱) e imersão (■).

Através do gráfico dos escores de PC1 versus PC2, que explicam 99,3% da variância, podemos notar que as amostras estão divididas em 4 grupos, sendo cada um deles referente a diferentes teores de umidade.

Pode-se observar que a separação dos quatro grupos com teores de umidade diferentes não é observada nos espectros da Figura 28, onde se observa apenas dois grupos diferentes de espectros. Tal fato ocorre devido à existência de dois grupos com teores de umidades semelhantes (5 - 10%) e também outro grupo de espectros cujo teor de umidade também é semelhante (50-70%). Isto pode ser observado através da proximidade dos grupos de amostras de vaporização e imersão e também da proximidade das amostras secas com as de aspersão na Figura 31. Tal fato nos permite concluir que provavelmente será necessária a construção de um modelo diferente para cada grupo de umidade diferente.

Através do gráfico dos pesos (Figura 32) foi possível observar que as bandas referentes aos números de onda responsáveis pela absorção de água (aproximadamente 5200 e 7000 cm^{-1}) apresentam grande importância na construção do modelo do PCA, pois o peso destas é grande tanto na primeira quanto na segunda componente principal.

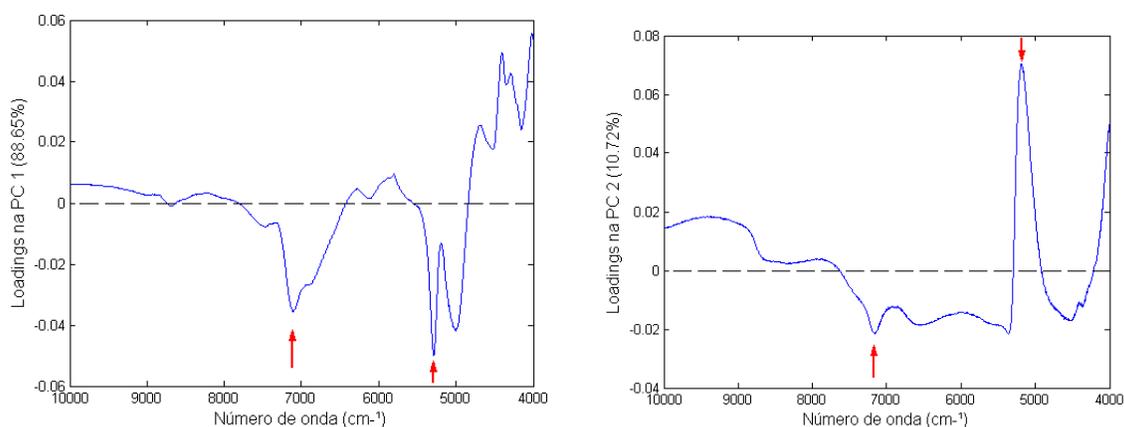


Figura 32. Gráfico dos pesos do PCA na PC1 (esquerda) e na PC2 (na direita).

7.2 Análise Discriminante com método de mínimos quadrados parciais (PLS-DA)

Diante dos resultados obtidos com o PCA, resolveu-se testar um método que fosse capaz de prever a classe pertencente de uma amostra desconhecida. Sendo assim, utilizou-se o PLS-DA com 6 variáveis latentes e como pré-processamento utilizou-se o MSC/centrado na média. A classe 1 representa amostras de aspensão (8-10% de umidade), a classe 2 de vaporização (aproximadamente 70% de umidade), a classe 3 de imersão (50-60% de umidade) e a classe 4 as amostras secas naturalmente ao ar (aproximadamente 5% de umidade). Os resultados obtidos para cada uma das 4 classes estão representados nas Figuras 33, 34, 35 e 36, respectivamente.

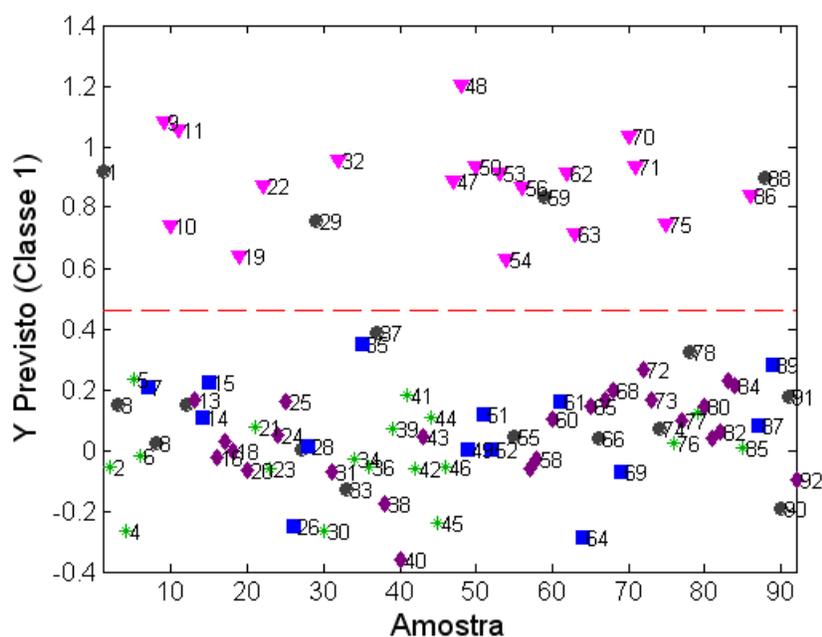


Figura 33. Classificação das amostras da classe 1 (aspensão ▼) e também de amostras de validação (●).

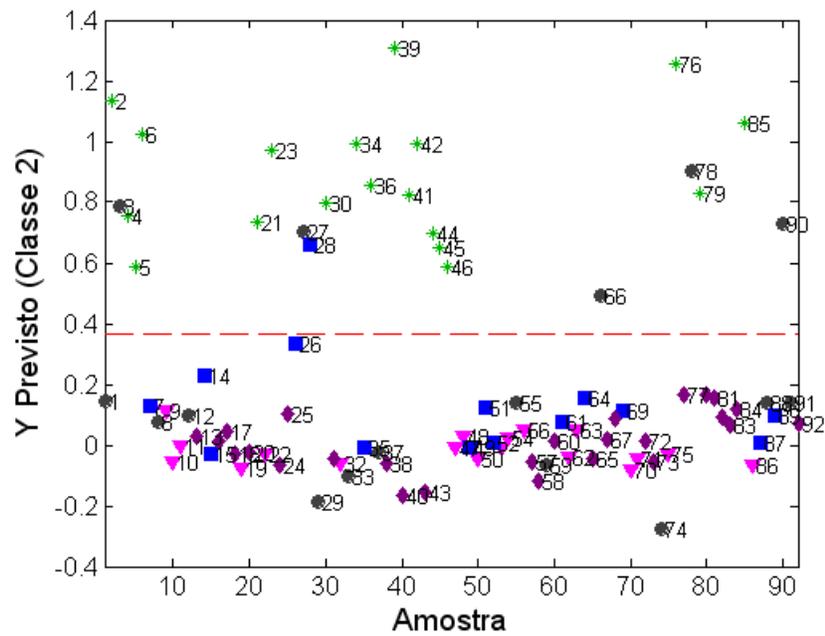


Figura 34. Classificação das amostras da classe 2 (vaporização ★) e também de amostras de validação (●).

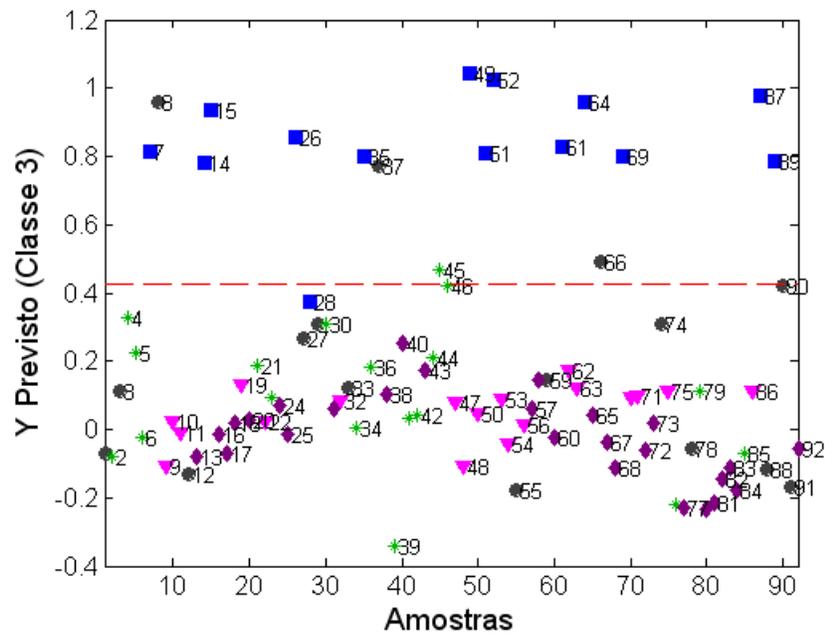


Figura 35. Classificação das amostras da classe 3 (imersão ■) e também de amostras de validação (●).

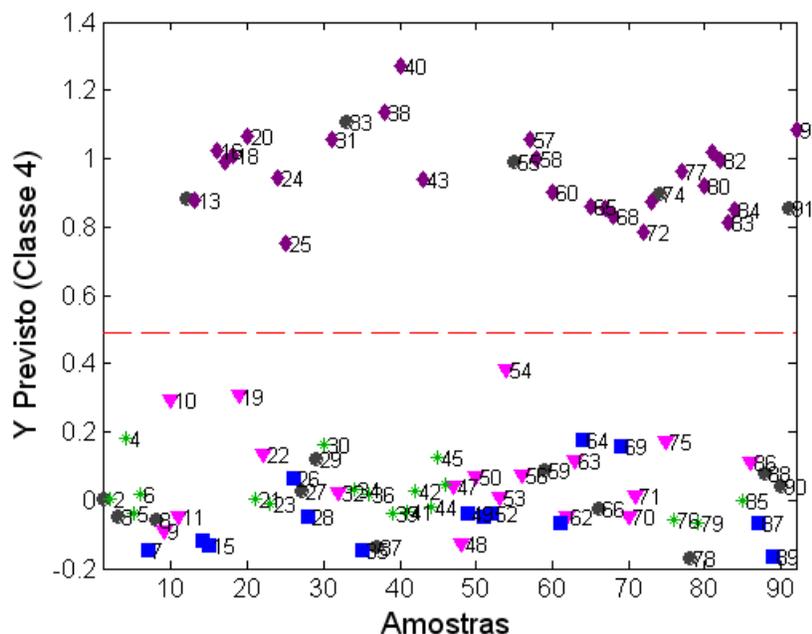


Figura 36. Classificação das amostras da classe 4 (secas naturalmente ao ar ◆) e também de amostras de validação (●).

Nos gráficos obtidos com o PLS-DA, as amostras são classificadas como pertencentes a uma determinada classe se estão localizadas acima da linha tracejada em vermelho. Esta linha é o valor limite, entre os valores previstos, que é calculado pelo algoritmo. Esse valor é estimado a partir das amostras de calibração e os valores de previsão pelo PLS.

Os gráficos apresentados nas Figuras 33, 34, 35 e 36 indicam as amostras pertencentes a cada uma das classes classificadas pelo PLS-DA. Na primeira classe (aspersão), todas as amostras de calibração foram classificadas corretamente, assim como as amostras de validação. As amostras de calibração pertencentes à segunda classe (vaporização) foram classificadas de maneira correta, exceto uma pertencente à imersão, que foi incluída como sendo da classe 2. Já as amostras de validação foram classificadas corretamente, porém com uma amostra de validação da terceira classe (imersão).

Houve uma amostra de calibração da segunda classe classificada como pertencente a terceira e as amostras de validação também foram classificadas

adequadamente. Por fim, tanto as amostras de calibração como as de validação da classe 4 (secas naturalmente ao ar), foram adequadamente classificadas.

Desta forma, o erro de classificação do modelo foi bem baixo, com erro da validação cruzada de 0% para as classes 1 e 4, 1,7% para a classe 2 e 6,0% para a classe 3. Portanto o método pode ser utilizado para a classificação das amostras com teores de umidade diferentes.

Diante dos resultados obtidos com o PCA e o PLS-DA, modelos individuais para cada classe com teor de umidade diferente foram desenvolvidos, pois como a umidade exerce grande influência no espectro dos cavacos, deve-se utilizar um modelo apropriado.

7.3 Modelos de LS-SVM para cada classe

7.3.1 Densidade básica

Após os resultados obtidos com o PCA e com o PLS-DA, construiu-se um modelo para cada classe, sendo o número de amostras por classe o seguinte: 22 amostras de aspersão, 22 amostras de vaporização, 17 amostras de imersão e 31 amostras secas naturalmente ao ar.

Um algoritmo que constrói bons modelos para dados não-lineares é a Máquina de Vetores de Suporte por Mínimos Quadrados (LS-SVM) ^{44,45,53}, e a presença de água nas amostras pode fazer com que os dados deixem de ser lineares uma vez que a quantidade de água pode variar entre duas amostras. Isto pode ocorrer devido ao procedimento experimental, onde algumas amostras de cavaco podem estar mais expostas a água e também a diferentes cavacos de um mesmo tipo de amostra (quando por exemplo, tem-se um cavaco da região central da madeira e outro da região periférica, eles também podem ter diferentes interações com a água). Desta forma, construíram-se quatro modelos (LS-SVM) com os espectros e os valores de densidade básica medidos por análise química convencional.

Também não foi possível com o LS-SVM construir um modelo com todas as amostras de diferentes graus de umidade juntos (modelo global), pois o erro foi de $RMSEP=25,90 \text{ Kg/m}^3$, o que foi considerado alto. Os modelos com o PLS para cada uma das 4 classes separadamente forneceram erros 10 vezes maiores que os obtidos com o LS-SVM, cerca de 2000 Kg/m^3 . Não foi possível com a utilização do PLS o desenvolvimento de um modelo global (utilizando as 4 classes simultaneamente), pois os erros de previsão foram considerados elevados ($RMSEP$ de $26,65 \text{ Kg/m}^3$).

Os gráficos dos valores reais *versus* previstos, assim como as superfícies obtidas nos modelos construídos com o algoritmo LS-SVM para cada classe com teor de umidade diferente estão representados nas Figuras 37, 38, 39 e 40.

Pode-se observar que ocorreu um sobreajuste dos dados de calibração para os gráficos dos valores reais *versus* os previstos. Porém, como tal fato ocorreu para todas as classes com teores de umidade diferentes e para todos os parâmetros estudados, resolveu-se apresentar os dados da validação.

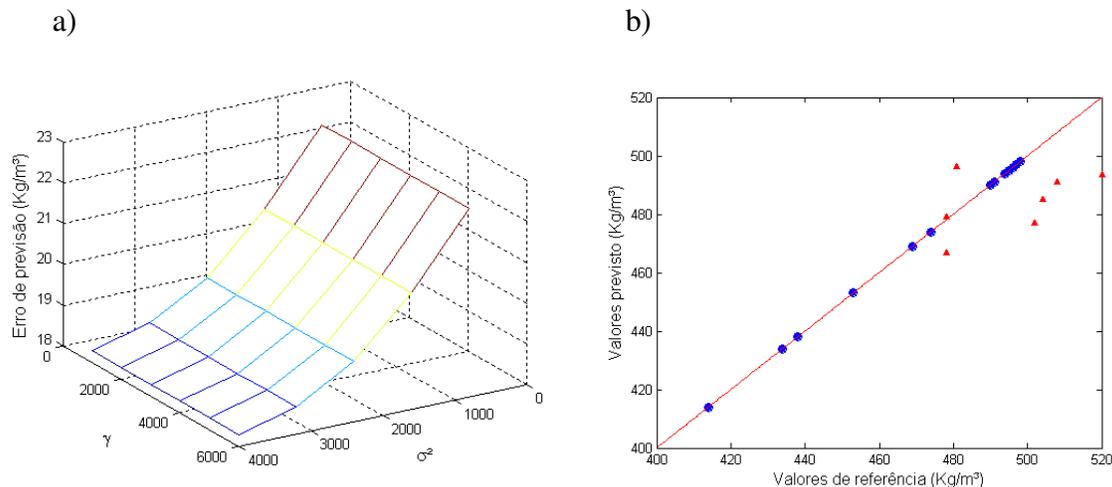


Figura 37. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de aspersão. b) Gráfico dos valores de densidade de referência *versus* previstos pelo modelo de LS-SVM para a classe de aspersão com as amostras de calibração (●) e validação (▲).

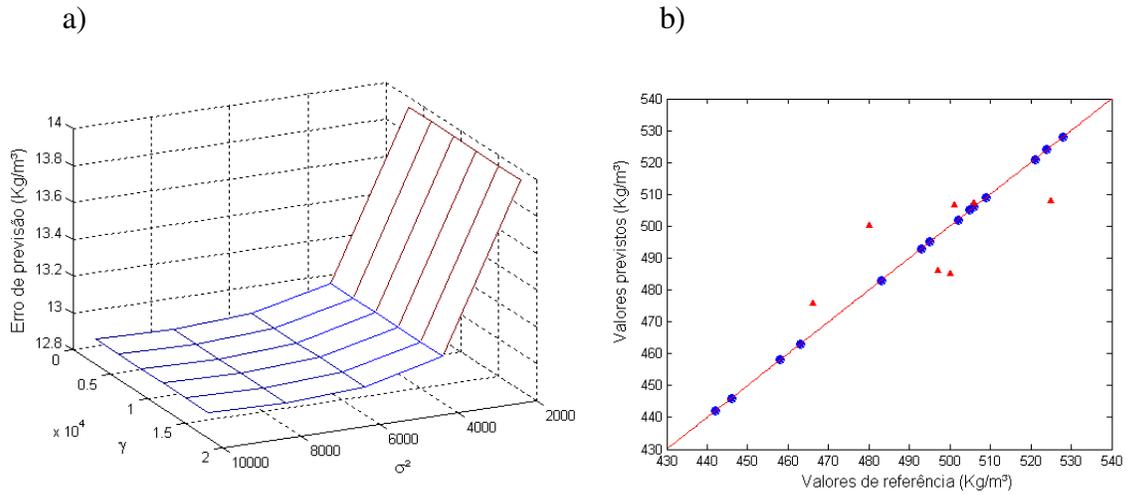


Figura 38. a) Superfície mostrando os valores dos parâmetros γ e σ^2 para amostras de vaporização. b) Gráfico dos valores de densidade de referência *versus* previstos pelo modelo de LS-SVM para a classe de vaporização para amostras de calibração (●) e validação (▲).

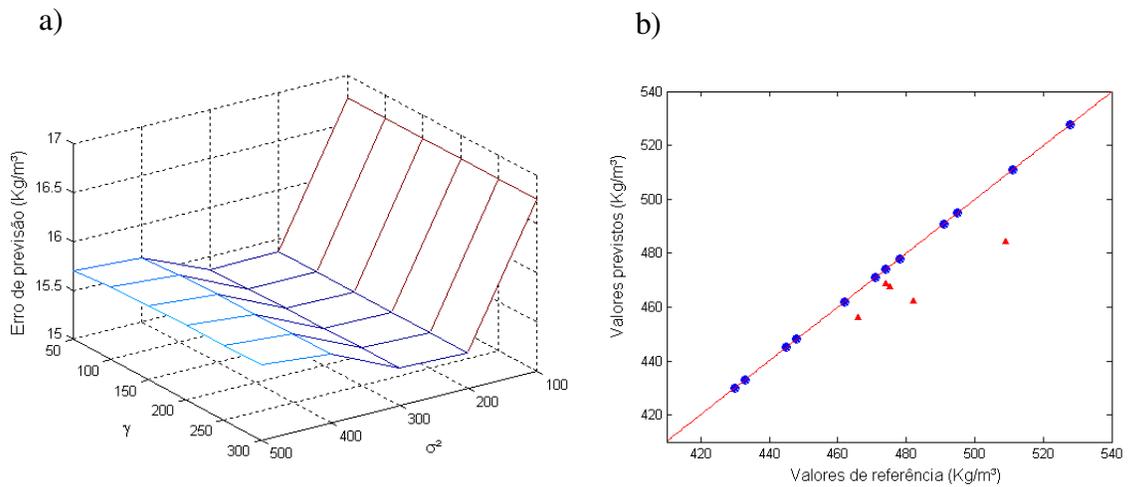


Figura 39. a) Superfície mostrando os valores para os parâmetros γ e σ^2 em amostras de imersão. b) Gráfico dos valores de densidade de referência *versus* previstos pelo modelo de LS-SVM para a classe de imersão para amostras de calibração (●) e validação (▲).

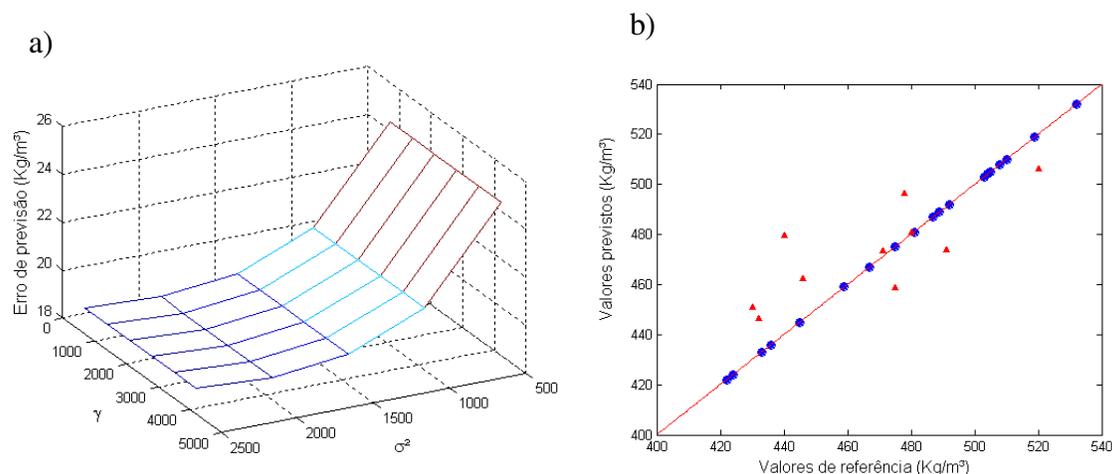


Figura 40. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras secas naturalmente ao ar. b) Gráfico dos valores de densidade de referência *versus* previstos pelo modelo de LS-SVM para a classe de amostras secas naturalmente ao ar para amostras de calibração (●) e validação (▲).

Os resultados obtidos para cada classe com teor de umidade para o parâmetro densidade básica e utilizando o algoritmo LS-SVM estão dispostos na Tabela 9, assim como os valores de γ e σ^2 utilizados para a construção do modelo.

Tabela 9. Características dos modelos construídos para cada classe com o LS-SVM.

Modelo	γ	σ^2	RMSEP (kg/m^3)	RMSEP PLS (kg/m^3)	Média Er (%) LS-SVM
Aspersão	1000	800	18,28	25,57	3,62
Imersão	50	100	15,36	26,21	3,11
Vapor	3000	2000	12,93	32,50	2,61
Secos	700	500	18,93	25,09	4,19

Os resultados apresentados na Tabela 9 mostram os resultados para cada modelo construído com cada classe de teor de umidade e os valores de RMSEP obtidos são inferiores aos obtidos com as amostras de cavacos secas. Além disto, estes valores de RMSEP são semelhantes à repetitividade do método padrão. Assim, o método pode ser utilizado como alternativa à análise química convencional.

Desta forma, pode-se utilizar o algoritmo do PLS-DA para a classificação de uma amostra desconhecida em uma das classes com teor de umidade diferente e então utilizar o modelo de LS-SVM específico para àquela classe para a previsão da densidade básica desta amostra desconhecida.

7.3.2 Lignina Total

Para as amostras de lignina total com diferentes teores de umidade também foram construídos modelos utilizando o LS-SVM. Os modelos construídos foram melhores que os globais e não foi possível com a utilização do PLS o desenvolvimento de um modelo global (utilizando as 4 classes simultaneamente), pois o RMSEP foi de 1,14 %).

Os resultados obtidos para cada uma das classes estão representados nas Figuras 41, 42, 43 e 44. Nestas figuras, podem-se observar as superfícies que foram utilizadas para a escolha dos melhores valores de γ e σ^2 (os valores que geram um menor erro de previsão) e também estão apresentados os gráficos com os valores de lignina de referência *versus* os valores previstos pelo modelo.

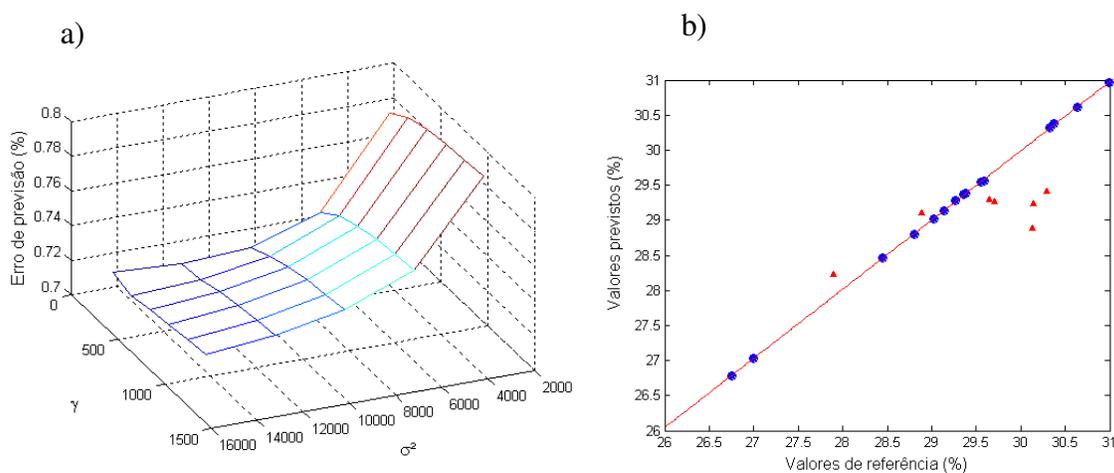


Figura 41. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de aspersão. b) Gráfico dos valores de lignina de referência *versus* previstos pelo modelo de LS-SVM para a classe de aspersão para amostras de calibração (●) e validação (▲).

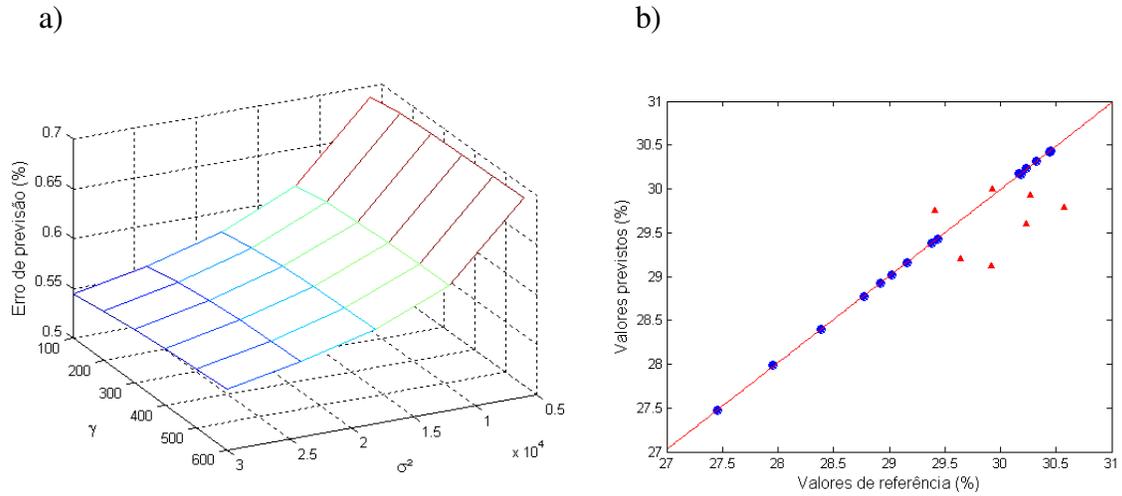


Figura 42. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de vaporização. b) Gráfico dos valores de lignina de referência *versus* previstos pelo modelo de LS-SVM para a classe de vaporização para as amostras de calibração (●) e validação (▲).

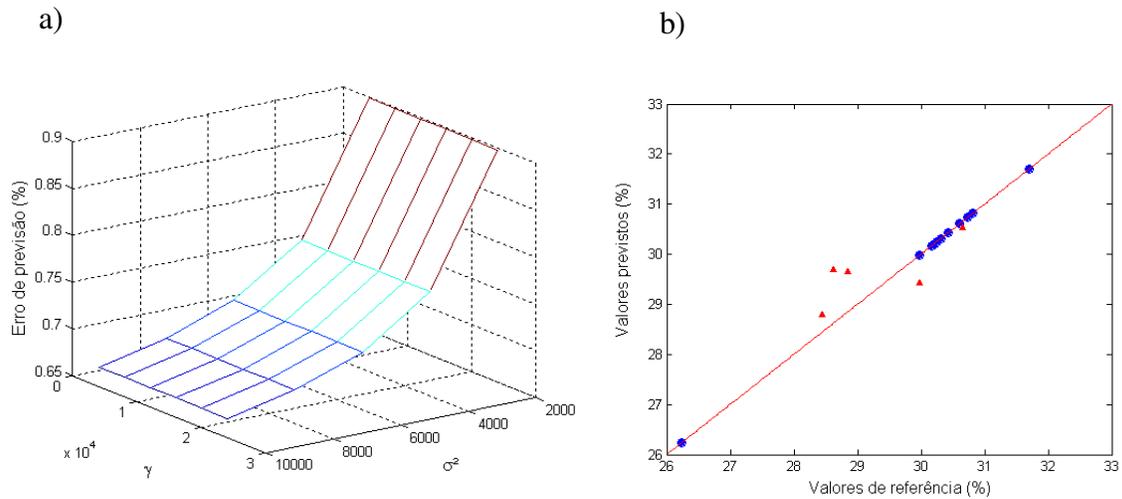


Figura 43. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de imersão. b) Gráfico dos valores de lignina de referência *versus* previstos pelo modelo de LS-SVM para a classe de imersão para as amostras de calibração (●) e validação (▲).

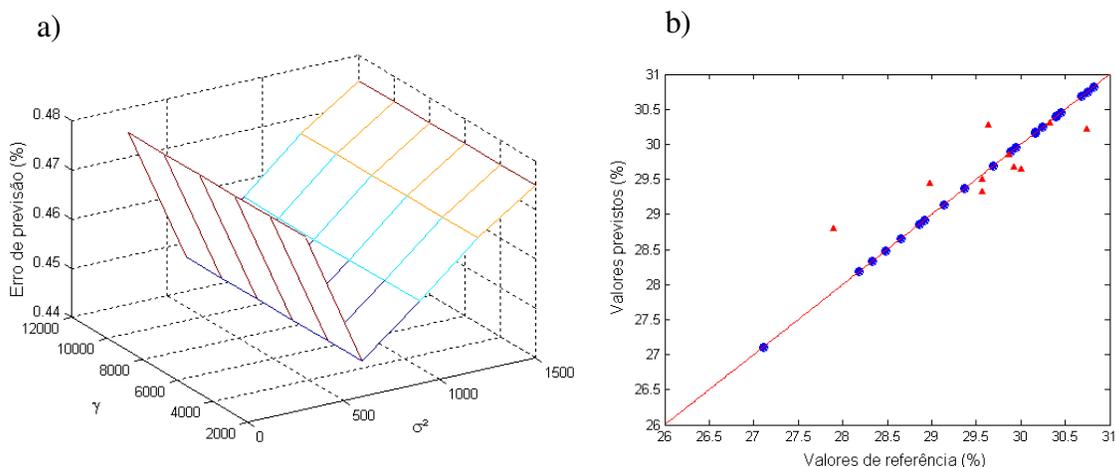


Figura 44. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras secas naturalmente ao ar. b) Gráfico dos valores de lignina de referência *versus* previstos pelo modelo de LS-SVM para a classe de amostras secas naturalmente ao ar para as amostras de calibração (●) e validação (▲).

Os resultados obtidos para cada classe com teor de umidade para o parâmetro lignina total e utilizando o algoritmo LS-SVM estão dispostos na Tabela 10, assim como os valores de γ e σ^2 utilizados para construção do modelo.

Tabela 10. Características dos modelos construídos para cada classe com o LS-SVM.

Modelo	γ	σ^2	RMSEP LS-SVM (%)	RMSEP PLS (%)	Média Er (%) LS-SVM
Aspersão	200	3000	0,72	1,17	2,39
Imersão	4000	2000	0,67	1,20	2,32
Vapor	100	6000	0,55	0,81	1,81
Secos	2000	300	0,45	0,71	1,54

Os resultados apresentados na Tabela 10 mostram os resultados para cada modelo construído com cada classe de teor de umidade e os valores de RMSEP obtidos são inferiores aos obtidos com as amostras de cavacos secas. Também foram obtidos valores de RMSEP semelhantes à repetitividade do método padrão. Sendo assim, os modelos podem ser utilizados como alternativa na determinação da lignina. Desta forma, pode-se utilizar o algoritmo do PLS-DA para a classificação de uma amostra desconhecida em uma das classes com teor de

umidade diferente e então utilizar o modelo de LS-SVM específico para àquela classe para a previsão da lignina total desta amostra desconhecida.

7.3.3 Extrativos em etanol/tolueno

Para as amostras de extrativos em etanol/tolueno com diferentes teores de umidade também foram feitos modelos utilizando o LS-SVM. Os modelos construídos foram melhores que os globais e também melhores que os modelos construídos para cada classe utilizando o PLS. Não foi possível com a utilização do PLS o desenvolvimento de um modelo global (utilizando as 4 classes simultaneamente), pois o RMSEP foi de 0,41 %).

Os resultados obtidos para cada uma das classes estão representados nas Figuras 45, 46, 47 e 48. Nestas figuras, podem-se observar as superfícies que foram utilizadas para a escolha dos melhores valores de γ e σ^2 (os valores que geram o valor mais baixo de erro de previsão) e também estão apresentados os gráficos com os valores de extrativos de referência *versus* os valores previstos pelo modelo.

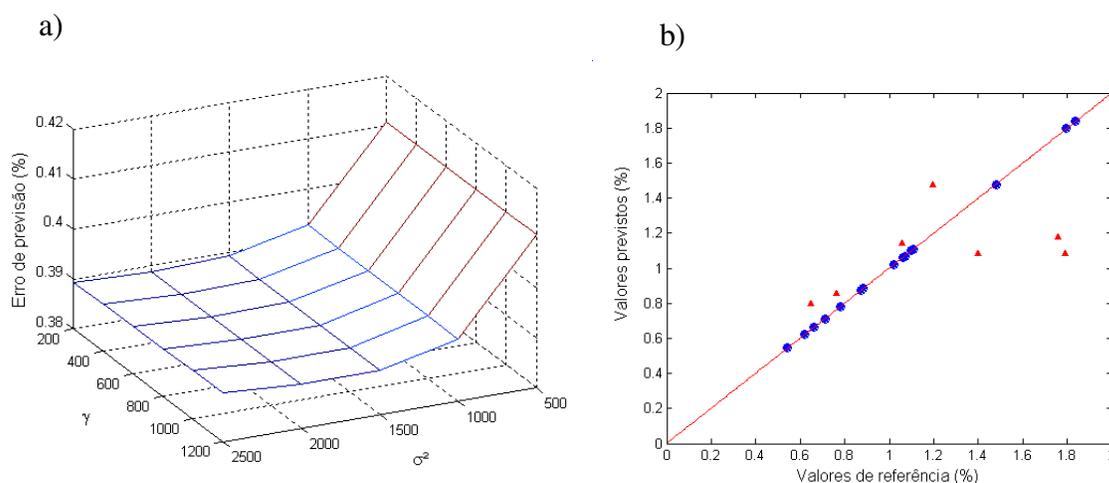


Figura 45. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de aspersão. b) Gráfico dos valores de extrativos de referência *versus* previstos pelo modelo de LS-SVM para a classe de aspersão com as amostras de calibração (●) e validação (▲).

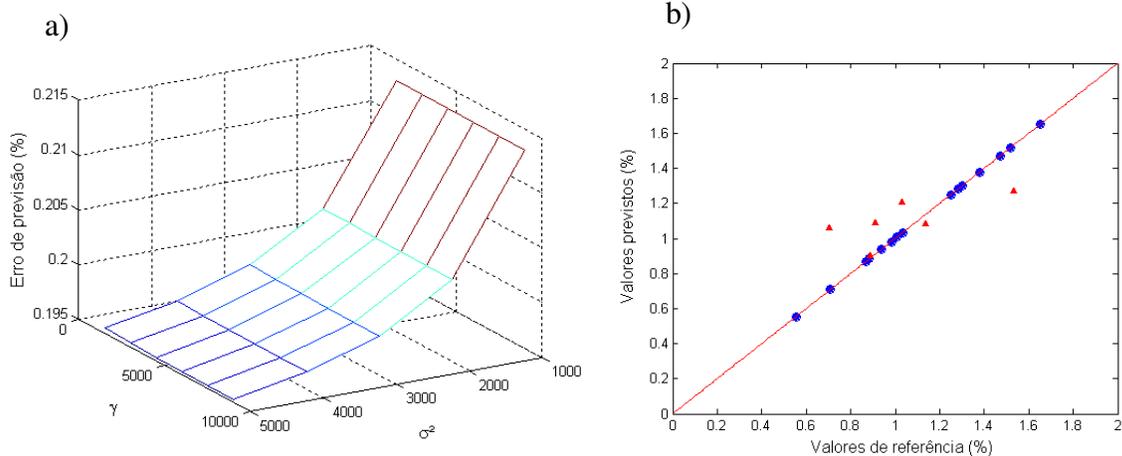


Figura 46. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de vaporização. b) Gráfico dos valores de extrativos de referência *versus* previstos pelo modelo de LS-SVM para a classe de vaporização para as amostras de calibração (●) e validação (▲).

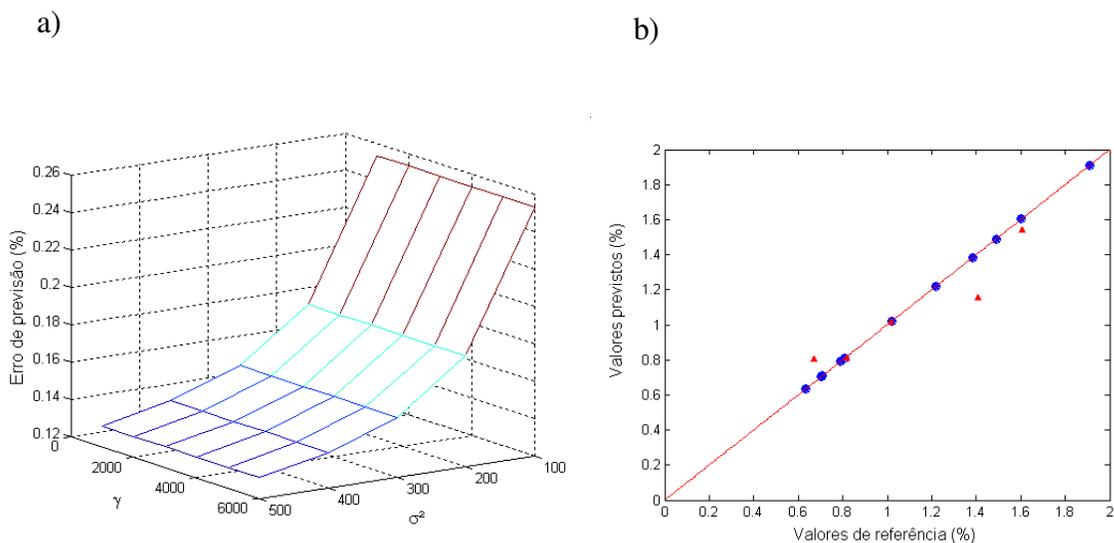


Figura 47. a) Superfície mostrando os valores para os parâmetros γ e σ^2 para amostras de imersão. b) Gráfico dos valores de extrativos de referência *versus* previstos pelo modelo de LS-SVM para a classe imersão com as amostras de calibração (●) e validação (▲).

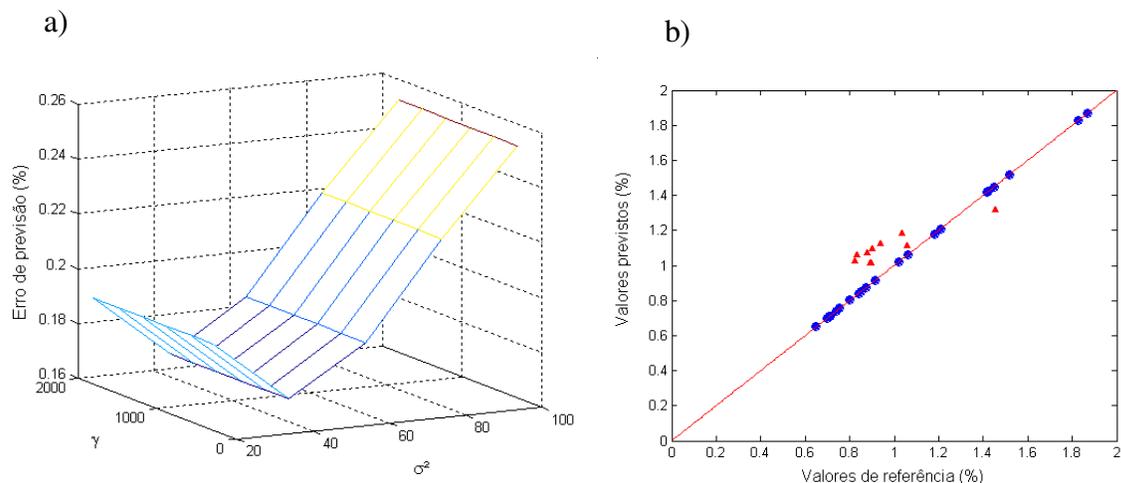


Figura 48. a) Superfície mostrando os valores obtidos para os parâmetros γ e σ^2 para amostras secas naturalmente ao ar. b) Gráfico dos valores de extrativos de referência *versus* previstos pelo modelo de LS-SVM para a classe de amostras secas naturalmente ao ar para as amostras de calibração (●) e validação (▲).

Os resultados obtidos para cada classe com teor de umidade para com parâmetro extrativos em etanol/tolueno e utilizando o algoritmo LS-SVM estão dispostos na Tabela 11, assim como os valores de γ e σ^2 utilizados para construção do modelo.

Tabela 11. Características dos modelos construídos para cada classe com o LS-SVM.

Modelo	γ	σ^2	RMSEP LS-SVM (%)	RMSEP PLS (%)	Média Er (%) LS-SVM
Aspersão	200	500	0,39	0,54	25,24
Imersão	1000	100	0,13	0,30	11,97
Vapor	1500	1000	0,20	0,35	23,21
Secos	300	20	0,17	0,30	18,68

Desta forma, novamente pode-se construir um modelo para cada teor de umidade diferente para o parâmetro extrativos em etanol/tolueno obtendo-se valores de RMSEP aceitáveis, principalmente se comparados aos obtidos com os cavacos secos. A média dos erros relativos também é aceitável, e portanto, o modelo pode ser utilizado para a determinação de extrativos após a classificação

de uma amostra desconhecida utilizando o PLS-DA. A comparação dos valores de RMSEP dos modelos construídos com o LS-SVM e o valor da repetitividade do método padrão nos mostra que são próximos e que, portanto, podem ser utilizados para a previsão dos extrativos em etanol/tolueno em uma amostra desconhecida.

CAPÍTULO 8

Conclusões

8. Conclusões

Através da realização deste trabalho, foi possível a construção de modelos de iPLS para 160 amostras de cavacos secas ao ar para cada parâmetro avaliado (densidade básica, lignina total e extrativos em etanol/tolueno) obtendo-se uma média de erros relativos compatíveis com os erros da análise química convencional.

Além disso, realizou-se um estudo sobre o efeito da umidade nos espectros dos cavacos, no qual se verificou um grande aumento na intensidade das bandas relativas à água no NIR com o aumento da umidade presente nas amostras de cavaco. Também foi possível separar os espectros dos cavacos úmidos (utilizando o PCA) em grupos de acordo com o teor de umidade apresentados, assim como classificá-los de acordo com o teor de umidade através do algoritmo PLS-DA, mesmo na presença de amostras de cavacos novas.

Por fim, foram construídos modelos com o algoritmo LS-SVM para cada uma das classes com teores de umidade diferentes e também para cada um dos parâmetros estudados (densidade, lignina e extrativos). Foi possível a obtenção de RMSEP próximos ao valor da repetibilidade do método padrão, mesmo com um número pequeno de amostras. Então estes modelos podem ser utilizados como uma alternativa ao método convencional.

Tais resultados permitirão a utilização do PLS-DA para classificação de uma amostra desconhecida e em seguida será possível utilizar um modelo de LS-SVM para cada teor de umidade diferente. Com isso, durante o processo produtivo, poderá ser feita a previsão dos parâmetros desejados.

Este trabalho permite então a obtenção de uma alternativa ao método tradicional de análise de alguns parâmetros de qualidade em cavacos de madeira, que poderia ser viabilizado com a implantação de um NIR *on-line*. Com este equipamento na correia de cavacos, os espectros poderiam ser obtidos sem interrupção do processo produtivo e os modelos construídos poderiam ser utilizados para previsão destes parâmetros, diminuindo o tempo consumido e o custo.

CAPÍTULO 9

Referências

9. Referências

1. Malkavaara, P.; Alén, R.; Kolehmainen, E. "Chemometrics: An important tool for the modern chemist, an example from wood-processing chemistry" *J. Chem. Inf. Comput. Sci.*, **40**, **2000**, 438-441.
2. Klock, U.; Muñiz, G.I.B.; Hernandez, J.; Andrade, A.S. *Química da Madeira*, 3ª Ed., UFP, **2005**.
3. D'Almeida, M.L.O. *Celulose e Papel- Tecnologia de fabricação da pasta celulósica*, IPT, 2ª Ed., Vol. 1, **1988**.
4. Silva Jr, E.C. *et al.*, "Eucalipto & meio ambiente em tempos de aquecimento global" – Aracruz Celulose SA – **2008**.
5. Foelkel, C.E.B.; Brasil, M.A.M.; Barrichelo, L.E.G. "Métodos para determinação da densidade básica de cavacos para coníferas e folhosas" – IPEF, 2/3, **1971**, 65-74.
6. Santos, C.A.S.A.; Laureano, O.F.P.; Bertolucci, R.C.P. Manual de ensaios do Laboratório de Cozimento e Branqueamento – Aracruz Celulose – Barra do Riacho E.S. MT/TEC – 005, **2002**.
7. Pereira, R.P.W.; Monteiro, M.B.O; Abreu, H.S.; Lelis, R.C.C. "Sistemas de diferenciação vascular em plantas", *Floresta e Ambiente*, **10**, **2003**, 106-115.
8. Santos, C.A.S.A.; Laureano, O.F.P.; Bertolucci, R.C.P. Manual de ensaios do Laboratório de Cozimento e Branqueamento – Aracruz Celulose – Barra do Riacho E.S. MT/TEC – 006, **2002**.
9. Santos, C.A.S.A.; Laureano, O.F.P.; Bertolucci, R.C.P. Manual de ensaios do Laboratório de Cozimento e Branqueamento – Aracruz Celulose – Barra do Riacho E.S. MT/TEC – 006 b, **2002**.
10. Avó, M.; Altmann, R. "O futuro do setor de papel e celulose a partir de uma ótica brasileira", *O Papel*, **2009**, 37-41.
11. Santos, C.P.; Reis, I.N.; Moreira, J.E.B.; Brasileiro, L.B. "Papel: como se fabrica?" *Quim. Nova Escola*, **14**, **2001**, 3-7.
12. Burns, D.A.; Ciurczak, E.W. *Handbook of Near-Infrared Analysis*. New York: Marcel Dekker, **2001**.
13. Pasquini, C. "Near Infrared Spectroscopy: Fundamentals, practical aspects and analytical applications". *J. Braz. Chem. Soc.*, **14**, **2003**, 198-219.
14. Skoog, D.A.; Holler, F.J.; Nieman, T.A. *Princípios de análise instrumental*. 5ª Ed. São Paulo: Bookman, **2002**.
15. Coates, J. "A review of current new technology: Used in instrumentation for industrial vibrational spectroscopy", *Spectroscopy*, **14**, **1999**, 21-34.
16. Rocha, W.F.C.; Sabin, G.P.; Março, P.H.; Poppi, R.J. "Quantitative analysis of piroxican polymorphs pharmaceutical mixtures by hyperspectral imaging and chemometrics". *Chem. Intell. Lab. Syst.*, **102**, **2010**.
17. Sikora, Z.; Salacki, W. "Use of near-infrared (NIR) spectroscopy to predict several physical and operating properties of oil fractions and diesel fuels", *Petroleum and Coal*, **38**, **1996**, 65-68.
18. Workman Jr, J.J. "Interpretative spectroscopy for near infrared", *Appl. Spectrosc. Review*, **31**, **1996**, 251-320.

19. Bokobza, L. "Near infrared spectroscopy", *J. Near Infrared Spectrosc.*, **6**, **1998**, 3-17
20. Schultz, T.P.; Burns, D. A; "Rapid secondary analysis of lignocellulose – Comparison of near-infrared (NIR) and Fourier-transform infrared (FTIR)", *TAPPI Journal*, **73**, **1990**, 209-212.
21. Easty, D.B.; Berben, S.A.; Dethomas, F.A.; Brimmer, P.J. "Near-infrared spectroscopy for the analysis of wood pulp – Quantifying hardwood softwood mixtures and estimating lignin content". *TAPPI Journal*, **73**, **1990**, 257-261.
22. Henriksen, H.C.; Naes, T.; Rødbotten, R.; Aastveit, A. "Simultaneous modeling of process variables and raw material properties as measured by NIR. A case study from cellulose production". *Chem. Intell. Lab. Syst.*, **77**, **2005**, 238-246.
23. Thygesen, L.G.; Lundqvist, S-O.; "NIR measurement of moisture content in wood under unstable temperature conditions. Part 2. Handling temperature fluctuations". *J. Near Infrared Spectrosc.*, **8**, **2000**, 191-199.
24. Poke, F.S.; Wright, J.K.; Raymond, C.A. "Predicting extractives and lignin contents in *Eucalyptus globulus* using near infrared reflectance analysis". *J. Wood Chem. Technol.*, **24**, **2004**, 55-67.
25. Magalhães, W.L.E.; Zangiski, F.; Klein, C.H.; Higa, A.R. "Uso da espectroscopia no infravermelho próximo (NIR) para predição não-destrutiva de densidade básica da madeira de *Pinus taeda*". *Comunicado Técnico Embrapa*, **159**, **2006**, 1-4.
26. Caldeira, A.F.; Rocha, A.P.; Santos, C.V.C.; Almeida, C.; Patelli, J.E.; Calvosa, P.S.P.; Sacon, V. "NIR *on-line*: an innovation in the VCP Wood yard" *Relatório interno VCP*.
27. Tsuchikawa, S. "A review of recent near infrared research for wood and paper". *Appl. Spectrosc. Reviews*, **42**, **2007**, 43-71.
28. Otto, M. *Chemometrics*. Weinheim:Wiley, **1999**.
29. Brereton, R.G. "Introduction to multivariate calibration in analytical chemistry", *Analyst*, **125**, **2000**, 2125-2154.
30. Martens, H.; Naes, T. *Multivariate calibration*. New York: Wiley, **1996**.
31. Sekulic, S.; Seasholtz, M.B.; Wang, Z.; Kowalski, B. R. "Nonlinear multivariate calibration methods in analytical chemistry", *Anal. Chem.*, **65**, **1993**, A835-A845.
32. Charne, R.; De Luna Freire, C. A; Charnet, E.M.R.; Bovino, H. "*Análise de modelos de regressão linear com aplicações*". Campinas: Unicamp, **1999**.
33. Miller, J.N.; Miller, J.C. "*Statistics and Chemometrics for Analytical Chemistry*", London: Prentice Hall, **2000**.
34. Chui, Q.S.H.; Zucchini, R.R.; Lichtig, J. "Qualidade de medições em química analítica. Estudo de caso: determinação de cádmio por espectrofotometria de absorção atômica com chama", *Quim. Nova*, **24**, **2001**, 374-380.
35. Simas, R. C. "*Determinação de proteína bruta e aminoácidos em farelo de soja por Espectroscopia no Infravermelho Próximo*". , UNICAMP, **2005**. Dissertação de mestrado.

36. Geladi, P.; Kowalski, B. R. "Partial least square regression: a tutorial", *Anal. Chim. Acta*, **185**, **1986**, 1-17.
37. Valderrama, P. "Avaliação de figuras de mérito em calibração multivariada na determinação de parâmetros de controle de qualidade em indústria alcooleira por Espectroscopia no Infravermelho Próximo", UNICAMP, **2005**. Dissertação de mestrado.
38. Oliveira, F.C.; Souza, A.T.P.C.; Dias, J.A.; Dias, S.C.L.; Rubim, J.C. "A escolha da faixa espectral no uso combinado de métodos espectroscópicos e quimiométricos", *Quim. Nova*, **27**, **2004**, 218-225.
39. Osborne, S.D.; Jordan, R.B.; Künnemeyer, R. "Method of wavelength selection for partial least square", *Analyst*, **122**, **1997**, 1531-1537.
40. Costa Filho, P.A.; Poppi, R.J. "Algoritmo genético em química", *Quim. Nova*, **22**, **1999**, 405-411.
41. Centner, V.; Massart, D. "Elimination of uninformative variables for multivariate calibration", *Anal. Chem.*, **68**, **1996**, 3851-3858.
42. Norgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J.P.; Munck, L.; Engelsen, S.B. "Interval partial least-square regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy", *Appl. Spectrosc.*, **53**, **2000**, 413-419.
43. Barker, M.; Rayens, W. "Partial least squares for discrimination" *J. Chemom.*, **17**, **2003**, 166-173.
44. Cogdill, R.P.; Dardenne, P. "Least-squares support vector machines for chemometrics: an introduction and evaluation" *J. Near Infrared Spectrosc.*, **12**, **2004**, 93-100.
45. Buydens, M.C. *et al.* "Multivariate calibration with Least-Squares Support Vector Machines", *Anal. Chem.*, **76**, **2004**, 3099-3105.
46. Wang, Y.D.; Veltkamp, D.J.; Kowalski, B.R. "Multivariate instrument standardization", *Anal. Chem.*, **63**, **1991**, 2750-2756.
47. Filho, H.A.D.; Galvão, R.K.H.; Araújo, M.C.U.; Silva, E.C.; Saldanha, T.C.B.; José, G.E.; Pasquini, C.; Raimundo Jr, I.M.; Rohwedder, J.J.R. "A strategy for selecting calibration samples for multivariate modelling", *Chemom. Intell. Lab. Syst.*, **72**, **2004**, 83-91.
48. Kennard, R.W.; Stone, L.A. "Computer aided design of experiments", *Technometrics*, **11**, **1969**, 137-148.
49. Vandeginste, B.M.G.; Massart, D.L.; Buydens, L.M.C.; Jing, S.; Lewi, P.J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics: Part B*. Amsterdam:Elsevier, **1998**.
50. Maesschalck, R.D.; Jouan-Rimbaud, D.; Massart, D.L. "The mahalanobis distance", *Chemom. Intell. Lab. Syst.*, **63**, **2002**, 27-39.
51. Fernández Pierna, J.A.; Wahl, F.; Noord, O.E.; Massart, D.L. "Methods for outlier detection in prediction" *Chemom. Intell. Lab. Syst.*, **63**, **2002**, 27-39.
52. Annual Book of ASTM Standards. *Standards practices for infrared, multivariate, quantitative analysis*, E1655, vol 03.06. ASTM International, West Conshohocken, Pennsylvania, USA, **2000**.
53. Thygesen, L.G.; Lundqvist, S-O. "NIR measurement of moisture content in wood under unstable temperature conditions. Part 2. Handling temperature fluctuations". *J. Near Infrared Spectrosc.*, **8**, **2000**, 191-199.