

MAGALE KARINE DIEL RAMBO

**CARACTERIZAÇÃO DE RESÍDUOS LIGNOCELULÓSICOS POR ESPECTROSCOPIA
NIR ALIADA À QUIMIOMETRIA PARA A
OBTENÇÃO DE INSUMOS QUÍMICOS**

CAMPINAS

2013

i



UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE QUÍMICA

MAGALE KARINE DIEL RAMBO

**CARACTERIZAÇÃO DE RESÍDUOS LIGNOCELULÓSICOS POR
ESPECTROSCOPIA NIR ALIADA À QUIMIOMETRIA PARA A
OBTENÇÃO DE INSUMOS QUÍMICOS**

ORIENTADORA: PROFA. DRA. MÁRCIA MIGUEL CASTRO FERREIRA

**TESE DE DOUTORADO APRESENTADA AO INSTITUTO DE
QUÍMICA DA UNICAMP PARA OBTENÇÃO DO TÍTULO DE
DOUTORA EM CIÊNCIAS**

**ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE
DEFENDIDA POR MAGALE KARINE DIEL RAMBO, E ORIENTADA PELA
PROFA. DRA. MÁRCIA MIGUEL CASTRO FERREIRA**

Assinatura da Orientadora

**Campinas
2013**

FICHA CATALOGRÁFICA ELABORADA POR DANIELLE DANTAS DE SOUSA - BIBLIOTECA DO INSTITUTO DE QUÍMICA DA UNICAMP

Rambo, Magale Karine Diel (1986-).
Caracterização de resíduos lignocelulósicos por espectroscopia NIR aliada à quimiometria para a obtenção de insumos químicos. -Campinas, SP: [s.n.], 2013.

Orientadora: Márcia Miguel Castro Ferreira
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Química.

1. Biomassa residual. 2. Espectroscopia no infravermelho próximo. 3. Calibração multivariada. 4. Figuras de mérito. I. Ferreira, Márcia Miguel Castro Ferreira. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.

Informações para Biblioteca Digital

Título em inglês: Characterization of lignocellulosic residues by NIR spectroscopy coupled with chemometrics to obtain chemical products

Palavras – chaves em inglês:

Biomass. Near infrared spectroscopy. Multivariate calibration. Figures of merit.

Área de concentração: Química Analítica

Titulação: Doutora em Ciências

Banca examinadora:

Márcia Miguel Castro Ferreira [Orientadora]

Reinaldo Francisco Teófilo

André Luis Ferraz

Matthieu Tubino

Jarbas José Rodrigues Rohwedder

Data de defesa: 06/08/2013

Programa de pós-graduação: Química

Dedicatória

Dedico esta Tese aos meus queridos

pais, Nelson e Iria,

e à minha irmã

Michele.

AGRADECIMENTOS

À Deus que, de várias formas, abre caminhos para novas descobertas.

À Prof. Márcia pela capacidade de nos estimular a enfrentar novos desafios dia após dia, proporcionando nosso crescimento profissional e intelectual; obrigado pela oportunidade e pelos incentivos.

Aos demais professores que, direta ou indiretamente, contribuíram para o desenvolvimento deste trabalho.

Aos meus pais, Nelson e Iria, e à minha irmã Michele, pelo estímulo e apoio incondicional e por fazerem da nossa família um verdadeiro lar, transmitindo toda experiência, incentivo e amor.

Ao meu namorado Henrique, pelo incentivo, amor e imensa ajuda e colaboração no trabalho. Obrigada meu amor, por sempre me direcionar para o caminho correto.

Aos meus amigos de todas as horas, Flamys, Eva, Angélica e Aparecido. Sem palavras para agradecer todo o apoio, principalmente as gurias Angélica e Flamys, que não mediram esforços em me ajudar. Vocês são verdadeiras irmãs para mim.

Aos amigos e colegas do LQTA: Márcia, Maíra, Samuel, Clécio, Euzébio, Celso Júnior, pelo convívio, incentivo e apoio durante a realização desta dissertação. Um agradecimento especial ao Samuel, pela imensa ajuda em todos os momentos.

Ao Centro de Tecnologia Canavieira (CTC), principalmente ao Wokimar, Paulino, Paulo Gouveia e Marcelo pela colaboração direta na execução deste trabalho. Um agradecimento muito especial ao meu sempre amigo Cidinho.

À EMBRAPA e IAC pelas amostras cedidas. À Fazenda Monte Ato, principalmente a Maria Helena, do Café Helena pelas amostras cedidas. Ao Laboratório de Frutas e Hortaliças, da Faculdade de Engenharia de Alimentos da UNICAMP, principalmente ao Professor Flávio Schmidt e a Ana pelo espaço cedido.

Ao DIBANET e ao Sétimo Programa da União Européia [FP7/2007-2013] sob subvenção n ° [227248] pelo financiamento do projeto. Ao Instituto de Química (IQ) da UNICAMP, e aos funcionários Bel, Miguel e Gabriela da secretaria do IQ pela atenção.

Curriculum Vitae

Magale Karine Diel Rambo

Formação Acadêmica/Titulação

2013- Professora da Universidade Federal do Tocantins- UFT

2010-2013 Doutorado em Química-UNICAMP

Universidade Estadual de Campinas, UNICAMP, Campinas, Brasil.

Título: Caracterização de resíduos lignocelulósicos por espectroscopia NIR aliada a quimiometria para a obtenção de insumos químicos.

Orientadora: Prof. Dr. Márcia Miguel Castro Ferreira.

2008-2010 Mestrado em Química-UFSM

Universidade Federal de Santa Maria, Santa Maria, Brasil.

Título: Utilização de cascas de arroz para a produção de xilitol e sílicas xerogéis

Orientador: Prof. Dr. Ayrton Figueiredo Martins.

2004-2008 Graduação em Química-UFSM

Química Licenciatura

Universidade Federal de Santa Maria, Santa Maria, Brasil.

Artigos completos publicados e submetidos em periódicos

1.**RAMBO, M. K.D.**, CARDOSO, A. L., BEVILAQUA, D. B., RIZZETTI, T. M., RAMOS, L. A., KORNDORFER, G. H., MARTINS, A. F. Silica from Rice Husk Ash as an Additive for Rice Plant. **Journal of Agronomy**, v.10, p.99 – 104, 2011.

2.**RAMBO, M. K. D.**, BEVILAQUA, D. B., BRENNER, C., ALVES, S.H, MARIO, D., MALLMANN, C., MARTINS, A.F. Xylitol from rice husks by acid hydrolysis and *Candida tropicalis* and *guilhermondii* yeast fermentation. **Química Nova**, v.36, p.1-6, 2013.

3.BEVILAQUA, D. B., **RAMBO, M. K. D.**, CARDOSO, A. L., RIZZETTITI, T.M, MARTINS, A. F. Cleaner Production: Levulinic Acid from Rice Husks. **Journal of Cleaner Production**, v.47, p. 96-101, 2013.

4.**RAMBO, M. K. D.**, AMORIM, E. P., FERREIRA, M. M. C. Potential of visible-near infrared spectroscopy combined with chemometrics for analysis of some constituents of

coffee and banana residues. **Analytica Chimica Acta**, v.775, p. 41-49, 2013.

5. **RAMBO, M. K. D.**, SCHMIDT, F., FERREIRA, M. M. C. Characterization of Brazilian biomasses for biorefineries opportunities. **Biomass Bioenergy**. Submetido.

6. **RAMBO, M. K. D.**, ALVES, A.R., TEIXEIRA, W.G., FERREIRA, M. M. C. Multivariate analysis of Coconut samples by Near infrared spectroscopy. Para submissão.

Principais Trabalhos apresentados em congressos

1. **RAMBO, M. K. D.**, FERREIRA, M. M. C. The new generation of feedstocks. Apresentado em DIBANET-Summer School, 2010, UFRJ, Rio de Janeiro.

2. **RAMBO, M. K. D.**, SGUASSABIA, E. C. J, SCHMIDT, F. L., FERREIRA, M. M. C. Análise da composição da casca de arroz e casca de soja para otimização da produção de insumos químicos. Apresentado em: Simpósio Latino americano de Ciência de Alimentos, 2011, Campinas, São Paulo. 9 Slaca.

3. **RAMBO, M. K. D.**, FERREIRA, M. M. C., SGUASSABIA, E. C.J, SCHMIDT, F. L. Determinação de carboidratos em hidrolisados de caroço de açaí e engaço de banana por cromatografia de troca iônica com detecção amperométrica. Apresentado em: Simpósio Latino americano de Ciência de Alimentos, 2011, Campinas. 9 Slaca.

4. **RAMBO, M. K. D.**, FERREIRA, M. M. C. Latin America Feedstocks Composition. Apresentado em: DIBANET, Networking Day, 2012, Tessaloniki, Grécia.

5. **RAMBO, M. K. D.**, FERREIRA, M. M. C. Prediction of sugar content of coconut samples using Near Infrared Spectroscopy coupled to multivariate data analysis. Apresentado em: DIBANET, Networking Day, 2012, Tessaloniki, Grécia.

6. **RAMBO, M. K. D.**, FERREIRA, M. M. C. Prediction of the extractives, moisture and ashes contents of banana using near infrared spectroscopy coupled to multivariate analysis. Apresentado em: XIII Chemometrics in Analytical Chemistry, 2012, Budapeste, Hungria. CAC.

Apresentação de trabalhos orais e palestras

1. **RAMBO, M. K. D.**, FERREIRA, M. M. C. Latin American Feedstocks. Apresentado em DIBANET-Summer School, 2010, UFRJ, Rio de Janeiro.

2. **RAMBO, M. K. D.**, FERREIRA, M. M. C. The New Generation of Biofuels. Apresentado em DIBANET, 2010, University of Buenos Aires, Buenos Aires, Argentina.

3. **RAMBO, M. K. D.**, FERREIRA, M. M. C. Feedstock Database. Apresentado em DIBANET Meeting, 2012, Fundacion Chile, Santiago, Chile.

RESUMO

Rambo, M.K.D, 2013. Caracterização de resíduos lignocelulósicos por espectroscopia NIR aliada à quimiometria para a obtenção de insumos químicos. Campinas: Tese de Doutorado – Instituto de Química- UNICAMP, 178p.

A geração de materiais residuais e o seu descarte no ambiente têm sido motivo de grande preocupação, tanto ambiental quanto econômica. O aproveitamento dessas matérias-primas residuais renováveis, não apenas para a produção de energia e calor, mas também para a geração de produtos tecnológicos, tornou-se uma prioridade. Neste trabalho, utilizaram-se caracterizações físico-químicas para o aproveitamento de resíduos agroindustriais e agroflorestais como caroço de açaí; serragem de eucalipto; cascas de arroz, café e soja; capim elefante; bambu; resíduos de banana e coco. Tais biomassas foram submetidas à caracterizações físico-químicas a fim de avaliá-las como potenciais matérias-primas no conceito de biorrefinarias. Essas foram caracterizadas e analisadas por meio de técnicas de difração de raios-X (XRD), espectrometria no infravermelho próximo (NIR), extração acelerada por solvente (ASE), cromatografia de troca iônica (HPAEC), análise térmica (TGA), poder calorífico, análise no ultravioleta-visível (UV-Vis), análise elementar e análise imediata. Com base nesses resultados, os resíduos de banana, café e coco foram selecionados como biomassas promissoras no contexto de biorrefinarias. Vários estágios de processamento dessas amostras, úmidas - não moídas (WU), secas – não moídas (DU) e secas moídas (DS) foram analisadas. Usando a espectroscopia NIR e análise multivariada foi possível construir modelos de regressão para prever os teores de lignina Klason, lignina ácida solúvel, resíduo ácido insolúvel, lignina total, extrativos, umidade, cinzas, glicose, xilose, manose, arabinose, ramnose, galactose, bem como o teor de açúcares totais de 104 amostras de resíduos de banana, 102 resíduos de café e 30 resíduos de coco oriundos de diferentes regiões do Brasil. Modelos utilizando o método de regressão por quadrados mínimos parciais (PLS) foram construídos para amostras de banana, café, coco e amostras combinadas de (Banana + Café) e de (Banana + Café + Coco). Os coeficientes de regressão multivariados de cada modelo foram interpretados a fim de garantir a obtenção de modelos confiáveis e não obtidos ao acaso. A precisão da metodologia NIR foi melhor ($p < 0,05$) que o método de referência para quase todos os parâmetros, sendo pior apenas para a umidade. Com exceção do teor de cinzas, lignina ácida solúvel, e açúcares (principalmente os minoritários) que foram preditos pobremente ($R^2_{\text{val}} < 0,80$) em alguns modelos de diferentes frações, os modelos para os demais analitos exibiram $R^2_{\text{val}} > 0,80$. Normalmente as frações DU e WU forneceram modelos um pouco inferiores, necessitando um maior número de variáveis latentes, provavelmente devido à heterogeneidade dessas amostras. Os erros relativos variaram de 2 a 25%, e os desvios padrões entre os valores de referência e os previstos foram todos menores que 10%. Os modelos foram validados utilizando figuras de mérito de calibração multivariada. Com base nos resultados da validação externa e figuras de mérito, a espectroscopia NIR mostrou-se útil para a predição química dos resíduos de banana, coco e café, podendo ser usada como uma alternativa rápida, econômica e com menor produção de resíduos químicos frente às metodologias padrões.

ABSTRACT

Rambo, M.K.D, 2013. Characterization of lignocellulosic residues by NIR spectroscopy coupled with chemometrics to obtain chemical products. Campinas: PhD Thesis–Institute of Chemistry - UNICAMP, 178p.

The generation of waste disposed in the environment has been of great concern, both environmentally and economically. The use of the wastes from renewable feedstocks, not only for the production of power and heat, but also for the generation of technology products, has become a priority. In this work, we investigated the physicochemical characteristics for the utilization of agro-industrial and agroforestry residues such as acai seed; eucalyptus sadwust; rice husks, coffee husks, soybean peel, elephant grass; bamboo; banana and coconut wastes. These biomasses were subjected to physicochemical characterizations in order to evaluate their as potential as raw materials in the biorefinery concept. They were characterized and analyzed using techniques of X-ray diffraction (XRD), Near-infrared spectroscopy (NIR), Accelerated Solvent Extraction (ASE), Anion-Exchange chromatography (HPAEC-PAD), Thermogravimetric analysis (TGA), Heating Value, ultraviolet-visible spectroscopy analysis (UV-Vis), proximate and elemental analysis. Based on the results, banana, coffee and coconut residues were selected for biorefineries opportunities. Several stages of processing of the samples, wet and unground (WU), dried and unground (DU), dried and sieving (DS) were analyzed. Using NIR spectroscopy, coupled with chemometric tools was possible to build regression models to predict the Klason lignin (KL), acid soluble lignin (ASL), acid insoluble residue (AIR), total lignin (TL), extractives, moisture, ash, glucose, xylose, manose, arabinose, rhamnose, galactose and the total sugar content, of 104 samples of banana residues, 102 samples of coffee residues and 30 samples of coconut residues, all from different regions of Brazil. Models using the Partial least squares regression (PLS) method were built for Banana, Coffee and Coconut samples and for pooled samples (Banana + Coffee) and (Banana + Coffee + Coconut). Regression Coefficients Interpretation of the model was made to help in the identification of spectral signals characteristics to the analyte and also to detect potential spurious correlations. The precision of the NIR methodology was better ($p < 0.05$) than the reference method for almost all the parameters, being worse for moisture. With the exception of ash, acid soluble lignin and sugars (especially the minority) content, which was predicted poorly ($R^2_{val} < 0.80$) in some biomass models from different fractions, the models for all the analytes exhibited $R^2_{val} > 0.80$. Normally WU and DU fractions provided somewhat worse models, requiring a greater number of latent variables (LV), probably due to the heterogeneity of the samples. The relative errors varied from 2.00 to 25.00%, and the standard deviations between the reference and predicted values were less than 10.00%. The quality of the models were evaluated by the determination of the multivariate figures of merit. Based on the results of external validation and figures of merit, NIR spectroscopy proved to be useful for chemical prediction of banana, coconut and coffee residues and can be used as a faster and more economical alternative to the standard methodologies.

Sumário

LISTA DE SIGLAS	xix
LISTA DE TABELAS	xxi
LISTA DE FIGURAS	xxiii
1 INTRODUÇÃO E JUSTIFICATIVA	1
1.1 Objetivos	5
2 REVISÃO BIBLIOGRÁFICA E CARACTERIZAÇÃO DO PROBLEMA	7
2.1 Biomassas	8
2.1.1 Cascas de arroz	8
2.1.2 Bambu	9
2.1.3 Serragem de eucalipto	9
2.1.4 Capim elefante	10
2.1.5 Carochos de açaí	11
2.1.6 Cascas de café	11
2.1.7 Caule e engaço da banananeira	12
2.1.8 Fibras de coco	13
2.1.9 Cascas de soja	13
2.2 Composição química das biomassas lignocelulósicas	14
2.2.1 Carboidratos	14
2.2.2 Lignina	20
2.2.3 Extrativos	22
2.3 Hidrólise da biomassa	22
2.4 Espectroscopia no Infravermelho Próximo (NIR)	26
2.5 Quimiometria	30
2.5.1 Análise por Componentes Principais (PCA)	31
2.5.2 Modelos de calibração	33
2.5.3 Regressão por quadrados mínimos parciais (PLS)	34
2.5.4 Escolha do número de variáveis latentes	36
2.5.5 Pré-tratamento dos dados espectrais	39
2.6 FIGURAS DE MÉRITO	45
3 PARTE EXPERIMENTAL	51
3.1 Seleção e coleta de amostras	51
3.2 Análises	55
3.2.1 Espectroscopia no Visível e Infravermelho-próximo (Vis-NIR)	56
3.2.2 Análise química aproximada	56
3.2.3 Análise elementar	57
3.2.4 Análise térmica	57
3.2.5 Difração de Raios-X	58
3.2.6 Poder Calorífico Superior (PCS)	59
3.2.7 Extração	60
3.2.8 Hidrólise ácida	61
3.2.9 Cromatografia	61
3.2.10 Espectroscopia no ultravioleta visível (UV-Vis)	62
3.2.11 Análise dos dados	63
3.2.12 Qualidade das análises e dos modelos	65

4 RESULTADOS E DISCUSSÕES.....	67
4.1 Escolha e caracterização das biomassas	67
4.1.1 Poder calorífico.....	68
4.1.2 Cinzas.....	71
4.1.3 Umidade	72
4.1.4 Análise elementar	72
4.1.5 Extrativos	73
4.1.6 Lignina	73
4.1.7 Carboidratos	74
4.1.8 Análise Térmica	78
4.1.9 Hidrólise ácida e Bioprodutos	82
4.1.10 Difração de raio-X.....	84
4.1.11 Espectroscopia Vis-NIR	88
4.1.12 Conclusões a respeito da seleção das biomassas	90
4.2 Análises de referência para construção de modelos de calibração multivariada ..	91
4.3 Análise por Componentes Principais (PCA)	92
4.4 Modelos de regressão multivariada.....	96
4.5 Interpretação dos Coeficientes de Regressão	128
4.6 Figuras de Mérito em calibração multivariada	134
5 CONCLUSÕES.....	145
6 REFERÊNCIAS BIBLIOGRÁFICAS.....	147

LISTA DE SIGLAS

1D – Primeira derivada

2D – Segunda derivada

AACC- Associação Americana dos Químicos Cerealistas (do inglês, American Association of Cereal Chemists)

AF – Ácido fórmico

AIA – Cinza ácida insolúvel (do inglês, Acid Insoluble Ash)

AIR – Resíduo ácido insolúvel (do inglês, Acid Insoluble Residue)

AL – Ácido levulínico

ASE – Extração acelerada por solvente (do inglês, Accelerated solvent extraction)

ASL – Lignina ácida solúvel (do inglês, Acid soluble lignin)

ASTM – Sociedade Americana de Testes e Materiais (do inglês, American Society for Testing and Materials).

Viés – Erro Sistemático

CONAB – Companhia Nacional de Abastecimento

CF – Carbono fixo

CI – Índice de cristalinidade (do inglês, crystallinity index)

CZ – Teor de cinzas

DIBANET – Desenvolvimento de uma Rede para Abordagens de Integração da Biomassa (do inglês, Development of Integrated Biomass Approaches NETwork)

DS – Fração seca e moída com tamanho de partícula entre 180 e 850 µm (do inglês, Dry Sieved fraction).

DTG – Análise termogravimétrica derivada (do inglês Derivative Thermogravimetric Analysis)

DU – Fração seca e não moída (do inglês, Dry unground fraction)

FAO – Organização das Nações Unidas para Agricultura e Alimentação (do inglês, Food and Agriculture Organization)

FAR- Infravermelho Distante (do inglês, Far Infrared)

FT – NIR – Infravermelho Próximo com Transformada de Fourier (do inglês, Fourier Transform Near Infrared)

GC-MS – Cromatografia gasosa acoplada à espectrometria de massas (do inglês, Gas chromatography–mass spectrometry)

HMF – Hidroximetilfurfural

HPAEC-PAD – Cromatografia de troca iônica com detector amperométrico (do inglês, High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection)

HPLC-RI – Cromatografia líquida de alta eficiência com detector por índice de refração (do inglês, High performance liquid chromatography – *refractive index*)

IBGE – Instituto Brasileiro de Geografia e Estatística

ICO – Organização Internacional do Café (do inglês, International Coffee Organization)

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

KL – Lignina Klason (do inglês, Klason Lignin)

LD – Limite de Detecção
LQ – Limite de Quantificação
MID – Infravermelho Médio (do inglês, Middle Infrared)
NAS – Sinal analítico líquido (do inglês, Net Analyte Signal)
NIR – Infravermelho Próximo (do inglês, Near Infrared)
OSC – Correção ortogonal do sinal (do inglês, Orthogonal Signal Correction)
PC – Componente Principal (do inglês, Principal Component)
PCA – Análise de Componentes Principais (do inglês, Principal Component Analysis)
PLS – Quadrados Mínimos Parciais (do inglês, Partial Least Squares)
R² – Coeficiente de determinação
RMSEC – Raiz Quadrada do Erro Quadrático Médio da Calibração (do inglês, Root Mean Square Error of Calibration)
RMSECV – Raiz Quadrada do Erro Quadrático Médio de Validação Cruzada (do inglês, Root Mean Square Error of Cross Validation)
RMSEP – Raiz Quadrada do Erro Quadrático Médio de Previsão (do inglês, Root Mean Squares Error of Prediction)
RSD – Erro relativo (do inglês Relative standard deviation)
SEL – Seletividade (do inglês, selectivity)
SEN – Sensibilidade (do inglês, sensitivity)
SEP – Erro padrão de previsão (do inglês, Standard error of prediction)
SNV – Padronização normal de sinal (do inglês, Standard Normal Variate)
TGA – Análise termogravimétrica (do inglês Thermo Gravimetric Analysis)
TL – Lignina Total (do inglês, Total Lignin)
TMV – Teor de matéria volátil
TS – Açúcares Totais (do inglês, Total sugars)
UE – União Européia
UV-Vis – Espectroscopia no Ultravioleta e Visível
VL – Variável Latente
WU – Fração úmida e não moída (do inglês, Wet unground fraction)

LISTA DE TABELAS

Tabela 1. Equações para o cálculo das figuras de mérito.....	49
Tabela 2. Identificação, origem e ano de colheita das amostras de banana, café e coco.	53
Tabela 3. Composição físico-química das biomassas de cana – de- açúcar e <i>miscanthus</i> (%).....	68
Tabela 4. Resultados da análise imediata, elementar e poder calorífico das biomassas.	69
Tabela 5. Composição química das matérias-primas (%).	70
Tabela 6. Composição de carboidratos das matérias-primas (%)*.....	75
Tabela 7. Composição química das biomassas (%) pelos métodos TG e DTG.	81
Tabela 8. Valores da literatura para rendimentos experimentais, em %, de ácido levulínico (AL), ácido fórmico (AF) e furfural (Furf.).....	83
Tabela 9. Valores dos índices de cristalinidade das biomassas (%).	86
Tabela 10. Parâmetros para avaliação e validação dos modelos PLS para a fração DS.	98
Tabela 11. Parâmetros para avaliação dos modelos PLS obtidos para a fração DU. ...	107
Tabela 12. Parâmetros para avaliação dos modelos PLS obtidos para a fração WU. .	108
Tabela 13. Resultados das figuras de mérito para os modelos PLS para a fração DS.	141
Tabela 14. Resultados das figuras de mérito para os modelos PLS para a fração DU.	143
Tabela 15. Resultados das figuras de mérito para os modelos PLS para a fração WU.	144

LISTA DE FIGURAS

Figura 1. Projeção de Fischer para a D-glicose e a D-frutose.	16
Figura 2. Estruturas do monossacarídeo D-glicose na forma cíclica da α -D-glicose e na forma cíclica da β -D-glicose.	16
Figura 3. Estrutura da celulose (Adaptado: Murphy e McCarthy, 2005).	17
Figura 4. Representação da estrutura das fibrilas (Adaptado: Wickholm, 2001).	19
Figura 5. Açúcares que compõe a hemicelulose (Fengel e Wegener, 1989).	19
Figura 6. Estrutura ilustrando as diferentes ligações existentes na lignina (Adaptado: Fengel e Wegener, 1989).	21
Figura 7. Bioprodutos obtidos a partir da hidrólise da biomassa (AL: ácido levulínico; e 5-HMF: 5-hidroximetilfurfural).	24
Figura 8. Estrutura química do ácido levulínico.	25
Figura 9. Estrutura química do ácido fórmico.	26
Figura 10. Estrutura química do furfural.	26
Figura 11. Diagrama da energia potencial, níveis vibracionais e transições para o oscilador: (A) harmônico e (B) anarmônico (Fonte: Sala, 2008).	29
Figura 12. Gráfico para escolha correta do número de variáveis latentes. (* indica o número de variáveis latentes selecionado).	37
Figura 13. Relações entre as estimativas de RMSECV e RMSEP em função do método de validação (* indica o número de variáveis latentes selecionado), sendo $n=1$ LOO e $n>1$ LNO.	38
Figura 14. (A) Espectro bruto, (B) espectro após submetido à primeira derivada, (C) espectro após submetido à segunda derivada, (D) espectro com ruído e linha de base linear, (E) espectro com ruído após submetido à primeira e segunda derivada e (F) espectro com ruído após submetido à segunda derivada por Savitzky-Golay.	42
Figura 15. Representação geométrica do vetor NAS para a amostra i	47
Figura 16. Frações botânicas da bananeira.	51
Figura 17. Representação das cascas de café.	52
Figura 18. Frações botânicas do coco.	52
Figura 19. Amostras da fração DS utilizadas para as análises de referência (A) cascas de soja, (B) cascas de arroz, (C) bambu, (D) fibras de coco, (E) engaço de banana (F) serragem de eucalipto, (G) capim elefante, (H) caule de banana, (I) caroços de açaí e (J) cascas de café.	54
Figura 20. Representação esquemática dos procedimentos experimentais.	55
Figura 21. Difrátograma de raios-X ilustrando os picos usados no cálculo da cristalinidade pela intensidade (Adaptado: Park, 2010).	58
Figura 22. Difrátograma de raios-X ilustrando os picos usados no cálculo da cristalinidade pelo método da deconvolução de picos gaussianos.	59
Figura 23. Extrator acelerado por solvente (ASE).	60
Figura 24. Correlação entre o poder calorífico (PC) e a composição química das biomassas (ASL: lignina ácida solúvel; KL: lignina Klason; AIR: resíduo ácido insolúvel; TL: Lignina total; AIA: cinza ácida insolúvel; CF: carbono fixo; Gli: glicose; Xil: xilose; Ara: arabinose; Rha: ramnose; Gal: galactose; Man: manose e TS: açúcar total).	71
Figura 25. Cromatogramas dos hidrolisados de (A) cascas de café (B) fibras de coco e (C) engaço de banana.	77

Figura 26. Análises termogravimétricas (TG) das biomassas.....	78
Figura 27. Análises termogravimétricas (DTG) das biomassas.....	80
Figura 28. Rendimentos estimados em ácido levulínico (AL), ácido fórmico (AF) e furfural.....	82
Figura 29. Difrátogramas de raios-X das biomassas.....	85
Figura 30. Difrátogramas de raios-x da (A) serragem de eucalipto (B) caroço de açaí (C) engaço de banana e (D) caule de banana.....	88
Figura 31. Espectros médios da região Vis-NIR das biomassas.....	89
Figura 32. Análise descritiva e estatística dos parâmetros (TL: lignina total; KL: lignina Klason; ASL: lignina ácida solúvel; AIR: resíduo ácido insolúvel; Todos grupos: todas as frações botânicas; TS: açúcares totais; Man: manose; Gli: glicose; Xil: xilose; Ram: ramnose; Arab: arabinose; Galac: galactose).....	91
Figura 33. (A) Gráfico de pesos da análise PCA. (B) Gráfico de escores das duas primeiras componentes principais. (C) Gráfico de escores da terceira e da quarta PC para os conjuntos de banana, coco e café.....	93
Figura 34. (A) Gráfico de pesos da análise PCA. (B) Gráfico de escores das duas primeiras componentes principais. (C) Gráfico de escores da terceira e da primeira PC para o conjunto de banana.....	95
Figura 35. Espectros brutos da região Vis-NIR das amostras de banana, coco e café da fração DS.....	96
Figura 36. Espectros pré-tratados por OSC para os modelos (A) da região NIR de lignina total do conjunto da Banana e (B) da região Vis-NIR de lignina Klason do conjunto Banana+Café, ambos da fração DS.....	97
Figura 37. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) lignina total (TL) com 4 VL, (B) lignina ácida solúvel (ASL) com 6 VL, (C) resíduo ácido insolúvel (AIR) com 4 VL, e (D) lignina Klason (KL) com 4 VL, para os modelos de Banana + Café da fração DS com o pré-tratamento indicado na Tabela 8: (2D (15) +OSC).....	100
Figura 38. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) lignina total (TL) com 5 VL, (B) lignina Klason (KL) com 5 VL, (C) resíduo ácido insolúvel (AIR) com 7 VL, e (D) lignina ácida solúvel (ASL) com 6 VL, para os modelos de Banana + Café da fração DS com o pré-tratamento 2D (15).....	101
Figura 39. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) lignina ácida solúvel (ASL) com 7 VL, (B) lignina total (TL) com 7 VL, (C) lignina Klason (KL) com 7 VL e (D) resíduo ácido insolúvel (AIR) com 6 VL para os modelos de Banana+Café+Coco da fração DS.....	102
Figura 40. Gráficos de valores de referência versus valores previstos da calibração e validação cruzada para (A) lignina total (TL) com 3 VL, (B) lignina Klason (KL) com 3 VL, (C) lignina ácida solúvel (ASL) com 3 VL, e (D) resíduo ácido insolúvel (AIR) com 4 VL para os modelos de banana da fração DS.....	103
Figura 41. Espectros brutos da região Vis-NIR das amostras de banana, coco e café da fração DU.....	105
Figura 42. Espectros brutos da região Vis-NIR das amostras de banana, coco e café da fração WU.....	106
Figura 43. (A) Espectros de banana da fração WU pré-tratados por 1D (15); (B) espectros de coco da fração WU pré-tratados por 2D (15); (C) espectros de banana da	

fração DU pré-tratados por SNV e (D) espectros de coco da fração DU pré-tratados por SNV+Detrend.	111
Figura 44. Espectros pré-tratados da fração DU por (A) SNV+1D(3), (B) 1D(3) + SNV, (C) 2D(15)+SNV e (D) SNV+2D(15), respectivamente.	112
Figura 45. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) lignina ácida solúvel (ASL) com 8 VL do modelo da Banana, (B) resíduo ácido insolúvel (AIR) com 8 VL do modelo da Banana, (C) lignina Klason (KL) do modelo do Café com 7 VL, e (D) gráfico dos valores de referência versus valores previstos da calibração e validação cruzada para a lignina ácida solúvel (ASL) com 6 VL para o modelo de Coco, todos da fração DU.	114
Figura 46. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) lignina ácida solúvel (ASL) com 7 VL do modelo da Banana, (B) lignina Klason (KL) com 4 VL do modelo de Café, (C) resíduo ácido insolúvel (AIR) do modelo do Café com 4 VL, e (D) gráfico dos valores de referência versus valores previstos da calibração e validação cruzada para lignina ácida solúvel (ASL) com 6 VL para o modelo de Coco, todos da fração WU.....	115
Figura 47. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) cinzas do modelo Banana + Café + Coco da fração DS com 6 VL, (B) extrativos do modelo Banana + Café da fração DS com 7 VL, (C) cinzas do modelo Banana + Café da fração DS com 7 VL (D) umidade do modelo de Banana + Café da fração DS com 7 VL(E) extrativos do modelo de Banana da fração DU com 7 VL(F) e extrativos do modelo do Coco da fração WU com 6 VL.	118
Figura 48. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) arabinose com 7 VL (B) ramnose com 7 VL (C) glicose com 5 VL e (D) açúcares totais (TS) com 4 VL da fração DS para os modelos de café.	121
Figura 49. Gráficos de valores de referência versus valores previstos da calibração e validação cruzada para (A) açúcares totais (TS) com 2 VL (B) glicose com 2 VL (C) xilose com 2 VL e (D) galactose com 4 VL para os modelos de coco da fração DS. ...	122
Figura 50. Gráficos de valores de referência versus valores previstos da calibração e validação cruzada para (A) glicose com 3 VL (B) açúcares totais (TS) com 4 VL (C) galactose com 4 VL e (D) xilose com 6 VL para os modelos de banana da fração.....	123
Figura 51. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) teor de cristalinidade calculado pelo método da intensidade com 7 VL e (B) teor de cristalinidade calculado pela área com 7 VL, para os modelos de banana da fração DS.	125
Figura 52. Coeficientes de regressão para os modelos PLS de ligninas (lignina Klason e lignina ácida solúvel). Os ordinais (1°; 2°; e 4°) indicam as bandas de sobretons; comb. se refere às bandas de combinação; e v são bandas de estiramento.	130
Figura 53. Coeficientes de regressão para os modelos PLS de extrativos (conjuntos combinados de B+C e conjunto apenas da banana). Os ordinais (1°; 2°; e 4°) indicam as bandas de sobretons; comb. refere-se as bandas de combinação; e v são bandas de estiramento.....	131
Figura 54. Coeficientes de regressão para o teor de açúcares totais (conjunto de coco e café) dos modelos PLS. Os ordinais (1°; e 2°) indicam os sobretons; comb. refere se as bandas de combinação; e v são bandas de estiramento.	132
Figura 55. Gráfico do número de amostras no conjunto de validação versus R^2_{val} . (A)	

Valores médios e desvio padrão de R^2_{val} para o teor de resíduo ácido insolúvel para o modelo de banana. (B) Valores médios e desvio padrão de R^2_{val} para o teor de lignina total para o modelo de Banana + Café.	134
Figura 56. Representação dos modelos nas suas formas (A) pseudo-univariada dos valores de sinal analítico líquido (NAS) versus valores de referência da calibração e validação externa para o teor de extrativos do modelo de Banana+Café+Coco,(B) pseudo-univariada dos valores de sinal analítico líquido (NAS) versus valores de referência da calibração e validação externa para o teor de lignina total do modelo de Banana+Café+Coco (C) multivariada de valores de referência versus valores previstos da calibração e validação externa para o teor de extrativos do modelo de Banana+Café+Coco, (D) multivariada de valores de referência versus valores previstos da calibração e validação externa para o teor de lignina total do modelo de Banana+Café+Coco.	136
Figura 57. Gráfico dos valores previstos <i>versus</i> os de referência para (A) resíduo ácido insolúvel (AIR) e (C) extrativos da fração DS dos modelos de Banana+ Café; Gráfico dos resíduos <i>versus</i> os valores previstos para (B) resíduo ácido insolúvel (AIR) e (D) extrativos da fração DS dos modelos de Banana+ Café.	137
Figura 58. Gráfico dos valores previstos <i>versus</i> os de referência para (A) extrativos da fração WU do modelo de Banana e (C) lignina ácida solúvel (ASL) da fração DU do modelo de Café; Gráfico dos resíduos <i>versus</i> os valores previstos para (B) extrativos da fração WU do modelo de Banana e (D) lignina ácida solúvel (ASL) da fração DU do modelo de Café.	138

1 INTRODUÇÃO E JUSTIFICATIVA

Os materiais lignocelulósicos (biomassa) representam a maior fonte de carbono renovável terrestre e, portanto, as mais prováveis alternativas para o suprimento de energia e insumos para as atividades humanas dos nossos dias (Rodrigues, 2011; Dhillon e Wuehlisch, 2013). Por outro lado, resíduos das biomassas necessitam destino adequado, evitando-se assim, prejuízos ambientais face aos desequilíbrios gerados pelo descarte inadequado desses materiais.

A utilização desses resíduos e materiais requer soluções criativas, que devem levar em conta não somente os aspectos econômicos, mas também ambientais. Para isso, torna-se fundamental conhecer a composição biomássica, a fim de que se possa ter um aproveitamento dessas fontes renováveis, com possibilidade de geração adicional de um leque de bioprodutos de apreciável valor agregado (Goldemberg, 2006).

Diversos resíduos de natureza lignocelulósica disponíveis podem gerar energia e insumos químicos. Muitas dessas biomassas, como cascas de café, de soja, arroz, caule e engaço da banana, resíduos de coco, caroço de açaí, serragem de eucalipto, dentre outras, apresentam em sua constituição um grande teor de carboidratos, que podem ser convertidos em produtos de alto valor agregado, entre os quais destacam-se o ácido levulínico, o ácido fórmico e o furfural. O ácido levulínico, um importante produto da degradação da celulose, possui grande importância para as indústrias alimentícia, farmacêutica e química (Bozell *et al.*, 2000; Yaini *et al.*, 2012; Girisuta *et al.*, 2013). Já o furfural, derivado das pentoses, pode ser usado como solvente ou utilizado na produção de álcool e ácido levulínico. O ácido fórmico, produto das hexoses, é usado extensivamente na fabricação de medicamentos, corantes, inseticidas e refrigerantes (Hayes *et al.*, 2006). Ainda se busca valorizar a lignina residual dessas biomassas, através do desenvolvimento de seus derivados (Lignin Institute, 2012).

Nesse contexto, o aproveitamento de tais rejeitos tem impulsionado o desenvolvimento de novos métodos para a determinação dos componentes das biomassas, principalmente os constituintes majoritários (celulose, hemicelulose e

lignina), que têm encontrado crescente aplicação no campo de biorrefinarias. O termo biorrefinaria refere-se à transformação da biomassa em insumos químicos, energia e combustíveis, com mínimo desperdício, baixa emissão de poluentes e maximizando o aproveitamento da biomassa renovável em substituição aos derivados do petróleo (Charlton *et al.*, 2009).

Métodos de análises robustos e precisos são cruciais para determinar a composição química de matérias-primas, devido à grande variabilidade das amostras. Esses compreendem desde as etapas de amostragem e preparo de amostra até às de separação e quantificação. A determinação da composição das biomassas depende da combinação de um conjunto de procedimentos, podendo oferecer assim uma análise quantitativa dos materiais lignocelulósicos, que em tese representam quase 100% do material original (Sluiter *et al.*, 2010).

A composição química de tais materiais pode ser determinada por métodos de referência, embora esses possam ser caros e demorados. E assim, métodos rápidos, robustos e precisos são necessários e de grande valia, particularmente se integrados de maneira *online* em biorrefinarias. A espectroscopia na região do infravermelho próximo (NIR) é uma técnica rápida, de simples aplicação e não destrutiva, muitas vezes apropriada para substituir alguns dos métodos tradicionais de caracterização (Liu *et al.*, 2010). No entanto, as aplicações da espectroscopia NIR são quase totalmente dependentes de métodos matemáticos multivariados (ferramentas quimiométricas). A regressão por mínimos quadrados parciais (PLS) é um método que pode ser diretamente aplicado aos espectros NIR, resultando em modelos multivariados utilizados para prever as propriedades de interesse (Pedro e Ferreira, 2007; Rambo *et al.*, 2013).

A fim de fornecer resultados confiáveis e comparáveis aos métodos de referência, assegurando que os parâmetros estejam de acordo com as exigências técnicas e legais, é necessário avaliar a qualidade analítica desses modelos, que pode ser obtida por meio da determinação das figuras de mérito de métodos multivariados (Olivieri *et al.*, 2006; Valderrama *et al.*, 2009).

Tendo-se em vista todos esses aspectos, estudos foram realizados com o objetivo de se encontrar alternativas, de baixo custo para a caracterização de

biomassas, de forma rápida e segura, a partir de materiais renováveis, contribuindo para a inovação tecnológica e, paralelamente, para a solução de problemas ambientais.

O presente estudo está inserido no projeto DIBANET (Development of Integrated Biomass Approaches NETwork) financiado pela União Européia (UE). O projeto é uma resposta à chamada “Energia” de 2008 - “Melhoria significativa na cooperação entre pesquisadores e indústrias da UE e América Latina na área de biocombustíveis”, e o tema central da pesquisa foi “A produção de biocombustíveis sustentáveis a partir de resíduos da Europa e da América Latina” (<http://www.dibanet.org>).

1.1 Objetivos

O objetivo geral deste trabalho foi o aproveitamento de biomassas residuais, por meio de caracterizações físico-químicas aliados às ferramentas quimiométricas, contribuindo, paralelamente, para a mitigação do crescente problema ambiental decorrente da disposição irregular de resíduos agroindustriais e agrofloretais na América Latina.

Os objetivos específicos visados foram os seguintes:

- ✓ Selecionar matérias-primas residuais próprias para a obtenção de monossacarídeos e seus derivados (bioprodutos);
- ✓ Analisar tais biomassas, caracterizando-as por métodos físico-químicos de referência;
- ✓ Desenvolver metodologias analíticas rápidas que possam ser utilizadas em processos *online*, com reduzido número de etapas e menor consumo de reagentes possível;
- ✓ Desenvolver um método analítico alternativo, de baixo custo, rápido e com baixa geração de resíduos (Química Verde), empregando a espectroscopia na região do NIR, aliada à quimiometria, para a construção de modelos de calibração da composição química das mais diversas biomassas;
- ✓ Validar os procedimentos desenvolvidos pela análise das figuras de mérito de calibração multivariada.

2 REVISÃO BIBLIOGRÁFICA E CARACTERIZAÇÃO DO PROBLEMA

Recursos energéticos e matérias primas alternativas e renováveis, dentre eles os materiais lignocelulósicos, têm se tornado prioridade para muitos países, pois provavelmente irão desempenhar um papel importante na indústria química num futuro próximo. Esses recursos renováveis rapidamente tornaram-se popularizados devido aos baixos riscos ambientais e à menor poluição, sendo assim alternativas favoráveis em substituição aos combustíveis fósseis e seus derivados (Xu *et al.*, 2013). Além desses fatores favoráveis, eles são, em geral, melhor distribuídos geograficamente do que os combustíveis fósseis. Estima-se que a produção mundial de biomassa é de 220 bilhões de toneladas por ano (Sasmal *et al.*, 2012).

Por conseguinte, eles aumentam as possibilidades de diversificação e descentralização do abastecimento de energia e favorecem a auto-suficiência energética no âmbito local, regional, e nacional. Nesse escopo, a biomassa lignocelulósica representa uma fonte renovável, barata e ainda, na maioria das vezes, inexplorada de matéria-prima bruta para a conversão em líquidos, gases e outros químicos relevantes para a indústria (Dhillon e Wuehlisch, 2013).

Para um país tropical, e de proporções continentais, como o Brasil, o substituto natural para o petróleo é a biomassa. Diante dessa possibilidade o Brasil vem buscando novas alternativas para aumentar a diversidade de sua atual matriz energética. Dentre as alternativas destacam-se o uso de resíduos florestais e agrícolas cujo processamento adquire grande interesse socioeconômico. Além do mais, tais resíduos não afetam a produção de alimentos, nem provocam devastação ou qualquer outra forma de agressão às florestas (Shuchardt *et al.*, 2001; Dhillon e Wuehlisch, 2013).

Além do bagaço e da palha da cana-de-açúcar, que são biomassas já bem estabelecidas, outros materiais renováveis a partir de resíduos agroindustriais e agroflorestais devem ser considerados para que não haja a dependência por uma única fonte (Sasmal *et al.*, 2012). Entre esses, destacam-se as casca de arroz, de soja e de café, caroços de açaí, serragem de eucalipto, fibras de coco, capim-elefante, bambu, engaço e caule de banana, entre outros.

2.1 Biomassas

2.1.1 Cascas de arroz

Entre as alternativas agrícolas como fonte energética encontram-se as cascas de arroz. O arroz é o terceiro cereal mais produzido no mundo, superado apenas pelo trigo e o milho; é consumido por 2/3 da população mundial (Binod *et al.*, 2010). A produção mundial desse grão é de 590 Mt/a (milhões de toneladas por ano) resultando em 118 Mt/a de casca residual (20%) (Ma *et al.*, 2011). Segundo a Companhia Nacional de Abastecimento (CONAB, 2013), a produção brasileira é de 12 Mt/a de arroz fazendo com que o país ocupe o décimo lugar na lista dos produtores mundiais. Após a colheita, secagem e limpeza, o arroz é transportado para indústrias beneficiadoras onde é submetido a processo de descascamento. As cascas produzidas possuem baixo valor comercial, pois a sílica e as fibras não possuem valor nutritivo e por isso não são usados na alimentação humana ou animal (Amick, 1982) sendo, portanto, na maioria das vezes, descartadas e muitas vezes ilegalmente incineradas, a céu aberto. Alternativas de aproveitamento desse resíduo vêm sendo apresentadas na literatura.

Bevilaqua *et al.*, (2013) submeteram as cascas de arroz a um processo de hidrólise ácida para a produção de ácido levulínico, obtendo dessa forma um rendimento de 59,40%. Também uma ampla utilização dos produtos hidrolisados da casca de arroz, tais como a glicose, sílica e sulfato de sódio cristalino foram obtidos (Ma *et al.*, 2011). Rambo *et al.*, (2011) e (2013) demonstraram a possibilidade de utilização das cascas de arroz na obtenção de sílicas xerogéis e de xilitol. Esses e demais trabalhos publicados (Karagoz *et al.*, 2005; Binod *et al.*, 2010; Xiujuan *et al.*, 2011) demonstram a possibilidade de aproveitamento das cascas de arroz na obtenção de insumos de grande valia para a indústria petroquímica.

2.1.2 Bambu

Outro tipo de biomassa de interesse energético é o bambu. Apesar de não ser exatamente um resíduo, é de extrema valia por ser uma gramínea C4, como o milho e a cana-de-açúcar, onde o primeiro produto da fixação do CO₂ é um composto de quatro carbonos, diferente do 3-fosfoglicerato da via bioquímica "tradicional". Plantas C4 apresentam uma alta taxa de fotossíntese líquida associado com um alto ponto de saturação de luz. Além disso, plantas C4 possuem uma elevada afinidade pelo CO₂ e a fotorrespiração é quase nula, aumentando em muito sua eficiência fotossintética, e assim é um vegetal que tem rápido crescimento, elevada produção de biomassa e fácil multiplicação vegetativa (Peng *et al.*, 2012). Algumas espécies de bambu tem taxas de crescimento muito altas, atingindo sua altura máxima em cerca de 30 dias para as espécies de pequeno porte (15 cm), e no máximo 180 dias para as espécies gigantes (15 m). Dessa forma pode-se programar o sistema de colheitas periódicas e efetuar de 2 a 12 cortes por ano. Adicionalmente o Brasil é o país com maior diversidade de espécies de bambu, com uma produção de 900 Mt/a, dos quais aproximadamente 300 Mt/a correspondem aos resíduos de bambu, na forma principalmente de pó. Trabalhos de utilização do bambu como biomassa lignocelulósica para produção de bioenergia, bioetanol e açúcares já tem sido propostos na literatura (Leenakul e Tippayawong, 2010; Tippayawong e Chanhom, 2011; Peng *et al.*, 2012; Sun *et al.*, 2011 e 2013).

2.1.3 Serragem de eucalipto

Outra cadeia produtiva brasileira de elevada relevância econômica que gera uma quantidade apreciável de resíduos é a madeireira. Segundo dados da Associação Brasileira de Produtores de Florestas Plantadas (ABRAF, 2012), há no Brasil 6,5 milhões de hectares de florestas plantadas e, desse total, 4,8 milhões de hectares são plantados com eucalipto. A produção de madeira serrada no Brasil gera aproximadamente 960 mil toneladas/ano de serragem de eucalipto, um resíduo descartado a céu aberto. Outros resíduos tais como folhas e cascas, também são gerados e na maioria das vezes são descartados ou reciclados parcialmente. No Brasil, o eucalipto leva aproximadamente sete anos até ser colhido e requer poucas ações do

homem sobre o solo. Pode ser cultivado em terrenos de baixa fertilidade natural, não exigindo muitos nutrientes e defensivos agrícolas em comparação com outras culturas (ABRAF).

Algumas aplicações de interesse econômico e ambiental vêm sendo dadas a estes resíduos da produção de eucalipto como destacado nos trabalhos de Martins *et al.*, (2007) que consistem na utilização da serragem para a geração adicional de produtos de apreciável valor econômico (insumos químicos e energéticos); Rafiqul e Sakinah, (2011 e 2012) que utilizam a serragem de eucalipto na hidrólise ácida para a produção de xilose.

2.1.4 Capim elefante

Por ser uma espécie de rápido crescimento e de alta produção de biomassa vegetal, o capim-elefante, também uma planta C4, apresenta um alto potencial para uso como fonte alternativa de energia. Da mesma forma que o bambu, o capim elefante não é um resíduo das produções agroindustrias e agroflorestais, mas dadas às suas inúmeras vantagens, representa uma biomassa promissora quando se trata de biorrefinarias.

Além disso, deve-se destacar que o capim elefante, por apresentar um sistema radicular bem desenvolvido, pode contribuir de forma eficiente para aumentar o conteúdo de matéria orgânica do solo, além de aumentar o sequestro de carbono (Monono *et al.*, 2013). Na Europa já se produz, com sucesso, energia elétrica a partir desse capim. Sendo que, em clima temperado, a produtividade média é de 10 toneladas de massa seca por hectare por ano (t/ha/a), enquanto no Brasil se chega a cerca de 60 t/ha/a de massa seca, uma vez que as condições climáticas são favoráveis ao capim no Brasil (Andrade e Canellas, 2012). Gómez *et al.*, (2012) apontaram o potencial do capim elefante na produção de etanol, enquanto Redding *et al.*, (2011) apresentaram os altos rendimentos de açúcares fermentescíveis, (94%) após a hidrólise enzimática do capim elefante.

2.1.5 Caroços de açaí

De acordo com a CONAB a projeção da produção nacional de frutos da palmeira açaí para 2013 é de 500.033 toneladas. Nutricionalmente trata-se de uma importante fonte de energia e têm sido reconhecidas certas propriedades funcionais no açaí (Gordon *et al.*, 2012; Inácio *et al.*, 2013). O principal produtor é o Estado do Pará, concentrando 87% da produção nacional. A polpa corresponde a 15% do peso total do fruto enquanto que o caroço (resíduo da produção) corresponde a 85%. Este pode ser utilizado na fabricação de móveis, placas acústicas, na indústria automobilística, na geração de vapor, entre outros (Homa, 2005). Devido o grande consumo do açaí, principalmente na região norte do país, existe uma grande quantidade de resíduos não aproveitados, com os caroços jogados pelas ruas e nos lixões, sem nenhum tratamento (Gordon *et al.*, 2012; Inácio *et al.*, 2013). O alto teor de celulose presente nos caroços de açaí é o que tem atraído interesse para o seu uso em biorrefinarias (Rodríguez-Zúñiga, 2008).

2.1.6 Cascas de café

É sabido que o Brasil é o maior produtor de café com uma produção de 2,6 Mt/a, dos 7,8 Mt/a produzidos mundialmente, segundo dados da Organização Internacional do Café (ICO, 2012). Da colheita ao processamento e consumo, diversos resíduos são gerados, onde as cascas são os principais subprodutos da indústria do café, já que para cada kg de frutos de café produzidos, aproximadamente 0,18 kg de cascas são geradas (Murthy e Naidu, 2012). Essas cascas podem ser utilizadas para várias finalidades, visto serem ricas em carboidratos e outros nutrientes (Murthy e Naidu, 2012; Mussatto *et al.*, 2012; Tsai *et al.*, 2012). No entanto, devido a enorme produção de café no país, um destino adequado e total desses resíduos não é alcançado.

Trabalhos de utilização destes resíduos são propostos. Mussatto *et al.*, (2012) submeteram os resíduos da indústria do café a processos de hidrólise ácida, fornecendo rendimentos em etanol com 50,20% de eficiência. Gouvea *et al.*, (2009), também avaliaram a possibilidade de obter etanol a partir das cascas de café, obtendo resultados bastante satisfatórios.

2.1.7 Caule e engaço da banananeira

Banana e seus resíduos - engaço e pseudocaule - são matérias-primas de potencial devido à sua elevada abundância. Anualmente cerca de 95,6 Mt/a de bananas são produzidas, com uma produção total de 6,9 Mt/a no Brasil (Organização das Nações Unidas para Alimentação e Agricultura, FAO), fazendo com que o país ocupe a quinta posição mundial. A partir dessa produção, aproximadamente, 20 Mt/a de pseudocaule e 1 Mt/a de engaço são gerados, uma vez que para cada 1 tonelada de banana colhida, aproximadamente 4 toneladas correspondem aos resíduos lignocelulósicos (3 toneladas de pseudocaule, 160 kg de engaço, 480 kg de folhas e 440 kg de cascas). Atualmente, essa grande quantidade de resíduos representa uma importante fonte de contaminação ambiental causado pelo descarte incorreto e decomposição, liberando gases tóxicos, tais como sulfeto de hidrogênio, amônia, entre outros (Tock *et al.*, 2010). Frente a esses problemas de disposição inadequada, alternativas aos resíduos da produção de banana devem ser procuradas.

Li *et al.*, (2012) submeteram folhas e pseudocaule de banana a um processo de hidrólise enzimática a fim de avaliar as melhores condições para aproveitamento da celulose desses resíduos. Arredondo *et al.*, (2010) realizaram uma avaliação do ciclo de vida energético para o processo de produção de etanol obtido a partir da hidrólise de amido e de material celulósico e hemicelulósico presente na banana e nas suas biomassas residuais. Foram consideradas na análise, o cultivo da banana, o tratamento dos resíduos, o transporte da matéria-prima, a hidrólise, a fermentação, a destilação e a utilidade da planta. A análise mostrou, no entanto, que os processos podem ser considerados energeticamente viáveis. Os frutos da banana e seus resíduos orgânicos são matérias-primas que podem ser usadas para produzir etanol e reduzir dessa forma seus impactos ambientais. Os resultados obtidos na avaliação apresentaram um balanço positivo de energia e, portanto, elas podem ser consideradas fontes de energia renováveis. Embora o material lignocelulósico de banana apresente resultados positivos, suas pesquisas em escala industrial, são ainda muito pobres, conclui Arredondo (2010).

2.1.8 Fibras de coco

No Brasil, outro resíduo apropriado para ser utilizado na substituição aos combustíveis fósseis são os resíduos do coco, devido ao alto potencial que o país apresenta na produção dessas fibras vegetais. São produzidos anualmente 1,5 bilhões de cocos, sendo que desse total apenas 15% são aproveitados (Wei e Gu, 2009; FAO, 2012). O coco não é apenas utilizado para fornecer alimentos e bebidas, mas pode ser usado também para a fabricação de cortinas, tapetes, cestos, chapéus, bandejas e cordas, e ainda como matéria orgânica do solo (Hemstock, 2013). No entanto, a maior parte das fibras e das cascas são abandonadas na natureza, consideradas um desperdício de recursos naturais e causadoras de poluição ambiental. Tendo isso em vista, Vaithanomsat *et al.*, (2011) e Yuan *et al.*, (2011) avaliaram o potencial de utilização das fibras de coco para produção de bioetanol, mostrando alta produtividade de etanol a partir dessa biomassa.

2.1.9 Cascas de soja

O Brasil é o segundo maior produtor mundial de soja, com uma produção de 83 Mt/a do grão. As cascas de soja representam o principal subproduto da indústria do seu processamento e constituem cerca de 8% do grão inteiro, o que resulta em 6,64 Mt/a de cascas de soja por ano (CONAB). Parte dessas cascas são utilizadas na alimentação animal, mas devido a enorme quantidade produzida, não é possível todo o aproveitamento desse resíduo, o qual passa a constituir um problema na esfera ambiental (Michel *et al.*, 2008). Portanto é de extrema importância, agregar valor a esses resíduos. Foi essa a ideia proposta por Michel *et al.*, (2008) e Cassales *et al.*, (2011) em seus trabalhos. Os primeiros autores produziram etanol e xilose com altos rendimentos a partir da hidrólise enzimática das cascas de soja. O segundo trabalho avaliou a composição química e a cinética da hidrólise ácida das cascas de soja, demonstrando o potencial de utilização dessas como substrato para diversos processos biológicos após a sua hidrólise. Seu elevado conteúdo de açúcares e o baixo teor de substâncias tóxicas é favorável ao seu uso em biorrefinarias. Também ácidos orgânicos de alto valor para a indústria petroquímica, como o ácido succínico, foram obtidos utilizando as cascas de soja como substrato (Thakker *et al.*, 2013). Até agora, esse

material produzido em excesso, mais de 6 Mt/a, tem sido amplamente subutilizado na alimentação animal ou simplesmente descartado no ambiente, sendo que inúmeras aplicações relevantes, como algumas já citadas, podem ser atribuídas a tal resíduo.

Frente a essas constatações e à diversidade dos constituintes entende-se importante a busca de soluções para o uso desses resíduos como insumos energéticos ou para outras finalidades. Além disso, como já mencionado, a geração de resíduos força o desenvolvimento de tecnologias para controlar sua geração, para dispor ou modificar os mesmos, ou para desenvolver outros produtos a partir deles.

2.2 Composição química das biomassas lignocelulósicas

Os materiais lignocelulósicos estão entre os recursos orgânicos renováveis mais abundantes da terra e possuem imenso potencial de uso como matérias-primas em processos industriais para a produção de alimentos, biocombustíveis, insumos químicos, enzimas, biofertilizantes e bens de consumo diversos (Charlton *et al.*, 2009; Sasmal *et al.*, 2012; Thakker *et al.*, 2013). Essa biomassa inclui materiais oriundos das atividades de exploração agro-industrial e florestal, tais como as citadas anteriormente, entre outras. Esses resíduos, em geral, apresentam uma composição básica, que consiste de cerca de 40 a 60% de celulose, 20 a 40% de hemicelulose, 15 a 25% de lignina e de 2 a 25% de extrativos (Limayem e Ricke, 2012).

Devido à importância dos carboidratos nas tecnologias de biorrefinarias, uma descrição mais detalhada desses será dada a seguir.

2.2.1 Carboidratos

Os carboidratos são os principais constituintes da maioria dos materiais lignocelulósicos. Esses são compostos poli-hidroxi orgânicos com uma composição elemental geral $(CH_2O)_n$, o que dá um teor de carbono geralmente uniforme de cerca de 40%. A natureza oxigenada dos carboidratos proporciona propriedades químicas importantes para processos de conversão e utilização, justificando porque eles são tão dominantes e essenciais na biota (Roberts, 1996).

Os carboidratos podem ser classificados em três grupos:

- Monossacarídeos – açúcares simples como glicose e xilose, que constituem as moléculas dos carboidratos, as quais são relativamente pequenas, solúveis em água e não hidrolisáveis.
- Dissacarídeos e Oligossacarídeos – são cadeias orgânicas constituídas de dois (dissacarídeos) ou entre 3 ou 10 resíduos de monossacarídeos unidos por ligações glicosídicas (oligossacarídeos). Os dissacarídeos e os trissacarídeos (3 unidades de monossacarídeos) são os oligossacarídeos mais importantes e abundantes na natureza, como por exemplo, a sacarose, celobiose e a rafinose. Devido ao baixo comprimento da cadeia, os oligossacarídeos são facilmente hidrolisados em seus componentes monoméricos e são muitas vezes solúveis em água. Eles são, portanto, importantes constituintes da fração dos extrativos (Seção 2.2.3).
- Polissacarídeos – geralmente considerados como polímeros de mais de 10 resíduos de monossacarídeos. Esses compostos apresentam uma massa molecular muito elevada que depende do número de unidades de monossacarídeos que se unem. Podem ser hidrolisados em polissacarídeos menores, assim como em dissacarídeos ou monossacarídeos. Os principais polissacarídeos são o amido, glicogênio e a celulose.

Os monossacarídeos, formados como os primeiros produtos da fotossíntese de CO_2 e água, são à base de construção para todos os carboidratos da planta. Eles raramente estão presentes na forma livre em plantas, mas ocorrem como unidades de oligossacarídeos e polissacarídeos - que estocam reserva alimentar para a planta (por exemplo, amido) ou fornecem suporte para a parede celular (por exemplo, celulose e hemicelulose).

Os monossacarídeos possuem um grupo carbonila podendo ser tanto aldeído como cetona. Os poli-hidroxi-aldeídos também conhecidos como aldoses, contém um grupo aldeído, já as poli-hidroxi-cetonas, (cetoses) são monossacarídeos com um grupo cetona. Também apresentam várias hidroxilas, geralmente uma em cada átomo de carbono que não fazem parte do grupo funcional carbonílico.

Entre as aldoses e cetoses mais frequentes na natureza pode-se destacar a D-glicose (aldose) e D-frutose (cetose). Esses são desenhados na projeção de Fischer na Figura 1. Em solução, verifica-se um equilíbrio entre a forma linear e as formas cíclicas, o qual se desloca para as estruturas cíclicas (Figura 2).

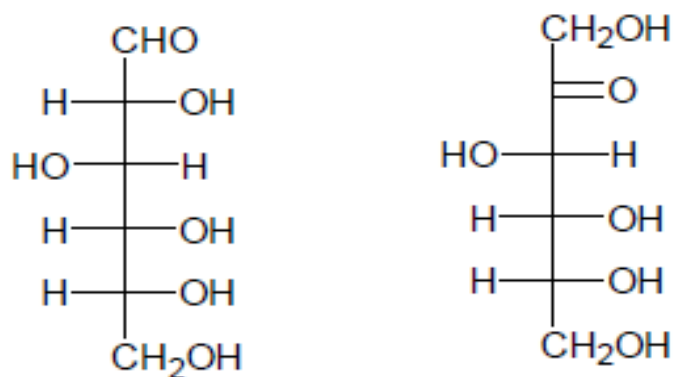


Figura 1. Projeção de Fischer para a D-glicose e a D-frutose.

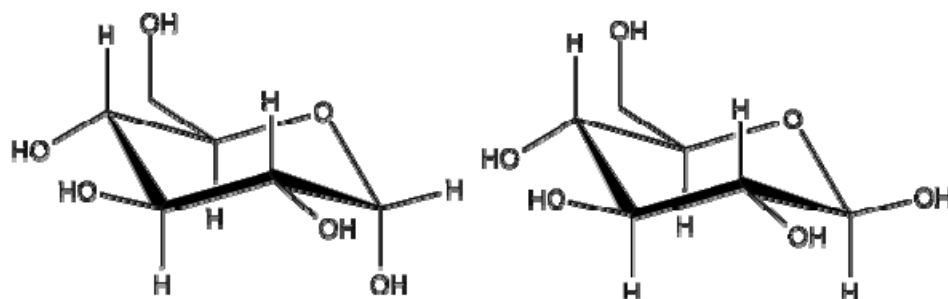


Figura 2. Estruturas do monossacarídeo D-glicose na forma cíclica da α-D-glicose e na forma cíclica da β-D-glicose.

Os açúcares são classificados de acordo com o número de átomos de carbono no monossacarídeo: por exemplo, trioses (3 átomos), tetroses (4), pentoses (5), hexoses (6), e heptoses (7). Os carboidratos na natureza mais importantes são as aldopentoses e as aldohexoses (Davis e Fairbanks, 2002).

A glicose é o mais abundante dos monossacarídeos na natureza, ocorrendo predominantemente como um componente da celulose e do amido. É também um importante precursor de outros monossacarídeos e carboidratos. Por exemplo, um, ou mais dos grupos CH₂OH das hexoses pode ser substituído por COOH resultando no

ácido glicurônico; por CH_3 (originando ramnose e fucose), ou por H (resultando na xilose).

Outras aldohexoses importantes na biomassa são a manose e a galactose - ambas presentes em determinados polissacarídeos hemicelulósicos. Já as aldopentoses mais frequentes são a xilose e arabinose.

A grande maioria dos açúcares existe nas biomassas como polissacarídeos. Homopolissacarídeos são aqueles que contêm uma única repetição de açúcar e são geralmente nomeados de acordo com o respectivo monossacarídeo, por exemplo, glucanas, mananas, xilanas, galactanas e arabinanas. Já os heteropolissacarídeos contêm mais do que um tipo de açúcar na cadeia. A celulose, por exemplo, é um homopolissacarídeo de glicose, sendo o constituinte principal encontrado na composição das biomassas.

A partir desses carboidratos citados acima, uma série de insumos químicos de alto valor agregado podem ser obtidos. Dessa forma é importante que os resíduos lignocelulósicos disponíveis apresentem altos teores de açúcares.

2.2.1.1 Celulose

A celulose é o constituinte normalmente encontrado em maior abundância nas biomassas vegetais. É a substância formadora da estrutura esquelética da planta, consistindo em um polímero (Figura 3) formado por unidades de anidro-glicose ($\text{C}_6\text{H}_{12}\text{O}_6$) conectadas através de ligações químicas envolvendo átomos de oxigênio (ligação denominada β -glicosídica).

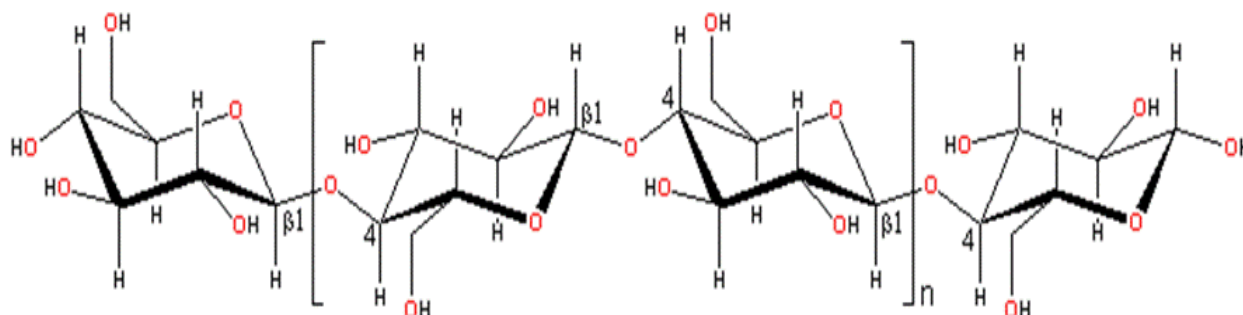


Figura 3. Estrutura da celulose (Adaptado: Murphy e McCarthy, 2005).

As cadeias de celulose agregam-se formando fibrilas e apresentam ligações de hidrogênio entre os grupamentos hidroxila intra e intercadeias, resultando na cristalinidade da celulose (Fengel e Wegener, 1989). Muitas vezes a região cristalina da celulose, com moléculas altamente ordenadas, dificulta a hidrólise devido às fortes ligações existentes entre as unidades de glicose. O que ocorre, na maioria das vezes, é uma hidrólise parcial, uma vez que a região amorfa, com uma estrutura desorganizada, é facilmente hidrolisada. Devido a isso, ácidos concentrados são empregados na etapa da hidrólise ácida a fim de reduzir ao máximo a cristalinidade da celulose, convertendo-a rapidamente para um estado totalmente amorfo (Yu *et al.*, 2008).

Embora materiais amiláceos requerem temperaturas de apenas 60 - 70 °C para serem convertidos de uma textura cristalina para uma amorfa, a celulose requer temperaturas de 320 °C, bem como uma pressão de 25 MPa para passar de uma estrutura rígida cristalina para uma estrutura amorfa em água (Limayem e Ricke, 2012). A cristalinidade da celulose é conhecida por variar em função da origem da amostra, por exemplo, a celulose de algodão é mais cristalina do que a celulose da madeira (Sjostrom, 1981). Também é provável que esse percentual varie entre a parede primária e secundária. Sete formas cristalinas foram identificadas para a celulose, sendo que cada uma dessas apresenta características físicas e químicas próprias (Ogeda e Petri, 2010).

Modelos, como o proposto por Wickholm (2001) depois de estudos de ressonância magnética nuclear (RMN), e ilustrado na Figura 4, descrevem a região cristalina como estando presente apenas no núcleo interno das microfibrilas e com uma estrutura não cristalina (celulose para-cristalina) entre o núcleo e a superfície. Superfícies acessíveis (expostas) e inacessíveis (devido à agregação das microfibrilas) também existem na estrutura.

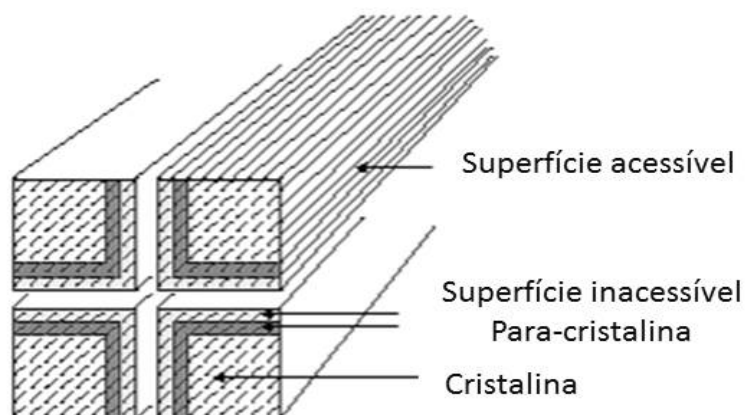


Figura 4. Representação da estrutura das fibrilas (Adaptado: Wickholm, 2001).

2.2.1.2 Hemicelulose

As hemiceluloses, por outro lado, são heteropolímeros de menor massa molar compostas pelos açúcares glicose, manose, ramnose e galactose (hexoses), e, em maior quantidade, xilose e arabinose (pentoses), podendo, ainda, apresentar quantidades variáveis de ácidos urônicos e grupos acetila (Figura 5).

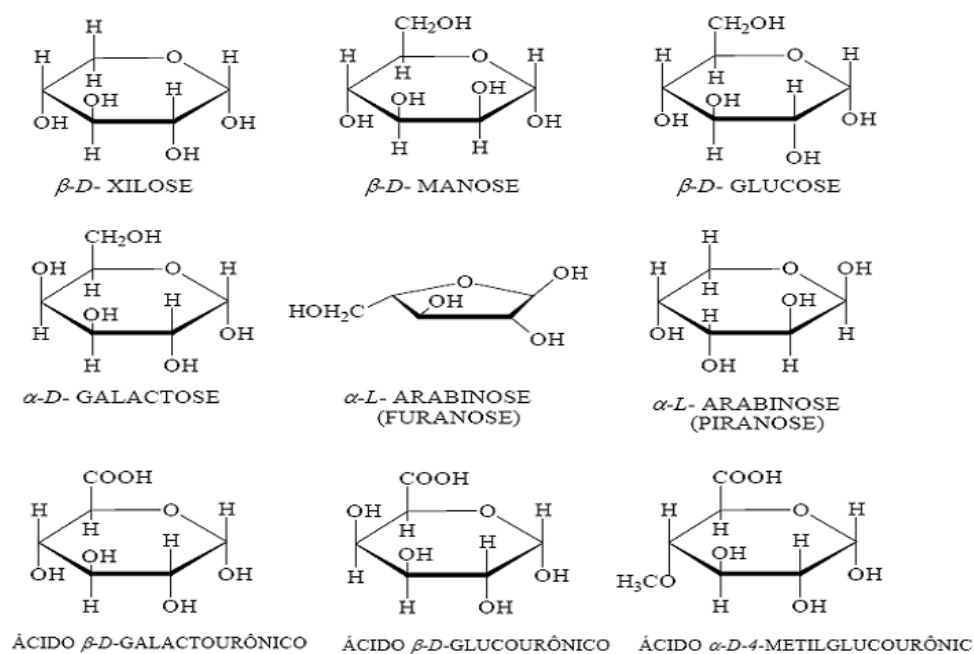


Figura 5. Açúcares que compõe a hemicelulose (Fengel e Wegener, 1989).

Devido à essa natureza diversificada, a hemicelulose apresenta uma estrutura amorfa (Pereira, 2008). Age como substância de reserva e de sustentação, servindo de conexão entre a lignina e a celulose. O grau de polimerização desse heteropolímero é geralmente inferior a 200 unidades (Fengel e Wegener, 1989). Por possuir massa molecular média menor, a hemicelulose é muito mais solúvel que a celulose e é mais rapidamente hidrolisada (Limayem e Ricke, 2012), pois suas ligações glicosídicas quebram mais facilmente que a celulose permitindo a utilização de seus açúcares, como a D-xilose, para a produção de produtos de interesse, como, por exemplo, o xilitol (Rambo *et al.*, 2013). O mais abundante polímero da hemicelulose é o xilano (um polímero da xilose).

São classificadas de acordo com os sacarídeos presentes na cadeia principal do polímero: xilanas, mananas, glucanas, galactanas e pectinas (Limayem e Ricke, 2012). As hemiceluloses são facilmente fracionadas e extraídas da biomassa – por essa razão são amplamente estudadas para sua utilização em biorrefinarias (Schuchardt e Ribeiro, 2001).

2.2.2 Lignina

Já a lignina é um polímero que atua como suporte das paredes celulares, é insolúvel na maioria dos solventes orgânicos, apresentando peso molecular elevado e muitas ligações cruzadas, características de polímeros termorrígidos. Possui natureza química bem distinta dos carboidratos, sendo caracterizada por uma estrutura preponderantemente aromática. A lignina é formada basicamente por três alcoóis (Fengel e Wegener, 1989) o p-cumarílico, coniferílico e sinapílico que levam respectivamente a formação da p-hidroxifenila, da guaiacila e da siringila.

Devido ao processo de polimerização ser aleatório, a macromolécula de lignina possui estrutura bastante complexa (Ralph *et al.*, 2004) com vários tipos de ligações químicas estáveis do tipo C-C, aril-éter e diaril-éter como apresentado na Figura 6. A lignina atua como ligante das fibras de celulose, contribuindo para a resistência e rigidez do tecido da madeira. A solubilidade da lignina em meio ácido, neutro ou alcalino depende dos seus precursores (p-coumaril, coniferil ou sinapil, ou então de uma combinação desses) (Limayem e Ricke, 2012). Muitos estudos já foram realizados

sobre a lignina, mas, mesmo assim, é uma das macromoléculas naturais mais difíceis de caracterizar quimicamente, dada a sua grande complexidade estrutural.

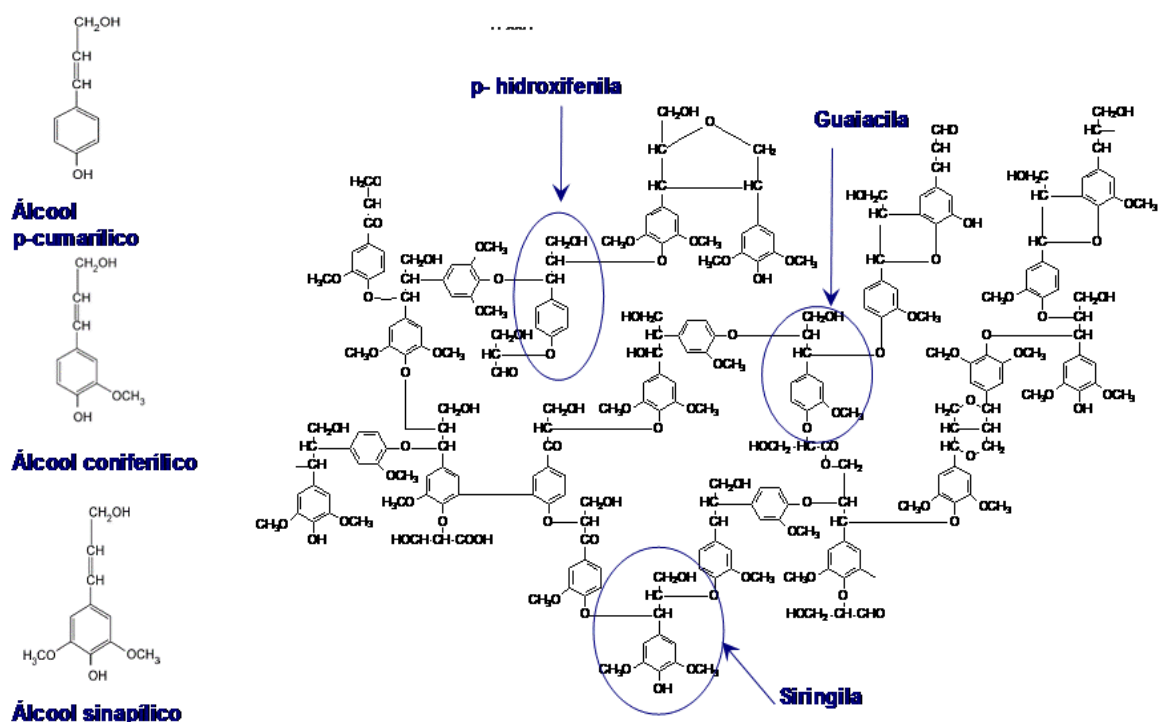


Figura 6. Estrutura ilustrando as diferentes ligações existentes na lignina (Adaptado: Fengel e Wegener, 1989).

A indústria começou a utilizar a lignina em 1880, quando lignosulfonatos foram utilizados em curtimento de couro. A partir desse momento, a lignina passou a ter aplicações em produtos alimentícios, servindo como emulsificantes na alimentação animal e como matéria-prima na produção de vanilina, que é amplamente utilizada como flavorizante em alimentos. Passou a ser utilizada ainda como componente de produtos de formulação farmacêutica e, também, como fragrância na indústria de perfumes. Os produtos derivados da lignina tendem a expandir-se, criando uma série de impactos em vários segmentos industriais (Lignin Institute, 2006). Apesar das centenas de utilidades da lignina, a sua principal utilização é como biocombustível sólido para substituir os combustíveis fósseis na geração de calor ou energia (Azadi *et al.*, 2013).

2.2.3 Extrativos

Existem ainda outros componentes que representam aproximadamente 2-25% da biomassa vegetal, que são os extrativos e consistem de gorduras, gomas, alcalóides, resinas, monossacarídeos, polissacarídeos, compostos aromáticos, óleos essenciais e outros constituintes citoplasmáticos. Extrativos são definidos como componentes estranhos aos principais constituintes estruturais supracitados, que podem ser separados do material insolúvel da parede celular pela sua solubilidade em água ou em solventes orgânicos neutros. Solventes de diferentes polaridades são necessários para remover diferentes tipos de extrativos. Por isso, os extrativos são geralmente classificados segundo qual solvente pode extraí-los, por exemplo, extrativos solúveis em etanol (Goldstein, 1991).

Os extrativos são de importância para o fracionamento das biomassas lignocelulósicas, uma vez que podem afetar a precisão dos procedimentos de análise. Têm sido demonstrado que os extrativos exercem influência significativa sobre o mecanismo de pirólise das biomassas (Guo *et al.*, 2010), liberando os principais produtos no início do termograma (ver seção 4.1.8). Além disso, os produtos extraíveis solúveis em água, como os carboidratos, constituem uma fonte adicional de açúcares. Estes compostos são comumente responsáveis por características como cor, sabor, resistência ao apodrecimento e propriedades abrasivas (Sjöström; Wetermark, 1999).

Tem-se ainda na biomassa uma classe de compostos conhecidos como os não extrativos, que incluem substâncias como sílica, carbonatos e oxalatos.

2.3 Hidrólise da biomassa

Os constituintes majoritários da biomassa geralmente encontram-se associados, dificultando assim seu ataque por agentes químicos, enzimáticos ou microbianos, que não possuem a capacidade de assimilar agregados tão grandes.

Dessa maneira, utilizam-se processos físico-químicos, que visam a separação dessas frações, como por exemplo, a hidrólise ácida pressurizada. Ácidos tais como sulfúrico, clorídrico e acético são comumente empregados como catalisadores nesses processos, hidrolisando preferencialmente a fração hemicelulósica, que é mais

susceptível ao tratamento ácido (Aguilar *et al.*, 2002). O que ocorre é que o catalisador tem sua difusão facilitada dentro da cadeia polimérica da hemicelulose, por essa apresentar uma estrutura aberta aliada à sua estrutura heterogênea e baixo grau de polimerização e cristalização, proporcionando um melhor rendimento em condições mais amenas e fazendo com que esse constituinte da biomassa seja bastante atrativo para uso em processos fermentativos (Parajó *et al.*, 1998) e demais processos no escopo de biorrefinarias.

Já em condições mais drásticas, ocorre também a hidrólise da celulose, pois essa necessita de temperaturas elevadas para romper suas ligações, uma vez que apresenta um alto grau de polimerização e regiões cristalinas resistentes à hidrólise (Parajó *et al.*, 1998).

Durante a hidrólise ácida, a concentração de ácido é o parâmetro mais importante que afeta a produção de monossacarídeos, enquanto que a temperatura é o fator de maior impacto na formação de produtos oriundos da degradação dos açúcares, tais como o ácido levulínico, ácido fórmico, furfural, entre outros (Rafiqul e Sakinah, 2012). Por exemplo, a hidrólise com ácidos diluídos proporciona um rendimento máximo de glicose de 50%. Já a hidrólise com ácidos concentrados rompe as ligações de hidrogênio entre as cadeias de celulose convertendo-as para um estado completamente amorfo. Nesse momento como a celulose é extremamente susceptível a hidrólise, pode-se diluir com água a temperaturas amenas fornecendo assim uma hidrólise completa (> 90% de glicose) e rápida (Yu *et al.*, 2008). Para o segundo parâmetro, altas temperaturas estão associadas a reações secundárias, levando à degradação a outros produtos, de interesse ou não (Hayes *et al.*, 2006; Yaaini *et al.*, 2012).

Os ácidos utilizados como catalisadores nos processos de hidrólise liberam prótons que atuam nas ligações glicosídicas entre os monômeros de açúcares nas cadeias poliméricas. O rompimento dessas ligações libera uma série de compostos, principalmente, monossacarídeos como xilose, glicose e arabinose. São liberados, também, produtos como furfural, da desidratação de pentoses, e hidroximetilfurfural (5-HMF), oriundo da desidratação de hexoses, havendo ainda formação de ácido fórmico e de ácido levulínico, produzidos a partir da desidratação de 5-HMF. Também é gerado ácido acético, dos grupos acetil da fração da hemicelulose, e ainda produtos de

degradação da lignina (Hayes *et al.*, 2006). A lignina permanece na forma de um resíduo sólido, o qual pode ser submetido a diversos processos, tais como pirolíticos (liquefação e gaseificação) e oxidativos, a fim de obter produtos de alto valor agregado, como vanilina, eugenol, resinas, gás de síntese, bioóleo, biocarvão (*biochar*) entre outros (Rodrigues, 2011).

Na Figura 7 tem-se um esquema simplificado dos produtos que podem ser obtidos a partir da hidrólise ácida de biomassas lignocelulósicas.

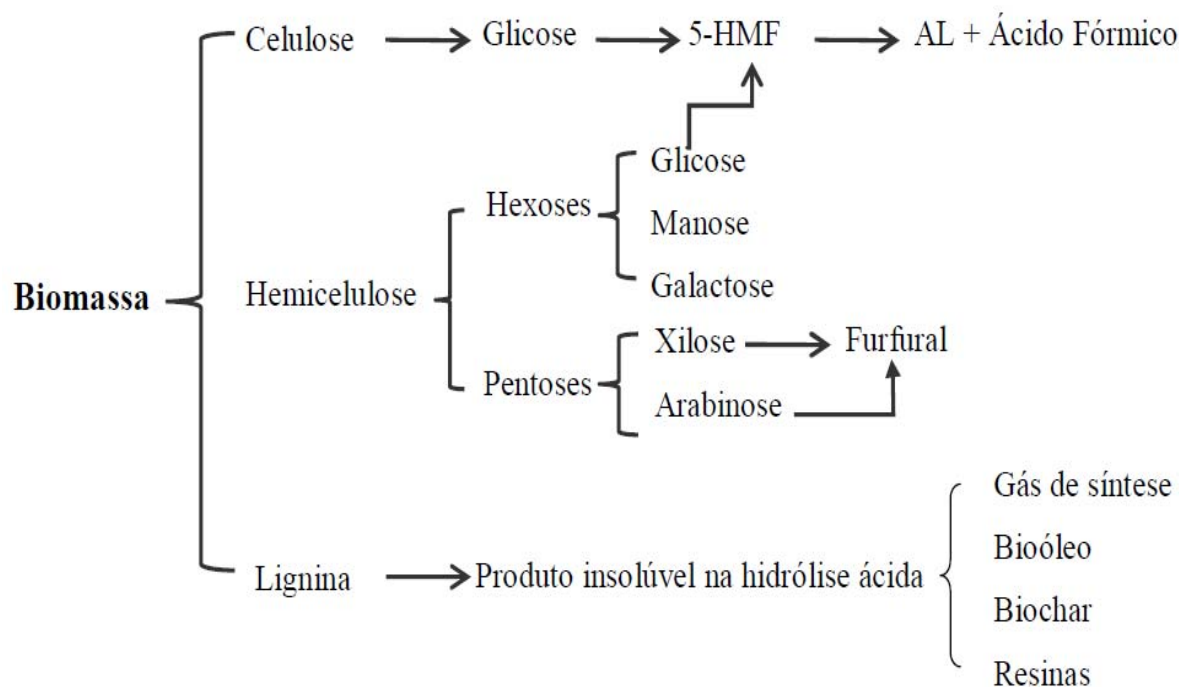


Figura 7. Bioprodutos obtidos a partir da hidrólise da biomassa (AL: ácido levulínico; e 5-HMF: 5-hidroximetilfurfural).

O rendimento teórico máximo de ácido levulínico, por exemplo, no processo de hidrólise ácida a partir de uma hexose é de 71% m/m e o restante é de ácido fórmico. No entanto o rendimento real depende das reações de degradação envolvidas na hidrólise ácida. Normalmente a conversão é de cerca de 50% da massa das hexoses para ácido levulínico, com 20% sendo convertido em ácido fórmico e os restantes 30% em alcatrão. O rendimento real em massa de furfural a partir das pentoses é cerca de 50%, sendo que os restantes 50% são de carvão (Hayes *et al.*, 2006), de acordo com

as equações empíricas abaixo:

Ácido levulínico (%): $0,5 \times \text{teor de hexoses}$

Ácido fórmico (%): $0,2 \times \text{teor de hexoses}$

Furfural (%): $0,5 \times \text{teor de pentoses}$

O ácido 4-oxopentanoico ($\text{C}_5\text{H}_8\text{O}_3$) ou ácido levulínico como também é conhecido, (Figura 8) tem sido alvo de muitos estudos por ser produto de fácil produção via sintética e, mais recentemente, via processos envolvendo biomassas (Yaaini *et al.*, 2012; Bevilaqua *et al.*, 2013; Girisuta *et al.*, 2013). É um intermediário químico muito versátil com diversas aplicações de grande interesse industrial, tais como: na síntese de solventes; pesticidas; polímeros; poliésteres; na síntese de aditivos de combustíveis; na indústria farmacêutica e na indústria de tabaco para reduzir o teor de alcatrão nas frações de nicotina (Rackemann e Doherty, 2011). Ainda uma importante reação envolvendo o grupo carboxila do ácido levulínico é a esterificação, pois ésteres levulínicos têm também vasta aplicação sintética e industrial (www.dibanet.org; Fernandes *et al.*, 2012).

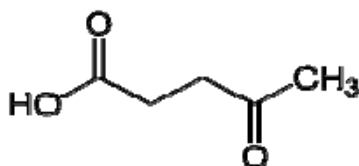


Figura 8. Estrutura química do ácido levulínico.

O ácido fórmico ou ácido metanóico (HCOOH), mostrado na Figura 9, é um subproduto da produção de ácido levulínico a partir da celulose. Ele é usado extensivamente como um descalcificante, como um agente acidulante no tingimento e acabamento de têxteis, e em curtimento de couro. É também utilizado para a preparação de ésteres orgânicos, corantes, inseticidas e refrigerantes. Além disso, os ésteres de ácido fórmico (por exemplo, formiato de metila e etila) têm valor como um componente de combustíveis e outros produtos químicos da indústria petroquímica (Hayes *et al.*, 2006).

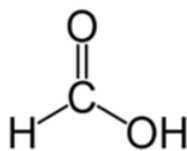


Figura 9. Estrutura química do ácido fórmico.

O furfural ($C_5H_4O_2$) é um composto orgânico heterocíclico aromático conhecido também como furaldeído (Figura 10). Pode ser utilizado como um solvente ou como material de partida para a preparação de outros solventes como álcool furfurílico, tetra-hidrofurano (THF) e ácido levulínico (Ribeiro *et al.*, 2012). Inúmeros produtos primários são obtidos por transformação direta do furfural.

É também muito utilizado na síntese de polímeros, em pastilhas para freios e no refino de óleos lubrificantes. O furfural é um produto químico orgânico produzido em larga escala, a partir de fontes de carboidratos. Atualmente o maior produtor de furfural no mundo é a China, sendo a biomassa residual de sabugo de milho o material mais utilizado para a produção desse químico (Zhang *et al.*, 2013).

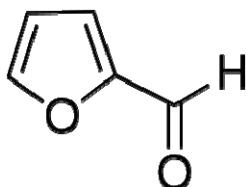


Figura 10. Estrutura química do furfural.

2.4 Espectroscopia no Infravermelho Próximo (NIR)

Infra no Latim significa, “abaixo”, uma vez que a radiação infravermelha apresenta energia mais baixa do que a parte visível (Vis) do espectro eletromagnético, tendo frequência menor que o vermelho do espectro eletromagnético. A região do infravermelho estende-se desde comprimentos de onda de 750 nm para comprimentos de onda de cerca de 1×10^6 nm. Essa faixa é dividida em radiação no infravermelho próximo - NIR (comprimentos de onda entre 750 nm - 2500 nm), médio - MID (comprimentos de onda entre 2500 nm e 10,000 nm) e distante - FAR (comprimentos de onda entre 10,000 nm e 1×10^6 nm).

Dentre essas, a região do MID foi a que recebeu maior atenção, dada a sua possibilidade de identificar grupos funcionais nas moléculas, permitindo assim obter uma

grande quantidade de informações. Por outro lado a região do NIR dava origem a bandas alargadas, superpostas e de fraca absorção, o que dificultava sua interpretação.

Independente da região utilizada, para que ocorra a absorção da radiação infravermelha é necessário que ocorra uma variação do momento de dipolo elétrico da molécula como consequência de seu movimento vibracional ou rotacional. Nessa condição, o campo elétrico da radiação eletromagnética pode interagir com aquele proveniente da vibração da molécula, aumentando a amplitude do movimento vibracional, originando os espectros de absorção (Lima *et al.*, 2009). Os dois principais tipos de movimento molecular, no escopo da espectroscopia na região do infravermelho, são as transições de energia vibracional e rotacional (Murray e Williams, 1987).

Consideremos o exemplo do cloreto de hidrogênio (HCl). Esse possui uma distribuição de carga assimétrica, uma vez que o cloro é mais eletronegativo do que o hidrogênio. Como consequência, o momento de dipolo da molécula varia durante a vibração, o que resulta em absorção de energia na região do infravermelho. Por outro lado, hidrogênio (H₂) e cloro (Cl₂) não absorvem energia nesta região do espectro, uma vez que o momento de dipolo destas moléculas não varia com a vibração.

As duas principais vibrações moleculares de interesse para a espectroscopia na região do infravermelho são os estiramentos e as deformações angulares. Quando a deformação ocorre na direção do eixo da molécula, podendo acontecer de forma simétrica ou assimétrica, a distância interatômica aumenta e diminui alternadamente, e o modo de vibração é denominado estiramento ou deformação axial. As vibrações de deformação angular correspondem à variação do ângulo entre duas ligações e recebem a denominação de deformação angular no plano e fora do plano da molécula.

A espectroscopia na região NIR tem se tornado uma das mais promissoras como técnica analítica rápida e de baixo custo. O baixo custo é justificado dada a aliança entre a espectroscopia NIR e a quimiometria, através de modelos de calibração multivariada, o que minimiza as análises por via úmida, uma vez já obtidos os dados de referência (laboratório). Essa grande atenção tem sido atribuída à possibilidade de se realizarem análises diretas, sem necessidade de pré-tratamento das amostras, de ser uma técnica não destrutiva e rápida, adequada para análises *online*, obtendo-se

resultados em intervalos de tempo da ordem de segundos. Dado o grande e rápido crescimento da espectroscopia na região NIR nas mais diversas áreas (agricultura, indústria petroquímica, indústria alimentícia, indústria farmacêutica, entre outros) é importante entender a fundamentação teórica básica dessa técnica (Pasquini, 2003).

O que ocorre durante a interação da energia (radiação eletromagnética na região do infravermelho próximo) com a matéria (amostra) é a conversão da energia da radiação incidente na amostra em energia vibracional das moléculas, resultando na absorção líquida de energia que pode ser detectada, promovendo movimentos que alteram os comprimentos e ângulos das ligações. Embora o fenômeno seja quântico, ele pode ser convenientemente representado por um modelo mecânico clássico conhecido como oscilador harmônico simples.

As características da vibração de estiramento podem ser representadas, a grosso modo, como um sistema massas-mola, o qual descreve a vibração de uma molécula consistindo de duas massas conectadas por uma mola. Uma perturbação na posição de equilíbrio das massas resulta em uma vibração denominada de estiramento em um movimento harmônico simples. Assim as massas representam os átomos e a mola representa a ligação química entre eles. Uma perturbação de uma das massas ao longo do eixo da mola resulta em uma vibração denominada de movimento harmônico simples e a força restauradora é proporcional ao deslocamento (lei de Hooke) e tende a restaurar as massas para sua posição original.

O modelo do oscilador harmônico descreve os níveis energéticos vibracionais como igualmente espaçados (Figura 11 (A)) e assim existe uma regra de seleção, na qual só são permitidas transições para $\Delta v = \pm 1$, onde v corresponde aos números quânticos vibracionais (Pasquini, 2003; Sala, 2008). No entanto, esse modelo funciona bem para moléculas diatômicas simples, enquanto que para moléculas poliatômicas ele falha, por não considerar que os elétrons sofrem influência de átomos ou grupos vizinhos, o que afeta a frequência da vibração das ligações (Pasquini, 2003). Embora o modelo do oscilador harmônico possa ajudar a compreender a espectroscopia vibracional, esse produz algumas restrições para espectroscopia NIR, uma vez que não permite transições onde Δv é maior do que 1. Transições com $\Delta v = 2$ ou superiores são proibidas pela quântica (modelo harmônico) e, desse modo, a maior parte dos

fenômenos observáveis na região NIR, como as bandas de sobretons, não deveriam existir. Além disso, as vibrações no modelo harmônico são independentes e suas combinações não existem de acordo com as restrições impostas pelo modelo. No entanto, tanto sobretons como bandas de combinação são observadas no NIR (Pasquini, 2003).

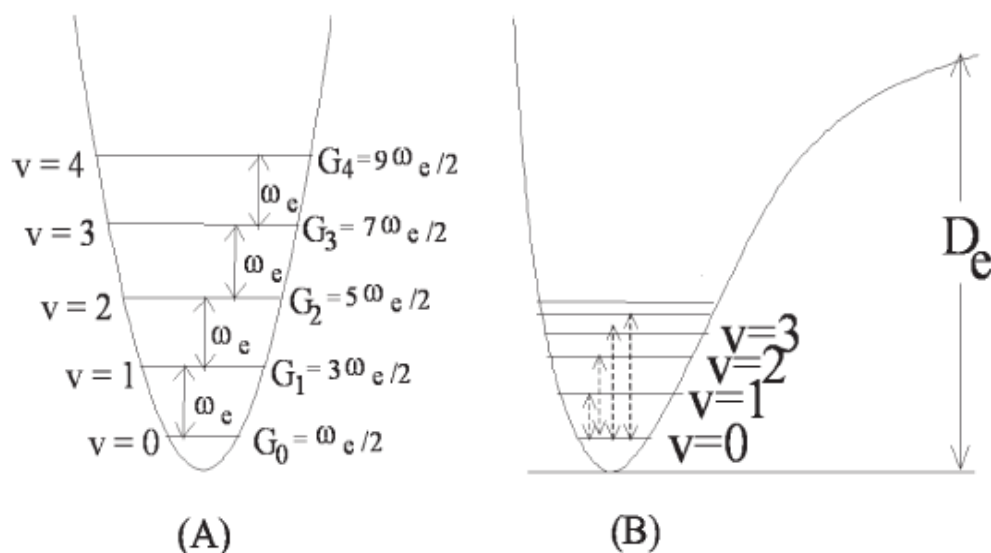


Figura 11. Diagrama da energia potencial, níveis vibracionais e transições para o oscilador: (A) harmônico e (B) anarmônico (Fonte: Sala, 2008).

Portanto o comportamento das moléculas pode ser melhor representado pelo modelo do oscilador anarmônico, no qual a separação entre os níveis não é igualmente espaçada, sendo que quando o número quântico v aumenta, a energia dos níveis diminui, como pode ser observado na Figura 11 (B). Neste modelo a regra de seleção muda para $\Delta v = \pm 2, \pm 3, \pm 4, \dots$, permitindo a observação de transições denominadas sobretons (Sala, 2008). Também são encontradas bandas de combinação entre as vibrações (Pasquini, 2003; Lima *et al.*, 2009).

Dessa forma as bandas de absorção no espectro na região do infravermelho próximo são associadas principalmente a sobretons e bandas de combinações de vibrações fundamentais de ligações N-H, C-H, O-H e S-H que ocorrem na região do infravermelho médio (Skoog *et al.*, 2006). Outros grupos funcionais importantes como C-O, grupos carbonila, C-N, C-C, entre outros também são observados.

Geralmente, as bandas de absorção no NIR apresentam intensidades muitas vezes inferiores às das do MID ou FAR (cerca de 10 a 100 vezes), devido ao fato de se tratarem de transições menos prováveis ($\Delta v > 1$ e bandas de combinação). Como a probabilidade de interação da amostra com a radiação incidente, no caso do NIR, é menor, isso pode garantir uma maior penetração dessa radiação, em relação ao MID, onde a superfície da amostra absorve quase que a totalidade da radiação incidente. No caso de amostras sólidas em medidas de reflectância difusa, a radiação pode chegar muitas vezes a atingir até 3 mm de profundidade.

O grande inconveniente associado ao NIR, que trata da difícil interpretação dos espectros, pode ser contornado tratando os dados espectrais com o uso de técnicas quimiométricas obtendo-se assim resultados precisos e confiáveis a partir das informações contidas nos espectros NIR (Rambo *et al.*, 2013). Ambos, quimiometria e a espectroscopia NIR evoluíram em uma espécie de simbiose, ou seja, uma relação mutuamente vantajosa, que permite obter mais informação da espectroscopia com possibilidade de gerar modelos de quantificação estendendo assim a sua aplicabilidade, enquanto novos desafios motivam a melhoria de muitas técnicas quimiométricas (Pasquini, 2003).

2.5 Quimiometria

A quimiometria pode ser definida como uma disciplina multidisciplinar introduzida na química e que emprega métodos matemáticos e estatísticos para planejar ou selecionar experimentos de forma otimizada e fornecer o máximo de informação química com a análise dos dados obtidos. Trata-se de uma disciplina da química, uma vez que os problemas que ela se propõe a resolver são de interesse e originados na química, ainda que as ferramentas de trabalho provenham principalmente da matemática, estatística e computação (Ferreira *et al.*, 1999; Cruz *et al.*, 2013). As ferramentas quimiométricas surgiram para auxiliar os químicos numa melhor interpretação dos dados, com um maior conhecimento a respeito da natureza dessas informações (Kowalski e Seasholtz, 1991).

Graças à quimiometria é possível a interpretação de dados que antes seriam impossíveis de serem analisados. Ela permite extrair informações relevantes de dados

multivariados complexos e de difícil interpretação, como o caso dos espectros NIR, citados anteriormente.

O crescimento da quimiometria se deu de tal forma que foram desenvolvidas novas ferramentas para tratamento de dados apresentando inúmeras aplicações conforme o objetivo do estudo, indo desde o planejamento e a otimização de experimentos, análises exploratórias, classificação de dados, até as determinações qualitativas e quantitativas.

2.5.1 Análise por Componentes Principais (PCA)

A análise por componentes principais (PCA) foi proposta em 1901 por Karl Pearson. Mais tarde Hotelling maximizando a variância dos dados originais incrementou a PCA. Há inúmeras excelentes descrições tratando a base matemática da PCA (Wold *et al.*, 1987; Brereton, 1992; Brereton, 1993) e este trabalho irá fornecer apenas uma visão geral.

A Associação Americana para Testes e Materiais (ASTM) define a PCA com um procedimento matemático de projeção que projeta a matriz de dados com muitas variáveis, em um espaço de dimensão menor, reduzindo a dimensionalidade do conjunto dos dados. Ou seja, um novo sistema de eixos é construído, que podem ser denominados fatores, componentes principais ou variáveis latentes, para representar a matriz de dados. Como resultado, as informações mais importantes e relevantes se tornam mais fáceis de serem interpretadas. Essas novas componentes são obtidas em ordem decrescente de variância, de forma que a primeira componente principal tem a maior variância possível (ou seja, representa o máximo da variabilidade nos dados). Cada componente seguinte, por sua vez, tem a máxima variância residual, não representada pelas componentes anteriores, sob a restrição de ser ortogonal aos componentes anteriores, dessa forma, as componentes principais são não-correlacionadas. O número máximo de componentes, que podem ser obtidos com essas condições, é menor do que ou igual ao número de variáveis ou de amostras, o que for menor (ASTM E131-10, 2010).

Matematicamente na PCA, as componentes principais são obtidas como uma combinação linear ortogonal que transforma os dados para um novo sistema de coordenadas, como mostra a equação 1¹:

$$\mathbf{X} = \mathbf{TP}^T \quad \text{Equação 1}$$

onde $\mathbf{X}$ é a matriz de dados; $\mathbf{T}$ é a matriz de escores que contém as informações sobre as amostras nas linhas; $\mathbf{P}$ é a matriz dos pesos (*loadings*), que contém informações sobre as variáveis nas colunas; e o sobrescrito T significa a transposta (Brereton, 2000).

Essa combinação linear nada mais é do que o agrupamento das variáveis altamente correlacionadas (ou com alta covariância), dando-lhes um maior peso, numa nova variável. Essa nova variável é a componente principal. Assim a porcentagem da variância total dos dados pode ser explicada em um número pequeno de componentes principais, reduzindo a dimensionalidade dos dados.

Geralmente, há uma enorme quantidade de informação que pode ser extraída a partir de um conjunto de dados após ter sido realizada a PCA. No entanto, nem sempre fornece um resultado de fácil interpretação, podendo ocorrer de algumas variáveis terem pesos elevados para mais de uma componente principal. Dessa forma diferentes alternativas são buscadas a fim de encontrar uma matriz mais facilmente interpretável. Isso pode ser feito com o auxílio de diferentes métodos de rotação. A varimax, uma rotação ortogonal, é o método mais usado e maximiza a soma dos quadrados entre os pesos de cada componente principal, resultando numa solução otimizada em que cada variável terá, na medida do possível, peso elevado para uma única componente principal (Kaiser, 1958).

A PCA tem sido muito utilizada, com inúmeras publicações na literatura, auxiliando no trabalho de análise exploratória dos dados, entre outras aplicações (Liu *et al.*, 2010; Cirlini *et al.*, 2011). Cheng *et al.*, (2013), Thyrel *et al.*, (2013) e Xiao *et al.*, (2013) conseguiram distinguir amostras de diferentes regiões geográficas, diferentes tipos de matérias-primas e diferentes fabricantes utilizando a PCA, demonstrando assim a habilidade dessa técnica.

¹ Neste trabalho se adotou a seguinte notação: letras maiúsculas e em negrito representam matrizes; letras em minúsculo e negrito representam vetores e letras não negritos são escalares.

2.5.2 Modelos de calibração

Dentre os métodos de calibração existentes, encontram-se os métodos de calibração univariada e os de calibração multivariada. Devido ao grande avanço nessa área, o uso de métodos de calibração multivariada tem se destacado nas determinações de uma série de parâmetros nas mais diversas áreas. Além disso esses métodos se tornaram uma alternativa quando os métodos univariados não podem ser aplicados, por exemplo, quando há colinearidade (Martens e Naes, 1996).

Calibração é o procedimento usado para modelar a correlação entre dados medidos ou calculados (por espectroscopia, por exemplo) ($\mathbf{X}$) e alguma propriedade (umidade, cinzas, atividade biológica) das amostras ($\mathbf{y}$).

Os espectros, um para cada amostra, são organizados numa matriz, $\mathbf{X}$ ($I \times J$), com I sendo o número de amostras e J o número de variáveis preditoras (no caso do NIR as variáveis são os comprimentos ou números de onda). A propriedade de interesse $\mathbf{y}$ ($I \times 1$), também conhecida como variável predita, é obtida experimentalmente por meio de padrões certificados ou algum método de referência. Obtidas as variáveis preditoras e a variável predita, o modelo apropriado será construído correlacionando $\mathbf{X}$ e $\mathbf{y}$ (Ferreira *et al.*, 1999).

O modelo de calibração é uma função matemática que relaciona a variável predita ($\mathbf{y}$) e as variáveis preditoras ($\mathbf{X}$), e tem por objetivo encontrar uma relação entre elas, representada pela equação 2:

$$\mathbf{y} = \mathbf{X} \mathbf{b} + \mathbf{e} \quad \text{Equação 2}$$

onde $\mathbf{e}$ é o vetor de resíduos do modelo, e $\mathbf{b}$ ($J \times 1$) é um vetor de constantes denominadas de coeficientes de regressão.

No entanto, visando verificar a qualidade do modelo para previsão de novas amostras, é prática corrente dividir o conjunto de amostras em dois subconjuntos, um de calibração ($\mathbf{X}_{\text{cal}}$) e outro de validação externa (conjunto teste - $\mathbf{X}_{\text{val}}$). Nesse caso ajusta-se o modelo (equação 2) utilizando apenas as amostras do conjunto de calibração. A próxima etapa é a validação, na qual as variáveis preditoras ($\mathbf{X}_{\text{val}}$), são utilizadas em conjunto com o vetor de regressão, $\mathbf{b}$, para calcular os valores previstos

para a variável predita, que serão comparados com os valores experimentais de y (Martens e Naes, 1996).

Os métodos multivariados de calibração mais utilizados, principalmente para fins quantitativos, tem sido a regressão linear múltipla (MLR), regressão por componentes principais (PCR) e regressão por quadrados mínimos parciais (PLS) (Geladi e Kowalski, 1986; Martens e Naes, 1996; Naes *et al.*, 2002; Berendt *et al.*, 2006; Rambo *et al.*, 2013). Esses são métodos lineares porque partem da premissa de efeitos aditivos entre as respostas e as concentrações dos componentes das amostras. Por exemplo, a lei de Lambert-Beer, onde, num determinado intervalo de concentração, há uma relação linear entre a concentração do analito e a absorbância.

PLS é o método de calibração mais utilizado em combinação com a espectroscopia NIR para quantificar propriedades de interesse em biomassas lignocelulósicas (Ye *et al.*, 2008; Xu *et al.*, 2013; Krapf *et al.*, 2013; Rambo *et al.*, 2013).

2.5.3 Regressão por quadrados mínimos parciais (PLS)

A regressão por quadrados mínimos parciais (PLS) é a técnica multivariada mais usada para construção de modelos usando dados espectrais (Geladi e Kowalski, 1986). PLS é um método de regressão linear que relaciona os dados espectrais X à propriedade de interesse y , como indicado na equação 3:

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \hat{b}_j x_{ji} \quad \text{Equação 3}$$

No método de regressão PLS, a covariância entre os escores e o vetor y é maximizada, ou seja, a informação de y é incorporada e se calcula a matriz de pesos (W) de dimensão $(J \times m)$ segundo a equação 4, sendo m o número de fatores selecionados, denominados de variáveis latentes. Um dos critérios para a determinação do número adequado de variáveis latentes é aquele no qual se obtêm o menor erro de previsão.

$$X = yw^T + E \quad \text{Equação 4}$$

onde o sobrescrito T significa a transposta e E a matriz de resíduos.

A seguir se obtém os escores (**T**) e loadings (**P**) de X:

$$\mathbf{T} = \mathbf{XW} \quad \text{Equação 5}$$

$$\mathbf{P} = \mathbf{X}^T \mathbf{T} (\mathbf{T}^T \mathbf{T})^{-1} \quad \text{Equação 6}$$

Com a matriz de escores **T**, calcula-se os loadings de **y** (**q**) e por fim as estimativas dos coeficientes de regressão (**b̂**):

$$\mathbf{q} = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{y} \quad \text{Equação 7}$$

$$\hat{\mathbf{b}} = \mathbf{W} (\mathbf{P}^T \mathbf{W})^{-1} \mathbf{q}^T \quad \text{Equação 8}$$

Inúmeros trabalhos são publicados a cada ano ressaltando a importante aliança entre a espectroscopia NIR e o método PLS na interpretação dos dados.

Por exemplo, a determinação rápida do teor de resina e borracha na biomassa de guaiúle (*Parthenium argentatum*), um arbusto do deserto da América Sententrional - por meio de espectroscopia no infravermelho próximo e modelos PLS foram avaliados por Suchat *et al.*, (2013). O bom desempenho das calibrações obtidas indica que os componentes da biomassa podem ser determinados com precisão utilizando NIR aplicada diretamente à biomassa em pó.

Hayes (2012) desenvolveu modelos utilizando a região do espectro do Vis-NIR em combinação com o método PLS para a predição quantitativa dos componentes lignocelulósicos de amostras úmidas não moídas de capim (*Miscanthus* sp.). Os melhores resultados foram obtidos na região do NIR. Os modelos para amostras úmidas de glicose, xilose e lignina Klason apresentaram precisões boas. Já os modelos úmidos para arabinose, lignina ácida solúvel, cinzas, extrativos, ramnose, e resíduo ácido insolúvel, mostraram valores de $R^2_{\text{cal,val}}$ menores, sendo, portanto úteis apenas na etapa de triagem.

A espectroscopia NIR juntamente com a análise multivariada (PLS), foi utilizada para prever as propriedades químicas e mecânicas de resíduos de madeira de pinheiro. As amostras foram selecionadas a partir de diferentes localizações radiais e alturas de três diferentes pinheiros. A composição química e as propriedades mecânicas foram medidas com as técnicas tradicionais de via úmida. E então essas propriedades foram correlacionadas com os espectros NIR usando modelos PLS. As correlações foram boas para a composição química, com os coeficientes de determinação geralmente

acima de 0,80. As propriedades mecânicas também foram previstas com $R^2_{\text{cal}} > 0,80$ utilizando o espectro reduzido, no intervalo de (650 nm a 1150 nm), o que permite medições em campo dessas propriedades usando espectrômetros NIR portáteis (Kelley *et al.*, 2004).

Um método rápido para analisar a composição química de matérias-primas renováveis foi desenvolvido usando a espectroscopia no infravermelho próximo com transformada de Fourier (FT-NIR), acoplado com a análise multivariada. O estudo apresentou um modelo de base ampla onde um único modelo preditivo pode ser desenvolvido para analisar diferentes tipos de matérias-primas e com diferentes frações botânicas. As duas fontes de biomassa foram a palha de milho e gramíneas, incluindo diversas frações de folhas, cascas, nós, internos e bainhas das plantas. Ambos os resultados de validação externa e de validação cruzada mostraram que o modelo amplo desenvolvido foi promissor para a previsão química futura de ambas as espécies de biomassa (Liu *et al.*, 2010).

2.5.4 Escolha do número de variáveis latentes

O número ideal de variáveis latentes é aquele que permite a obtenção de um modelo com boa capacidade de previsão de amostras externas. Um excesso de variáveis latentes leva a um sobreajuste modelando ruídos, o que poderá resultar em modelos com baixa capacidade preditiva para novas amostras, enquanto que um modelo com variáveis latentes de menos pode deixar informações importantes de fora.

É necessário encontrar o melhor número de variáveis latentes para os modelos construídos, o que normalmente pode ser feito usando diferentes procedimentos.

Um dos procedimentos mais comuns para essa finalidade é a chamada validação cruzada (Martens *et al.*, 1983; Martens e Naes, 1996). A validação cruzada é baseada na avaliação da magnitude dos erros de previsão comparando as concentrações das amostras do conjunto de calibração com as respectivas previsões quando as mesmas não participam na construção do modelo de regressão. A validação cruzada pode ser realizada em blocos, ou seja, remove-se n amostras do conjunto de calibração (*leave-n-out*) e constrói-se o modelo PLS e usa-se esse novo modelo para prever os valores de y das amostras removidas. Também a validação cruzada pode ser

feita quando uma única amostra é deixada de fora no processo de construção do modelo e a seguir y dessa amostra é prevista pelo modelo, processo conhecido como “*leave one out*” (LOO). Em ambos os casos, o processo é repetido até que todas as amostras tenham sido deixadas de fora do conjunto de calibração e previstas pelo respectivo modelo respectivo e a raiz quadrada do erro quadrático médio de validação cruzada (RMSECV) é calculada (Brereton, 2000), de acordo com a Tabela 1. Escolhe-se o número de variáveis latentes associadas ao menor valor de RMSECV e/ou quando esse já não se altera significativamente (Martens e Naes, 1989; Faber, 1999) como mostra a Figura 12. Nesse caso, por exemplo, o número de variáveis latentes escolhidas deverá ser 5.

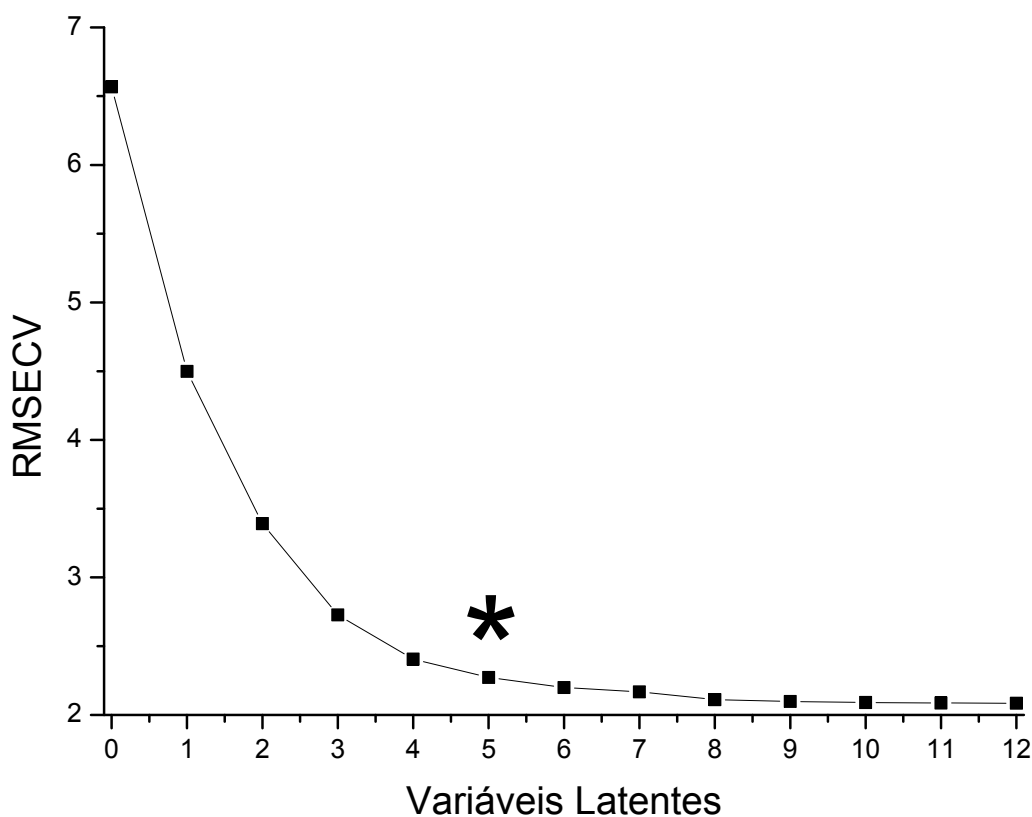


Figura 12. Gráfico para escolha correta do número de variáveis latentes. (* indica o número de variáveis latentes selecionado).

O conjunto de validação externa também pode ser usado para determinar o número de variáveis latentes, e é muitas vezes citado como a estimativa mais realista, particularmente do erro de previsão. No entanto requer uma grande quantidade de dados (Faber e Rajkó, 2007; Esbensen e Geladi, 2010).

No entanto as questões sobre como avaliar quantos componentes devem ser usados em um modelo PLS, e quão bom é o modelo, estão entre os temas mais discutidos em quimiometria (Faber e Rajkó, 2007; Wiklund *et al.*, 2007; Esbensen e Geladi, 2010). Nem todos quimiometristas acreditam que a validação cruzada seja o método mais eficiente para estimar o número ideal de componentes de um modelo (Esbensen e Geladi, 2010).

O que importa mesmo, é que o número de variáveis latentes identificados como o ideal, tanto na validação cruzada (LNO ou LOO) como na validação externa seja o mesmo (Figura 13).

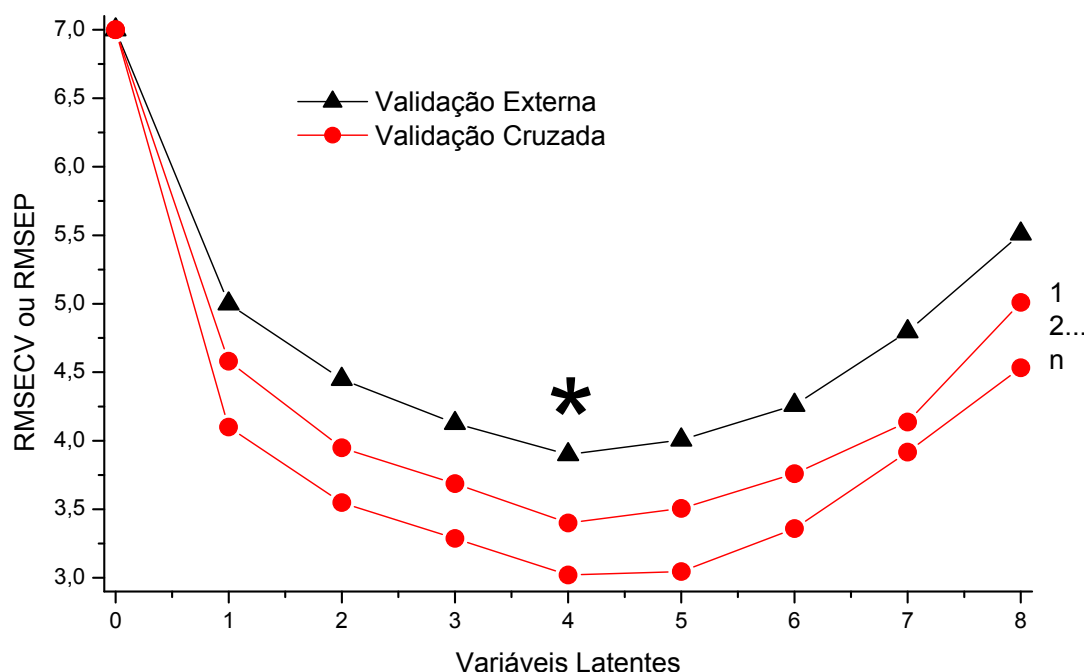


Figura 13. Relações entre as estimativas de RMSECV e RMSEP em função do método de validação (* indica o número de variáveis latentes selecionado), sendo $n=1$ LOO e $n>1$ LNO. (Adaptado: Esbensen e Geladi, 2010).

Como conclui Faber e Rakjó (2007), cada abordagem de validação tem vantagens, mas também graves desvantagens. É difícil uma escolha correta diante de tantas discussões a respeito do assunto.

2.5.5 Pré-tratamento dos dados espectrais

Os dados experimentais são preparados para análise, removendo matematicamente fontes de variação indesejáveis que não serão removidas naturalmente durante a análise dos dados. Podem haver fontes de variação sistemática ou aleatória que mascarem uma variação de interesse e, assim, uma técnica de pré-processamento deve ser empregada.

Deve-se tomar cuidado, pois um pré-processamento inadequado pode remover informação útil e, por isso, um re-exame deve sempre ser realizado após o pré-processamento. E melhor do que isso, ao invés de se fazer por "tentativa e erro", é conhecer o que cada pré-processamento faz com os dados e quais as variações são relevantes e quais devem ser removidas.

Diversos pré-tratamentos podem ser aplicados nos espectros antes do desenvolvimento dos modelos de calibração, sendo que o método adequado irá depender do tipo de variação presente nos dados espectrais (Martens e Naes, 1996). As transformações são aplicadas às amostras (orientadas nas linhas), enquanto que os pré-processamentos são aplicados às variáveis (orientadas nas colunas). Técnicas de alisamento, normalização, correção da linha de base e ponderação podem ser aplicados às amostras. As técnicas de pré-processamento nas variáveis consistem basicamente em centrar os dados na média ou autoescalar. Serão discutidos a seguir apenas os pré-tratamentos utilizados neste trabalho. E os resultados desses pré-tratamentos são todos mostrados na seção Resultados e Discussão.

2.5.5.1 Centragem dos dados na média

Calcula-se o valor médio para cada variável e subtrai-se esse valor para cada elemento de uma variável. Dessa forma cada variável passa a ter média zero movendo-se as coordenadas para o centro dos dados. Com isso os dados de entrada serão os desvios em relação à média, de forma que se elimina esse parâmetro dos modelos (a média) (Boysworth e Booksh, 2001).

2.5.5.2 Padronização normal de sinal (SNV)

A Padronização normal de sinal (SNV) é uma transformação geralmente aplicada a dados espectroscópicos, para remover os efeitos de dispersão, centrando e escalando cada espectro individual. A transformação por SNV é mostrada na equação 9 abaixo (Naes *et al.*, 2007):

$$x_{ij}^{SNV} = (x_{ij} - m_i) / s_i \quad \text{Equação 9}$$

onde m_i é a média das j variáveis para a amostra i (média do espectro i), e s_i é o seu desvio padrão. O efeito do SNV é que cada espectro é centrado em zero na escala vertical, apresenta desvio padrão unitário e varia (mínimo e máximo) de cerca de - 2 a + 2. Considera-se que esse tratamento pode muitas vezes ser mais adequado quando o conjunto de amostras compreende espectros diversos, ou quando o tamanho do conjunto é pequeno, visto que ele independe do conjunto de dados, sendo feito individualmente para cada espectro (Esbensen, 2004).

2.5.5.3 Derivadas

Esse método é muito útil em casos onde a linha de base é difícil de ser identificada, uma vez que as derivadas são capazes de remover efeitos sobre a linha de base. Ao aplicarmos as operações de derivação - primeira (1D) ou segunda derivada (2D) aos espectros brutos, as informações contidas ao longo dos diferentes comprimentos de onda são geralmente acentuadas (a variação da inclinação de bandas espectrais é mais evidente que variações na sua intensidade, resultando em bandas mais estreitas, porém menos intensas) e se tem uma aparente melhora na "resolução" espectral (Figura 14A–C). Isso pode contribuir para resolver problemas com bandas sobrepostas, enfatizando pequenas variações espectrais não evidentes nos dados brutos (Esbensen, 2002). Porém, os ruídos tornam-se acentuados em relação aos sinais espectrais, piorando a relação sinal/ruído (Figura 14D-E), portanto, deve-se ter cuidado com a qualidade dos espectros com os quais se deseja aplicar derivadas (Naes *et al.*, 2007).

A 1D de um espectro é simplesmente uma medida da inclinação da reta tangente à curva. A inclinação não é afetada por deslocamentos verticais (*offset*) e, portanto, a

primeira derivada é um método muito eficaz para a remoção de tais deslocamentos. Enquanto que a linha de base linear (espectro com uma inclinação linear) é convertida em um offset no espectro de 1D (Figura 14E). No entanto, o ponto máximo das bandas espectrais normalmente passa por zero nos espectros de primeira derivada, visto que a inclinação da reta tangente à curva nessa região é zero, o que pode dificultar a interpretação (Figura 14B). Já a 2D é uma medida da mudança na inclinação infinitesimal da curva. Além de remover o efeito aditivo (*offset*), também não é afetada por qualquer inclinação linear da linha de base como um todo, que possam existir nos dados, e é, portanto, um método muito eficaz para a remoção de ambos. A 2D pode ajudar a resolver picos próximos e aguçar características espectrais. Entretanto os máximos das bandas agora passam por um mínimo (Figura 14C). A Figura 14 demonstra os diversos efeitos da aplicação das derivadas em um espectro simulado formado por duas bandas gaussianas parcialmente sobrepostas.

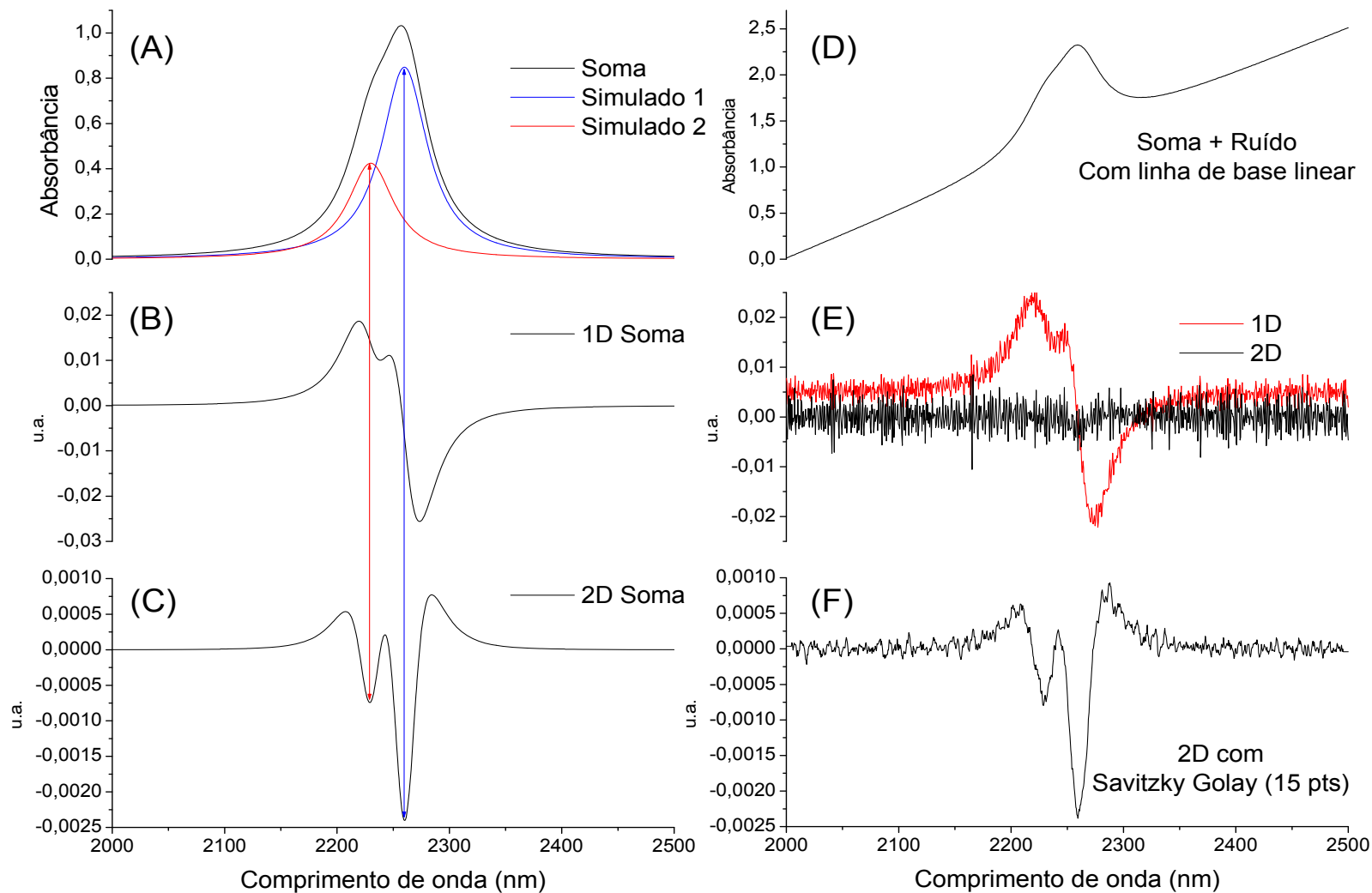


Figura 14. (A) Espectro bruto, (B) espectro após submetido à primeira derivada, (C) espectro após submetido à segunda derivada, (D) espectro com ruído e linha de base linear, (E) espectro com ruído após submetido à primeira e segunda derivada e (F) espectro com ruído após submetido à segunda derivada por Savitzky-Golay.

Como já mencionado, a derivação provoca a piora da relação sinal/ruído, uma alternativa é utilizar métodos de alisamento, uma dessas, e a que foi empregada neste trabalho, é calcular os espectros derivados usando o tratamento Savitzky Golay (Savitzky e Golay, 1964; Naes *et al.*, 2007). No filtro de alisamento (suavização) de Savitzky Golay um segmento do espectro (janela) é escolhido e um polinômio é ajustado a esses pontos usando quadrados mínimos. O valor experimental central do segmento é substituído pelo valor ajustado pelo polinômio e esse processo é repetido ponto a ponto com o deslocamento do segmento. Quando combinado com derivadas, coeficientes apropriados para cada ordem de derivação são empregados. A Figura 14F mostra claramente a melhora do espectro derivado (2D) quando submetido ao tratamento por Savitzky Golay, no qual bandas que não eram observadas, confundidas com o ruído, passam agora a ser identificadas. O tamanho do segmento e a ordem do polinômio são importantes, uma vez que se o segmento for muito grande, pode suavizar demasiadamente o espectro descartando informação importante, assim como distorcer (alargar) as linhas e se for muito baixo e/ou com ordem polinomial elevada, a vantagem do filtro é perdida. Deve-se encontrar a melhor combinação de tamanho da janela, a ordem polinomial, e a ordem da derivada (Naes *et al.*, 2003). Segundo Madden (1978), os melhores parâmetros dependem do objetivo, visto que, para o mínimo alargamento da linha, sem perda de resolução, um segmento de 0,7 da largura a meia altura da linha mais estreita é o mais indicado (Edwards e Willson, 1974), enquanto que o melhor ganho na relação sinal/ruído é obtido com um segmento igual ao dobro da largura a meia altura da linha mais estreita (Enke e Nieman, 1976).

2.5.5.4 Detrend (DT)

Detrend é uma técnica de correção da linha de base que utiliza polinômios e pode ser usada para remover tendências lineares e de maior ordem nos espectros (Mobley *et al.*, 1996). Ela faz isso através do ajuste, por quadrados mínimos, de uma função polinomial para cada espectro, função essa (linha de base) que será subtraída de cada espectro.

Um polinômio de ordem zero irá remover o offset enquanto um polinômio de primeira ordem irá remover o offset e a inclinação, e um polinômio de segunda ordem

irá remover o offset, inclinação e a curvatura parabólica. O procedimento é geralmente utilizado em conjunto com o pré-tratamento SNV (Barnes *et al.*, 1989) desde que os dados corrigidos por SNV possam ainda ser afetados por efeitos de curvatura. É importante notar que esse processo irá remover deslocamento de linha de base e curvatura sem alterar excessivamente a forma do espectro, o que pode ser uma vantagem em relação aos espectros derivados quando se tenta interpretar os espectros resultantes (Krasznai *et al.*, 2012).

2.5.5.5 Correção ortogonal do sinal (OSC)

A maioria dos espectros NIR são geralmente pré-processados por derivadas ou outros pré-processamentos. No entanto, esses pré-tratamentos podem remover também informação do espectro ligada à variável predita (**y**). A correção por OSC assegura que a informação removida (filtrada) não esteja relacionada a **y** (Sjöblom *et al.* 1998; Fearn, 2000; Gabrielsson e Trygg, 2006).

Muitas vezes os espectros NIR contêm variação sistemática que não está relacionada com as respostas **y**. Por exemplo, para amostras sólidas essa variação sistemática é devida, entre outras, ao espalhamento de luz e diferenças espectroscópicas no caminho ótico, além de que o analito de interesse pode absorver apenas em pequenas partes da região espectral. Nesses casos as primeiras variáveis latentes explicam bastante da variação em **X** e pouca em **y** (Wold *et al.*, 1998).

A fim de remover essa variação sistemática indesejável nos dados, é possível fazer a sua modelagem matemática e subtraí-la dos espectros, por exemplo, utilizando a correção OSC. Essa técnica remove do espectro (matriz **X**) apenas a parte que definitivamente não é relacionada ao vetor **y**. No caso do OSC isso é feito com a garantia de que a parte removida é matematicamente ortogonal a **y** (Trygg e Wold, 2003).

No presente trabalho utilizou-se o algoritmo de Fearn (2000), que consiste em calcular, a partir da matriz de espectros **X** e do vetor **y**, a matriz **M_⊥**, que é ortogonal à **y**, de acordo à equação 10:

$$\mathbf{M}_{\perp} = \mathbf{I} - \mathbf{X}^T \mathbf{y} (\mathbf{y}^T \mathbf{X} \mathbf{X}^T \mathbf{y})^{-1} \mathbf{y}^T \mathbf{X} \quad \text{Equação 10}$$

onde **M_⊥** é a matriz ortogonal a **y** e **I** é a matriz identidade.

A seguir os pesos e escores das componentes ortogonais a $\mathbf{y}$ ($\mathbf{P}_\perp$ e $\mathbf{T}_\perp$) são obtidos pelo cálculo das componentes principais de $\mathbf{M}_\perp \mathbf{X}^T \mathbf{X}$ e a matriz OSC é obtida por $\mathbf{X}_{\text{OSC}} = \mathbf{X} - \mathbf{T} \mathbf{P}^T$.

Há diversos outros algoritmos para a obtenção dos espectros corrigidos OSC (vide Westerhuis *et al.*, 2001), mas em essência o princípio é o mesmo: encontrar a solução que ortogonaliza $\mathbf{X}$ em $\mathbf{y}$, e remover esses componentes da matriz $\mathbf{X}$ antes da construção do modelo PLS.

Esse procedimento tem como vantagem a obtenção de modelos PLS com menor número de variáveis latentes, visto que a variância sistemática indesejável de $\mathbf{X}$ é removida anteriormente ao ajuste do modelo PLS e não se elimina variância em $\mathbf{X}$ relacionada a $\mathbf{y}$. Porém, é comum um baixo desempenho desses modelos em predições futuras, provavelmente por envolver o emprego de dois modelos em sequência (OSC seguido por PLS) ajustados por métodos de estimativa tendenciosos, onde pequenas variações nos espectros das novas amostras pode ter uma grande influência no resultado final. Por isso é fundamental um conjunto de amostras para a validação externa, inclusive da transformação OSC (Wold *et al.*, 1998).

2.6 FIGURAS DE MÉRITO

Um resultado sem uma declaração de confiabilidade não pode ser publicado uma vez que não é (ainda) um resultado (Bièvre, 1997). A validação desses resultados pode ser atestada através da determinação de parâmetros conhecidos como figuras de mérito, que certificam se o método utilizado apresenta um desempenho adequado nas condições em que ele será aplicado.

Na calibração univariada a aplicação das figuras de mérito como processo de validação já está bem estabelecida. Já na calibração multivariada a validação pode ser considerada relativamente complexa (Valderrama *et al.*, 2009).

Com relação à aplicação das figuras de mérito, não há diferença conceitual entre os métodos de calibração univariada e multivariada. Consequentemente, os métodos multivariados também devem incluir as correspondentes figuras de mérito (Olivieri *et al.*, 2006), uma vez que o processo de validação deve ser realizado sempre que um procedimento analítico é proposto ou desenvolvido.

A diferença é que para os métodos univariados, a União Internacional de Química Pura e Aplicada (IUPAC) introduziu um conjunto bem definido de regras para o cálculo das figuras de mérito. Para a calibração multivariada, a IUPAC recomenda a utilização da teoria do sinal analítico líquido (NAS) para calcular as figuras de mérito, uma vez que a sua estimativa não é facilmente realizada através dos métodos tradicionais (Olivieri *et al.*, 2006). O vetor NAS permite a avaliação das figuras de mérito em calibração multivariada em uma abordagem semelhante aos métodos univariados (Silva *et al.*, 2012; Ferreira *et al.*, 2013; Rocha *et al.*, 2012).

Segundo Morgan (1977) e Lorber (1986) o sinal analítico líquido é definido como a parte do sinal medido que é ortogonal aos interferentes. Lorber introduziu o NAS como uma ferramenta para obter as figuras de mérito. Em sua definição, as figuras de mérito são relacionadas a um modelo específico, permitindo a comparação entre diferentes modelos. Posteriormente uma nova abordagem surgiu em que se buscava estimar o “verdadeiro” sinal analítico líquido (se é que ele existe), dessa forma ele seria independente do modelo de calibração e seria definido apenas pelo conjunto de dados, e comparações não seriam possíveis (Bro e Andersen, 2003). Embora essa abordagem tenha sua utilidade, ela não faz parte do escopo deste trabalho e adotou-se a proposta de Bro e Andersen (2003), que retornou aos conceitos originais de Lorber (1986), avaliando modelos de calibração específicos, ou seja, buscando um sinal analítico líquido que descreve especificamente cada modelo de calibração, possibilitando comparações entre modelos competitivos.

A idéia do método para o cálculo do NAS é de separar as contribuições dos dados da matriz $\mathbf{X}$ em duas partes ortogonais entre si (Figura 15). Uma parte que corresponde ao analito de interesse ($\mathbf{X}_k$) e outra correspondente aos interferentes ($\mathbf{X}_{-k}$).

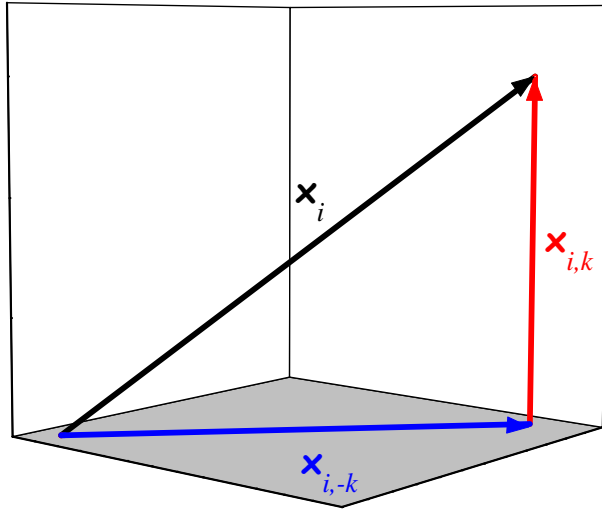


Figura 15. Representação geométrica do vetor NAS para a amostra i (Adaptado: Faber, 1998).

Dessa forma o vetor NAS para uma determinada amostra, $\mathbf{x}_i$, é calculado como a projeção dos dados em direção ao espaço gerado pelo vetor de regressão ($\mathbf{b}$) como mostra a equação 11:

$$\hat{\mathbf{x}}_{k,i}^{\text{nas}} = \mathbf{b}(\mathbf{b}^T \mathbf{b})^{-1} \mathbf{b}^T \mathbf{x}_i \quad \text{Equação 11}$$

onde $\hat{\mathbf{x}}_{k,i}^{\text{nas}}$ é o vetor NAS estimado para o analito de interesse k e para a amostra i , a partir do vetor dos coeficientes de regressão $\mathbf{b}$. A seguir, pode-se calcular o análogo do NAS univariado, que é um escalar, como a norma euclidiana do vetor NAS (Lorber, 1986):

$$\text{n\hat{a}s}_{k,i} = \|\hat{\mathbf{x}}_{k,i}^{\text{nas}}\| \quad \text{Equação 12}$$

Entretanto, quando os dados são centrados na média, para a construção do modelo de calibração multivariada, o $\text{n\hat{a}s}$ precisa ser corrigido pelo sinal do $\hat{y}$ para evitar o erro de sinal que é introduzido pelo uso da norma Euclidiana (Faber, 1998). Na verdade isso é válido também para os casos em que os dados não são centrados na média, mas que a predição resulte em valores negativos (Bro e Andersen, 2003) e assim esses autores propuseram outra forma de calcular o escalar $\text{n\hat{a}s}$ que contorna esse problema:

$$\hat{n}_{k,i} = \frac{\hat{\mathbf{x}}_{k,i}^{\text{nasT}} \mathbf{b}}{\|\mathbf{b}\|}$$

Com isso o modelo de regressão é obtido, de forma análoga à calibração univariada, ou seja, utiliza-se o **nâs** para um ajuste $\mathbf{y} = f(\mathbf{nâs})$, por quadrados mínimos ordinários, e a partir desse modelo pseudounivariado, é possível quantificar as figuras de mérito para modelos específicos.

A determinação das figuras de mérito em modelos de calibração multivariada como, a exatidão, precisão, robustez, ajuste e viés são estimadas de maneira bastante similar aos métodos de calibração univariada. Por outro lado, a estimativa de parâmetros como linearidade, sensibilidade, razão sinal/ruído e seletividade difere um pouco dos modelos univariados (Valderrama *et al.*, 2009). As equações usadas para o cálculo das figuras de mérito estão representadas na Tabela 1.

A exatidão expressa o grau de concordância entre o valor de referência (verdadeiro) e o valor estimado pelo modelo multivariado. É geralmente expressa como sendo a raiz quadrada do erro quadrático médio de calibração (RMSEC), a raiz quadrada do erro quadrático médio de validação cruzada (RMSECV), e a raiz quadrada do erro quadrático médio de previsão (RMSEP), de acordo com as equações da Tabela 1. Já a precisão (diferença entre várias medidas) é expressa pelo erro padrão da previsão (SEP). Uma vez que o SEP inclui uma estimativa do viés, um grau de liberdade é perdido, assim $I-1$ é usado no denominador ao invés de I (Tabela 1). O valor do viés é incluído nessa equação para assegurar resultados corretos e evitar uma capacidade preditiva excessivamente otimista dos modelos (Naes *et al.*, 2007).

A seletividade (SEL) pode ser definida como a parte do sinal medida unicamente devida ao analito de interesse. É a medida do grau de sobreposição entre o sinal da espécie de interesse e os interferentes presentes na amostra indicando, a parte do sinal que não é perdida por essa sobreposição. A sensibilidade (SEN) mede as mudanças na resposta, pelo acréscimo unitário na concentração de um analito particular.

A sensibilidade analítica (γ), normalmente expressa como o inverso desse parâmetro γ^{-1} representa a menor diferença de concentração entre amostras, que pode ser distinguida pelo método.

Os limites de detecção (LD) e de quantificação (LQ) expressam respectivamente, as menores quantidades da espécie de interesse que podem ser detectadas e determinadas quantitativamente.

A linearidade pode ser verificada qualitativamente pela análise dos gráficos de resíduos e/ou também pelos gráficos dos valores previstos versus os valores de referência. Esses gráficos devem apresentar comportamento aleatório no primeiro caso e uma distribuição linear no segundo gráfico.

A significância estatística do viés é verificado através de um teste t, usado para determinar se os modelos de calibração multivariada apresentam erros sistemáticos (viés) estatisticamente significativos. Inicialmente é calculado um viés para o conjunto de validação, e em seguida um valor de $t_{\text{viés}}$ é calculado (Tabela 1). Caso o valor de $t_{\text{viés}}$ apresentar resultado maior do que o valor tabelado de t crítico para $I-1$ graus de liberdade (sendo I o número de amostras do conjunto de validação) a 95% de confiança, isso significa que os erros sistemáticos presentes no modelo multivariado são significativos e não devem ser negligenciados. Ou seja, a validade do modelo multivariado é suspeita. Ocorrendo o contrário, os erros sistemáticos incluídos nos modelos podem ser desprezados (ASTM E1655-00).

Tabela 1. Equações para o cálculo das figuras de mérito.

SEL (média) (Short <i>et al.</i>, 2007)	$SEL_k = \frac{\sum_{i=1}^I \frac{ \hat{n}as_{k,i} }{\ \mathbf{x}_i\ }}{I}$ onde $$ é o módulo do escalar $\hat{n}as_{k,i}$; e $$ é a norma euclidiana do vetor $\mathbf{x}_i$ que representa as respostas instrumentais da amostra i; e I é o número de amostras.
SEN (Ferré <i>et al.</i>, 2001; Olivieri <i>et al.</i>, 2006; Short <i>et al.</i>, 2007)	$SEN_{k,i} = \frac{\hat{x}_{k,i}^{nas}}{y_{k,i}}$ onde $y_{k,i}$ é o valor de referência do analito k da amostra i. $SEN_k = \ \mathbf{SEN}_{k,i}\ $ e SEN_k é a norma euclidiana de $SEN_{k,i}$

γ (Short <i>et al.</i> , 2007)	$\gamma = \frac{SEN_k}{\delta x}$ onde δx é uma estimativa para o ruído instrumental.
LD (Rocha <i>et al.</i> , 2012)	$LD = 3,3 \delta x \frac{1}{SEN_k}$
LQ (Rocha <i>et al.</i> , 2012)	$LQ = 10 \delta x \frac{1}{SEN_k}$
	Figuras de mérito para expressar a precisão e a exatidão
SEP (Esbensen, 2004; ASTM:E1655–05)	$SEP = \sqrt{\sum_i^I \frac{(y_i - \hat{y}_i - \text{viés})^2}{I-1}}$ sendo o viés definido como: $VIÉS = \frac{\sum_{i=1}^I (y_i - \hat{y}_i)}{I}$ onde $\hat{y}_i$ e y_i são os valores estimados e os valores de referência da propriedade de interesse, respectivamente. E $t_{\text{viés}}$ é calculado da seguinte forma: $t_{\text{viés}} = \frac{ \text{viés} \sqrt{I}}{SEP}$
RMSEC (Esbensen, 2004)	$RMSEC = \sqrt{\sum_{i=1}^I \frac{(y_i - \hat{y}_i)^2}{I-m-1}}$ onde m é o número de variáveis latentes do modelo. A subtração de m na equação é uma tentativa de corrigir o sobreajuste que pode ocorrer quando um excesso de variáveis latentes são utilizadas nos modelos.
RMSECV (Esbensen, 2004; Esbensen e Geladi, 2010)	$RMSECV = \sqrt{\sum_{i=1}^I \frac{(y_i - \hat{y}_i)^2}{I}}$
RMSEP (Esbensen, 2004)	$RMSEP = \sqrt{\sum_{i=1}^I \frac{(y_i - \hat{y}_i)^2}{I}}$

3 PARTE EXPERIMENTAL

3.1 Seleção e coleta de amostras

Dez diferentes biomassas correspondentes a fração úmida e não moída foram devidamente coletadas de diversas regiões do Brasil no período de março 2010 - março 2012. Essas amostras foram mantidas congeladas em sacos plásticos até as etapas seguintes do processo. Após secas e moídas foram armazenadas em recipientes herméticos e então caracterizadas com o intuito de selecionar as biomassas mais promissoras dentro do contexto de biorrefinaria e do projeto DIBANET. As matérias-primas avaliadas foram às seguintes: caroço de açaí, casca de café, serragem de eucalipto, casca de arroz, casca de soja, casca de coco, capim elefante, bambu, engaço e pseudocaule de banana.

Os resíduos de banana abrangeram diferentes frações botânicas como mostra a Figura 16 (pseudocaule, engaço, folha, e menor número de rizoma e ráquis) coletadas durante o período de 2010-2011 em diferentes regiões do Brasil: 54 dessas são oriundas das regiões Norte e Nordeste, enquanto que 50 procederam do Sul e Sudeste. O conjunto de amostras incluiu banana prata, maçã, nanica, ouro, terra, entre outras.

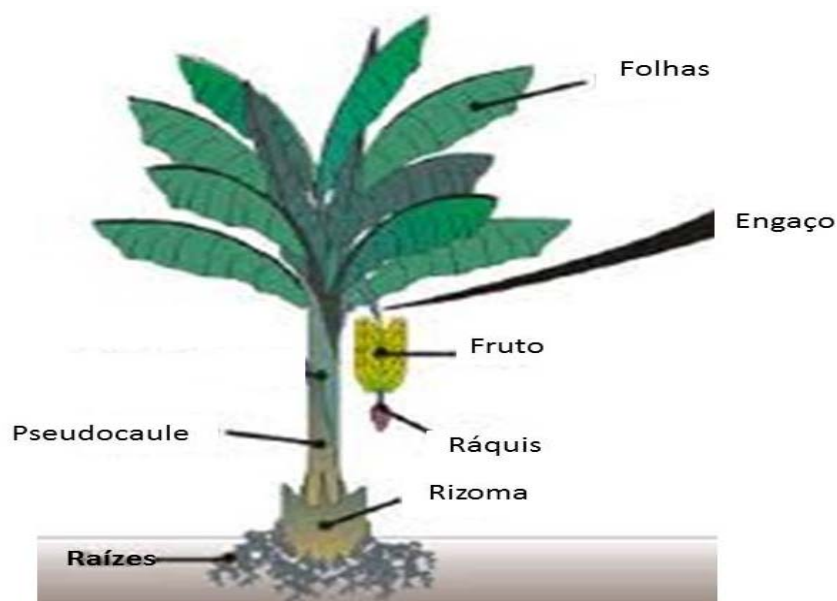


Figura 16. Frações botânicas da bananeira.

As 102 amostras de café (cascas e folhas) foram coletadas em 2010 e todas são oriundas da região Sudeste do Brasil e incluíram espécies de café arábica e robusta de diversos cultivares (mundo novo, caturra, catuaí, conilon, e outros). A Figura 17 mostra o fruto de café, indicando a fração botânica (cascas) usadas no trabalho.

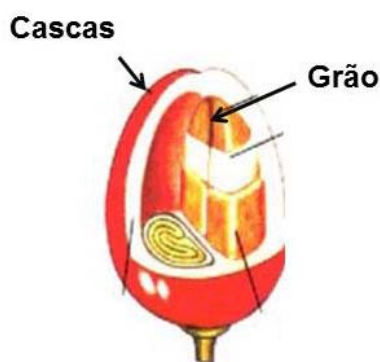


Figura 17. Representação das cascas de café.

As amostras de coco foram todas procedentes do Norte e Nordeste do país, divididas entre 15 amostras de casca e 15 amostras de fibras de coco, como mostra a Figura 18. Na maioria das vezes não foi possível identificar a variedade dos cocos, visto que elas foram obtidas do comércio.

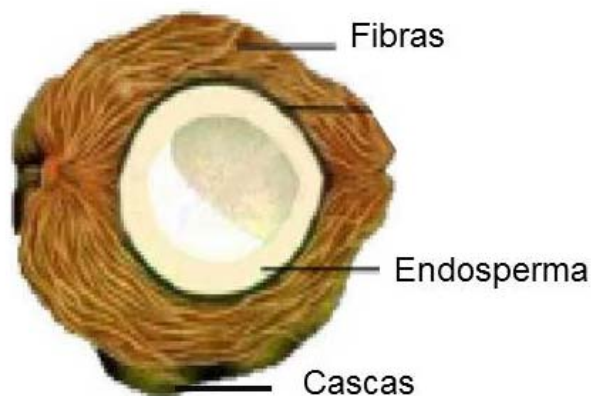


Figura 18. Frações botânicas do coco.

A identificação das amostras de banana, café e coco, com relação à variedade e a fração, bem como o local e o ano de coleta é mostrada na Tabela 2.

Tabela 2. Identificação, origem e ano de colheita das amostras de banana, café e coco.

Banana									
Variedade/Fração	Pseudoc	Engaço	Folha	Rizoma	Ráq	Casca	Total	Região	Ano
Maça	6	14	2	5		1		NE	2010
Nanica	2	4						SE	2011
Ouro	2	3				1		SE	2010
Prata	2	4						NE	2011
Caturra	1	2	2		1			S	2010
Bucaneiro	1	1						NE	2012
Calipso	1	1						NE	2011
Prataanã	1	1						NE	2011
Grande Neine	1	1						NE	2011
Terra	2	1	2					NE	2010
Pacovam	1	1						NE	2011
Preciosa	1	1						NE	2011
Maravilha	2	1						NE	2012
*	17	13			2			SE	2011
Total	40	48	6	5	3	2	104		
Café									
Variedade/Fração	Cascas	Folhas							
<i>C. arabica</i>	50	6						SE	2011
<i>C. canephora</i>	46							SE	2011
							102		
Coco									
Variedade/Fração	Fibras	Cascas							
*	2	2						N	2011
*	13	13					30	NE	2012

Pseudoc: pseudocaule; Ráq: Ráquis; NE: nordeste; SE: Sudeste; N: Norte; S: Sul*: não identificada

Todas as amostras coletadas foram secas em estufa (315 Se, Fanem) a 105 $\pm$ 5 °C por um tempo mínimo de 12 horas. Experimentos anteriores demonstraram que esse é o tempo suficiente para se atingir massa constante. Em seguida as amostras foram moídas em um moinho tipo faca (MA 920, Marconi) e classificadas em agitador de peneiras automático (VP-01, Bertel) para análises granulométricas e para redução do tamanho de partícula. Foram utilizadas as frações secas e moídas com partículas de diâmetro entre 180 μ m e 850 μ m (fração conhecida como DS, do inglês *Dried and Sieved*) para os trabalhos de investigação pelos métodos de referência (Figura 19).

Também as amostras não moídas secas conhecidas como - DU (*Dried and Unground*) e não moídas úmidas - WU (*Wet and Unground*) foram utilizadas a fim de avaliar o potencial de análises *online* em biorrefinarias, sem a necessidade de preparo de amostra.



Figura 19. Amostras da fração DS utilizadas para as análises de referência (A) cascas de soja, (B) cascas de arroz, (C) bambu, (D) fibras de coco, (E) engaço de banana (F) serragem de eucalipto, (G) capim elefante, (H) caule de banana, (I) caroços de açai e (J) cascas de café.

3.2 Análises

O diagrama a seguir refere-se à parte experimental desenvolvida neste trabalho.

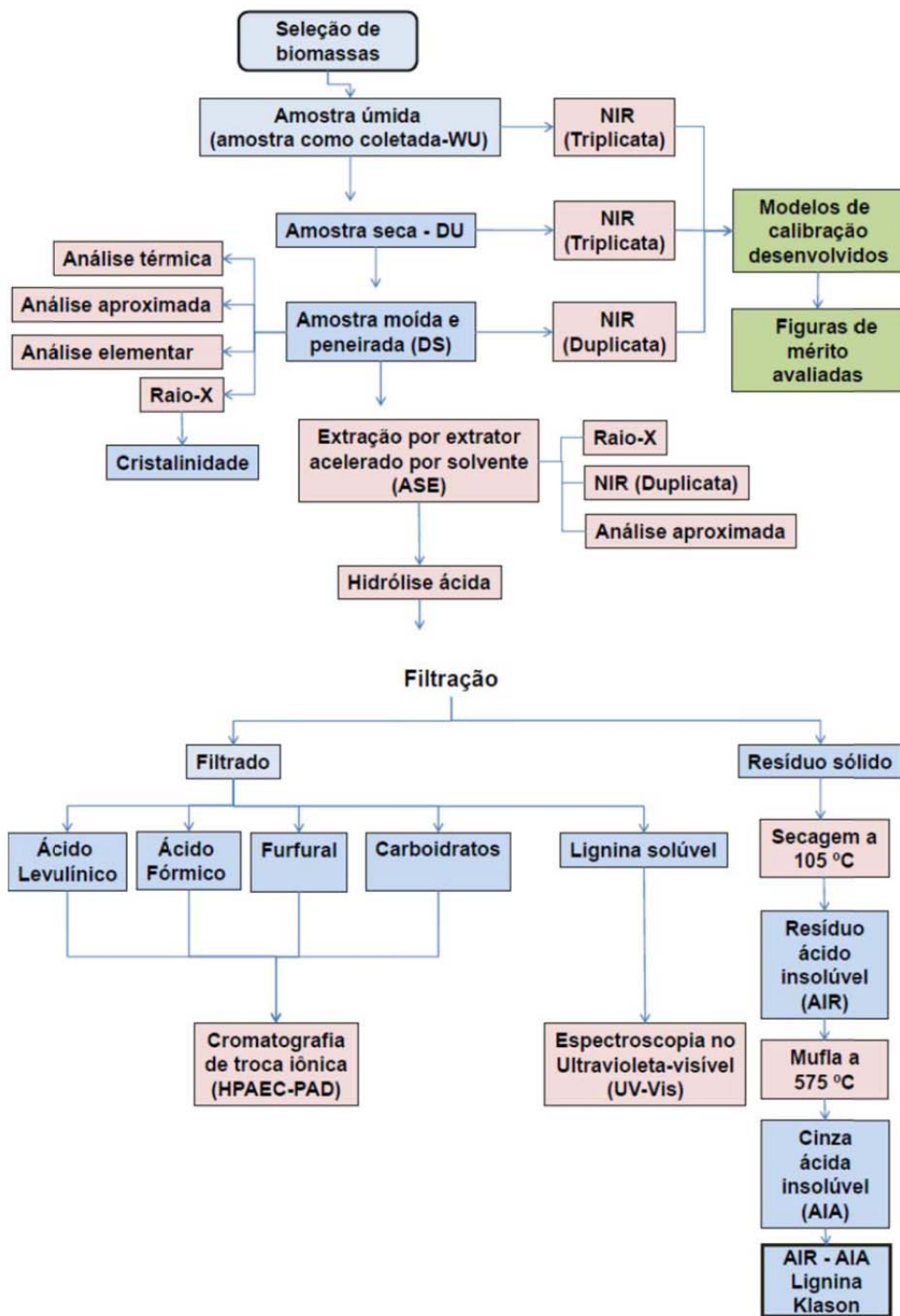


Figura 20. Representação esquemática dos procedimentos experimentais.

3.2.1 Espectroscopia no Visível e Infravermelho-próximo (Vis-NIR)

Os espectros na região do infravermelho próximo foram obtidos em um espectrômetro FOSS (XDS FOSS, Hillerød) de feixe único, no modo de reflectância difusa, na faixa de 400 a 2500 nm (completo). Os dados de reflectância (R) foram transformados para absorbância (A) usando $A = \log_{10} (1/R)$. Cada espectro foi obtido pela média de 32 varreduras, com incremento de 0,5 nm. Cada fração de amostra foi analisada em triplicata, exceto para a fração DS que foram feitas em duplicata, e então o espectro médio foi utilizado para análise posterior.

As amostras úmidas (retiradas do freezer) e as amostras secas foram ambientadas a temperatura ambiente por alguns minutos e em seguida seus espectros foram coletados.

3.2.2 Análise química aproximada

Os procedimentos utilizados para a análise química aproximada, ou imediata como também é conhecida, foram realizados segundo as normas ASTM D 3173-87, ASTM D 3174-02 e ASTM D 3175-02.

A umidade foi determinada após a amostra ser aquecida a $105^{\circ} \pm 5^{\circ} \text{C}$ em estufa (315 Se, Fanem) durante 12 horas, ou até obter massa constante. O teor de umidade foi calculado da seguinte forma:

$$\text{Umidade (\%)} = \frac{M_i - M_v}{M_i} \times 100 \quad \text{Equação 14}$$

onde M_i é a massa em gramas da amostra usada;

M_v é a massa da amostra em gramas após aquecimento.

O teor de matérias voláteis (TMV) foi determinado de acordo com a ASTM (D 3175 - 02) usando 1 grama da amostra, previamente seca, em mufla a $950 \pm 10^{\circ} \text{C}$ por 8 minutos. A amostra foi retirada e colocada em dessecador para ser resfriada, sendo finalmente pesada e seu teor de TMV calculado de acordo com a equação 15:

$$\text{TMV (\%)} = \frac{M_i - M_v}{M_i} \times 100 \quad \text{Equação 15}$$

onde M_i é a massa em gramas da amostra usada;

M_v é a massa da amostra em gramas após aquecimento.

Vale lembrar que a obtenção do TMV na ASTM é para carvões e neste estudo foi adaptada para biomassas, o que pode levar a valores superiores de voláteis (Naik *et al.*, 2010; Kleinlein, 2010; Sasmal *et al.*, 2012).

O teor de cinzas (CZ) foi determinado de acordo com a ASTM (D 3174 – 02) que envolveu a remoção de constituintes orgânicos através de altas temperaturas em mufla (Q 318M, Quimis) por um período de 6 horas a 600 ± 10 °C. A determinação de cinzas é expressa como a percentagem de resíduo (material inorgânico) que permanece após a oxidação a 550 - 600 °C.

O teor de carbono fixo (CF) é uma medida indireta e foi calculado pela equação 16:

$$CF (\%) = 100 - (CZ + TMV) \quad \text{Equação 16}$$

onde CZ e TMV são dados em %.

Todos os resultados são relatados em relação à massa seca da amostra em estufa a 105 °C.

3.2.3 Análise elementar

As análises elementares de carbono, hidrogênio e nitrogênio (CHN) foram realizadas num analisador elementar da Perkin Elmer modelo PE-2400, no Instituto de Química da UNICAMP. A porcentagem de oxigênio (O) foi determinada por diferença. Os dados são apresentados em base seca.

3.2.4 Análise térmica

A análise térmica foi avaliada por meio das técnicas de análise termogravimétrica (TGA) e suas derivadas (DTG). A termogravimetria foi realizada em um equipamento TG 209 Netzsch, na faixa de temperatura de 20 °C a 800 °C, em atmosfera de nitrogênio, sob fluxo de $15 \text{ mL} \cdot \text{min}^{-1}$ e taxa de aquecimento de $20 \text{ }^{\circ}\text{C} \cdot \text{min}^{-1}$. As amostras utilizadas para a análise termogravimétrica foram 4 a 6 mg de biomassa na forma de pó. Essa técnica é utilizada a fim de monitorar a perda de massa de uma matéria-prima durante o aumento da temperatura, quer na ausência ou na presença de ar. As perdas de massa associadas com certos intervalos de temperatura são consideradas um

indicativo da quantidade de componentes químicos diferentes presentes na biomassa (por exemplo, umidade, extrativos, carboidratos, lignina).

3.2.5 Difração de Raios-X

As medidas de difração de Raios-X foram realizadas em um difratômetro Shimadzu modelo XRD 7000, trabalhando com uma diferença de potencial no tubo de 30 kV e uma corrente elétrica de 20 mA. A varredura foi feita na faixa de $5^\circ < 2\theta < 50^\circ$ e a fonte de radiação utilizada foi a Cu K α . Todas as medidas foram realizadas através do método do pó. O índice de cristalinidade (CI) da celulose foi calculado por dois métodos diferentes. No primeiro, o CI foi calculado subtraindo a intensidade máxima da difração ($22^\circ < 2\theta < 23^\circ$) que representa o material cristalino (I_c), da intensidade mínima da difração ($18^\circ < 2\theta < 19^\circ$) que representa o material amorfo (I_{am}) (Sasmal *et al.*, 2012; Pereira *et al.*, 2012), conforme a equação 17 e a Figura 21.

$$CI = \frac{I_c - I_{am}}{I_c} \times 100 \quad \text{Equação 17}$$

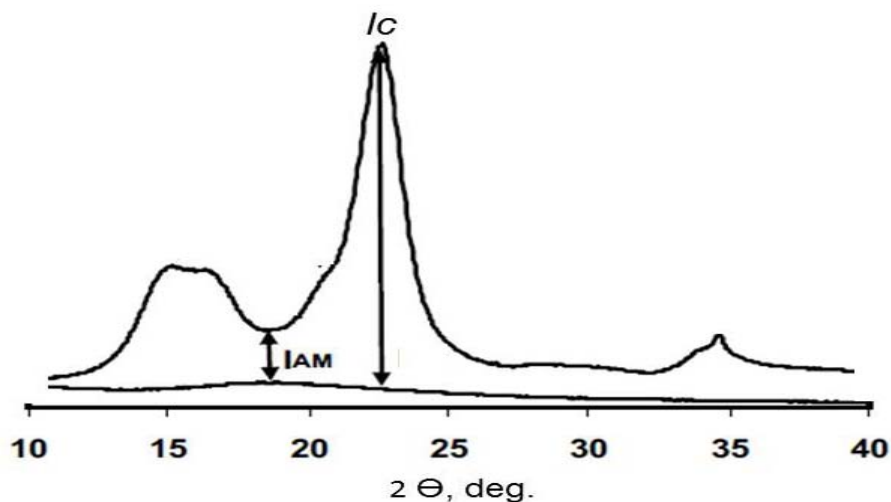


Figura 21. Difratograma de raios-X ilustrando os picos usados no cálculo da cristalinidade pela intensidade (Adaptado: Park, 2010).

No segundo método, picos gaussianos individuais foram ajustados a partir dos difratogramas (Figura 22), usando o programa peak fitting (PeakFit; www.sygmaplot.com, trial version). Uma banda larga centrada ao redor de $21,5^\circ$ foi

atribuída a contribuição amorfa (Figura 22). Iterações foram repetidas até a convergência de χ^2 (14 dígitos significativos), o que correspondeu a valores de R^2 maiores que 0,94 para todos os ajustes. A soma das áreas dos picos cristalinos (101; $\overline{101}$ e 002, Figura 22) e da banda larga amorfa foram usadas para calcular o CI (Figueiredo *et al.*, 2010; Park *et al.*, 2010), de acordo com a equação 18:

$$CI = \frac{I_c}{I_c + I_{am}} \times 100 \quad \text{Equação 18}$$

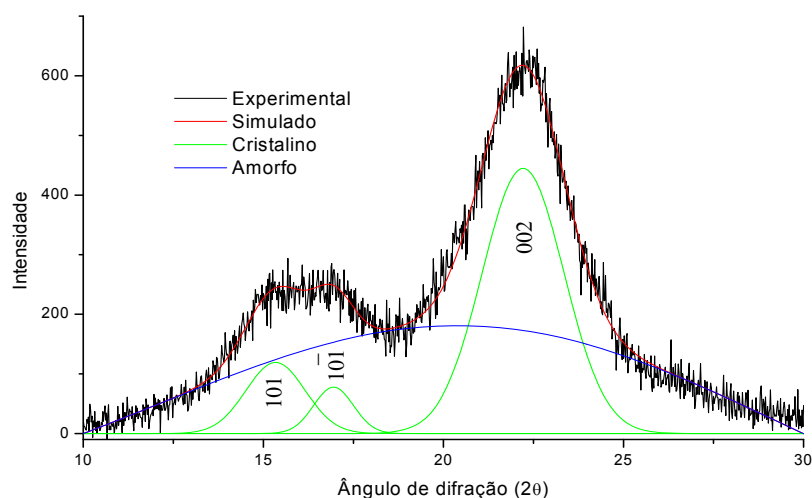


Figura 22. Difratoograma de raios-X ilustrando os picos usados no cálculo da cristalinidade pelo método da deconvolução de picos gaussianos.

3.2.6 Poder Calorífico Superior (PCS)

O poder calorífico superior (PCS) foi determinado numa bomba calorimétrica adiabática (IKA C2000) usando a metodologia da ASTM D5865-04. A combustão dos materiais foi feita em presença de O_2 e sob pressão de 20 a 30 atm. Desta forma o poder calorífico foi calculado a partir das diferenças de temperatura coletadas antes e após a combustão.

$$\chi^2 = \frac{(v_o - v_e)^2}{v_e}$$

onde: v_o é o valor observado (experimental) e v_e o valor esperado (ajustado).

3.2.7 Extração

Para esse fim utilizou-se um dispositivo de extração acelerada por solvente (ASE-200, Dionex) ilustrado na Figura 23. A parte superior do dispositivo é uma bandeja circular com 24 posições. Em cada uma dessas posições (células) pode ser colocado uma amostra. Na parte inferior do equipamento, frascos de coleta, são usados para recolher os extratos. Há um total de 30 posições nessa bandeja inferior, sendo 26 dessas usadas para recolher os extrativos a partir das células, enquanto que as outras quatro posições restantes são usadas para lavar os frascos e para limpar o sistema durante as extrações de cada amostra. O sistema é dito "acelerado" porque a extração ocorre em temperaturas (100 °C) e pressões (1500 psi) elevadas o que permite que o processo seja concluído mais rapidamente.



Figura 23. Extrator acelerado por solvente (ASE).

A determinação do teor de extrativos das amostras foi feita de acordo com a perda de massa depois de ter passado pelo método de extração do ASE 200 com etanol a 95% como solvente extrator. Três extrações (lavagens) por amostra foram necessárias para a maioria das biomassas.

3.2.8 Hidrólise ácida

A etapa da hidrólise ácida foi realizada de acordo com a metodologia do Laboratório Nacional de Energia Renovável (NREL, <http://www.nrel.gov/biomass/>). Esta consistiu na utilização de tubos de ensaio, nos quais foram colocados 0,300 g $\pm$ 0,05 g da biomassa extraída, adicionados 3,0 mL (4,91 g) de H₂SO₄ a 72% (m/m), e então esses foram levados a banho-maria (TE-056, Tecnal) por 60 min a 30 °C e agitados a cada 10 min. Essa é a etapa da hidrólise primária. Em seguida, 84 mL de água deionizada foram adicionados para que a concentração de ácido sulfúrico da solução fosse reduzida para 4%. A etapa de hidrólise secundária ocorre com essa solução em autoclave (AV 18, Phoenix) por 1 hora a 121 °C. Após essa etapa, a solução hidrolisada é filtrada em cadinhos filtrantes de porosidade média (10 a 16 micras) e o hidrolisado é retido para análises posteriores de cromatografia. Essa solução contém todos os açúcares hidrolisados, além da lignina ácida solúvel (ASL), e quaisquer outros componentes ácido solúveis da biomassa. Os sólidos retidos nos cadinhos filtrantes foram levados à estufa a 105 °C para posterior análise do teor de resíduo ácido insolúvel (AIR), e em seguida mantidos na mufla a 4 horas a 575 °C, para obtenção da cinza ácida insolúvel (AIA). Por diferença de AIR e AIA, obtém-se o teor de lignina insolúvel, conhecida também como Lignina Klason (KL). Também soluções de recuperação de açúcares foram hidrolisados juntamente com as amostras, a fim de se avaliar a perda de açúcares na hidrólise e corrigir para as amostras reais.

A hidrólise ácida para a obtenção de ácido levulínico, ácido fórmico e furfural foram realizadas em parceria com a Universidade de Limerick (UL) de acordo com a metodologia proposta por Girisuta *et al.*, (2013).

3.2.9 Cromatografia

A cromatografia líquida de alta eficiência por troca iônica com detector amperométrico pulsado (HPAEC-PAD), um sistema *Dionex modelo* ICS 3000, foi utilizada para as análises dos carboidratos nas seguintes condições: Coluna CarboPac PA1 de 4 x 250 mm e uma pré-coluna CarboPac PA1 de 4 x 50 mm; detector eletroquímico (ED 40, Dionex); eletrodo de ouro; fluxo de 0,4 mL.min⁻¹ usando gradiente

na seguinte seqüência: 0 a 23,25 min - 70% água e 30% de 10 mM NaOH; 24,50 a 26 min - 69% água, 11% 10 mM NaOH e 20% 500 mM NaOH; 26 min – 45 min 48,4% água, 6,9% 10 mM NaOH, 20% 500 mM NaOH e 24,7% 1mM NaOAc; 45 a 51 min – 50% água e 50% 500 mM NaOH. Um padrão interno de fucose foi utilizado, e os hidrolisados, soluções de recuperação de açúcares e padrões externos foram diluídos 5 vezes com uma solução conhecida desse padrão interno para as análises cromatográficas. Um volume de 10 µL foi injetado a cada análise. As condições cromatográficas permitiram, em uma única injeção de 25 min, a resolução entre fucose, arabinose, galactose, ramnose, glicose, xilose e manose. Amostras de açúcares padrão foram injetados em intervalos regulares na seqüência analítica (a cada 4 amostras, um padrão era injetado) a fim de checar a estabilidade do aparelho e dos tempos de retenção.

As análises cromatográficas para a quantificação de ácido levulínico, ácido fórmico e furfural foram realizadas em parceria com a Universidade de Limerick (UL) de acordo com a metodologia proposta por Girisuta *et al.*, (2013).

3.2.10 Espectroscopia no ultravioleta visível (UV-Vis)

Um espectrofotômetro UV-Vis (DR 5000, HACH) foi utilizado para a análise dos hidrolisados da biomassa a fim de determinar o teor de ASL das amostras.

Uma pequena porção da lignina de muitos materiais torna-se solúvel em processos de hidrólise ácida, a lignina ácida solúvel (ASL). Isso significa que a lignina Klason (KL) não é um verdadeiro reflexo do teor de lignina total da amostra. Métodos de determinação da ASL são, portanto, necessários para calcular o teor de lignina total (Lignina total = KL + ASL). Tipicamente, a ASL é medida usando métodos espectrofotométricos de absorção na região do ultravioleta no hidrolisado diluído. O hidrolisado foi colocado em uma cubeta de quartzo de capacidade de 3 mL e diluído quando necessário. O branco dessas análises foi uma solução de H₂SO₄ a 4% e o comprimento de onda selecionado foi de 205 nm a uma absorvidade molar entre 88-113 M⁻¹ cm⁻¹, baseados nos picos de máxima absorção, dependendo da biomassa analisada (Bland e Menshun, 1971; Sluiter *et al.*, 2010).

3.2.11 Análise dos dados

As análises estatísticas e multivariadas foram realizadas utilizando os softwares Unscrambler 10.3 para construção dos modelos de calibração multivariada e análise por componentes principais (Camo, Noruega), Matlab 7.2 para a determinação das figuras de mérito (MathWorks, Estados Unidos), Origin 6.0 (Originlab) e Statistica 7.0 para construção dos gráficos e figuras ilustrativas (StatSoft, Estados Unidos, 2005).

Uma análise estatística descritiva (média, amplitude de variação e desvio padrão) foi realizada para os teores de cada analito. A análise por componentes principais (PCA) foi realizada utilizando-se os espectros Vis-NIR com os dados centrados na média. A PCA foi utilizada para análise exploratória dos dados verificando possíveis agrupamentos das amostras, ou seja, a relação entre as amostras e variáveis. Para melhorar a interpretabilidade do modelo global, o qual inclui as três biomassas, a rotação varimax foi aplicada utilizando 5 fatores. Para a PCA, utilizando apenas as amostras de banana, nenhuma rotação foi necessária e 5 fatores foram utilizados.

Para os modelos preditivos dos componentes químicos foi utilizado o método de calibração multivariada PLS. Diferentes pré-processamentos foram testados, e os melhores resultados foram obtidos com um único ou uma combinação dos seguintes pré-processamentos: Padronização normal do sinal (SNV), primeira (1D) e segunda derivadas (2D), detrend, correção ortogonal do sinal (OSC). No uso das derivadas, variou-se o tamanho das janelas, usando de 3 até 31 pontos, sendo que o valor máximo empregado (31 pontos) não ultrapassou o valor do dobro da largura a meia altura da banda mais estreita do espectro, como sugerido por Enke e Niernan (1976), o que resulta no maior ganho na relação sinal/ruído. O polinômio usado foi sempre de ordem 2 para todos os modelos. A transformação empregando OSC, pode gerar modelos menos robustos (Fearn, 2000). De modo a verificar a sua robustez o conjunto completo de dados foi aleatoriamente dividido, deixando 20%, 25% e 30% das amostras de validação de fora. Os testes foram repetidos 10 vezes para cada porcentagem de amostras excluídas, de modo a confirmar que, de fato, os modelos são robustos.

A validação dos modelos foi realizada utilizando um conjunto externo de amostras, escolhido aleatoriamente, isso é: o conjunto de espectros foi dividido em dois

subconjuntos: um de calibração (70% das amostras) e outro de validação externa com os restantes 30% das amostras (Esbensen, 2004). Para determinar o número ótimo de variáveis latentes nos modelos finais, a validação cruzada (LOO), no caso de conjuntos com pequenas quantidades de amostras ($I < 40$), foi usada (Ferreira *et al.*, 1999), e em casos de conjuntos grandes de dados, apenas a validação externa foi utilizada para determinar o melhor número de variáveis latentes, sem o passo adicional da validação cruzada (Esbensen e Geladi, 2010), embora em nenhum dos casos estudados houve diferença no número ótimo de variáveis latentes selecionadas por ambos os métodos (validação externa e cruzada).

Modelos utilizando o método PLS foram construídos para diferentes analitos que foram: extrativos; umidade; cinzas; lignina total; lignina ácida solúvel; resíduo ácido insolúvel; lignina Klason; cristalinidade da celulose; glicose; xilose; manose; arabinose; ramnose; galactose, bem como os açúcares totais. Os conjuntos de dados foram organizados em matrizes, conforme resumido na seção Resultados e Discussões. As amostras foram agrupadas em cinco conjuntos, um para amostras de cada biomassa separadamente, banana, café e coco, e os outros dois de amostras combinadas de banana+café e de banana+café+coco. Os espectros foram utilizados considerando-se duas regiões: Vis-NIR (400 a 2500 nm) e apenas NIR (1100 e 2500 nm). Regiões de comprimento de onda específicos foram selecionados para os modelos da fração WU, a fim de verificar se excluindo as regiões onde a água absorve mais fortemente poderia permitir modelos mais simples e robustos resultando numa melhor previsão dos analitos de interesse. Dessa forma modelos foram desenvolvidos usando apenas a região de 1100-1800 nm, e também modelos combinando as regiões de 1100 -1800 nm e 2100-2500 nm (isso é, excluindo a região de 1801-2099 nm). Também nos modelos de açúcares, a fim de facilitar a construção dos modelos de calibração entre os seis monossacarídeos, o método de seleção de variáveis foi empregado, com base nas regiões características do espectro de carboidratos.

Os modelos de regressão obtidos foram avaliados utilizando os seguintes parâmetros: coeficiente de determinação (R^2), raiz quadrada do erro quadrático médio de calibração (RMSEC), raiz quadrada do erro quadrático médio de validação cruzada (RMSECV), raiz quadrada do erro quadrático médio de previsão (RMSEP), viés,

número de variáveis latentes, erro relativo ($RE\% = RMSEP * 100 / \text{média dos valores estimados}$) como proposto por Versari *et al.*, 2008 e Liu *et al.*, 2010; e a fração de amostras anômalas excluídas, sendo essas últimas identificadas pela análise dos gráficos de influência e resíduos de Student (Ferreira *et al.*, 1999). Também, a razão da amplitude de variação dos dados em relação ao erro de previsão ($RER = \text{amplitude de variação dos dados de referência, } y_{\text{máximo}} - y_{\text{mínimo}} / SEP$) foi determinado. De acordo com a Associação Americana de Químicos Cerealistas (AACC, 1999) - Método 39-00 qualquer modelo que tenha $RER \geq 4$ é qualificado para calibração de triagem. Quando $RER \geq 10$ o modelo é aceitável para o controle de qualidade, e se $RER \geq 15$ o modelo é muito bom para quantificação (Liu *et al.*, 2010).

Ainda, o sinal analítico líquido (NAS); e as figuras de mérito; sensibilidade (SEN); seletividade (SEL); sensibilidade analítica (γ); inverso da sensibilidade analítica (γ^{-1}); limite de detecção (LD) e limite de quantificação (LQ) foram calculadas. A precisão foi expressa como RMSEC e RMSEP, a significância estatística do viés foi testado pelo teste t de Student, conforme descrito na norma ASTM E1655-00, e a linearidade foi avaliada graficamente pela inspeção dos gráficos de resíduos, e o gráfico de valores previstos *versus* valores de referência. Além disso, o desvio padrão das duplicatas dos métodos de referência e NIR foram calculados e comparados estatisticamente pelo teste F de Snedecor. Todas as equações das figuras de mérito citadas acima, já foram definidas na Tabela 1 (ver página 48).

3.2.12 Qualidade das análises e dos modelos

Critérios foram definidos para as análises químicas via úmida de modo que a precisão do método de referência fosse alta. Esses foram baseados no desvio padrão das duplicatas. Se o desvio padrão de uma amostra fosse acima do limite atribuído, então a análise da amostra de referência deveria ser repetida. Em alguns casos, os dados analíticos foram mantidos, mesmo que o desvio padrão violasse o limite aceitável. Isso geralmente ocorreu quando a quantidade de material era insuficiente para permitir uma nova análise da amostra e a inclusão dessa amostra nos modelos PLS foi considerada importante. No entanto, nenhuma amostra com desvio padrão excessivo foi mantida nos modelos. Os limites de desvio padrão para cada constituinte

foram de 0,25% para umidade, cinza e lignina ácida solúvel; de 0,5% para extrativos, lignina Klason, e resíduo ácido insolúvel; de 0,55% para carbono fixo, material volátil, poder calorífico e análise elementar, e 1% para os açúcares.

4 RESULTADOS E DISCUSSÕES

4.1 Escolha e caracterização das biomassas

Com o intuito de selecionar potenciais candidatos a partir de dez diferentes biomassas no conceito de biorrefinaria, suas características físicas e químicas foram estudadas. O conhecimento das características físico-químicas de matérias-primas lignocelulósicas ajuda na seleção e personalização das biomassas para o processamento adicional na obtenção de produtos de alto valor agregado (Sasmal *et al.*, 2012).

Há vários outros fatores que também devem ser considerados quando se utiliza a biomassa como uma alternativa aos combustíveis fósseis. Esses incluem: baixo custo, matérias-primas que não competem com a produção agrícola voltada para alimentos, facilidade de coleta e transporte das amostras, biomassa de plantas abundantes, quantidade de resíduo produzido e finalidade de uso desse resíduo (Charlton *et al.*, 2009; Rodrigues, 2011; Sasmal *et al.*, 2012).

Os detalhes das propriedades físico-químicas com respeito à obtenção de bioprodutos será discutida a seguir. A interpretação dos resultados das diversas propriedades estudadas será feita comparando as relações entre elas nas diferentes biomassas.

Ainda, as biomassas selecionadas serão comparadas com biomassas residuais já promissoras na América Latina e na Europa, tais como os resíduos da produção da cana-de-açúcar e o *miscanthus*, respectivamente. Dessa forma, os resultados das caracterizações físico-químicas obtidos neste trabalho serão comparados aos dados da Tabela 3, a fim de verificar se as novas biomassas apresentam o mesmo potencial daquelas já bem estabelecidas.

Tabela 3. Composição físico-química das biomassas de cana – de- açúcar e *miscanthus* (%).

Biomassa	Gli	Xil	Arab	Ram	Gal	Man	TS	Extr	KL	ASL	Cinza	PC*	Referência
Bagaço da cana	43,3	22,4	2,3	-	0,6	0,4	69,0	3,6	18,0	2,3	6,0	19,0	Girisuta <i>et al.</i> , (2013)
Bagaço da cana	-	-	-	-	-	-	59,7		22,2		20,9		Rezende <i>et al.</i> , (2011)
Bagaço da cana							72,5	4,3	21,1		2,2		Rocha <i>et al.</i> , (2011)
Palha da cana							67,4	7,0	18,4		6,2		Szczerbowski <i>et al.</i> , (2009)
Palha da cana	30,1	12,7	4,5	-	-	-	-	-	34,2	1,2	12,5	20,3	Georges (2011)
<i>Miscanthus</i>	40,3	19,4	2,1	0,21	0,6	0,2	62,9	1,8	21,8	-	-	-	Haverty <i>et al.</i> , (2012)
<i>Miscanthus</i>	38,0	14,9	1,1	-	0,3	-	54,3	2,2	20,8	0,9	0,8	17-20	Brosse <i>et al.</i> , (2012)

Gli: glicose; Xil: xilose; Arab: arabinose; Ram: ramnose; Gal: galactose; Man: manose; TS: açúcares totais; Extr: Extrativos; KL: lignina Klason; ASL: lignina ácida solúvel; PC*: Poder calorífico (em MJ/Kg).

4.1.1 Poder calorífico

O poder calorífico é uma das mais importantes qualidades da biomassa quando se pensa na combustão direta da mesma como fonte de energia em substituição aos combustíveis derivados do petróleo. O poder calorífico, além do teor de umidade e cinzas (Everard *et al.*, 2012), é muito influenciado pela constituição química da matéria-prima, principalmente a lignina e extrativos (resinas, óleos, materiais graxos) (Tabelas 4 e 5).

Tabela 4. Resultados da análise imediata, elementar e poder calorífico das biomassas.

Biomassa	Análise imediata (%)				Análise elementar (%)				Poder Calorífico (MJ/Kg)
	Cinzas	Umidade (DS)	Matéria Volátil	Carbono fixo	C	H	N	O	
Caroço de açaí	0,69 ±0,00	13,27 ±0,18	80,77 ±0,20	18,50 ±0,20	47,60 ±0,84	6,41 ±0,41	0,78 ±0,03	45,12 ±0,19	18,60 ±0,30
Caule de banana	8,01 ±0,13	12,56 ±0,03	80,27 ±0,15	9,96 ±0,19	39,01 ±0,36	5,44 ±0,05	0,82 ±0,01	54,84 ±0,21	16,13 ±0,35
Engaço de banana	8,04 ±0,01	8,04 ±0,20	73,25 ±0,02	22,13 ±0,02	37,95 ±0,98	4,73 ±0,44	1,46 ±0,07	55,85 ±0,89	15,53 ±0,15
Bambu	1,71 ±0,01	9,01 ±0,03	81,08 ±0,05	17,20 ±0,05	44,60 ±0,09	5,55 ±0,10	0,91 ±0,00	48,93 ±0,00	18,33 ±0,20
Coco	2,95 ±0,25	9,87 ±0,40	77,01 ±0,10	20,05 ±0,27	47,40 ±0,20	5,41 ±0,16	0,55 ±0,01	46,64 ±0,19	18,70 ±0,20
Cascas de café	2,03 ±0,15	10,01 ±0,20	75,40 ±0,20	22,70 ±0,25	43,34 ±0,55	5,55 ±0,13	2,25 ±0,07	48,86 ±0,25	18,06 ±0,30
Serragem	0,11 ±0,01	10,10 ±0,08	83,88 ±0,10	16,01 ±0,10	50,30 ±0,11	6,08 ±0,00	0,15 ±0,06	43,43 ±0,09	20,01 ±0,40
Capim	13,01 ±0,05	8,10 ±0,04	76,50 ±0,05	10,81 ±0,07	42,01 ±0,99	5,21 ±0,18	2,03 ±0,12	50,95 ±0,70	16,77 ±0,13
Cascas de arroz	12,49 ±0,00	8,01 ±0,10	71,24 ±0,10	16,27 ±0,10	35,86 ±0,95	4,40 ±0,50	0,28 ±0,09	59,46 ±0,86	16,35 ±0,18
Cascas de soja	4,01 ±0,01	9,16 ±0,18	88,81±0,20	7,23 ±0,20	45,04 ±0,65	6,70 ±0,22	2,91 ±0,68	45,35 ±0,78	17,90 ±0,20

Tabela 5. Composição química das matérias-primas (%).

Biomassa	Extrativos	ASL	KL	AIR	TL	AIA
Caroço de açaí	9,51 ±0,04	1,62 ±0,07	15,63 ±0,26	15,79 ±0,26	17 ±0,33	0,16 ±0,00
Caule de Banana	25,10 ±0,06	1,33 ±0,06	7,06 ±0,23	7,58 ±0,28	8,38 ±0,29	0,52 ±0,18
Engaço de Banana	22,35 ±0,50	2,20 ±0,00	8,48 ±0,01	10,78 ±0,09	10,68 ±0,01	2,31 ±0,08
Bambu	12,80 ±0,13	1,48 ±0,02	16,17 ±0,08	16,31 ±0,16	17,65 ±0,10	0,15 ±0,08
Coco	1,41 ±0,44	1,28 ±0,07	30,15 ±0,37	30,48 ±0,24	32,51 ±0,44	0,35 ±0,13
Cascas de café	4,27 ±0,49	2,12 ±0,06	22,33 ±0,25	22,38 ±0,25	24,46 ±0,31	0,05 ±0,03
Serragem	8,12 ±0,00	2,32 ±0,00	30,55 ±0,04	30,57 ±0,02	32,87 ±0,04	0,02 ±0,03
Capim	11,05 ±0,09	4,36 ±0,03	11,29 ±0,13	19,66 ±0,19	15,61 ±0,16	8,4 ±0,07
Cascas de arroz	2,31 ±0,06	1,01 ±0,01	22,85 ±0,16	34,82 ±0,21	23,90 ±0,17	12,26 ±0,04
Cascas de soja	6,85 ±0,11	4,93 ±0,40	2,67 ±0,11	2,76 ±0,03	7,58 ±0,50	0,09 ±0,08

ASL: lignina ácida solúvel; KL: lignina Klason; AIR: resíduo ácido insolúvel; TL: Lignina total; AIA: cinza ácida insolúvel.

Para verificar a influência da constituição química no poder calorífico, correlacionou-se (correlação de Pearson) este com os demais parâmetros analisados (Figura 24). E assim, a serragem de eucalipto, que apresenta um conteúdo de lignina maior que as demais biomassas (Tabela 5) ostenta conseqüentemente um poder calorífico superior. O poder calorífico é mais alto quanto maior o teor de lignina e extrativos, porque os mesmos contêm menos oxigênio que os polissacarídeos presentes na holocelulose (celulose e hemicelulose). O extremo oposto é evidenciado no caule e engaço da banana, com poder calorífico inferior às demais biomassas e baixos teores de lignina. Para o teor de extrativos foi encontrada uma relação negativa com o poder calorífico, não mostrando um efeito tão significativo quanto para a lignina. Já para as cinzas e teor de oxigênio uma correlação negativa e bastante significativa foi encontrada com o poder calorífico. Ainda dentre os parâmetros que se correlacionam positivamente com o poder calorífico, o maior resultado de carbono fixo foi encontrado para as cascas de café com 22% e o menor resultado foi na casca de soja com 7,2%, indicando que o café poderá ter melhor aproveitamento na produção de carvão vegetal (Tabela 3). Com relação aos voláteis, a casca de soja apresentou o maior valor com

88%, e o menor valor encontrado foi na casca de arroz com 71%, indicando que a casca de soja poderá ser melhor aproveitada na produção de energia na combustão.

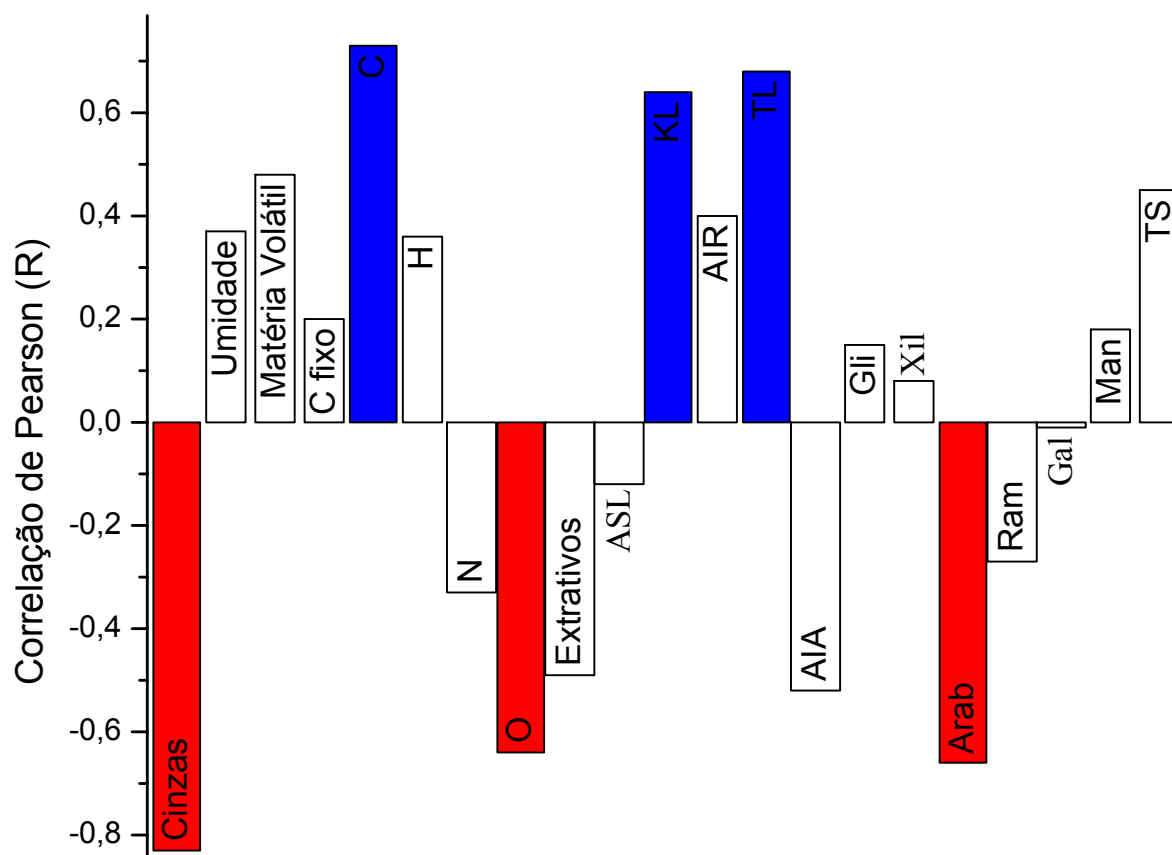


Figura 24. Correlação entre o poder calorífico (PC) e a composição química das biomassas (ASL: lignina ácida solúvel; KL: lignina Klason; AIR: resíduo ácido insolúvel; TL: Lignina total; AIA: cinza ácida insolúvel; CF: carbono fixo; Gli: glicose; Xil: xilose; Ara: arabinose; Rha: ramnose; Gal: galactose; Man: manose e TS: açúcar total).

4.1.2 Cinzas

As cinzas não têm valor energético, e sendo compostas de elementos inorgânicos da biomassa, não apresentam nenhum valor em tecnologias de aproveitamento de materiais lignocelulósicos. Altos conteúdos de cinzas podem causar problemas em reações de pirólise e na combustão da biomassa. No que diz respeito aos procedimentos de hidrólise ácida citados anteriormente, um alto teor de cinzas pode implicar num maior consumo de ácido, devido à sua natureza alcalina (Bruun *et al.*, 2010), e além disso, na hidrólise enzimática, os cátions presentes nas cinzas afetam

a atividade das enzimas (Bin e Hongzhang, 2010). Como observado na Tabela 3 apenas a casca de arroz e o capim elefante apresentaram teores mais elevados de cinzas, o que pode comprometer as suas hidrólises ácida/enzimática.

No entanto as cinzas da casca de arroz (CCA), que representam 12,5% da composição das cascas, geralmente contêm de 80-90% de sílica (Rambo *et al.*, 2011), que podem ser aproveitadas em diversas áreas. Entre as principais aplicações da CCA pode-se citar o uso na agricultura, em que é utilizada como fertilizante e corretivo de solos para aumentar a resistência à várias doenças (Rambo *et al.*, 2011); na indústria farmacêutica e de cosméticos (Kalapathy *et al.*, 2000); no cimento, em substituição à areia, melhorando o custo e as propriedades físico-mecânicas (Cisse *et al.*, 2000). Dessa forma a grande quantidade de cinza residual encontrada nas cascas de arroz, pode ser bem aproveitada.

4.1.3 Umidade

O teor de umidade da biomassa é um parâmetro de eficiência crucial quando a combustão é levada em consideração. Também, o excesso de umidade dificulta e aumenta os custos de transporte, secagem, armazenamento e manipulação da biomassa. Biocombustíveis sólidos são muitas vezes desvalorizados em função do seu teor de umidade devido à sua influência negativa no poder calorífico (Charlton *et al.*, 2009; Everard *et al.*, 2012). Os teores de umidade observados (Tabela 3) variam de 8-13%, e correspondem aos percentuais da biomassa seca ao ar. Vale ressaltar que a umidade no campo é superior.

4.1.4 Análise elementar

Uma alternativa ao uso de resíduos ricos em nitrogênio (N) seria o seu aproveitamento como fertilizante, em aplicação direta no solo. O baixo teor de nitrogênio nas estruturas químicas da serragem, cascas de arroz e coco (Tabela 4), por exemplo, é um fator desfavorável para a sua degradação microbiológica. Por outro lado, essa é a razão pela qual resíduos de banana, podem ser utilizados como fertilizantes em plantações após a colheita (Bilba *et al.*, 2007). Com relação ao teor de hidrogênio (H) amostras de caroço de açaí e cascas de soja contêm os percentuais mais elevados.

Isso provavelmente ocorre no caso das cascas de soja, devido à menor quantidade de duplas ligações que as outras amostras, o que pode estar associada a estrutura aromática da lignina (Bilba *et al.*, 2007) (Tabela 5).

4.1.5 Extrativos

A determinação do teor de extrativos é importante, pois permite que o balanço de massa total para os constituintes analisados possa se aproximar de 100%, indicando uma metodologia completa e confiável da análise. Além disso, os teores de extrativos devem ser cuidadosamente avaliados, uma vez que podem causar interferências nas etapas posteriores da análise, precipitando durante a hidrólise e sendo falsamente considerados como lignina, gerando resultados imprecisos para os teores de KL e polissacarídeos (Naik *et al.*, 2010; Tamaki *et al.*, 2010). Os teores totais de extrativos foram maiores nos resíduos de banana, seguidos de bambu e capim elefante. As demais biomassas apresentaram baixos teores, com percentuais inferiores a 10% (Tabela 5).

4.1.6 Lignina

A lignina é um subproduto do processo de hidrólise que pode ser utilizado como fonte de energia térmica (Hassuani, 2005), além da sua utilização em processos de pirólise para obtenção de bio-óleo e biochar (carvão residual), uma vez que apresenta o maior depósito de estruturas químicas aromáticas, constituindo-se em uma fonte potencial de valiosos insumos para a indústria química (Melligan *et al.*, 2012). Apesar de lignina apresentar grande potencial como fonte renovável para obtenção de combustíveis e produtos químicos aromáticos, as tecnologias de valorização de lignina são substancialmente ainda menos desenvolvidas do que aquelas para os polissacarídeos. Isso ocorre devido a inúmeras dificuldades encontradas no processamento catalítico. Rotas aproveitando as enormes quantidades de lignina disponíveis já estão sendo desenvolvidas (Azadi *et al.*, 2013).

Dessa forma, matérias-primas como cascas de coco e serragem de eucalipto apresentam grande potencial de uso da lignina para biorrefinarias no aproveitamento de compostos aromáticos, principalmente. Valores relativamente altos de lignina total

foram encontrados para a serragem de eucalipto, uma vez que teores na faixa de 15-25% para essa biomassa são esperados (Hillis e Brown, 1978). No entanto outros trabalhos já publicados, também encontraram valores superiores a 30% de lignina total para espécies de eucalipto (Trugilho *et al.*, 2003; Santos, 2008). Segundo Severo *et al.*, (2006) esta variação nos componentes químicos da madeira dá-se devido a vários fatores, tais como: espécie, tratos silviculturais e, principalmente, à sua estrutura anatômica.

Já biomassas com baixo percentuais de lignina, tais como o caule e engaço de banana, apresentam grande potencial nos processos de hidrólise ácida e/ou enzimática (Tabela 5).

4.1.7 Carboidratos

Com relação aos teores de carboidratos apresentados na Tabela 6, o caroço de açaí, o bambu e as cascas de café apresentaram os maiores percentuais de açúcares totais (TS), superiores a 60% da composição total. Todas as biomassas demonstraram ter potencial como matérias-primas em processos de biorrefinarias, apresentando teores totais de açúcares superiores a 40%, sendo o mínimo exigido de 25%.

Avaliando e comparando a composição química em carboidratos das mesmas biomassas estudadas neste trabalho: pseudocaule de banana (Aziz *et al.*, 2011), engaço de banana (Soffner, 2001), caroço de açaí (Rodríguez-Zúñiga *et al.*, 2008), cascas de arroz (Binod *et al.*, 2010), bambu (Tippayawong e Chanhom, 2011), serragem de eucalipto (García *et al.*, 2012), capim elefante (Crespo *et al.*, 2013), resíduos de coco (Bilba *et al.*, 2007), cascas de café (Tsai *et al.*, 2012) e cascas de soja (Cassales *et al.*, 2011) - os máximos teores de TS após a hidrólise ácida encontrados na literatura foram respectivamente de 57%, 44%, 65%, 56%, 67%, 77%, 44%, 42%, 47% e 66%. Apenas para as biomassas de cascas de soja, serragem de eucalipto, caule de banana e bambu os resultados da literatura foram superiores aos dados obtidos neste trabalho, e para as cascas de arroz e engaço os resultados foram similares. Para as demais amostras, os teores de carboidratos foram maiores neste trabalho. Esses resultados demonstram as concordâncias dos processos físico-químicos desenvolvidos para obtenção dos açúcares e consequentemente para os

processos de biorrefinarias. No entanto, apesar da satisfatória otimização do processo experimental, deve-se levar em conta que apesar das matérias-primas serem as mesmas, outros fatores, tais como diferentes espécies, cultivares, solos e locais de plantação podem afetar a composição química dessas biomassas, levando a diferenças significativas.

Tabela 6. Composição de carboidratos das matérias-primas (%)*.

Biomassa	Gli	Xil	Arab	Ram	Gal	Man	TS
Caroço de açaí	8,66 ±0,03	3,18 ±0,40	0,69 ±0,10	0,17 ±0,03	1,43± 0,03	53,60± 0,05	67,71 ±0,53
Caule de Banana	36,32 ±0,13	5,36 ±0,01	2,37 ±0,01	3,16 ±0,00	0,72 ±0,02	0,61± 0,27	48,53 ±0,44
Engaço de Banana	28,83 ±0,05	6,94 ±0,15	2,89 ±0,05	1,27 ±0,05	1,18± 0,00	1,46± 0,14	42,56 ±0,35
Bambu	44,65 ±0,98	14,78 ±0,43	0,81 ±0,03	0,06 ±0,07	0,32± 0,02	0,07± 0,01	62,57 ±1,0
Coco	32,41 ±0,14	14,37 ±0,26	1,79 ±0,12	0,30 ±0,03	0,71± 0,08	0,35± 0,07	50,01 ±0,11
Cascas de café	35,33 ±0,16	22,89 ±0,09	1,62 ±0,15	0,51 ±0,08	1,54± 0,08	1,68± 0,01	62,55 ±0,56
Serragem	38,80 ±0,20	9,72 ±0,05	0,26 ±0,27	0,25 ±0,01	1,23± 0,00	0,35± 0,12	50,60 ±0,11
Capim	27,52 ±0,21	16,12 ±0,24	3,56 ±0,06	0,10 ±0,10	1,22± 0,01	0,24± 0,27	49,84 ±0,55
Cascas de arroz	36,17 ±0,15	16,65 ±0,38	1,71 ±0,04	0,13 ±0,01	0,83 ±0,02	0,49± 0,21	56,01 ±0,66
Cascas de soja	35,05 ±0,99	9,85 ±0,01	4,64 ±0,04	0,92 ±0,00	3,13± 0,02	4,31± 0,02	58,90 ±1,0

Gli: glicose; Xil: xilose; Arab: arabinose; Ram: ramnose; Gal: galactose; Man: manose; TS: açúcares totais; *:valores incluídos os fatores de hidratação da hidrólise e a amostra estando livre de extrativos.

Cromatogramas típicos dos carboidratos presentes nos hidrolisados de cascas de café, fibras de coco e engaço de banana são mostrados nas Figuras 25 (A), (B) e (C). Tempos de separação de apenas 25 minutos foram necessários para separação dos seis monossacarídeos. Percebe-se uma boa resolução entre os seis monossacarídeos demonstrando a eficiência da cromatografia de troca iônica na identificação e quantificação da composição química de amostras vegetais, sem a necessidade de derivatização ou outras etapas longas de preparação de amostras. A cromatografia com detecção por índice de refração (HPLC-IR) e a cromatografia gasosa acoplada a massas (GC-MS) também foram testadas, no entanto, a HPLC-IR apresenta

baixa sensibilidade para monossacarídeos presentes em baixas concentrações tal como manose e ramnose, enquanto a GC-MS necessita de etapas laboriosas de preparação de amostra, uma vez que são usados agentes derivatizantes para converter os açúcares em compostos voláteis. Desta forma a HPAEC-PAD mostrou-se uma técnica de grande valia, rápida e com resultados satisfatórios para determinar todos os constituintes químicos derivados da hidrólise ácida.

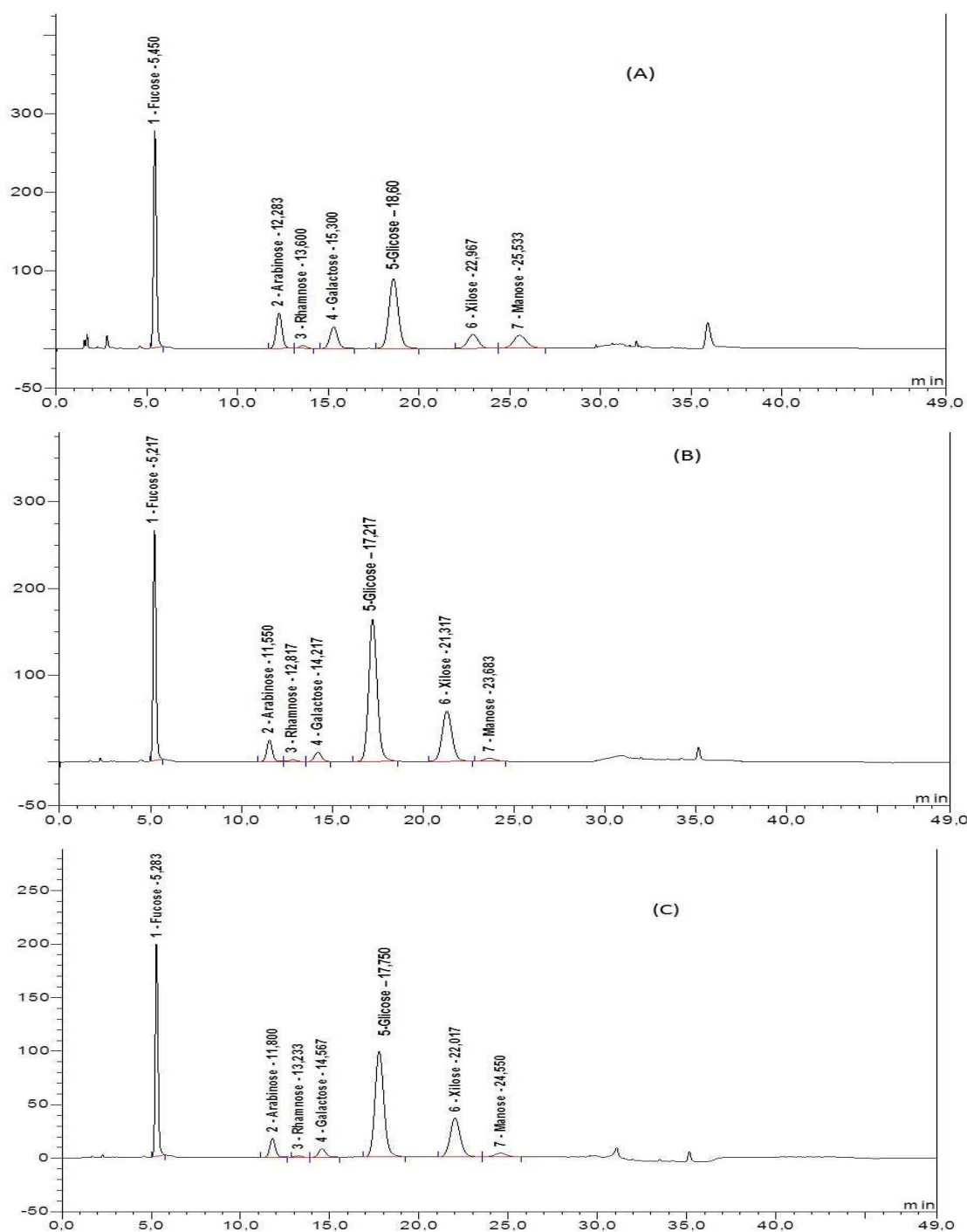


Figura 25. Cromatogramas dos hidrolisados de (A) cascas de café (B) fibras de coco e (C) engaço de banana.

4.1.8 Análise Térmica

Para que as biomassas sejam mais bem aproveitadas é de extrema importância o conhecimento do comportamento térmico dessas durante o processo de termoconversão. O conhecimento de tal processo é fundamental para o monitoramento das condições de processamento desses materiais. As técnicas de análise térmica (TGA), particularmente a termogravimetria (TG) e a termogravimetria derivada (DTG) possibilitam a obtenção dessa informação de maneira simples e rápida (Tomczak *et al.*, 2007). A utilização das técnicas TG e DTG serviu para a caracterização térmica das dez diferentes amostras analisadas.

O comportamento entre as diversas biomassas apresentou algumas diferenças (Figura 26).

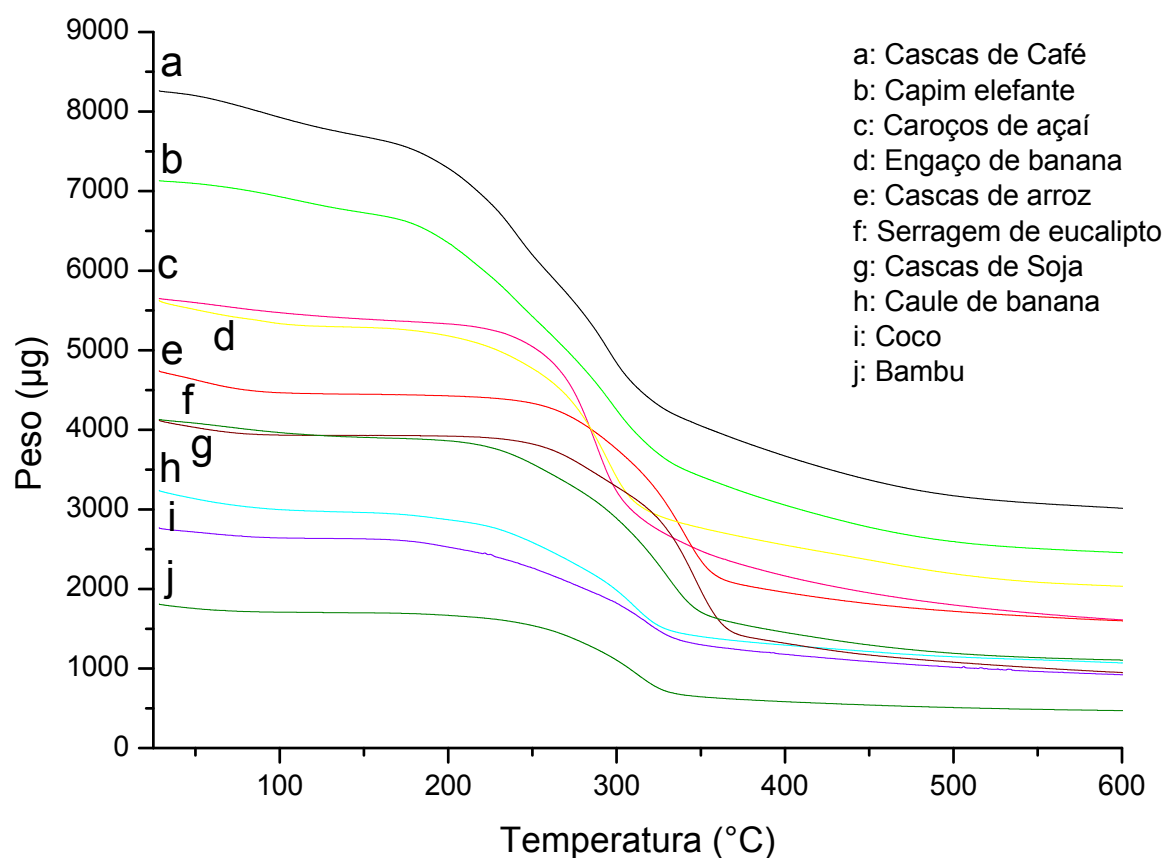


Figura 26. Análises termogravimétricas (TG) das biomassas.

Características semelhantes foram encontradas entre o capim elefante, engaço de banana e as cascas de café; depois entre o bambu, coco, e caule de banana uma certa similariedade foi encontrada; enquanto que as quatro biomassas restantes, caroço de açaí, serragem de eucalipto, cascas de soja e arroz, apresentaram comportamento semelhante.

O primeiro estágio de decomposição ocorre em uma temperatura entre 50 – 100 °C, atribuída principalmente a perda de água (Seye *et al.*, 2003; Naik *et al.*, 2010). Compostos de baixa massa molar e outros compostos orgânicos voláteis (extrativos) evaporam entre 120 – 250 °C, (Naik *et al.*, 2010). No intervalo de 100-200 °C as biomassas estudadas apresentaram uma relativa estabilidade térmica (Figura 26). A temperatura de pirólise para a maioria das amostras ocorre no intervalo de 250-320 °C, correspondendo a aproximadamente de 20 a 40% de perda de massa, sendo relacionada aos componentes orgânicos presentes nos materiais. Nessa etapa do processo a perda de massa geralmente é menor para espécies com alto teor de lignina. As amostras de coco, serragem de eucalipto, cascas de café e cascas de arroz, apresentam os maiores percentuais de lignina total (Tabela 5). Dessa forma essa relação foi verdadeira apenas para as cascas de arroz, serragem e coco, sendo que as cascas de café foram uma das biomassas com maiores perdas de massa nesse intervalo, mesmo associadas com um maior teor de lignina.

A perda de massa principal ocorre entre 250-400 °C para todas as biomassas (Figura 27), associada à decomposição dos açúcares, sendo o primeiro pico atribuído a decomposição da hemicelulose (250-300 °C) e o segundo pico à celulose (300 - 400 °C) (Khalil *et al.*, 2011). Acima de 400 °C, os materiais seguem com uma perda de peso contínua e lenta, atribuída a lignina (Guimarães *et al.*, 2009). As cinzas são consideradas como o resíduo remanescente ao final do processo (800 °C). Uma queda acentuada na massa é geralmente um indicativo de uma grande quantidade de um determinado componente, enquanto que um declínio lento indica a presença de vários componentes voláteis na região.

Uma estimativa da perda de massa das biomassas até a temperatura de 450 °C, indicada como a ideal para a produção de carvão vegetal, mostra que serragem de eucalipto, caule de banana e as cascas de soja tiveram mais de 60% de perda,

enquanto que as demais amostras apresentaram uma perda total de no máximo 53%. Isto demonstra a potencialidade destas últimas biomassas de oferecer uma quantidade maior de resíduo sólido (Seye *et al.*, 2003).

Os resultados da análise da termogravimetria derivada (DTG) são mostrados na Figura 27. Apenas uma etapa do processo é bem nítida, na região de 250 - 400 °C. Na região próxima a 100 °C observa-se um estágio de decomposição o qual está associado à eliminação de umidade, como já citado anteriormente.

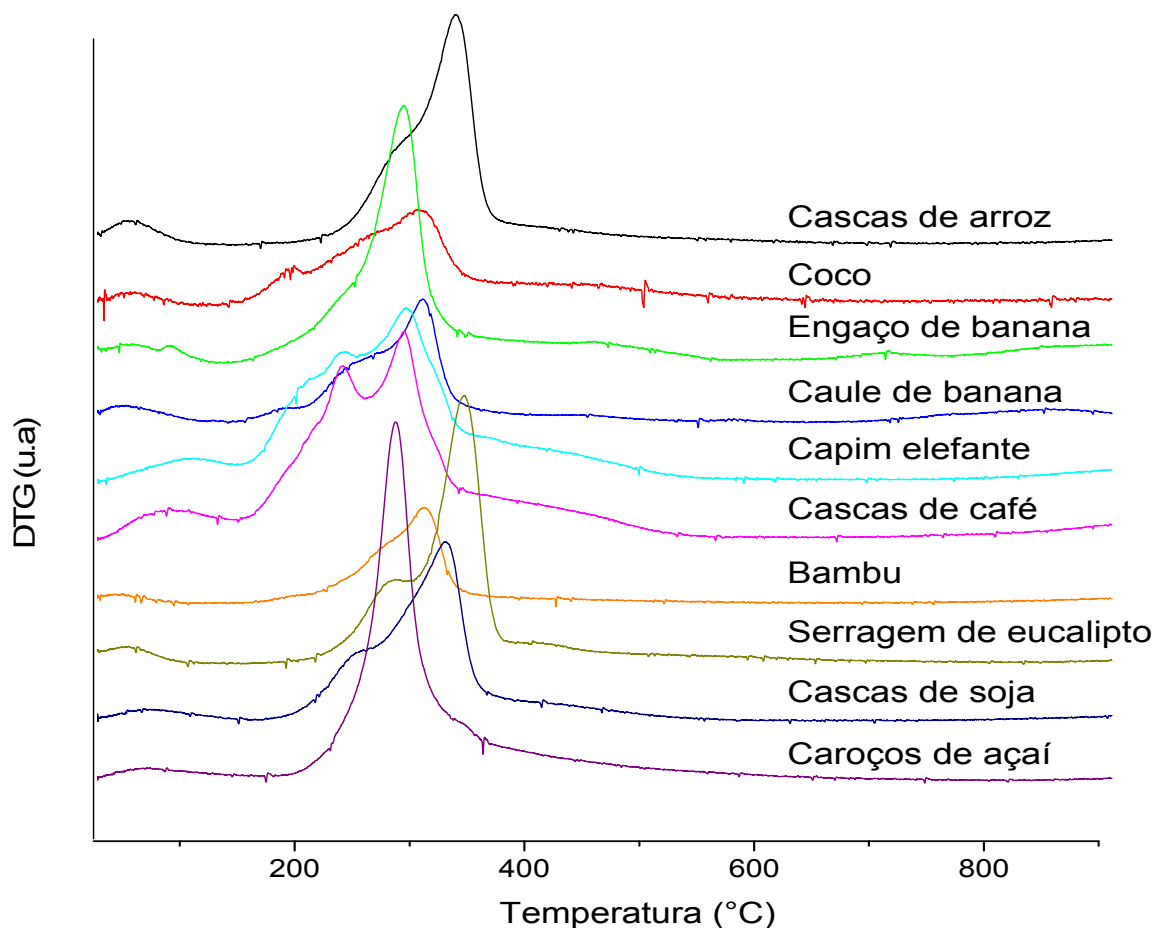


Figura 27. Análises termogravimétricas (DTG) das biomassas.

Os picos entre as diferentes biomassas diferem em posição e amplitude. As perdas de peso nessas regiões (250 - 400 °C), significam as quantidades de hemicelulose e celulose respectivamente, o que está de acordo com os resultados da composição química em carboidratos (veja seção 4.1.7).

Usando as perdas de massa associadas com as regiões identificadas nas Figuras 26 e 27, é possível obter uma aproximação da composição química (Naik *et al.*, 2010, Hayes, 2011) das matérias-primas em estudo (Tabela 7).

Tabela 7. Composição química das biomassas (%) pelos métodos TG e DTG.

Biomassa	Umidade	Extrativos	Hemicelulose	Celulose	Lignina
Caroço de açaí	5,10	12,11	23,09	10,01	35,31
Caule de banana	11,23	20,09	16,36	43,14	23,01
Engaço de banana	7,08	14,31	10,01	20,11	36,07
Bambu	6,12	11,53	16,29	39,08	33,46
Coco	6,23	12,03	15,45	27,01	44,67
Cascas de café	6,23	6,67	12,98	36,02	42,01
Serragem de eucal.	7,56	5,23	10,32	39,04	32,98
Capim elefante	5,07	13,98	22,01	36,12	42,2
Cascas de arroz	8,45	10,12	10,06	26,08	36,11
Cascas de soja	9,30	8,09	17,12	28,23	29,13

Comparando os resultados da TG/DTG com os resultados analíticos, para todas as amostras, os primeiros valores parecem ser um tanto imprecisos, com teores de lignina maiores e teores de hemicelulose inferiores aos analíticos (Tabelas 5 e 6). Os teores de celulose, extrativos e umidade praticamente não se alteram.

O que ocorre é que com as técnicas de TG/DTG a perda de massa entre uma dada faixa de temperatura provavelmente não seja exclusiva de algum composto simples, ainda mais por se tratar de amostras tão complexas e de composição tão heterogênea, e assim, esses componentes individuais podem ser superestimados. Além disso, os mesmos componentes químicos, mas em diferentes matérias-primas podem não ser sempre degradados sob as mesmas faixas de temperatura. Por essas razões a TGA pode ser usada apenas como uma estimativa qualitativa da quantidade de açúcares e lignina presente nas matérias-primas (Hayes, 2011). Sua principal utilização é em procedimentos de rastreio/triagem onde as matérias-primas são em um primeiro momento analisadas por TGA a fim de avaliar se a composição química merece uma análise mais detalhada (principalmente via-química).

4.1.9 Hidrólise ácida e Bioprodutos

As unidades de glicose liberadas pela hidrólise ácida da celulose sofrem desidratação intermolecular catalisada pelo ácido e com isso são convertidas a 5-hidroximetilfurfural (5-HMF). Esse, em uma segunda etapa, devido às condições do meio reacional, se converte em ácido levulínico (principal produto da hidrólise) e também em ácido fórmico (Girisuta *et al.*, 2013). A manose, hexose encontrada em maior quantidade no caroço de açaí, sofre o mesmo processo que a glicose, fornecendo dessa forma ácido levulínico e ácido fórmico em quantidades satisfatórias de 28% e 13%, respectivamente (Figura 28).

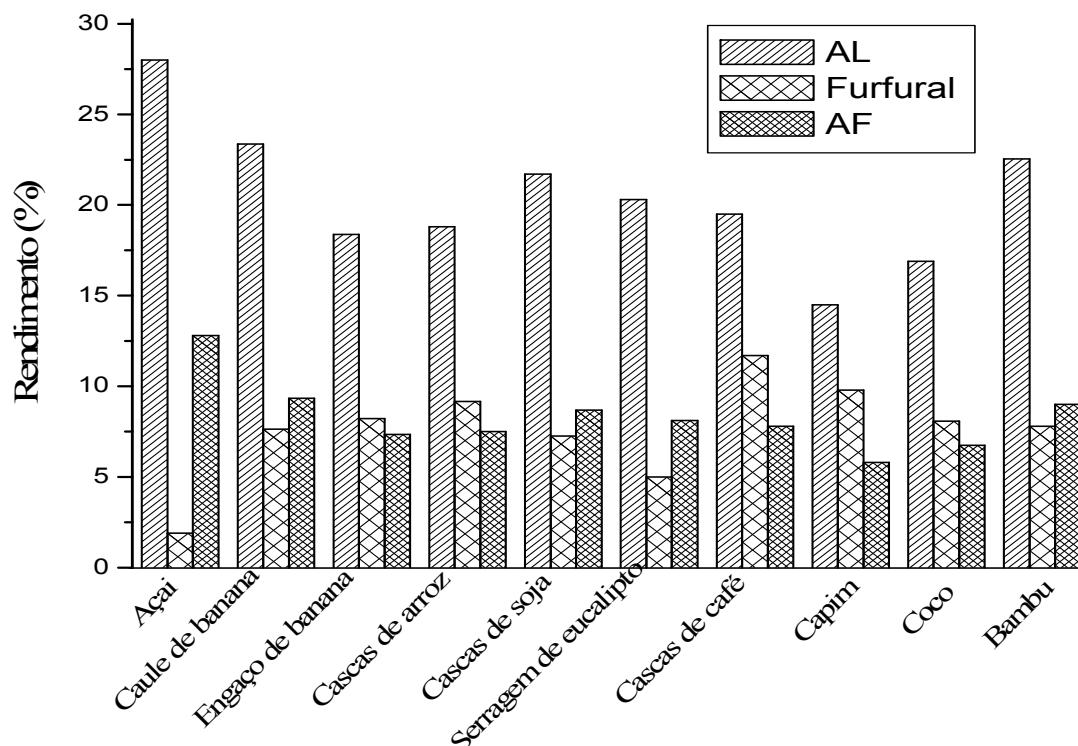


Figura 28. Rendimentos estimados em ácido levulínico (AL), ácido fórmico (AF) e furfural.

Quanto às pentoses, o maior percentual foi encontrado nas cascas de café, com alto teor de xilose, aproximadamente 22%, favorecendo o rendimento em furfural (12%), uma vez que esse é oriundo da ação de ácidos sobre as pentoses e seus polímeros (hemiceluloses). Os menores rendimentos em ácido levulínico e ácido fórmico foram

encontrados para o capim elefante, o que pode ser associado ao seu menor teor de hexoses, enquanto que o menor rendimento em furfural (2,0%) foi encontrado para o caroço de açaí, associável ao seu menor percentual em pentoses.

Uma fração da lignina também se dissolve durante as reações de hidrólise ácida e são identificadas como lignina ácida solúvel, e parte da lignina remanescente é definida como lignina insolúvel. Também reações paralelas indesejáveis podem ocorrer durante a hidrólise ácida com os materiais macromoleculares insolúveis, oriundos da lignina, o que pode afetar o rendimento de bioprodutos durante o processo (Girisuta *et al.*, 2013). Alguns resultados recentes da literatura sobre os rendimentos experimentais de ácido levulínico, ácido fórmico e furfural de diferentes substratos estão listados na Tabela 8.

Tabela 8. Valores da literatura para rendimentos experimentais, em %, de ácido levulínico (AL), ácido fórmico (AF) e furfural (Furf.).

Biomassa	Teor Cel.	Teor Hem.	Rendimento			Referências
			AL	AF	Furf	
Bagaço de cana-de-açúcar	43,3	25,7	63,0*	-	-	Girisuta <i>et al.</i> , 2013
Cascas de arroz	31,0	24,3	59,4*	-	-	Bevilaqua <i>et al.</i> , 2013
Espiga de milho	37,5	30,0	-	-	3,20	Sánchez <i>et al.</i> , 2013
Serragem de eucalipto	43,0	20,9	-	-	4,41	Domínguez <i>et al.</i> , 2013
Sorgo	42,6	26,2	-	1,65	3,12	Banerji <i>et al.</i> , 2012
Serragem de eucalipto	37,8	13,8	-	-	4,30	Gütsch, <i>et al.</i> , 2012
Algodão	42,6	27,2	9,51	3,77	-	Yang <i>et al.</i> , 2013
Sabugo de milho	30,2	31,7	-	-	19,1	Zhang <i>et al.</i> , 2013
Capim	35,2	19,8	-	-	31,4*	Zhang <i>et al.</i> , 2013
Madeira	36,5	18,7	-	-	33,6*	Zhang <i>et al.</i> , 2013
Quenafe	32,0	9,92	15,0	-	-	Ya'aini <i>et al.</i> , 2012
Palha de milho	40,4	25,6	19,9	-	-	Chang <i>et al.</i> , 2007
Cascas de arroz	46,0	10,2	22,8	-	-	Hongzhang <i>et al.</i> , 2011
Sorgo	73,8	-	32,0	-	-	Fang, 2002.

*eficiência do processo; Cel.: Celulose; Hem.: Hemicelulose

Os rendimentos experimentais obtidos de ácido levulínico (da literatura) estão próximos dos valores teóricos esperados para cada biomassa, considerando o teor de celulose. Isso confirma que todas as biomassas residuais analisadas neste trabalho apresentam potencial para a produção de ácido levulínico com rendimentos maiores a 15%. Inclusive, tais rendimentos são superiores a vários trabalhos já publicados como promissores, confirmando assim que todos os resíduos em questão tem potencial para fornecer insumos de grande valia. Já os resultados para furfural baseados nos teores das pentoses de cada biomassa, apresentaram rendimentos significativamente inferiores aos estimados. Provavelmente isso tenha ocorrido devido à degradação facilitada da cadeia da hemicelulose, frente à celulose, levando possivelmente a perda de açúcares da hemicelulose durante a hidrólise ácida, diminuindo assim os rendimentos em bioprodutos oriundos das pentoses.

4.1.10 Difração de raio-X

Ainda avaliando as dez diferentes biomassas, medidas de difração de raios-x foram importantes para identificar a cristalinidade no material antes e após a extração, a fim de avaliar se a extração com etanol tem efeito na remoção da celulose e/ou hemicelulose (Figura 29A-B).

A cristalinidade do material lignocelulósico é considerada como um dos principais fatores que influencia a eficiência da hidrólise ácida/enzimática (Li *et al.*, 2010), sendo que a cristalinidade da celulose pode impedir significativamente a hidrólise se uma concentração muita alta de ácido/enzimas não for utilizada. Dessa forma é de extrema importância conhecer o teor de cristalinidade antes de iniciar as etapas subsequentes, a fim de avaliar as quantidades de reagentes necessárias no meio reacional.

Os picos de difração de raios X mais intensos foram detectados em $2\theta = 22^\circ/23^\circ$ e $15^\circ/17^\circ$, para todas as matérias-primas, antes e após a extração, o que indica a existência de regiões cristalinas. Essas bandas são características de celulose (ICDD; Pereira *et al.*, 2005). Em $18^\circ \leq 2\theta \leq 19^\circ$ tem-se um vale mínimo associado as regiões semicristalinas da celulose (amorfa). Adicionalmente, caroço de açaí, cascas de soja, cascas de arroz e serragem de eucalipto apresentaram também um pico em $2\theta = 34,7^\circ/35^\circ$, associada com a celulose. Os difratogramas para os resíduos de banana

antes da extração mostraram picos finos no intervalo de $2\theta = 28^\circ - 40^\circ$, atribuídos provavelmente a substâncias inorgânicas (cinzas) ou outras possíveis contaminações (Guimarães *et al.*, 2009). Após a extração, foi observado que as amostras extraídas foram semelhantes às matérias-primas brutas, exceto para os resíduos de banana, com um difratograma muito mais limpo, provavelmente devido à remoção das impurezas pelo processo de extração com etanol (Figura 29B).

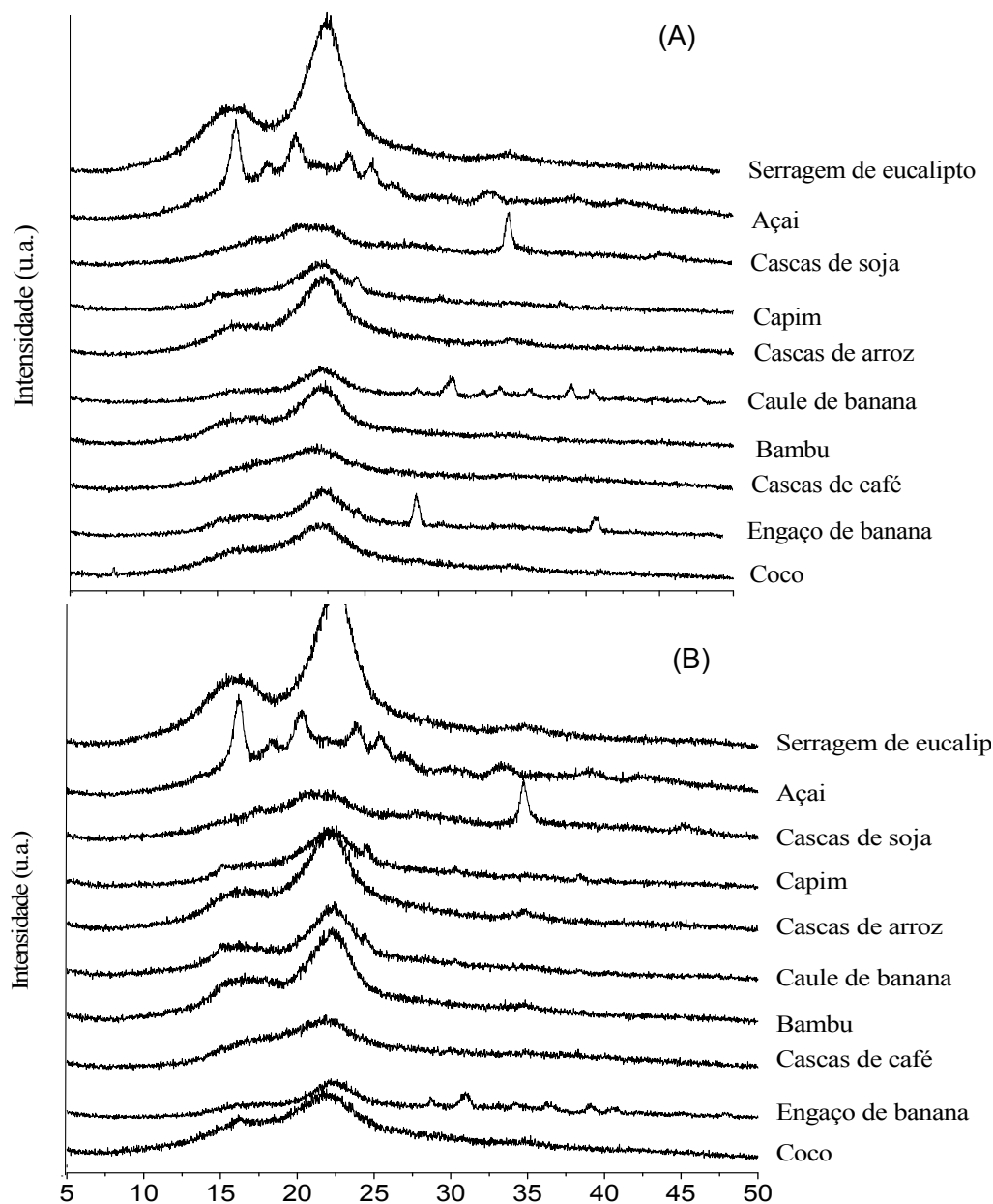


Figura 29. Difratogramas de raios-X das biomassas (A) antes e (B) após a extração.

Como as regiões cristalinas e amorfas da celulose já estão bem estabelecidas na literatura (ICDD; Foner e Adan, 1983; Figueiredo *et al.*, 2010; Park *et al.*, 2010) não houve a necessidade da obtenção de difratogramas de componentes inorgânicos, constituintes possíveis de serem encontrados nas biomassas, e que poderiam de alguma forma ser atribuídos falsamente a celulose cristalina. Além do mais o teor de cinzas totais, não ultrapassou 13% do total da composição das biomassas (Tabela 4), o que corrobora que tais constituintes não poderiam apresentar uma contribuição tão significativa nos difratogramas de raios-X. Outro fato que pode confirmar a ausência desses compostos, é que normalmente cinzas são associadas a picos finos e não bandas alargadas, como as bandas típicas de celulose (Guimarães *et al.*, 2009).

Os teores de cristalinidade foram calculados (Tabela 9), utilizando os métodos da intensidade (Figura 20) e da área (Figura 30). Para fins de comparação com resultados da literatura os teores que seguem na discussão abaixo são com base no cálculo das intensidades.

Tabela 9. Valores dos índices de cristalinidade das biomassas (%).

Biomassa	Cristalinidade pelo método da Intensidade	Cristalinidade pelo método da Área
Caroço de açaí	38	15
Caule de banana	51	36
Engaço de banana	48	36
Bambu	48	43
Coco	42	20
Cascas de café	29	14
Serragem de eucal.	59	46
Capim elefante	47	27
Cascas de arroz	57	44
Cascas de soja	22	14

A cristalinidade máxima foi observada na serragem de eucalipto seguida pelo arroz e caule de banana com 57% e 51%, respectivamente. Na sequência têm-se engaço de banana e bambu com 48% de cristalinidade. Esses resultados são similares

aos apresentados na literatura (Guimarães *et al.*, 2009). Ainda nessa análise, capim e coco apresentaram, respectivamente, 47% e 42% de cristalinidade, estando também de acordo com a literatura (Satyanarayana *et al.*, 2007). O caroço de açaí e as cascas de café apresentaram 38% e 29% de cristalinidade, respectivamente, e o valor mínimo foi obtido para as cascas de soja (22%).

A cristalinidade das biomassas é diretamente afetada pela quantidade de lignina, polioses e ceras (hidrocarbonetos de elevado peso molecular), além da natureza da ligação química entre celulose, hemicelulose e lignina (Zhang *et al.*, 2008). Isto pode ser confirmado, uma vez que após a remoção dos extrativos, o índice de cristalinidade de todas as biomassas foi superior, aumentando de 2% a 11%, exceto para a amostra de coco. Isto ocorre uma vez que a redução de constituintes amorfos das fibras vegetais favorece o aumento da cristalinidade, potencializando a intensidade dos picos da celulose, em função da maior exposição das estruturas celulósicas, e, portanto, refletindo no aumento da cristalinidade (Johar *et al.*, 2012; Jaimes *et al.*, 2013).

As Figuras 30 (A-D) apresentam os difratogramas deconvoluídos de algumas biomassas estudadas neste trabalho.

Os índices de cristalinidade foram também calculados por essa metodologia, apresentando sempre teores inferiores aos do método da intensidade, o que era esperado visto que pela intensidade, o pico (vale) atribuído a estruturas amorfas, é severamente subestimado.

Os difratogramas deconvoluídos sem a presença de extrativos, da mesma forma que pelo método da intensidade, apresentaram resultados superiores quando comparados aos difratogramas na presença de tais constituintes.

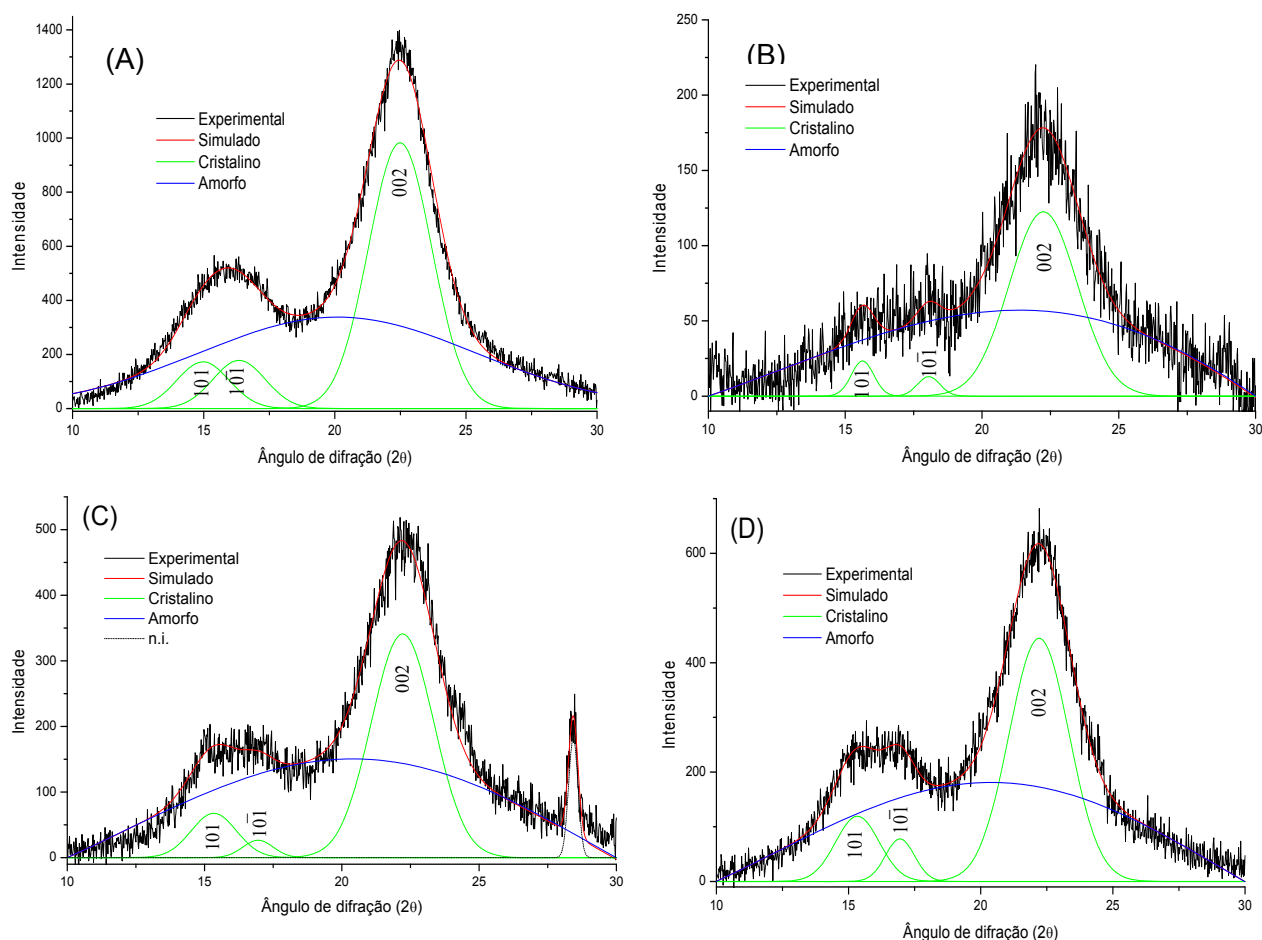


Figura 30. Difrátogramas de raios-x da (A) serragem de eucalipto (B) caroço de açaí (C) engaço de banana e (D) caule de banana. Onde “n.i.” corresponde a picos não identificados e os picos 101; $\overline{101}$ e 002 correspondem aos planos cristalográficos das áreas dos picos cristalinos.

4.1.11 Espectroscopia Vis-NIR

Os espectros médios (média de cada biomassa) completos (Vis-NIR) das biomassas são mostrados na Figura 31. As bandas principais estão na região do visível: uma banda larga na região de 460 nm e outra banda em 670 nm. Essas bandas estão associadas com as ligações π conjugadas dos sistemas cromóforos (Silverstein, 1981), onde a contribuição da lignina (quarto sobretudo do estiramento aril C-H) (Pastore *et al.*, 2008) e principalmente da transição eletrônica da clorofila (Gitelson *et al.*, 2003) são importantes. A banda centrada na região de 670 nm, também atribuída à clorofila, apresenta-se como um “ombro” nas amostras de café e de forma mais evidente para as biomassas de capim-elefante, bambu e coco.

Na região do NIR, os espectros são dominados pelos sinais de água, com uma banda de combinação intensa em 1940 nm e uma banda de estiramento O-H do primeiro sobretom em 1450 nm. No entanto, outros compostos com ligações O-H, tais como os carboidratos e a lignina, também podem contribuir para as referidas regiões. O sinal dos carboidratos pode ser encontrado como uma banda larga em torno de 2100 nm (banda de combinação O-H), mas essa região provavelmente apresenta também sobreposição com as bandas vibracionais de alquil ($C-H_n$), evidenciando a dificuldade da análise direta de espectros NIR sem o auxílio de ferramentas multivariadas. Esses grupos alquílicos apresentam também um par de bandas em 2310/2350 nm (banda de combinação C-H), 1720/1760 nm e uma banda em 1190 nm (primeiro e segundo sobretoms do estiramento C-H, respectivamente). Para a lignina, na região do NIR, bandas específicas (CH ou OH aromático), em 950, 1100 e 1670 nm não estão visíveis, apenas uma banda mal resolvida aparece na região de 2270 nm (Shenk *et al.*, 2008).

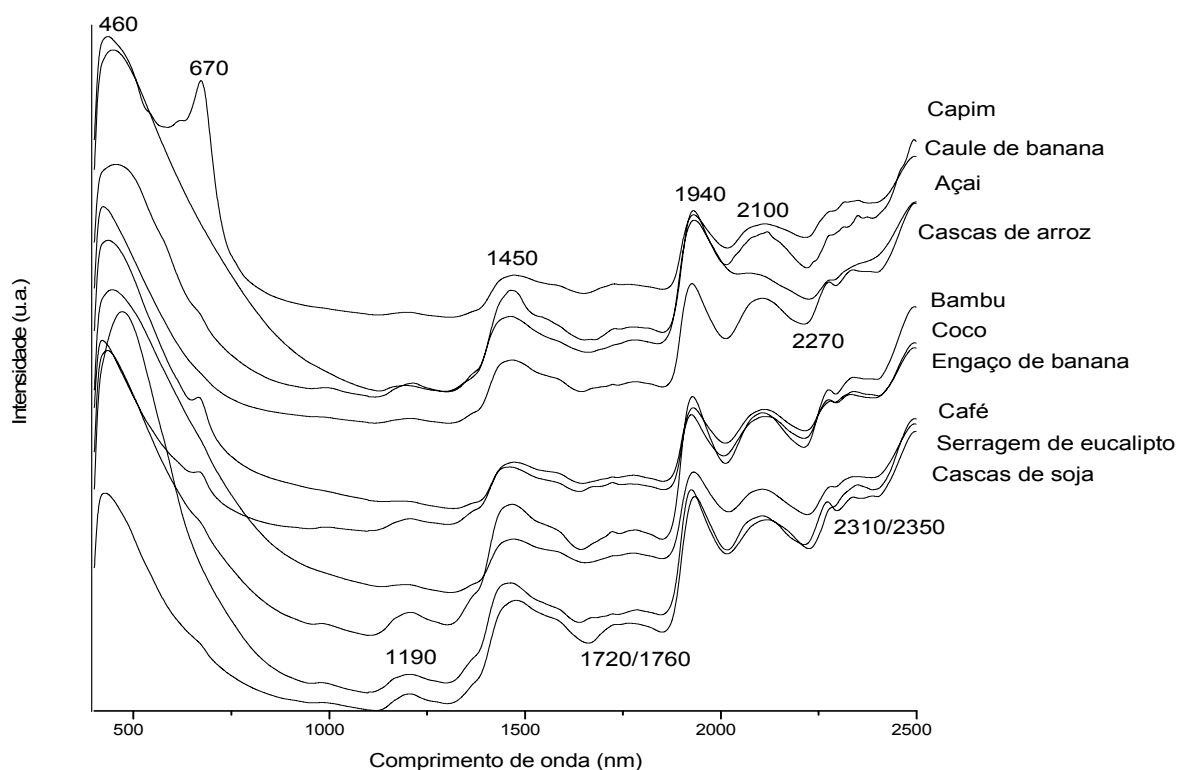


Figura 31. Espectros médios da região Vis-NIR das biomassas.

4.1.12 Conclusões a respeito da seleção das biomassas

Tendo em vista a avaliação físico-química exposta, alguns critérios devem ser levados em conta para a definição das biomassas ideais para a produção de bioenergia. Essas devem apresentar alto poder calorífico juntamente com alto teor de celulose e hemicelulose; e baixo teor de cinzas, extrativos e umidade. No entanto, a seleção de matérias-primas sustentáveis não deve se basear apenas na caracterização físico-química, uma vez que vários fatores como já citado anteriormente devem ser levados em conta. Os resultados indicaram que as cascas de arroz e o capim-elefante apresentam percentuais de cinza acima de 10%, sendo por tal motivo inapropriadas. O caroço de açaí apresenta um percentual de pentoses muito baixo, gerando baixo rendimento em furfural ($< 5\%$). Do ponto de vista da caracterização, o bambu e a serragem de eucalipto se encaixam perfeitamente no contexto de biorrefinarias, porém, eles possuem inúmeras utilidades, em indústria de móveis, etc, quando comparados a outros resíduos sem fim adequado e específico. Por sua vez, as cascas de soja encontram aplicação na alimentação animal.

Portanto, a partir de todas as biomassas analisadas, as cascas de café e as cascas de coco, com baixos percentuais de extrativos e cinzas, e elevado teor de açúcar e poder calorífico, são os mais apropriados resíduos para biorrefinarias. Resíduos de banana, devido ao seu baixo teor de lignina e significativo percentual de carboidratos, são também consideradas matérias-primas em potencial. Com relação ao seu descarte, para cada tonelada de banana industrializada, aproximadamente 3 toneladas de pseudocaule, 160 kg de engaços, 480 kg de folhas e 440 kg de cascas são gerados (Arredondo *et al.*, 2010). Assim essa cultura gera muitos resíduos, além de ser uma matéria-prima produzida em todos os estados da federação (Embrapa, 2012), o que pode beneficiar o país como um todo.

Após a avaliação físico-química, econômica e ambiental (realizada pela Fundação Chile, integrante do projeto DIBANET) das dez biomassas, três foram selecionadas dentro do contexto de biorrefinarias e atendendo as exigências do projeto DIBANET. Essas biomassas (cascas de café, cascas de coco e resíduos de banana) foram utilizadas para o desenvolvimento de modelos de calibração multivariada da composição química.

4.2 Análises de referência para construção de modelos de calibração multivariada

Os conjuntos de amostras apresentaram uma ampla faixa de variação dos parâmetros estimados (Figura 32).

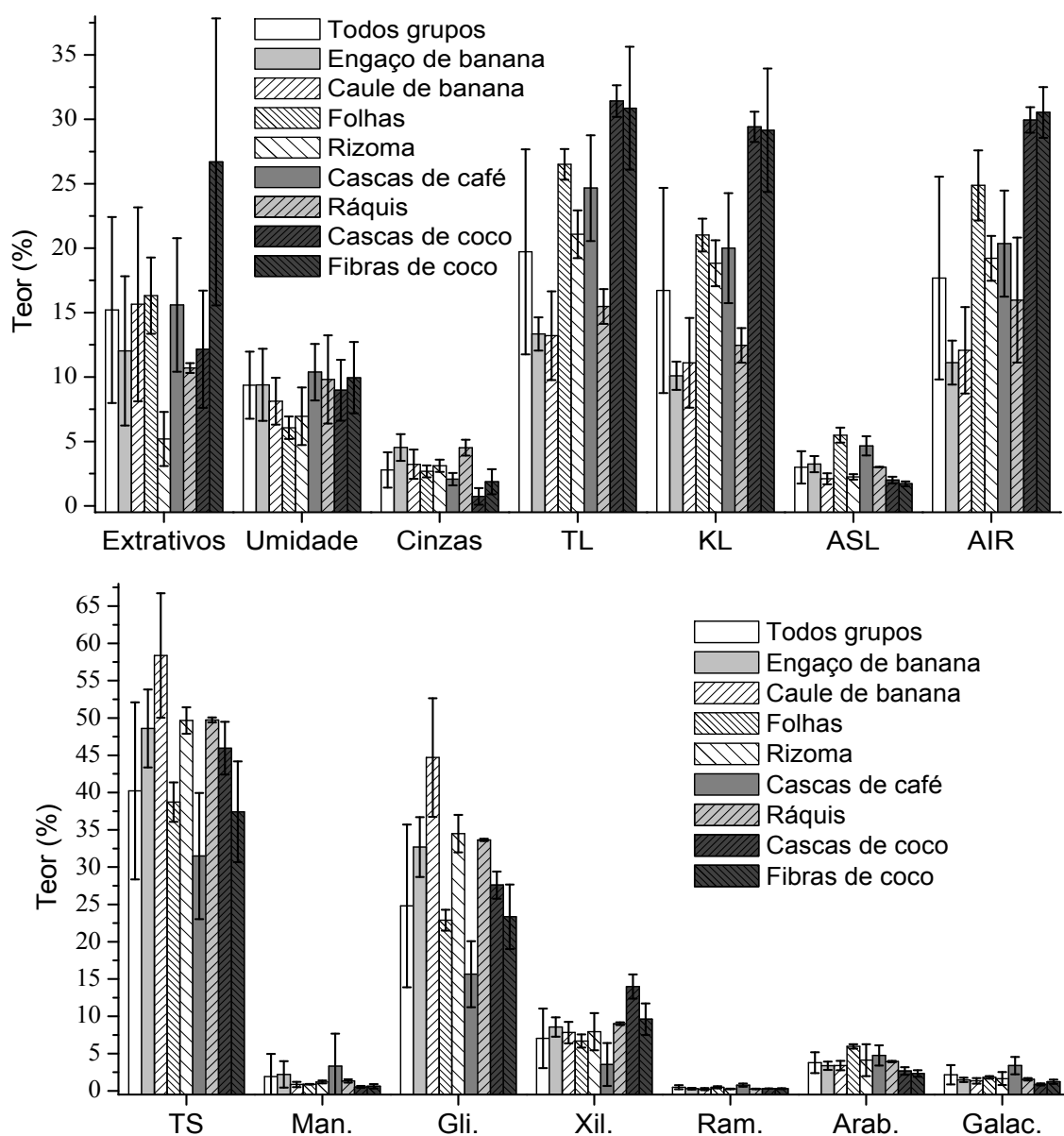


Figura 32. Análise descritiva e estatística dos parâmetros (TL: lignina total; KL: lignina Klason; ASL: lignina ácida solúvel; AIR: resíduo ácido insolúvel; Todos grupos: todas as frações botânicas; TS: açúcares totais; Man: manose; Gli: glicose; Xil: xilose; Ram: ramnose; Arab: arabinose; Galac: galactose).

O maior intervalo, em valores absolutos, foi observado para os extrativos (1,41 – 41,6%), seguido de açúcares totais (19,84 – 54,71%) e lignina total (8,15 – 35,8%). Por outro lado, o menor intervalo foi para os açúcares ramnose (0,14-1,2%) e arabinose (1,40-7,60%), seguidos de lignina ácida solúvel (0,68 – 6,3%) e cinzas (0,33 – 7,1%). Essa alta variabilidade de composição foi obtida devido ao fato de que populações muito heterogêneas, como diversas biomassas, café, coco e banana, de diferentes frações botânicas, espécies, cultivares, local de coleta etc, foram amostradas.

Dentre as biomassas estudadas, as cascas de café apresentam composição química com menor variabilidade, tendo em vista os menores desvios padrões mostrados na Figura 32. Além disso, as variações das frações botânicas na banana foram observadas como sendo maiores do que das outras cultivares, sugerindo que a composição da banana é mais heterogênea.

Uma grande variação foi observada entre as diferentes frações botânicas de cada espécie, e assim a mera seleção manual dessas diferentes frações pode fornecer eficientemente um conjunto de amostras com uma grande variabilidade. Essa ampla variação é extremamente desejável para a modelagem quimiométrica na construção de modelos de calibração.

4.3 Análise por Componentes Principais (PCA)

Os resultados da PCA dos espectros brutos de banana, coco e café, centrados na média, com rotação varimax são mostrados na Figura 33.

As duas primeiras PC explicam 39 e 13% da variância total, respectivamente. A região do espectro visível foi importante para todas as PC, com pesos elevados nessa região, exceto para PC1 (Figura 33A). Isso ocorre devido a uma característica intrínseca da espectroscopia utilizada, com uma maior absorvidade molar nas transições eletrônicas do que nas transições vibracionais proibidas da região NIR. Dessa maneira, a mesma variabilidade na composição das amostras resulta em uma maior variabilidade na região Vis do que no NIR (Silverstein, 1981), como ocorreu para a PC2, PC3, PC4 e PC5, com 51% da variância acumulada. A PC1 é caracterizada por pesos negativos em torno de 1900 nm, típico de estiramento OH do primeiro sobretom, provavelmente de celulose (Ali *et al.*, 2001), indicando que as cascas de café,

apresentam menores percentuais de celulose, do que as frações de coco e banana, o que é corroborado pelos dados analíticos (método de referência) apresentados na Figura 32, onde se observa um maior valor médio de glicose para o coco e a banana, do que para o café. Por outro lado, a PC2 foi caracterizada por pesos positivos em 677 nm, típico de bandas de clorofila (Gitelson *et al.*, 2003) e, conforme esperado, essa PC diferencia amostras de folhas, ricas nesses pigmentos fotossintéticos (Figura 33B). A PC3 apresenta pesos positivos em 562, 730 e 2100 nm, associados respectivamente à clorofila, estiramento ROH do terceiro sobretom e banda de combinação de ROH. Esses últimos estão associados aos carboidratos, o que é confirmado pela análise de escores positivos para a maior parte das amostras de coco e banana, as quais apresentam os percentuais mais elevados de celulose e hemicelulose, e também maior quantidade de clorofila (Figura 33C).

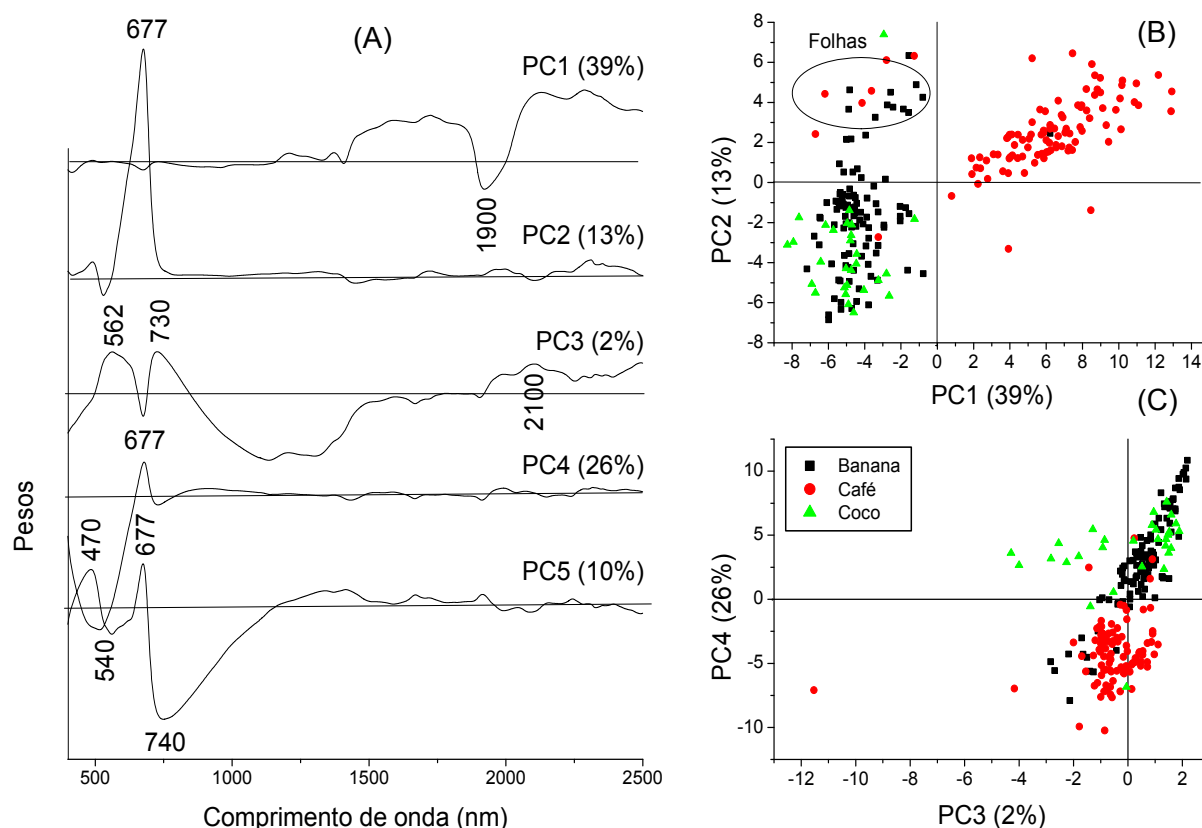


Figura 33. (A) Gráfico de pesos da análise PCA. (B) Gráfico de escores das duas primeiras componentes principais. (C) Gráfico de escores da terceira e da quarta PC para os conjuntos de banana, coco e café.

A PC4 apresenta uma banda negativa em 540 nm, típica de pigmentos acessórios de plantas (antocianinas, carotenóides, etc) (Gitelson *et al.*, 2003). Essa PC diferencia amostras de café com escores negativos (Figura 33C), indicando que elas apresentam altos teores de carotenóides quando comparado às amostras de banana e coco. A PC5 é caracterizada por pesos positivos em 470 nm e 677 nm, bandas típicas de clorofila A e B, novamente associada às amostras de folhas. Essa PC foi capaz também de diferenciar o teor de clorofila para as amostras de coloração verde como a banana e coco, enquanto o café com escore negativo apresenta baixo percentual de clorofila, o que era esperado, uma vez que não apresenta essa coloração típica. Esses resultados mostram a habilidade da PCA no auxílio na atribuição de bandas e na interpretação dos dados.

Além do mais, esse estudo demonstra a possibilidade da combinação das diferentes biomassas em um único modelo de calibração. É possível observar através dos escores que é difícil uma distinção entre banana e coco, o que justifica a semelhança química entre esses dois grupos, permitindo assim a construção de um único modelo de calibração para banana e coco. Além disso, as amostras de café se distanciaram significativamente dos demais grupos, e então modelos foram desenvolvidos combinando o café com as demais matérias-primas a fim de avaliar se mesmo assim essa combinação é possível, sem prejudicar a estabilidade dos modelos. Dessa forma modelos de calibração foram obtidos com uma maior amplitude de variação e os mais genéricos possíveis.

Já os resultados da PCA dos espectros brutos somente das amostras de banana (diversas frações botânicas), são mostrados na Figura 34, com os espectros centrados na média.

As três primeiras PC's explicam 94% da variância total (Figura 34A). A PC1 é caracterizada por pesos positivos em torno de 1906 e 1406 nm, típicos de frações cristalinas (Zidan *et al.*, 2012), e uma banda negativa em 1488 nm associada a celulose semicristalina (Tsuchikawa e Siesler, 2003b). Dessa forma a PC1 diferenciou amostras de bananas (geralmente caule e engaço) com maiores percentuais de celulose cristalina (Figura 34B). Por outro lado, a PC2 foi caracterizada por pesos positivos em 1704 e 2064 nm, típicos de componentes amorfos presentes na biomassa, atribuídos

principalmente às amostras de folhas e rizomas. Ainda uma banda negativa em 1406 nm típica de celulose cristalina mostra a relação inversa da PC2 com o teor de cristalinidade. Também a PC3 da mesma forma que a PC2 apresentou bandas amorfas características em 1340, 1720, 1914 e 2262 nm (Zidan et al., 2012), indicando que ambas separaram amostras com perfil mais amorfo do que cristalino (Figura 34C).

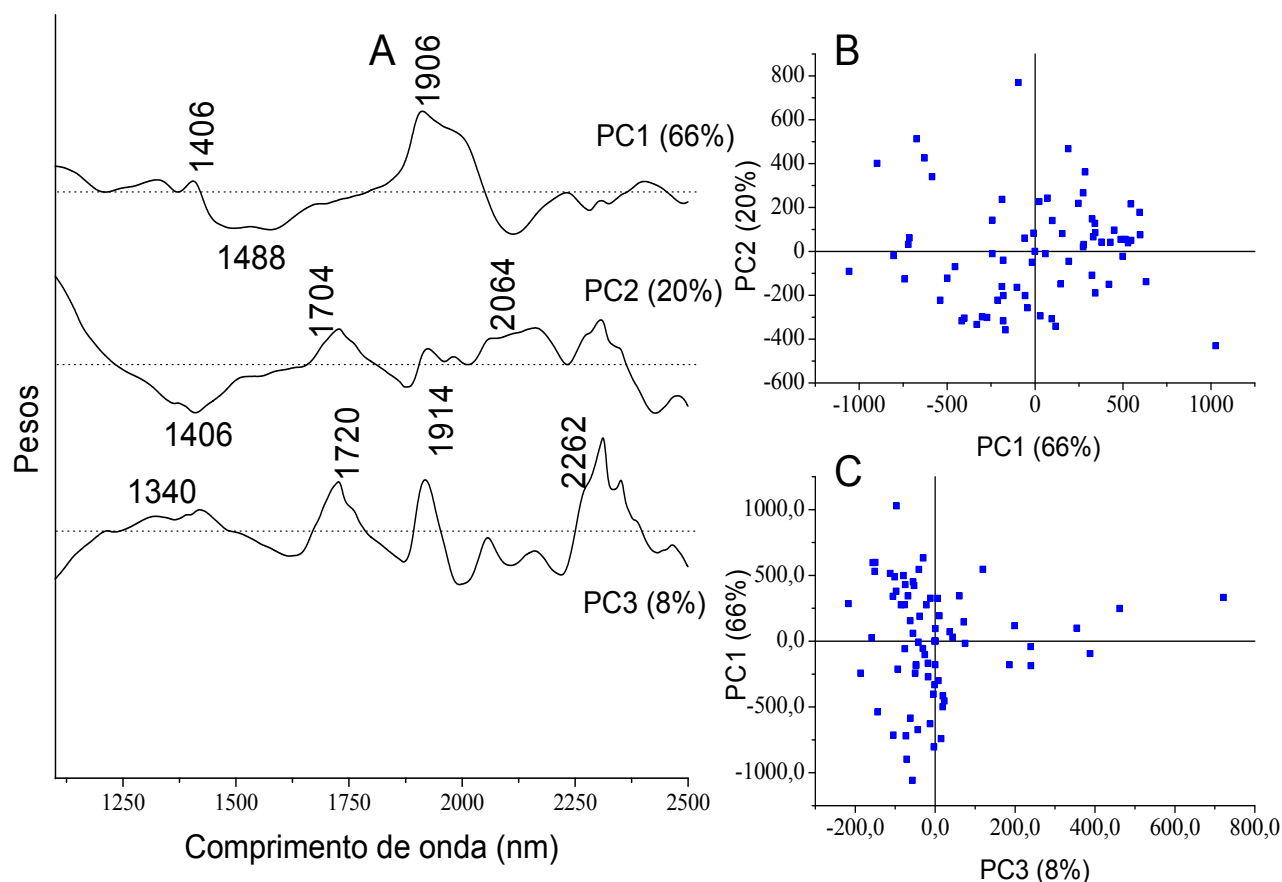


Figura 34. (A) Gráfico de pesos da análise PCA. (B) Gráfico de escores das duas primeiras componentes principais. (C) Gráfico de escores da terceira e da primeira PC para o conjunto de banana.

Dessa forma a análise por PCA mostrou ser uma ferramenta muito útil e simples para a identificação das amostras com altos teores de cristalinidade, contribuindo para as etapas seguintes do processo, nas quais o conhecimento da cristalinidade é fundamental para obtenção de bons resultados.

4.4 Modelos de regressão multivariada

Os resultados obtidos para os modelos construídos com o método PLS para a fração DS são mostrados na Tabela 10, e os espectros brutos da região Vis-NIR contendo todas as amostras de banana, café e coco, e suas diferentes frações botânicas, são mostrados na Figura 35.

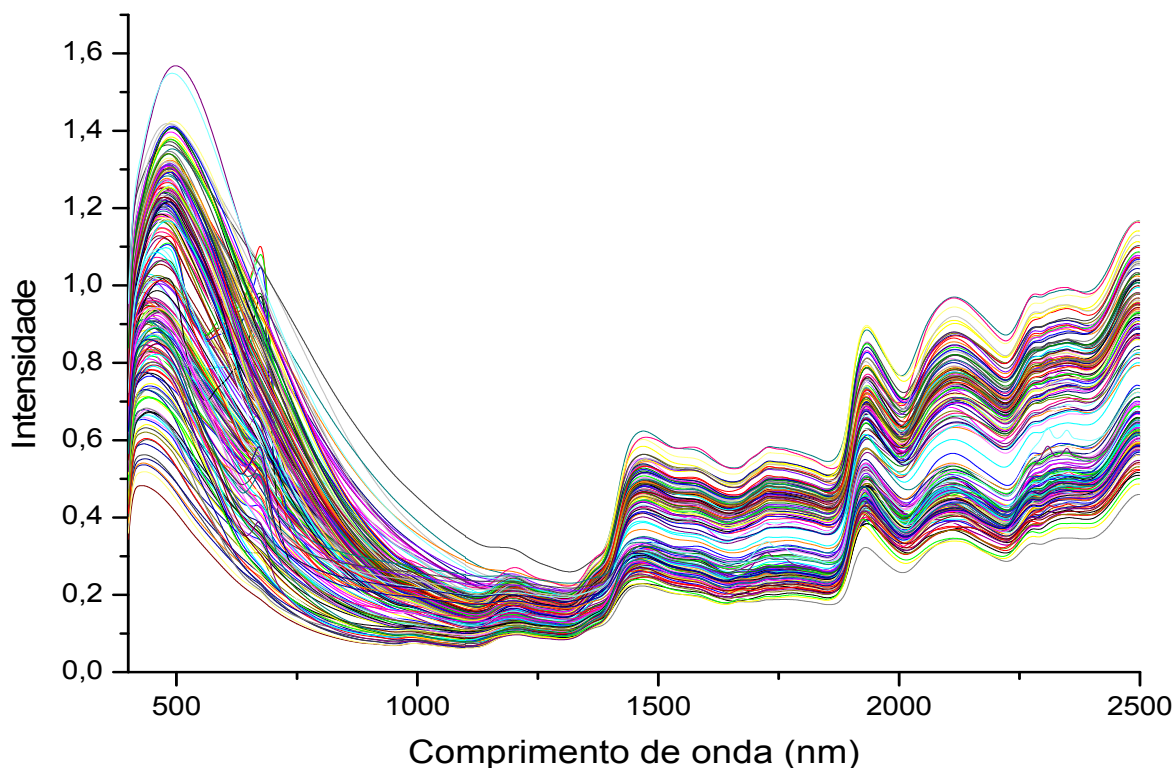


Figura 35. Espectros brutos da região Vis-NIR das amostras de banana, coco e café da fração DS.

O uso do espectro inteiro forneceu melhores resultados para as ligninas (lignina total, lignina klason, resíduo ácido insolúvel e lignina ácida solúvel) nos conjuntos de Banana+Café e de Banana+Café+Coco. Isso ocorreu, provavelmente devido à ocorrência de ligações π conjugadas no sistema de cromóforos (auxocromos e cromógenos) na estrutura da lignina que absorvem na região do visível (Pastore *et al.*, 2008). Outra característica comum à modelagem da lignina foi o uso de transformação OSC nos modelos de Banana e Banana+Café, no entanto, ao incluir as amostras de coco, outras transformações mais simples, que são independentes do conjunto de

dados, tais como derivadas, forneceram resultados equivalentes e portanto foram utilizadas. Esse pré-tratamento (OSC) tem o objetivo de eliminar a variação sistemática da matriz **X** (espectros) que não está relacionada, ou seja, que é ortogonal, a **y** (teor de ligninas) (Fearn, 2000; Wold *et al.*, 1998). Dessa forma, quando o primeiro fator(s) captura grande porcentagem de variação em **X** (> 80%) e menor em **y** (< 15%), o uso do pré-processamento por OSC é aconselhável. Apesar da transformação OSC frequentemente gerar modelos menos robustos, no presente caso isso não foi observado (veja mais adiante no item 4.6, a Figura 55).

A Figura 36 apresenta os espectros da região NIR (A) e Vis-NIR (B) dos modelos da fração DS de lignina total e lignina Klason para o conjunto da Banana e o conjunto da Banana+Café, respectivamente, ambos submetidos à transformação por OSC. A filtração, ou seja, removendo componentes do espectro bruto (modelando e removendo 3 e 5 componentes principais ortogonais a **y**, respectivamente), de cada modelo, fornece significativamente diferentes aparências para os espectros corrigidos por OSC quando comparados aos espectros brutos (Figura 35).

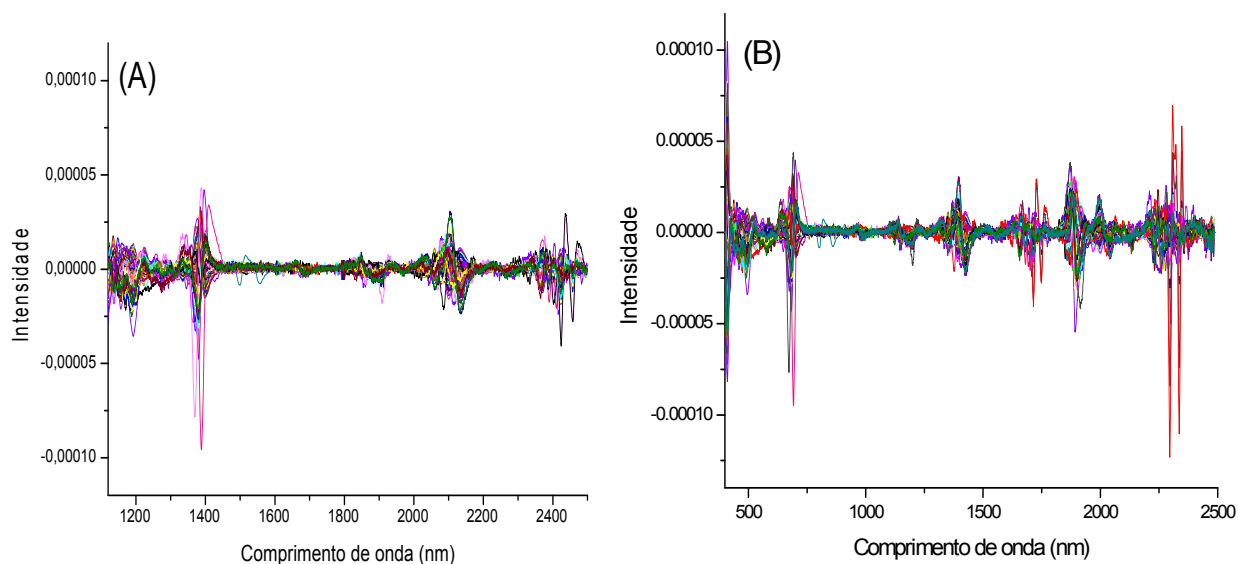


Figura 36. Espectros pré-tratados por OSC para os modelos (A) da região NIR de lignina total do conjunto da Banana e (B) da região Vis-NIR de lignina Klason do conjunto Banana+Café, ambos da fração DS.

Tabela 10. Parâmetros para avaliação e validação dos modelos PLS para a fração DS.

y	Conjunto de dados	Pré-Tratamento	Região Espectral	Tamanho Matriz	VL	Amostras Anômalas	R ²		RMSE		RE	RER
							Cal	Val	Cal	Pred		
TL	B	2D(15)+OSC	NIR	83x2800	3	-	0,96	0,93	0,8830	1,142	7,80	12,3
	C	2D(15)	NIR	42X2800	4	5	0,89	0,80	1,637	1,904	7,00	10,0
	COC*	2D(15)	NIR	30X2800	4	2	0,86	0,82	1,301	1,403	5,70	5,00
	B+C	2D(15)+OSC	Todo	125x4200	4	-	0,90	0,87	1,957	2,095	11,8	10,0
	B+C+COC	2D(15)	Todo	155X4200	7	5	0,94	0,94	1,499	1,585	9,00	12,5
KL	B	2D(15)+OSC	NIR	83x2800	3	1	0,96	0,92	0,7850	1,170	10,0	12,0
	C	2D(15)	NIR	42X2800	6	6	0,91	0,90	1,177	1,321	6,00	7,80
	COC*	2D(15)	NIR	30X2800	7	2	0,92	0,94	1,124	1,047	4,00	15,0
	B+C	2D(15)+OSC	Todo	125x4200	4	-	0,88	0,85	1,918	2,103	11,8	10,0
	B+C+COC	2D(15)	Todo	155X4200	7	5	0,93	0,92	1,622	1,814	12,9	12,0
ASL	B	2D(15)+OSC	NIR	83x2800	3	2	0,92	0,95	0,2551	0,2850	10,0	14,2
	C	2D(15)	NIR	42X2800	5	5	0,81	0,75	0,4772	0,8063	14,0	7,40
	COC*	2D(15)	NIR	30X2800	3	2	0,84	0,86	0,1513	0,1193	7,70	10,0
	B+C	2D(15)+OSC	Todo	125X4200	6	-	0,83	0,80	0,5283	0,5423	12,0	7,00
	B+C+COC	2D(15)	Todo	155X4200	7	4	0,90	0,85	0,4153	0,4751	17,0	7,50
AIR	B	2D(15)+OSC	NIR	83x2800	4	-	0,97	0,93	0,7980	1,267	6,00	16,0
	C	2D(15)	NIR	42X2800	5	4	0,89	0,87	1,568	1,498	6,90	11,0
	COC*	2D(15)	NIR	30X2800	7	2	0,93	0,90	1,044	1,405	5,00	11,0
	B+C	2D(15)+OSC	Todo	125x4200	4	-	0,86	0,83	2,041	2,263	13,5	8,00
	B+C+COC	2D(15)	Todo	155X4200	6	4	0,92	0,89	1,608	2,101	14,0	10,0
CINZAS	B	2D(5)	NIR	104x4200	7	4	0,76	0,70	0,5510	0,7166	18,0	5,00
	C	2D(15)	NIR	102X2800	6	6	0,77	0,60	0,2072	0,3529	17,0	7,50
	COC*	2D(3)	NIR	30X2800	4	1	0,86	0,86	0,3412	0,4002	20,0	6,20
	B+C	2D(3)+SNV	NIR	206x2800	7	3	0,83	0,80	0,5260	0,5561	18,0	3,00
	B+C+COC	2D(15)	NIR	236X2800	6	5	0,80	0,75	0,5870	0,6190	22,0	9,00
UMID	B	2D(3)	NIR	104x2800	6	-	0,86	0,86	0,7940	0,9850	12,0	8,10
	C	2D(15)	NIR	102X2800	5	1	0,84	0,83	0,9341	1,0980	11,0	8,50
	COC*	2D(3)	NIR	30X2800	3	3	0,91	0,85	0,6611	0,8111	8,00	7,30
	B+C	2D(3)+SNV	NIR	206x2800	7	1	0,81	0,79	0,9478	1,063	11,0	10,5
	B+C+COC	2D(15)	NIR	236X2800	7	5	0,82	0,80	1,038	1,335	13,0	13,0
EXTR	B	SNV+1D(3)	NIR	104X2800	4	2	0,90	0,91	2,060	1,869	15,0	8,30
	C	SNV+1D(3)	NIR	102X2800	7	6	0,94	0,87	0,9630	1,961	11,0	10,4
	COC*	SNV+1D(3)	NIR	30X2800	6	2	0,95	0,97	2,364	2,094	12,0	17,0
	B+C	SNV+1D(3)	NIR	206X2800	7	-	0,85	0,86	2,351	2,255	15,0	9,00
	B+C+COC	SNV+1D(3)	NIR	236X2800	7	7	0,86	0,82	2,515	3,203	19,0	12,0

GLIC	B*	2D(15)+SNV	NIR	30X2800	3	3	0,89	0,89	2,880	2,849	7,70	8,50
	C**	2D(15)	NIR	42X279	5	4	0,81	0,78	1,467	1,594	10,5	7,00
	COC**	2D(15)	NIR	30X1653	4	2	0,87	0,83	1,358	1,344	5,40	8,60
XIL	B*	2D(15)+SNV	NIR	30X2800	6	3	0,88	0,88	0,4500	0,4210	4,80	9,50
	C**	2D(15)+SNV	NIR	42X279	7	2	0,87	0,89	0,4241	0,4114	13,0	11,2
	COC**	2D(15)	NIR	30X1653	2	2	0,85	0,84	1,027	1,132	9,00	6,00
GALA	B*	2D(15)+SNV	NIR	30X2800	4	1	0,86	0,85	0,1362	0,1631	11,0	7,05
	C**	2D(15)+SNV	NIR	42X279	5	2	0,91	0,88	0,3530	0,7180	16,0	7,07
	COC**	SNV+1D(1)	NIR	30X623	4	5	0,85	0,80	0,0887	0,0971	9,00	8,00
RAM	B*	2D(15)+SNV	NIR	30X2800	4	4	0,86	0,80	0,0412	0,0649	2,00	7,00
	C**	SNV+1D(3)	NIR	42X279	7	1	0,86	0,83	0,0512	0,0765	7,00	8,00
	COC	-	-	-	-	-	-	-	-	-	-	-
ARAB	B*	2D(15)+SNV	NIR	30X2800	7	3	0,93	0,90	0,2390	0,2940	8,00	12,0
	C**	SNV+DT	NIR	42X279	7	4	0,76	0,72	0,6285	0,7368	17,0	6,00
	COC**	SNV+1D(1)	NIR	30X623	6	4	0,86	0,83	0,1721	0,1742	7,00	8,00
MAN	B*	2D(15)+SNV	NIR	30X2800	7	4	0,86	0,87	0,1224	0,1456	14,5	8,00
	C**	2D(15)+SNV	NIR	42X279	3	1	0,88	0,91	1,302	0,9670	25,0	10,0
	COC	-	-	-	-	-	-	-	-	-	-	-
TS	B*	2D(15)+SNV	NIR	30X2800	4	2	0,88	0,91	2,873	2,346	4,70	12,0
	C**	2D(15)	NIR	42X279	4	7	0,86	0,83	3,628	3,822	10,0	10,5
	COC**	2D(15)	NIR	30X1653	2	2	0,86	0,84	2,277	2,507	6,00	10,5
IC	B ¹	2D(31)	NIR	69x2800	7	3	0,89	0,86	2,557	2,752	6,00	12,4
	B ²	2D(31)	NIR	69x2800	7	4	0,82	0,75	3,195	4,553	10,0	7,00

VL: número de variáveis latentes; RE: erro relativo (RMSEP/média); RER: $y_{\text{máximo}} - y_{\text{mínimo}} / \text{SEP}$; o número entre parenteses indica o tamanho das janelas quando do uso das derivadas; B:Banana; C:Café; Coc: Coco; B+C: Banana+Café; B+C+Coc: Banana+Café+Coco; TL: lignina total; KL: lignina klason; ASL: lignina ácida solúvel; AIR: resíduo ácido insolúvel; Umid: umidade; Extr: extrativos; Glic: glicose; Xil: xilose; Gala:galactose; Ram: ramnose; Arab:arabinose; Man: manose; TS: açúcares totais; IC: Índice de cristalinidade; 1: Índice de cristalinidade pela intensidade; 2: Índice de cristalinidade pela área; *: validação cruzada; **: validação cruzada com seleção de variáveis para os carboidratos (as principais regiões utilizadas foram: 2222-2500 nm; 2000-2083 nm; 1923-1818 nm; 1470-1380 nm).

As Figuras 37 (A-D) mostram as respectivas curvas de calibração e validação externa, para os modelos da fração DS dos conjuntos de Banana+Café, para as ligninas – lignina total, lignina Klason, resíduo ácido insolúvel e lignina ácida solúvel.

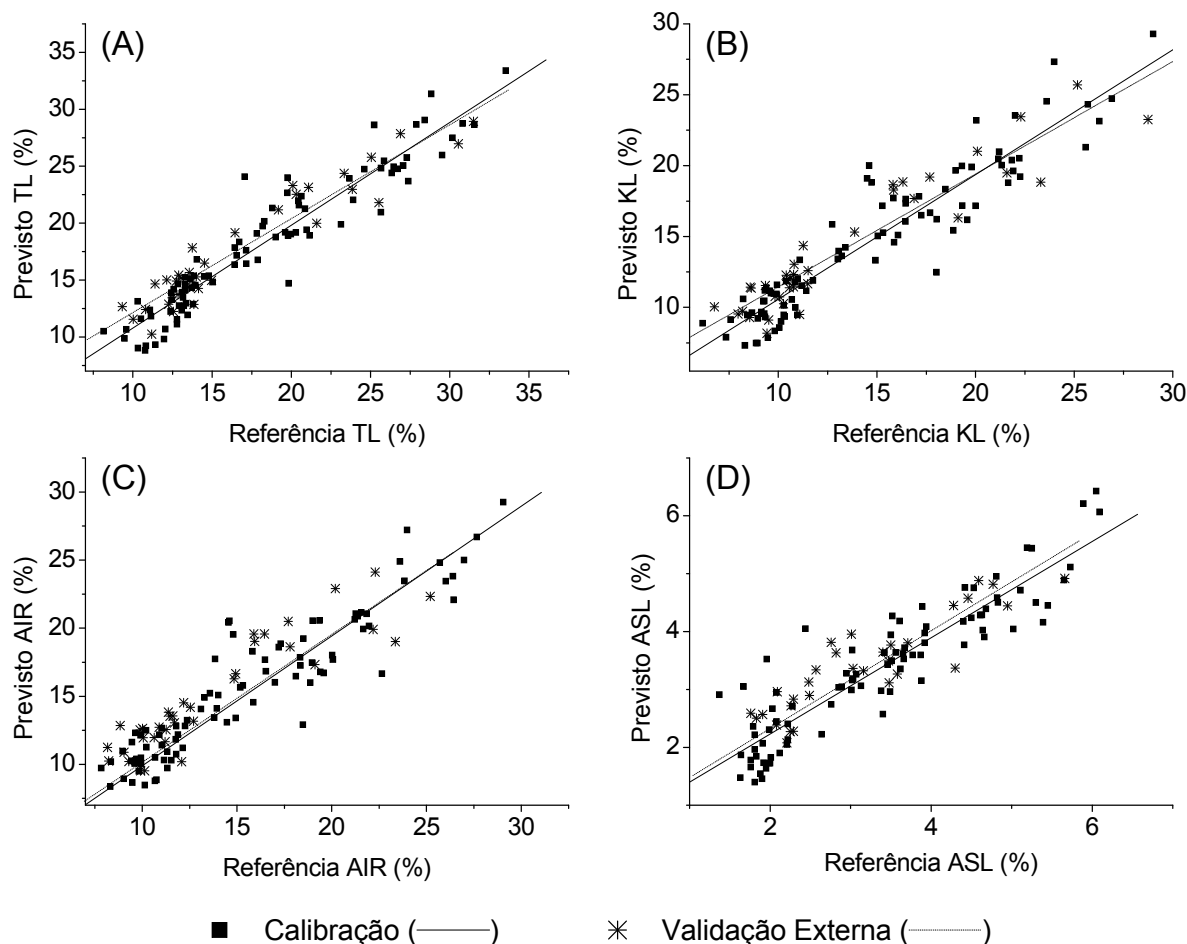


Figura 37. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) lignina total (TL) com 4 VL, (B) lignina ácida solúvel (ASL) com 6 VL, (C) resíduo ácido insolúvel (AIR) com 4 VL, e (D) lignina Klason (KL) com 4 VL, para os modelos de Banana + Café da fração DS com o pré-tratamento indicado na Tabela 8: (2D (15) +OSC).

As concordâncias entre os valores de referência e os valores previstos, tanto para os conjuntos de calibração e de validação externa indicam que os modelos são bons. Assim sendo, evita-se o sobreajuste com muitas VL, situação na qual se obtêm baixos erros no conjunto de calibração e erros elevados no conjunto de previsão.

Os mesmos modelos das ligninas do conjunto de Banana + Café pré-processados agora apenas com 2D (15) (Figura 38), apresentaram R^2 de validação e

calibração inferiores aos modelos usando OSC (Figura 37), e ainda um maior número de VL. Dessa forma percebe-se a necessidade desse pré-tratamento na melhoria dos modelos, ao menos na modelagem das ligninas dos modelos de Banana+Café, bem como do modelo de Banana.

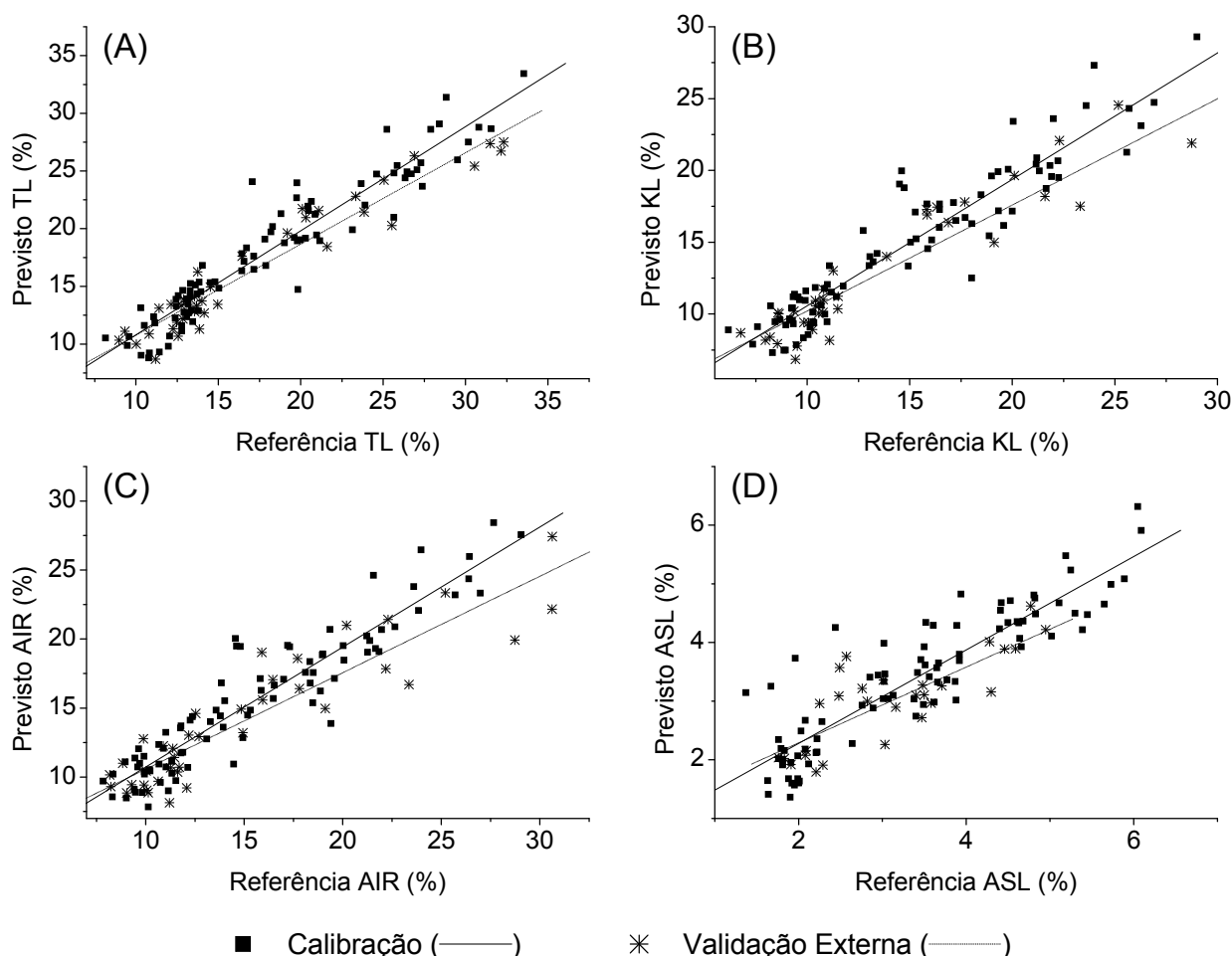


Figura 38. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) lignina total (TL) com 5 VL, (B) lignina Klason (KL) com 5 VL, (C) resíduo ácido insolúvel (AIR) com 7 VL, e (D) lignina ácida solúvel (ASL) com 6 VL, para os modelos de Banana + Café da fração DS com o pré-tratamento 2D (15).

Modelos de Banana+Café e Banana+Café+Coco (Figura 39) utilizando apenas a região do NIR forneceram resultados um pouco inferiores quando comparados aos modelos utilizando o espectro todo, ou seja incluindo a região visível. Normalmente nos valores de R^2 de validação e calibração não foram observadas diferenças significativas,

mas valores maiores de RER foram obtidos com um menor número de VL, o que justificou o uso do espectro todo quando comparado ao da região NIR apenas.

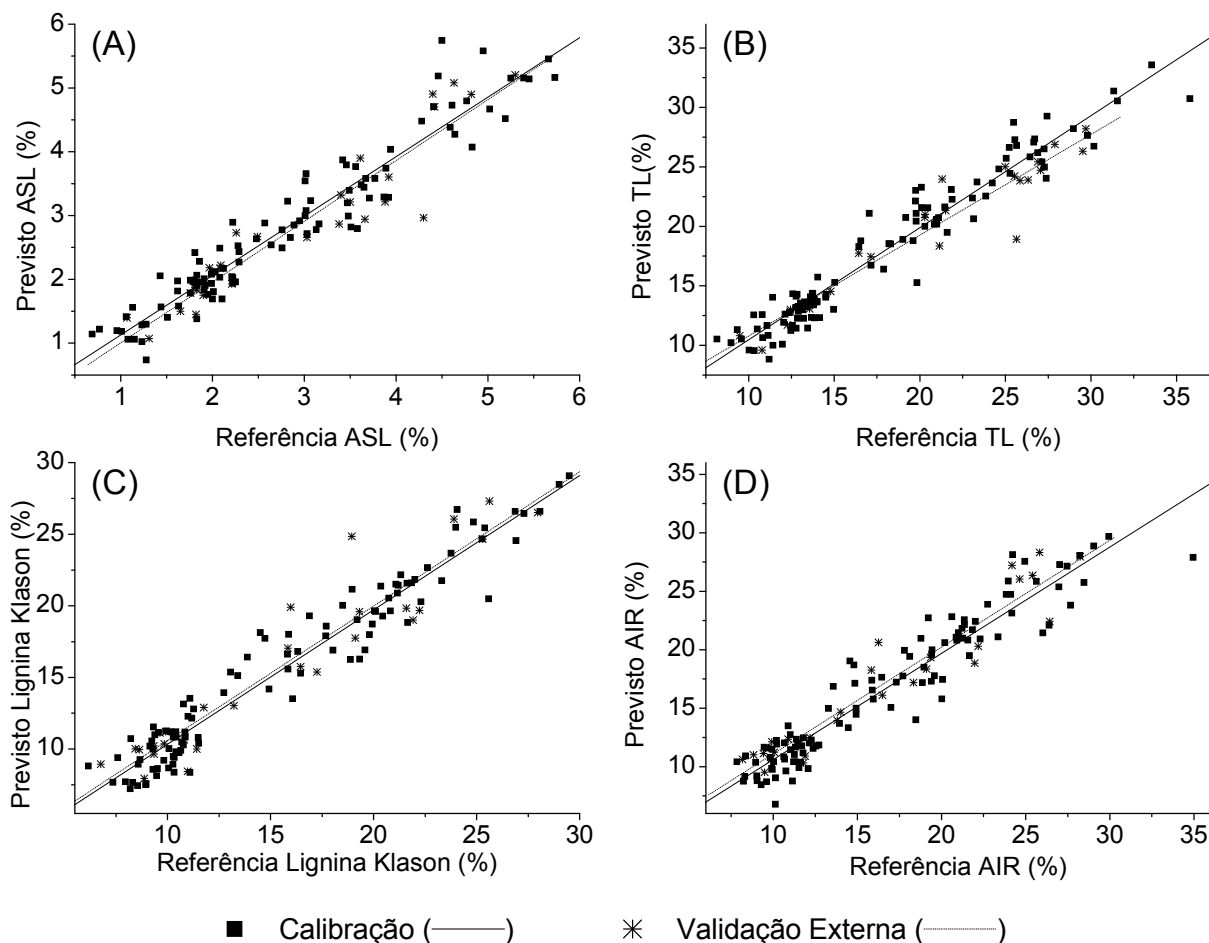


Figura 39. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) lignina ácida solúvel (ASL) com 7 VL, (B) lignina total (TL) com 7 VL, (C) lignina Klason (KL) com 7 VL e (D) resíduo ácido insolúvel (AIR) com 6 VL para os modelos de Banana+Café+Coco da fração DS.

Já para os modelos das ligninas constituídos apenas pelo conjunto das bananas a região do NIR forneceu resultados superiores quando comparados à região completa do Vis-NIR do espectro. Boas curvas de calibração para as previsões dos teores de ligninas foram obtidas, como mostra a Figura 40.

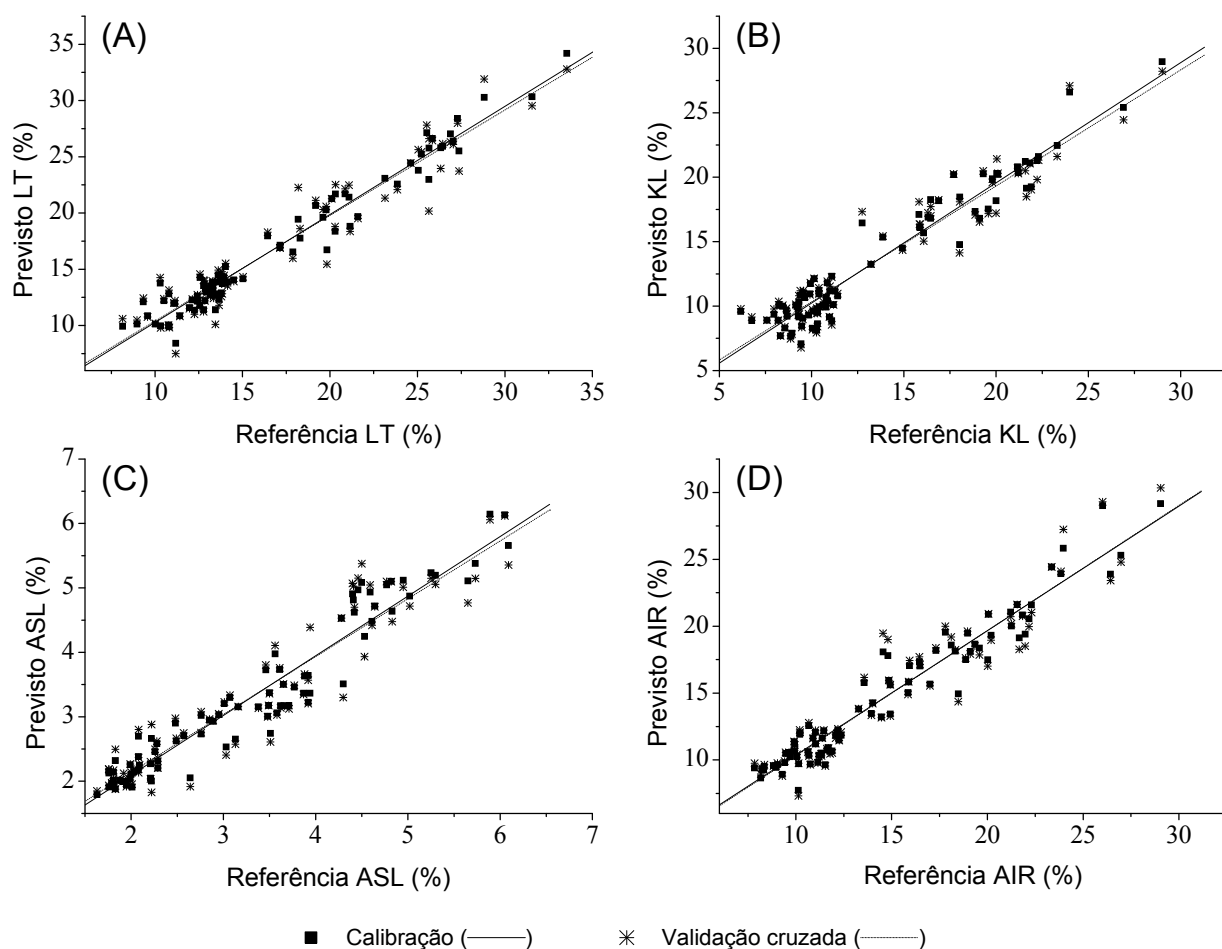


Figura 40. Gráficos de valores de referência versus valores previstos da calibração e validação cruzada para (A) lignina total (TL) com 3 VL, (B) lignina Klason (KL) com 3 VL, (C) lignina ácida solúvel (ASL) com 3 VL, e (D) resíduo ácido insolúvel (AIR) com 4 VL para os modelos de banana da fração DS.

Para as ligninas; lignina total, lignina Klason, resíduo ácido insolúvel e lignina ácida solúvel os modelos apresentados (Tabela 10) são melhores que os previamente reportados na literatura para outras biomassas. Por exemplo, Liu *et al.* (2010) modelando lignina total em biomassas de capim, obtiveram um $R^2_{cal} = 0,74$ e um $RER = 7,97$ para um modelo utilizando 9 VL. Dessa forma, melhores resultados foram obtidos no presente trabalho com modelos mais parcimoniosos. Ou seja, da análise da Tabela 10, os mesmos parâmetros obtidos neste trabalho para os modelos de Banana+Café e apenas Banana, por exemplo, foram respectivamente, $R^2_{cal} = 0,90$ e $0,96$ e RER de $10,00$ e $12,30$ (ambos aceitáveis para controle de qualidade) para

modelos com apenas 4 e 3 VL, respectivamente. Também os modelos de Café, Coco e Banana+Café+Coco foram superiores. No entanto, para a palha de milho, os mesmos autores (Liu *et al.*, 2010) obtiveram um RER = 18,00, adequado para quantificação, mas com um modelo de 9 VL, enquanto Sanderson *et al.* (1996) obtiveram um R^2_{cal} similar, mas sem informar o número de VL utilizados no modelo.

Os melhores resultados encontrados na literatura para lignina klason foram obtidos por Hodge e Woodbridge (2010), trabalhando com biomassa de pinheiro, com $R^2_{cal} = 0,97$ e $R^2_{val} = 0,95$ e RER = 25,00. Esses resultados são melhores do que os obtidos neste trabalho (Tabela 8), no entanto em seus modelos Hodge e Woodbridge utilizaram 11 VL contra no máximo 7 VL, no presente caso. Apesar do elevado R^2_{val} , o número de VL (11) é excessivo para a modelagem de um analito, mesmo ao se considerar a complexidade deste, evidenciando que, a análise cega de alguns parâmetros do ajuste, sem se considerar questões fundamentais e compreensão real do problema, é uma frequente armadilha em calibração multivariada. Krongtaew *et al.* (2010), obtiveram resultados semelhantes aos obtidos nesse trabalho, utilizando os mesmos números de fatores (3-4 VL) para a biomassa de palha de aveia.

Para lignina ácida solúvel, os resultados da literatura são piores do que os relatados aqui (Tabela 10). Poke (2006) trabalhando com toras de eucalipto obteve valores de $R^2_{cal, val}$ iguais a 0,72 e 0,12, respectivamente, para os modelos com 6 VL, e Jones *et al.* (2006) trabalhando com toras de pinheiro obtiveram valores de $R^2_{cal, val}$ iguais a 0,66 e 0,36, respectivamente, para modelos com 2 VL. Apenas o trabalho de Hayes (2012) obteve para a biomassa de *Miscanthus* valores de $R^2_{cal, val}$ de 0,96 e 0,94 e RER de 16,77, superiores aos obtidos nesse trabalho. No entanto utilizou 10 VL, contra 3, 5 e 3 VL para modelos de uma única biomassa (Banana, Café, e Coco) e no máximo 7 VL, em um modelo amplo combinado 3 espécies (Banana+Café+Coco) neste trabalho. Pode-se inferir que ocorreu um evidente sobreajuste, o que levou aos “bons” resultados do modelo no trabalho do autor citado.

De forma geral, os modelos de ligninas para a fração DS (Tabela 10) foram melhores para o conjunto de Banana e Banana+Café, seguidos então dos modelos de Banana+Café+Coco, Coco e Café. No entanto o modelo incluindo as três biomassas (Banana+Café+Coco) sempre utilizou um número maior de VL para a construção dos

modelos PLS. Esse fato está associado a uma maior variabilidade estrutural dos analitos em questão, o que justifica a necessidade de mais VL.

Os resultados obtidos para os modelos construídos com o método PLS para as frações DU e WU são mostrados nas Tabelas 11 e 12. Os espectros brutos da região Vis-NIR contendo as amostras de banana, café e coco, e suas diferentes frações botânicas, das frações DU e WU são mostrados nas Figuras 41 e 42, respectivamente.

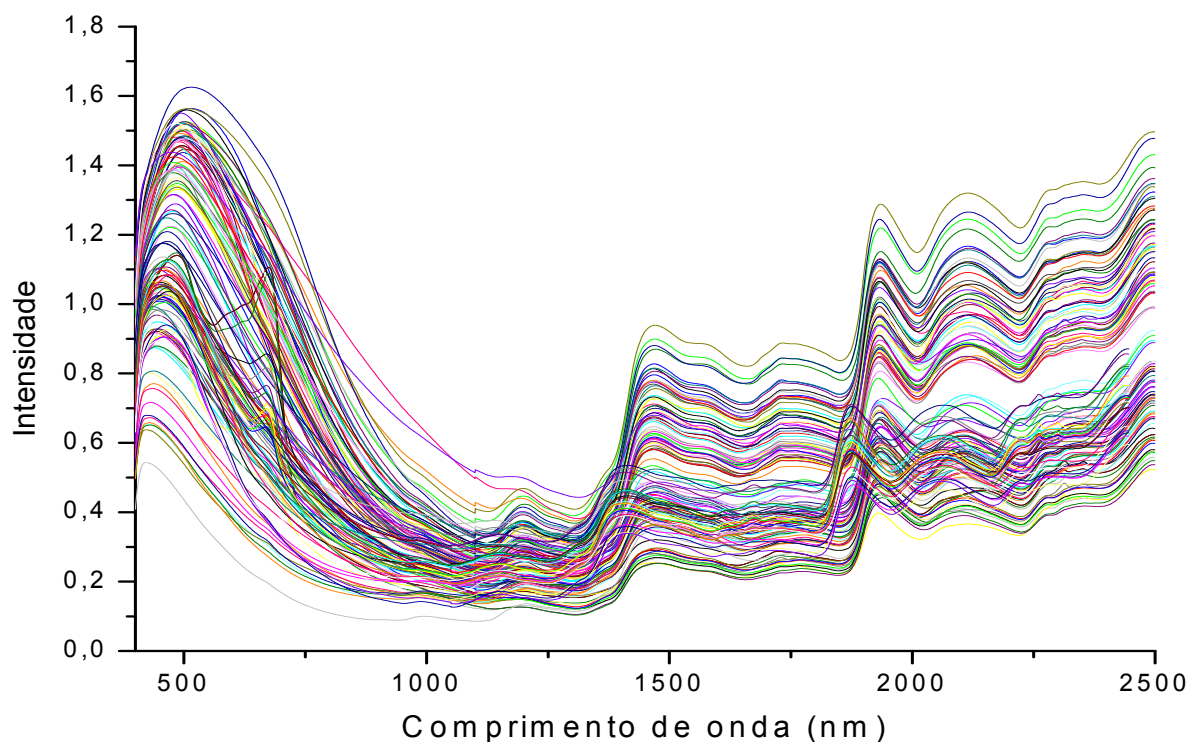


Figura 41. Espectros brutos da região Vis-NIR das amostras de banana, coco e café da fração DU.

A clara diferença entre os espectros da fração WU (Figura 42) e os espectros dos conjuntos da fração DS (Figura 35) e DU (Figura 41), está nas absorvâncias fortes nas regiões em torno de 1390 nm (associado com o primeiro sobretom da vibração simétrica de O-H) e 1895 nm (associado com a banda de combinação de deformação angular de OH), associadas provavelmente a água (Arnold *et al.*, 2002; Shenk *et al.*, 2008).

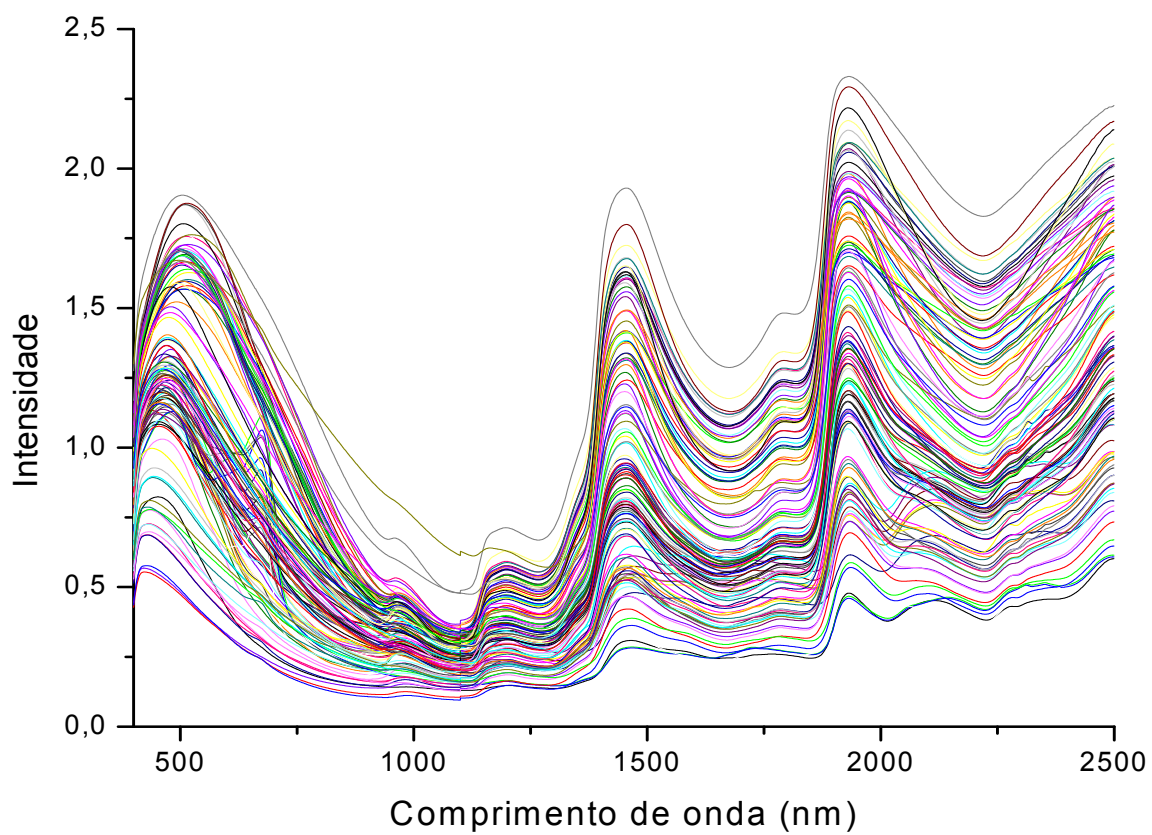


Figura 42. Espectros brutos da região Vis-NIR das amostras de banana, coco e café da fração WU.

Tabela 11. Parâmetros para avaliação dos modelos PLS obtidos para a fração DU.

y	Conjunto de dados	Pré-Tratamento	Região Espectral	Tamanho matriz	VL	Amostras Anômalas	R ²		RMSE		RE	RER
							Cal	Val	Cal	Pred		
TL	B	SNV+DT	NIR	55X2800	8	3	0,86	0,85	1,633	1,945	14,7	8,20
	C	SNV+2D(15)	NIR	42x2800	7	5	0,97	0,90	0,7000	1,831	8,00	6,20
	COC*	2D(3)	NIR	28x2800	5	4	0,97	0,97	0,5500	0,5600	2,00	16,0
KL	B	SNV+DT	NIR	55X2800	8	4	0,84	0,82	1,607	1,942	18,0	7,20
	C	SNV+2D(3)	NIR	42x2800	7	-	0,97	0,90	0,7142	1,693	9,00	6,70
	COC*	2D(3)	NIR	28x2800	5	4	0,98	0,96	0,4765	0,6765	3,00	11,2
ASL	B	SNV+1D(3)	NIR	55X2800	8	5	0,97	0,96	0,1702	0,1813	6,00	17,0
	C	SNV+1D(3)	NIR	42x2800	5	5	0,80	0,73	0,0853	0,1058	2,00	10,0
	COC*	2D(15)	NIR	28x2800	6	3	0,95	0,82	0,0610	0,1690	12,0	8,30
AIR	B	SNV+1D(3)	NIR	55x2800	8	2	0,94	0,92	1,039	1,227	13,0	10,0
	C	SNV+1D(3)	NIR	42x2800	7	5	0,97	0,90	0,7430	1,835	9,00	5,40
	COC*	2D(3)	NIR	28x2800	5	3	0,96	0,91	0,6511	0,9112	4,00	6,60
CINZAS	B	2D(15)	NIR	55X2800	6	4	0,89	0,75	0,4001	0,4803	12,0	5,00
	C	SNV+1D(3)	NIR	42x2800	5	5	0,60	0,60	0,2308	0,2301	11,0	5,60
	COC*	2D(3)	NIR	28X2800	4	5	0,79	0,77	0,4317	0,5720	25,0	4,00
EXTRAT	B	SNV+1D(3)	NIR	55x2800	7	3	0,95	0,87	1,683	2,609	17,0	8,00
	C	SNV+1D(3)	NIR	42x2800	7	5	0,84	0,77	1,513	1,870	11,0	7,00
	COC*	SNV+DT	NIR	28X2800	4	3	0,95	0,94	2,721	3,153	20,0	14,0
GLIC	B*	2D(15)+SNV	NIR	30X2800	3	5	0,91	0,87	3,035	3,373	8,60	9,00
	C**	2D(15)	NIR	42x195	4	7	0,83	0,75	1,266	1,623	11,0	5,50
	COC**	SNV+DT	NIR	28x300	3	4	0,85	0,83	1,428	1,672	6,00	5,80
XIL	B*	-	-	-	-	-	-	-	-	-	-	-
	C**	2D(15)	NIR	42x195	4	4	0,83	0,85	0,4850	0,4931	15,0	5,00
	COC**	SNV+DT	NIR	28X300	4	4	0,90	0,87	0,9610	1,094	9,00	5,50
GALA	B*	2D(15)+SNV	NIR	30X2800	4	6	0,85	0,81	0,1051	0,1394	9,00	6,50
	C**	2D(15)	NIR	42x195	4	6	0,88	0,84	0,3510	0,4770	13,5	8,65
	COC**	SNV+DT	NIR	28x300	4	4	0,80	0,78	0,1198	0,1348	13,0	5,00
RAM	B	-	-	-	-	-	-	-	-	-	-	-
	C	-	-	-	-	-	-	-	-	-	-	-
	COC	-	-	-	-	-	-	-	-	-	-	-
ARAB	B*	2D(15)+SNV	NIR	30X2800	3	5	0,81	0,75	0,4055	0,6609	18,0	6,00
	C**	2D(15)	NIR	42x779	4	7	0,94	0,90	0,3001	0,3198	7,50	10,0
	COC.	-	-	-	-	-	-	-	-	-	-	-
MAN.	B	-	-	-	-	-	-	-	-	-	-	-
	C	-	-	-	-	-	-	-	-	-	-	-
	COC.	-	-	-	-	-	-	-	-	-	-	-

TS	B*	2D(15)+SNV	NIR	30X2800	5	6	0,91	0,87	2,814	3,032	6,00	8,00
	C**	2D(15)+SNV	NIR	42x195	4	1	0,89	0,83	2,508	2,612	9,00	7,10
	COC**	SNV+DT	NIR	28x300	4	4	0,86	0,87	2,454	2,418	5,00	7,00

VL: número de variáveis latentes; RE: erro relativo (RMSEP/média); RER: $y_{\text{máximo}} - y_{\text{mínimo}} / \text{SEP}$; o número entre parenteses indica o tamanho das janelas quando do uso das derivadas; B:Banana; C:Café; Coc: Coco; TL: lignina total; KL: lignina klason; ASL: lignina ácida solúvel; AIR: resíduo ácido insolúvel; Umid: umidade; Extr: extrativos; Glic: glicose; Xil: xilose; Gala:galactose; Ram: ramnose; Arab:arabinose; Man: manose; TS: açúcares totais; *: validação cruzada; **: validação cruzada com seleção de variáveis para os carboidratos (as principais regiões utilizadas foram: 2222-2500 nm; 2000-2083 nm; 1923-1818 nm; 1470-1380 nm).

Tabela 12. Parâmetros para avaliação dos modelos PLS obtidos para a fração WU.

y	Conjunto de dados	Pré-Tratamento	Região Espectral	Tamanho da matriz	VL	Amostras Anômalas	R ²		RMSE		RE	RER
							Cal	Val	Cal	Pred		
TL	B	1D(15)	NIR	62x4200	8	5	0,89	0,74	1,452	1,916	15,0	7,00
	C***	2D(3)	NIR	42x1400	4	5	0,90	0,84	1,611	1,745	7,00	8,00
	COC*	2D(15)	NIR	28X2800	4	4	0,81	0,80	1,937	2,036	8,00	6,50
KL	B	1D(15)	NIR	62x2800	7	4	0,85	0,76	1,590	1,772	17,0	6,60
	C***	2D(3)	NIR	42X1400	4	4	0,90	0,84	1,549	1,733	8,00	8,00
	COC*	2D(15)	NIR	28X2800	4	2	0,81	0,80	1,842	1,973	9,00	5,00
ASL	B	1D(15)	NIR	62x2800	7	3	0,82	0,81	0,3620	0,4241	16,0	8,50
	C	SNV+1D(3)	NIR	42X2800	5	5	0,80	0,76	0,4401	0,5911	16,0	4,00
	COC*	SNV+1D(3)	NIR	28X2800	6	3	0,80	0,82	0,1581	0,1390	9,00	5,60
AIR	B	1D(15)	NIR	62x2800	8	5	0,89	0,75	1,470	1,788	16,0	7,00
	C***	2D(3)	NIR	42X1400	4	4	0,91	0,85	1,448	1,673	7,50	7,70
	COC*	2D(15)	NIR	28X2800	7	3	0,93	0,92	1,132	1,320	5,50	5,50
CINZAS	B	SNV+2D(3)	NIR	62x2800	5	5	0,72	0,65	0,6281	0,7720	19,0	6,00
	C	2D(3)	NIR	42X2800	6	6	0,89	0,80	0,0980	0,1990	10,0	6,20
	COC*	1D(3)	NIR	28X2800	4	3	0,78	0,76	0,4231	0,4592	25,0	3,05
EXTRAT	B	SNV+1D(3)	NIR	62X2800	7	6	0,88	0,86	2,499	2,515	20,0	8,30
	C	SNV+1D(3)	NIR	42X2800	6	4	0,78	0,71	2,009	2,021	13,0	5,20
	COC*	SNV+1D(3)	NIR	28X2800	6	3	0,97	0,97	1,766	1,995	10,0	13,0
GLIC	B*	2D(15)+SNV	NIR	28X2800	5	4	0,77	0,72	4,938	5,218	14,5	4,30
	C**	2D(15)	NIR	42X262	6	5	0,78	0,73	1,587	1,761	11,0	5,90
	COC**	2D	NIR	30X326	5	3	0,91	0,90	1,187	1,240	5,00	9,00
XIL	B	-	-	-	-	-	-	-	-	-	-	-
	C**	1D(3)+SNV	NIR	42X262	5	5	0,73	0,76	0,5581	0,3742	13,0	5,00
	COC**	SNV+1D(3)	NIR	28X326	4	3	0,84	0,86	1,148	1,070	9,00	9,30

GALA	B	-	-	-	-	-	-	-	-	-	-	-
	C**	-	-	-	-	-	-	-	-	-	-	-
RAM	COC**	2D(15)	NIR	28X326	3	3	0,80	0,76	0,1237	0,1448	13,0	4,00
	B	-	-	-	-	-	-	-	-	-	-	-
ARAB	C	-	-	-	-	-	-	-	-	-	-	-
	COC**	2D(3)	NIR	30X150	2	3	0,96	0,96	0,0061	0,0060	1,70	20,0
	B	-	-	-	-	-	-	-	-	-	-	-
	C	2D(15)+SNV	NIR	42X262	6	5	0,88	0,80	0,4591	0,4700	10,0	6,00
MAN	COC	-	-	-	-	-	-	-	-	-	-	-
	B	-	-	-	-	-	-	-	-	-	-	-
	C	-	-	-	-	-	-	-	-	-	-	-
	COC**	2D(3)	NIR	28X150	3	3	0,97	0,98	0,0237	0,0185	4,00	25,3
TS	B*	2D(15)+SNV	NR	30X2800	5	6	0,92	0,91	2,744	2,851	6,00	6,00
	C**	2D(15)	NIR	42X262	5	1	0,87	0,85	2,799	3,206	10,0	7,00
	COC**	2D(15)	NIR	28X326	5	3	0,89	0,86	2,283	2,821	7,00	7,00

VL: número de variáveis latentes; RE: erro relativo (RMSEP/média); RER: $y_{\text{máximo}} - y_{\text{mínimo}} / \text{SEP}$; o número entre parenteses indica o tamanho das janelas quando do uso das derivadas; B: banana; C: café; Coc: coco; TL: lignina total; KL: lignina klason; ASL: lignina ácida solúvel; AIR: resíduo ácido insolúvel; Umid: umidade; Extr: extrativos; Glic: glicose; Xil: xilose; Gala:galactose; Ram: ramnose; Arab:arabinose; Man: manose; TS: açúcares totais; *: validação cruzada; **: validação cruzada com seleção de variáveis para os carboidratos (as principais regiões utilizadas foram: 2222-2500 nm; 2000-2083 nm; 1923-1818 nm; 1470-1380 nm); ***: Região espectral retirando a água.

Com relação ao uso de diversos pré-tratamentos, os modelos de calibração para as frações DU e WU que utilizaram a transformação SNV+DT (Figura 43) mostraram erros relativos significativamente maiores do que aqueles baseados em tratamentos por derivadas. Essa tendência quanto aos tratamentos espectrais foi observada para grande parte dos constituintes. Aparentemente algumas informações espectrais importantes nos espectros são removidas quando as técnicas de correção de dispersão são aplicadas, e as derivadas aparecem como uma melhor alternativa.

A Figura 43 apresenta os espectros NIR das amostras de banana e coco das frações DU e WU submetidos às transformações por (A) 1D(15); (B) 2D(15), (C) SNV e (D) SNV+Detrend. Ambos derivadas e Detrend, utilizaram polinômio de ordem 2.

Utilizando pré-tratamentos combinados para correção dos dados espectrais, tais como SNV seguido de derivadas, ou o contrário, obtém-se na maioria dos casos, ganho com as informações contidas nos espectros pela remoção de maior quantidade de variabilidade não relacionada ao analito de interesse. Derivadas e SNV removem diferentes tipos de efeitos, e pode muito bem haver alguma vantagem de usar ambos os processamentos. O que importa é em que ordem eles são aplicados (Fearn, 2003).

A Figura 44 apresenta os espectros submetidos aos pré-tratamentos combinados utilizando (A) SNV + 1D(3), (B) 1D(3) + SNV, (C) SNV + 2D(15) e (D) 2D(15) + SNV, nos quais é possível observar as diferenças entre a ordem e o grau dos tratamentos. Uma pequena melhoria nos espectros de 1D(3)+SNV e 2D(15)+SNV (Figuras B e C, respectivamente) foi observada. Aplicando SNV após as derivadas, remove-se mais variabilidade, ou seja informação irrelevante aos espectros é retirada. Se observarmos a região na faixa entre 1900-2000 nm para os espectros (A) e (B) percebe-se um espectro mais limpo para este último, aplicando a primeira derivada antes do SNV. Mas a diferença não chega a ser significativa.

Nos espectros (C) e (D), observa-se uma tendência maior quanto à melhoria do espectro derivado (2D) seguido de SNV, praticamente em toda a faixa espectral.

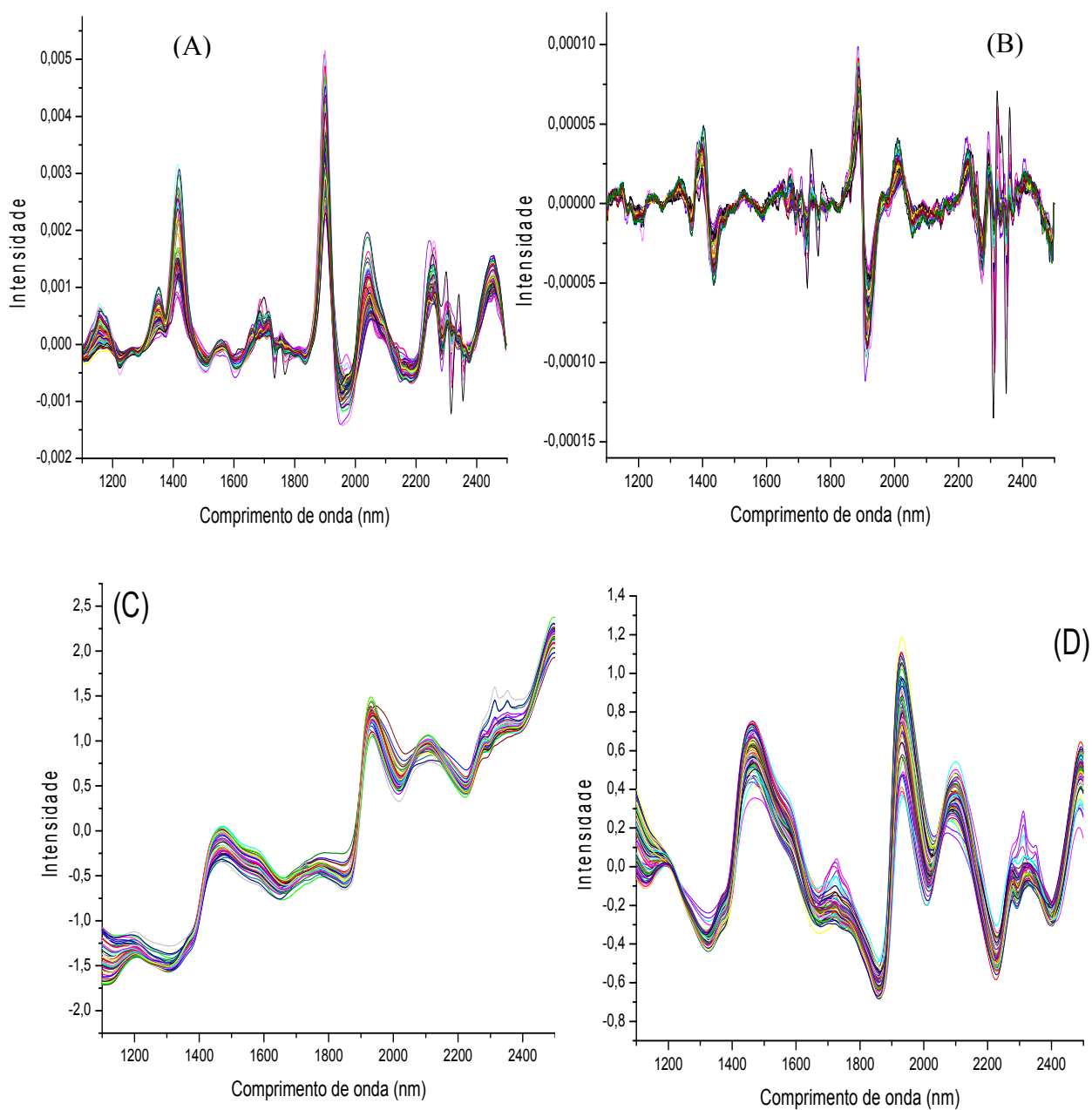


Figura 43. (A) Espectros de banana da fração WU pré-tratados por 1D (15); (B) espectros de coco da fração WU pré-tratados por 2D (15); (C) espectros de banana da fração DU pré-tratados por SNV e (D) espectros de coco da fração DU pré-tratados por SNV+Detrend.

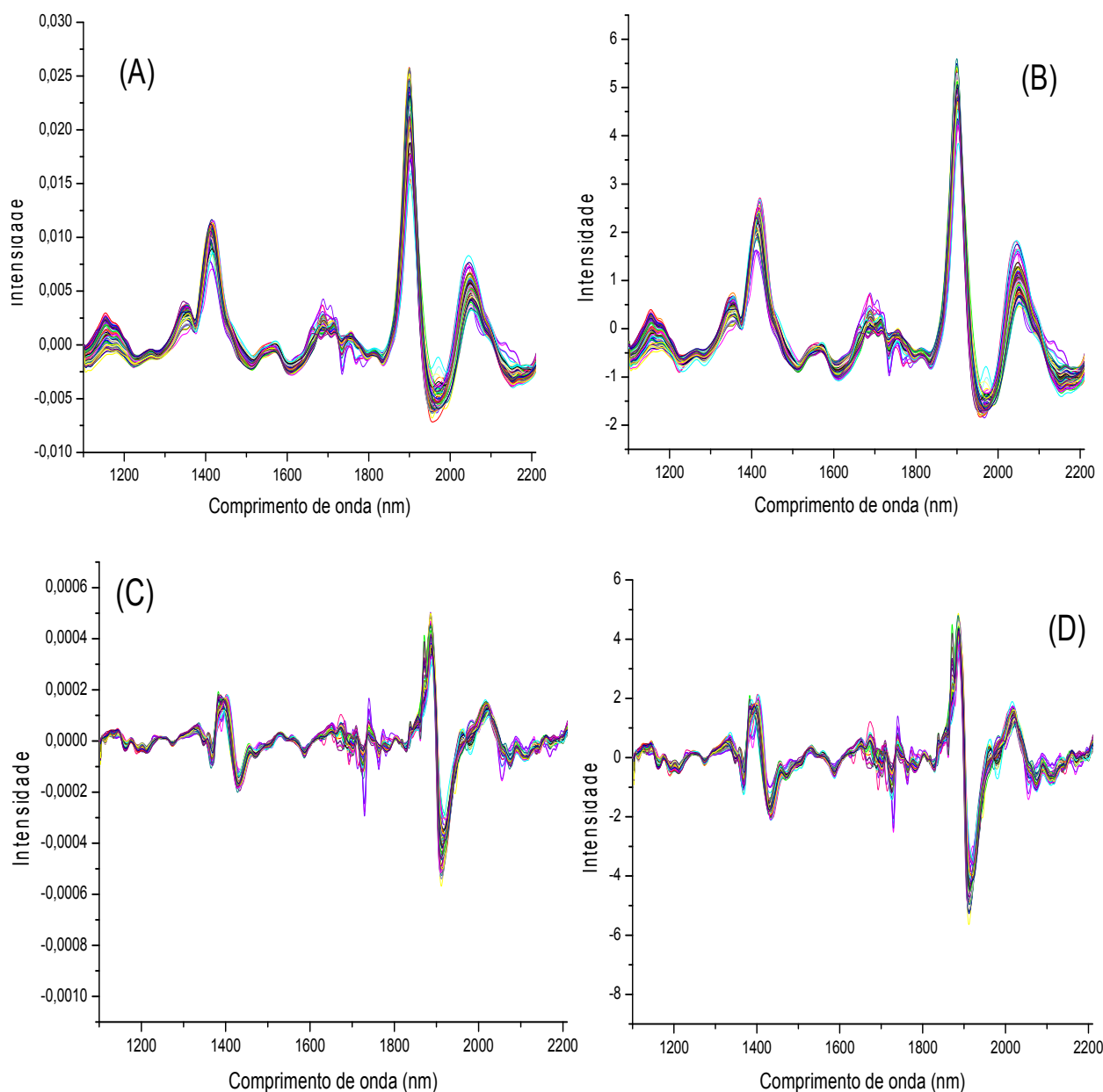


Figura 44. Espectros pré-tratados da fração DU por (A) SNV+1D(3), (B) 1D(3) + SNV, (C) 2D(15)+SNV e (D) SNV+2D(15), respectivamente.

Hayes (2012) utilizou as frações úmidas (WU) e secas (DU) não moídas da biomassa de *Miscanthus* para modelar a composição em ligninas. Para o teor de lignina Klason, obteve $R^2_{cal,val}$ de 0,95 e 0,88 com 10 VL e RER= 10,86; e $R^2_{cal,val}$ de 0,95 e 0,94 com 8 VL com RER= 17,70, para as frações WU e DU, respectivamente. Comparando com os resultados obtidos neste trabalho (Tabelas 11 e 12), o melhor

modelo da fração WU foi para o conjunto de café, com 4 VL e $R^2_{cal,val}$ de 0,90 e 0,84, e RER= 8,00 mostrando semelhanças entre os parâmetros obtidos, mas com menos da metade de VL utilizadas quando comparado com o trabalho de Hayes (2012). Para a fração DU, o melhor modelo foi o da biomassa de coco, com $R^2_{cal,val}$ de 0,98 e 0,96, RER= 11,20 e 5 VL. Esse modelo foi superior ao da biomassa de *Miscanthus*, quando avaliando todos os parâmetros do modelo. Da mesma forma Vergnoux *et al.*, (2009) modelando lignina Klason para a fração WU obteve valores de $R^2_{cal,val}$ de 0,99 e 0,80, usando 10 VL, num total de apenas 14 amostras (9 amostras no conjunto de calibração e 4 amostras no conjunto de validação externa). Aparentemente, apesar dos excelentes resultados, tais modelos não podem ser usados nem ao menos como forma de triagem já que seu valor de RER é inferior a 4,00.

Axrup *et al.*, (2000) utilizando um pequeno número de VL para modelagem da lignina Klason da fração WU, apresentou valores de RER de 2,17 (para 6 VL) e 6,40 (para 5 VL), variando o pré-tratamento espectral. Todos os valores de RER obtidos aqui foram superiores, exceto para a biomassa de coco (RER= 5,00), e também com um pequeno número de VL (4-7).

Jones *et al.*, (2006) modelaram a composição em ligninas (lignina total, lignina klason e lignina ácida solúvel) da fração DU da biomassa de pinus. Um número baixo de VL foi utilizado (2-4). As ligninas total e lignina klason (ambas com 4 VL) obtiveram os mesmos valores de $R^2_{cal,val}$ (0,85 e 0,51); a lignina ácida solúvel com apenas 2 VL obteve $R^2_{cal,val}$ de 0,66 e 0,36. Os resultados apresentados nesse estudo novamente mostraram-se superiores aos reportados na literatura.

Da mesma forma para os modelos de lignina ácida solúvel das frações WU e DU, a literatura reporta (Hayes, 2012) valores de R^2_{cal} superiores a 0,96, no entanto utilizando um número excessivo de fatores para a modelagem PLS, com 17 e 12 VL, respectivamente. Nesse trabalho os valores de $R^2_{cal,val}$ foram todos superiores a 0,80 e 0,70, mas utilizando no máximo 7 e 8 VL, para as frações WU e DU. Na opinião da autora, essa ênfase exacerbada em $R^2_{cal,val}$, RSD, RER etc, sem levar-se em consideração princípios básicos da calibração (número de variáveis preditoras), é uma das sérias armadilhas da modelagem cega "*blind push the botton*" soberbamente discutida na literatura (Esbensen e Geladi, 2010; Kjeldahl e Bro, 2010).

A Figura 45 apresenta as curvas de calibração para as ligninas da fração DU, dos modelos de banana, café e coco.

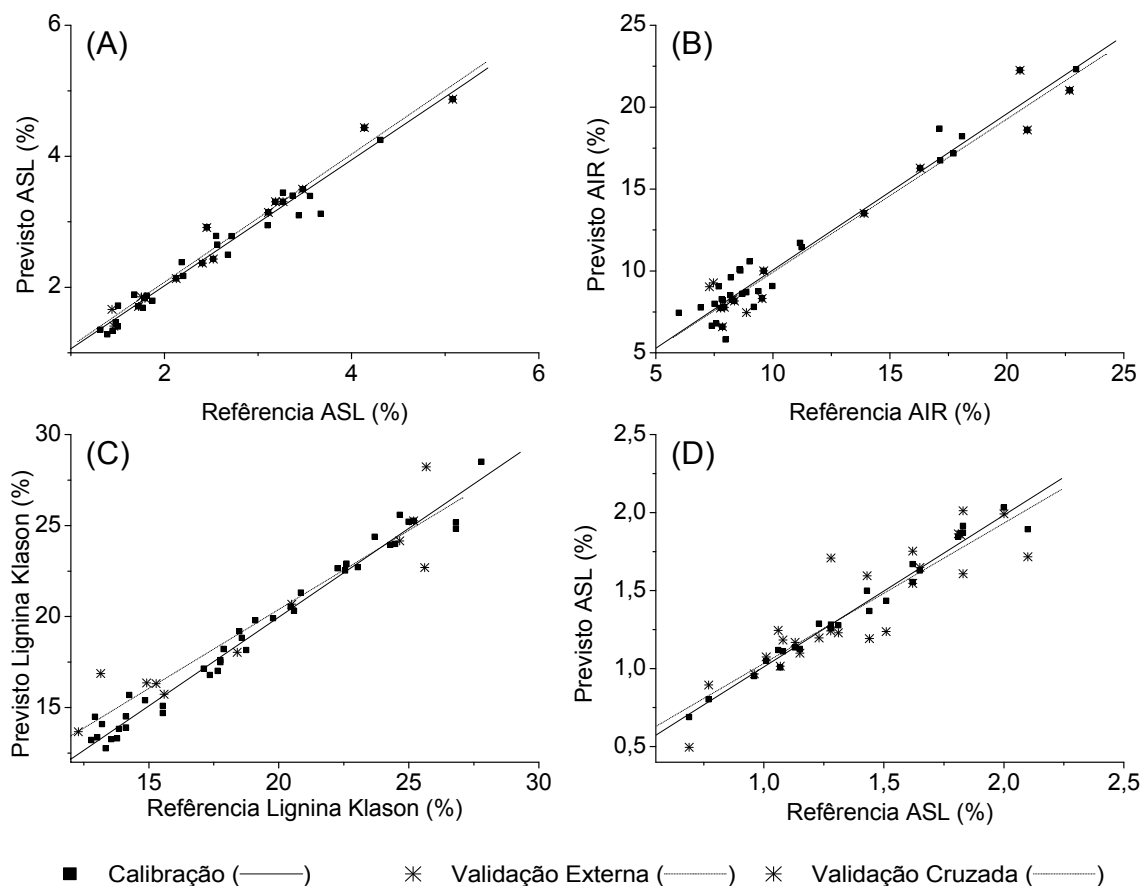


Figura 45. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) lignina ácida solúvel (ASL) com 8 VL do modelo da Banana, (B) resíduo ácido insolúvel (AIR) com 8 VL do modelo da Banana, (C) lignina Klason (KL) do modelo do Café com 7 VL, e (D) gráfico dos valores de referência versus valores previstos da calibração e validação cruzada para a lignina ácida solúvel (ASL) com 6 VL para o modelo de Coco, todos da fração DU.

A Figura 46 apresenta as curvas de calibração para as ligninas da fração WU, dos modelos de banana, café e coco.

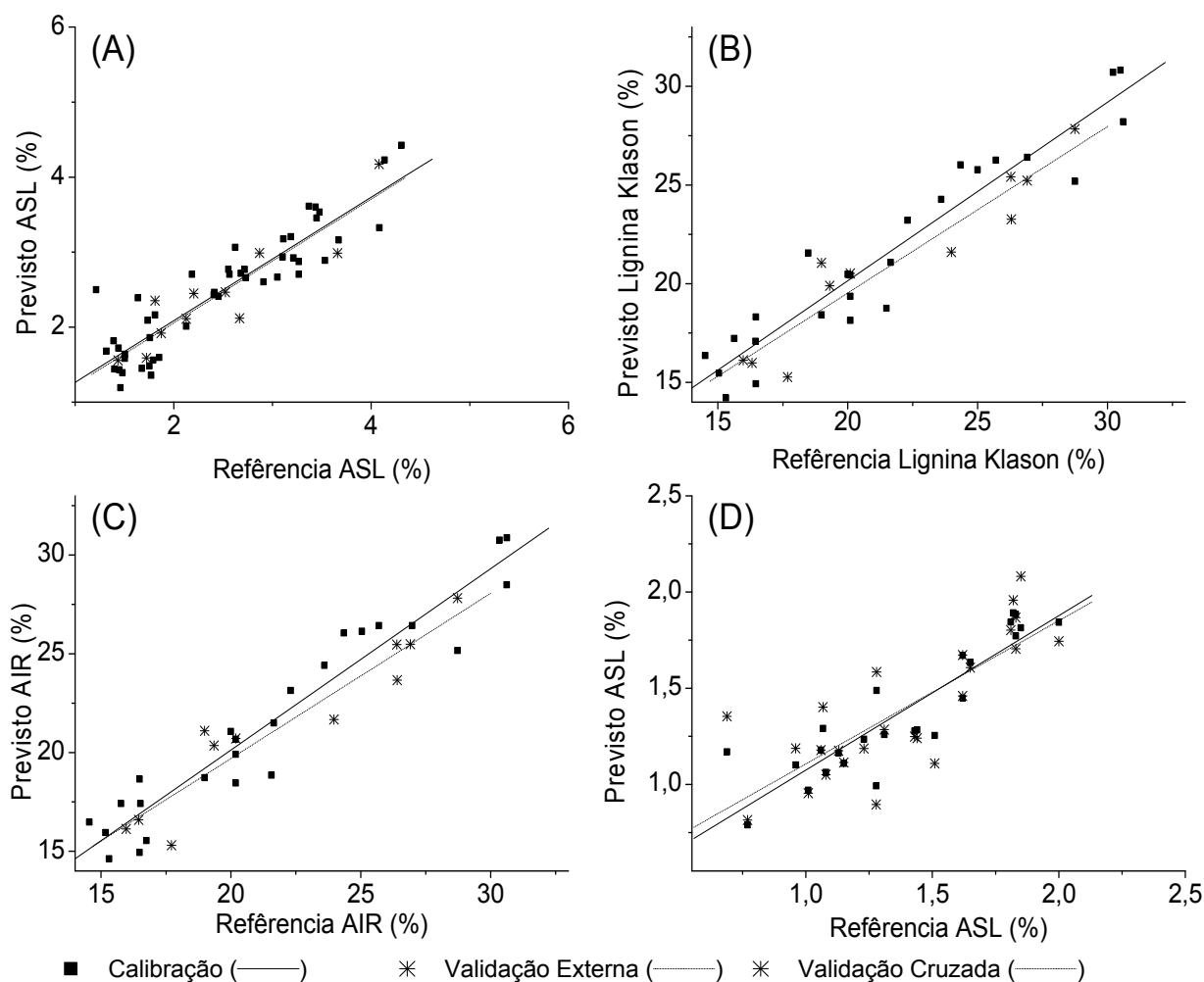


Figura 46. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) lignina ácida solúvel (ASL) com 7 VL do modelo da Banana, (B) lignina Klason (KL) com 4 VL do modelo de Café, (C) resíduo ácido insolúvel (AIR) do modelo do Café com 4 VL, e (D) gráfico dos valores de referência versus valores previstos da calibração e validação cruzada para lignina ácida solúvel (ASL) com 6 VL para o modelo de Coco, todos da fração WU.

De maneira geral, os modelos de ligninas da fração DS foram superiores aos da fração DU e WU, com menores RSD e VL, menor quantidade de amostras anômalas removidas e maiores valores de RER, com exceção da lignina ácida solúvel (dos modelos de Banana e Café) e lignina total (modelo de Coco) que apresentaram valores ligeiramente superiores para a fração DU. Os modelos da fração WU foram inferiores praticamente para todos os parâmetros de ligninas modeladas. Outra observação para fração WU é de que algumas ligninas foram mais bem modeladas para as regiões do espectro sem água. Essas observações podem ser explicadas tanto pela forte interferência da umidade nos espectros, como também pela maior heterogeneidade das amostras DU e WU.

Para os dados de extrativos da fração DS, os melhores modelos obtidos utilizaram 6 VL para os conjuntos de coco, 4 VL para a banana, 7 VL para os conjuntos de café e também 7 VL para os modelos amplos combinando as biomassas (Banana+Café e Banana+Café+Coco) (Tabela 10). Comparando as frações DU e WU com a fração DS, os números de VL utilizadas em cada modelo e os valores de RSD não diferem significativamente, no entanto os valores de RER foram maiores para a fração DS, demonstrando assim melhores resultados. O alto número de VL utilizadas para a modelagem dos extrativos, independente da fração, pode ser explicado pelo fato de que os extrativos são uma mistura de múltiplos compostos que provavelmente absorvem em diferentes regiões do espectro, e assim, um único fator (VL) não é capaz de explicar toda a variância associada aos extrativos. Esse fato justifica a necessidade de várias VL para explicar a variabilidade de y , especialmente quando diversas biomassas (café, banana e coco) estão incluídas no conjunto de dados (Várrová, 2008). Apesar disso, os modelos obtidos foram semelhantes aos apresentados na literatura (Nkansah *et al.*, 2010) ou inclusive, melhores, quando se compara com o elevado número de VL utilizadas na modelagem de extrativos, com 15, 17 e 13 fatores, respectivamente, para modelar *Miscanthus* da fração DS, WU e DU (Hayes, 2012). Para todos os modelos de extrativos foram obtidos valores de $R^2_{cal, val}$ superiores a 0,84 e 0,77, independente da biomassa e tamanho da partícula, com exceção apenas para o modelo de café da fração WU (com R^2_{cal} e R^2_{val} de 0,78 e 0,71, respectivamente).

Para a umidade, apesar de o analito alvo ser um composto único e a espectroscopia NIR ter bandas significativas para a água, os modelos requereram de 3 a 7 VL. Isso ocorreu devido à existência de diferentes interações moleculares entre a água e moléculas adjacentes da biomassa, como por exemplo, ligações de H (Lestander e Geladi, 2005), o que provoca o deslocamento das bandas de vibração, ou seja, na prática o analito alvo é heterogêneo, havendo diversos tipos de água, cada uma necessitando de sua VL para ser corretamente modelada. Comparando com resultados da literatura (Jin e Chen, 2007) os modelos obtidos aqui necessitaram menor quantidade de VL (6 *versus* 12 VL), apresentado R^2_{val} semelhante (0,86 *versus* 0,88), para o modelo de Banana, por exemplo.

Valores de baixa qualidade de $R^2_{cal, val}$; RSD e RER foram encontrados para as cinzas, independente da fração e da biomassa (Tabelas 10, 11 e 12), sendo que em alguns casos o valor de RER não permitiu o uso do modelo nem como forma de triagem, com valores inferiores a 4,00 (modelo de Banana+Café da fração DS, e modelo do Coco da fração WU). Os únicos modelos de cinzas, que provavelmente podem ser considerados razoáveis, com $RER > 6,00$; $R^2_{cal} > 0,89$ e $R^2_{val} > 0,80$, e erros relativos menores que 10,00%, são os do conjunto do café da fração WU ($RER= 6,20$, $R^2_{cal}= 0,89$ e $R^2_{val}= 0,80$, e $RSD= 10,00$), e o conjunto do modelo do Coco, da fração DS (Tabela 8) com $R^2_{cal, val}$ de 0,86 e $RER > 6,00$. O que ocorre no caso das cinzas é que os espectros NIR não se relacionam diretamente com o teor de substâncias inorgânicas, ou seja, os compostos inorgânicos presentes na biomassa, que se convertem em óxidos pela calcinação, de maneira geral não absorvem na região do NIR. Entretanto, a tentativa de criação de modelos para o teor de cinzas se justifica, pois, os compostos inorgânicos podem se correlacionar com outros compostos químicos que absorvem no NIR, e assim por “correlação indireta” serem modelados (Esbensen, 2002; Lestander e Rhen, 2005).

A Figura 47 apresenta as curvas de regressão para alguns dos modelos citados acima de extrativos, umidade e cinzas, para diferentes fração (DS, DU e WU). Exceto para as cinzas (Figura 47 A e C) que apresentaram resultados razoáveis, os demais modelos apresentaram concordâncias entre os valores de referência e os valores preditos indicando bons modelos.

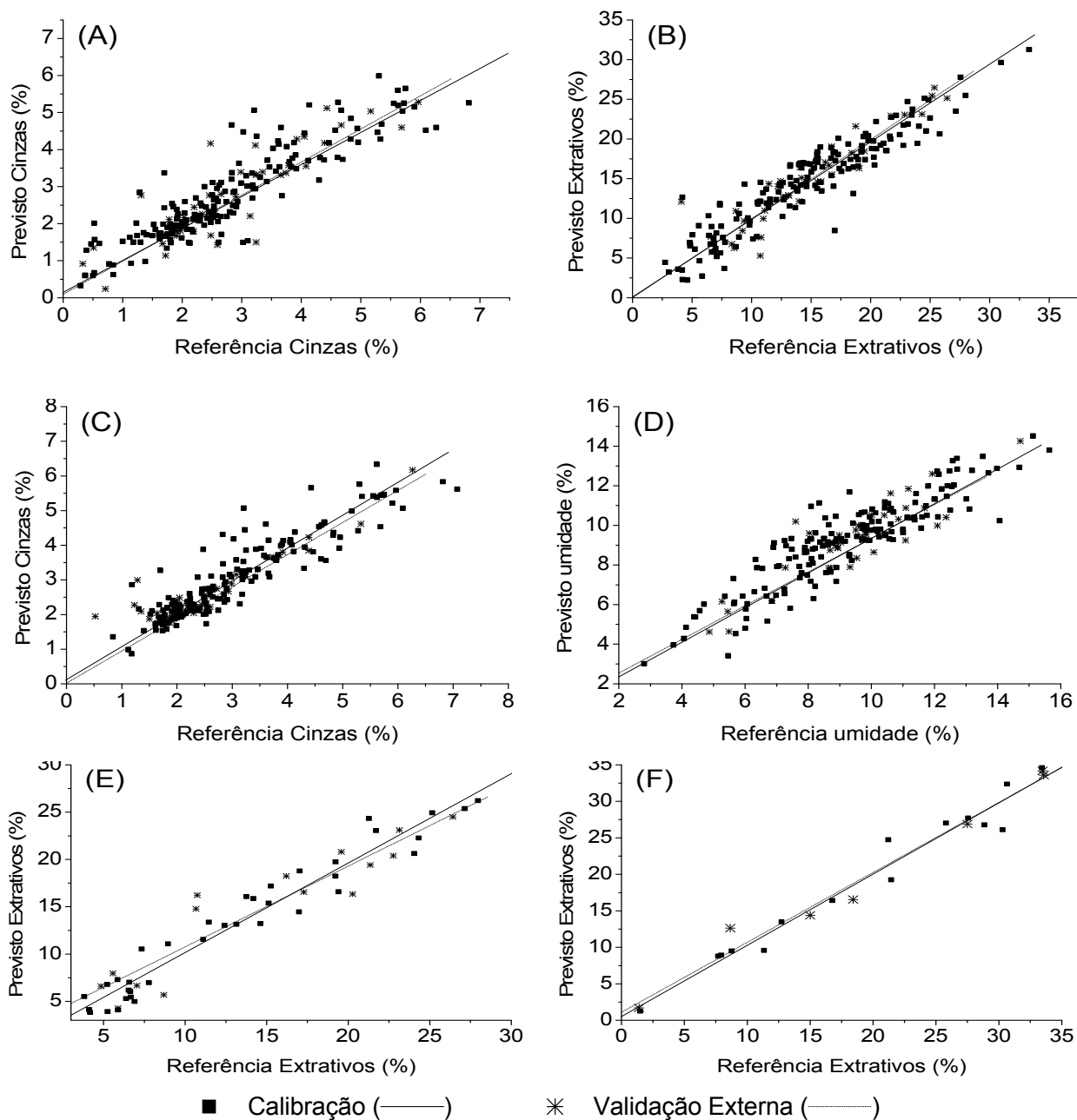


Figura 47. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) cinzas do modelo Banana + Café + Coco da fração DS com 6 VL, (B) extrativos do modelo Banana + Café da fração DS com 7 VL, (C) cinzas do modelo Banana + Café da fração DS com 7 VL (D) umidade do modelo de Banana + Café da fração DS com 7 VL (E) extrativos do modelo de Banana da fração DU com 7 VL (F) e extrativos do modelo do Coco da fração WU com 6 VL.

Para a modelagem dos açúcares, a maioria dos modelos necessitou de seleção de variáveis (Lomborg *et al.*, 2010). Essa seleção se deu com base nas regiões mais características do espectro NIR para os carboidratos. Essas regiões foram: 222-2500 nm (região de banda de combinação CH₃, CH₂, CH); 2083-2000 nm (banda de combinação de OH); 1923-1818 nm (estiramento C=O); 1470-1380 nm (sobretom de OH e região de combinação de CH). Esses modelos foram aplicados a regiões limitadas, ou seja apenas de 5%-50% do espectro original.

Os erros relativos para os maiores e mais importantes açúcares, glicose, xilose, e açúcares totais foram menores que 14,5%, 15% e 10%, respectivamente, mostrando que eles podem ter boa predição, independente da fração e do conjunto de dados usado. O desempenho dos modelos apresentados neste trabalho (Tabela 10) são similares aos previamente reportados na literatura para outras biomassas. Baum *et al.*, (2012), modelando glicose e xilose após a hidrólise enzimática de farelo de milho, obtiveram respectivamente valores de R^2_{cal} de 0,84 com 5 VL e R^2_{cal} de 0,95 para 3 VL. Lomborg *et al* (2010) obtiveram R^2_{cal} para glicose e xilose de 0,83 e 0,82, respectivamente, ambos com 5 VL. No entanto Hayes (2012) apresentou valores de R^2_{cal} para glicose, xilose e TS muito superiores aos encontrados nesse trabalho, e também a todos os demais já publicados até o momento. No caso desses açúcares, os valores obtidos de R^2_{cal} foram todos superiores a 0,98, mas utilizando 16 VL para construção dos modelos. Vale ressaltar novamente que os resultados das calibrações desses modelos construídos por Hayes (2012) podem indicar a ocorrência de sobreajuste. Fica a dúvida sobre a escolha correta do número de VL escolhidos para esses modelos.

Em comparação com os açúcares majoritários, os resultados preditivos para os açúcares minoritários (manose e ramnose) foram relativamente inferiores, provavelmente devido às suas menores porcentagens em peso, resultando em sinais menos intensos, passíveis de serem encobertos por outros. Outro fator que pode justificar o pior ajuste reside no fato de que pequenas proporções resultam em baixas concentrações no hidrolisado e, assim, a medição na cromatografia pode não ser tão precisa quanto para glicose e xilose, por exemplo. No entanto para alguns conjuntos, manose e ramnose apresentam resultados extraordinários, como no caso dos modelos

de coco da fração WU com $R^2_{cal, val}$ maiores que 0,96 e 0,96 e $RER > 20,00$ para manose e ramnose; e para ramnose modelos de banana e café da fração DS, apresentaram $R^2_{cal, val}$ maiores que 0,86 e 0,80, $RER > 7,00$ e $RSD < 8\%$ (Tabelas 10 e 11). Mas de maneira geral, os açúcares minoritários apresentam resultados pobres para a calibração, principalmente na fração DU, como também já relatados em outros trabalhos (Liu *et al.*, 2010; Hayes, 2012).

Os açúcares arabinose e galactose, considerados monossacarídeos de concentrações intermediárias, foram bem modelados para todos os conjuntos da fração DS. Com exceção dos modelos de banana e coco da fração WU para arabinose, e modelos de Banana e Café também da fração WU para galactose, e ainda modelos de coco da fração DU para arabinose, o restante dos conjuntos foram preditos satisfatoriamente para esses dois carboidratos, apresentando inclusive em alguns conjuntos (Banana, DS) resultados superiores ($R^2_{cal, val}$ de 0,93 e 0,90 para arabinose com 7 VL) aos já encontrados na literatura para outras biomassas. Lomborg *et al.*, (2010) e Baum *et al.*, (2012) obtiveram respectivamente R^2_{cal} de 0,82 (7 VL) e 0,77 (4 VL) modelando arabinose em palha de trigo e farelo de milho.

As Figuras 48 e 49 apresentam as curvas de calibração para alguns dos modelos de açúcares obtidos da fração DS do conjunto de café e de coco, respectivamente.

A não especificidade da espectroscopia NIR para diferenciar entre os 6 monossacarídeos, já que todos absorvem basicamente nas mesmas regiões do espectro, também justifica piores modelos para alguns desses açúcares, principalmente os presentes em baixas concentrações.

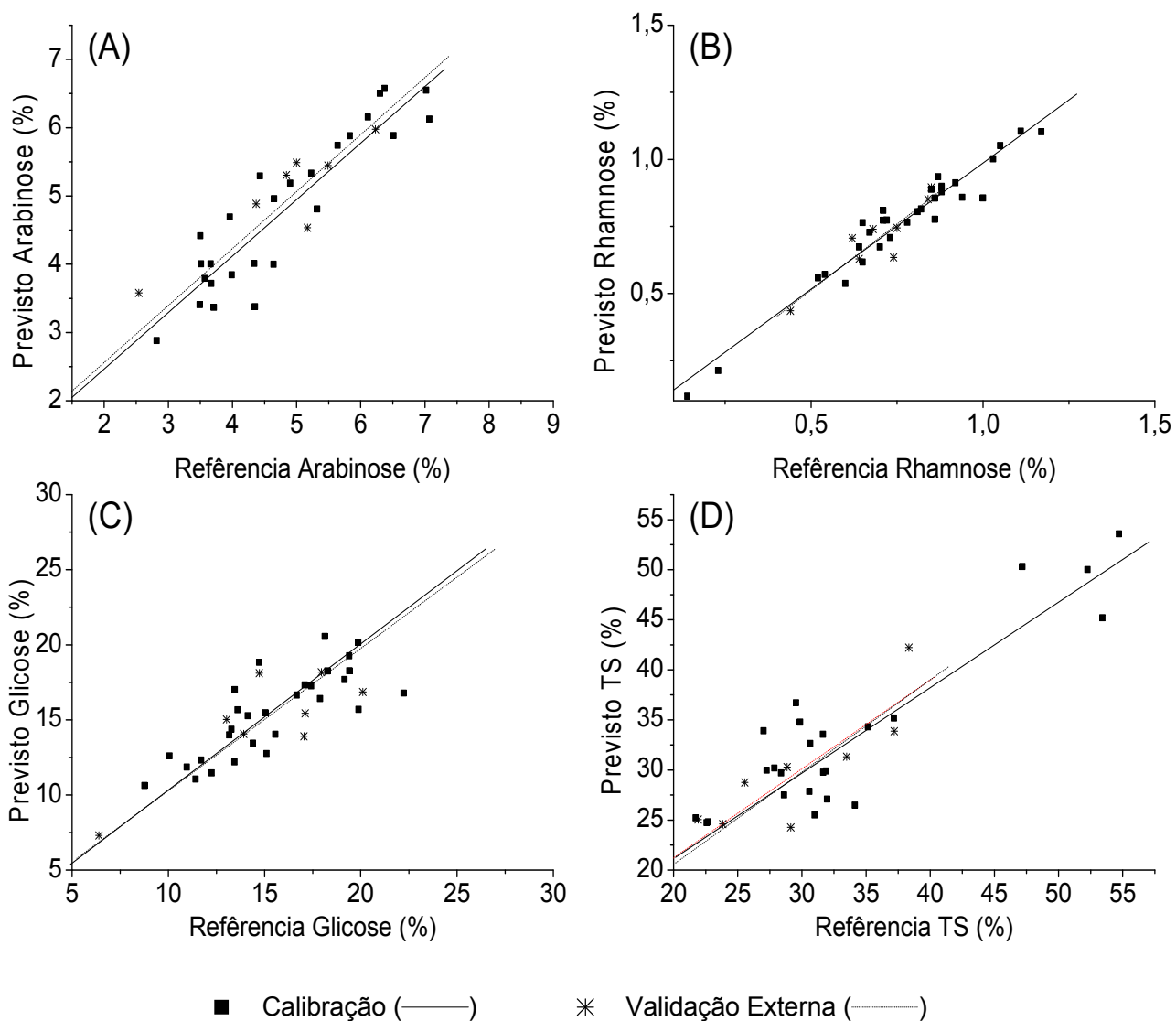


Figura 48. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) arabinose com 7 VL (B) ramnose com 7 VL (C) glicose com 5 VL e (D) açúcares totais (TS) com 4 VL da fração DS para os modelos de café.

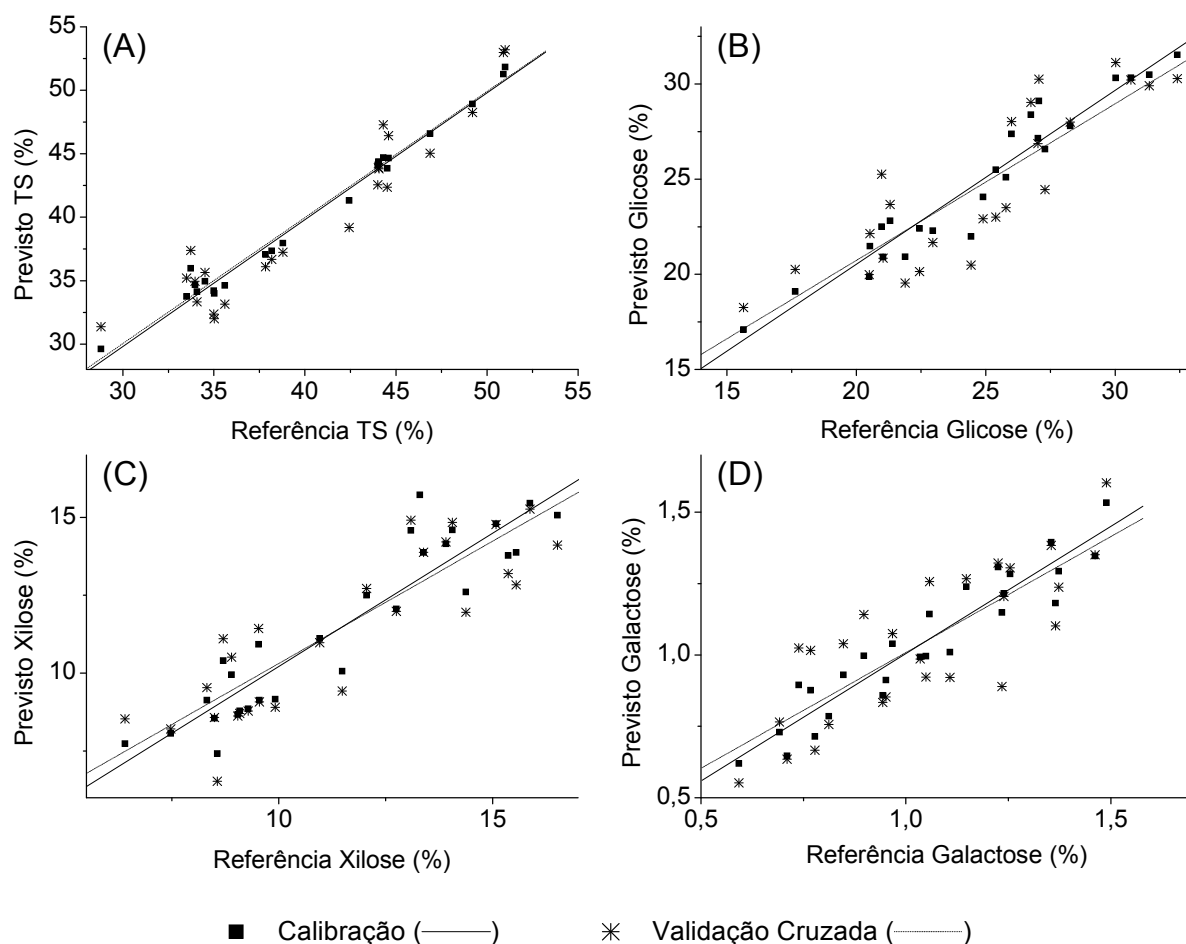


Figura 49. Gráficos de valores de referência versus valores previstos da calibração e validação cruzada para (A) açúcares totais (TS) com 2 VL (B) glicose com 2 VL (C) xilose com 2 VL e (D) galactose com 4 VL para os modelos de coco da fração DS.

Os açúcares do modelo de banana da fração DS (Figura 50), foram modelados satisfatoriamente, enquanto que a maioria dos modelos de banana da fração DU e WU não obtiveram calibrações razoáveis. Provavelmente esse problema tenha sido mais proeminente nas frações não moídas, devido a dificuldade na obtenção dos espectros de tais amostras. Como são partes da planta não maleáveis (como o engaço, por exemplo) a disposição na cubeta levava a espectros não representativos da amostra como um todo. Isso resulta em pobres modelos preditivos para tais frações, quando comparado às amostras moídas e peneiradas.

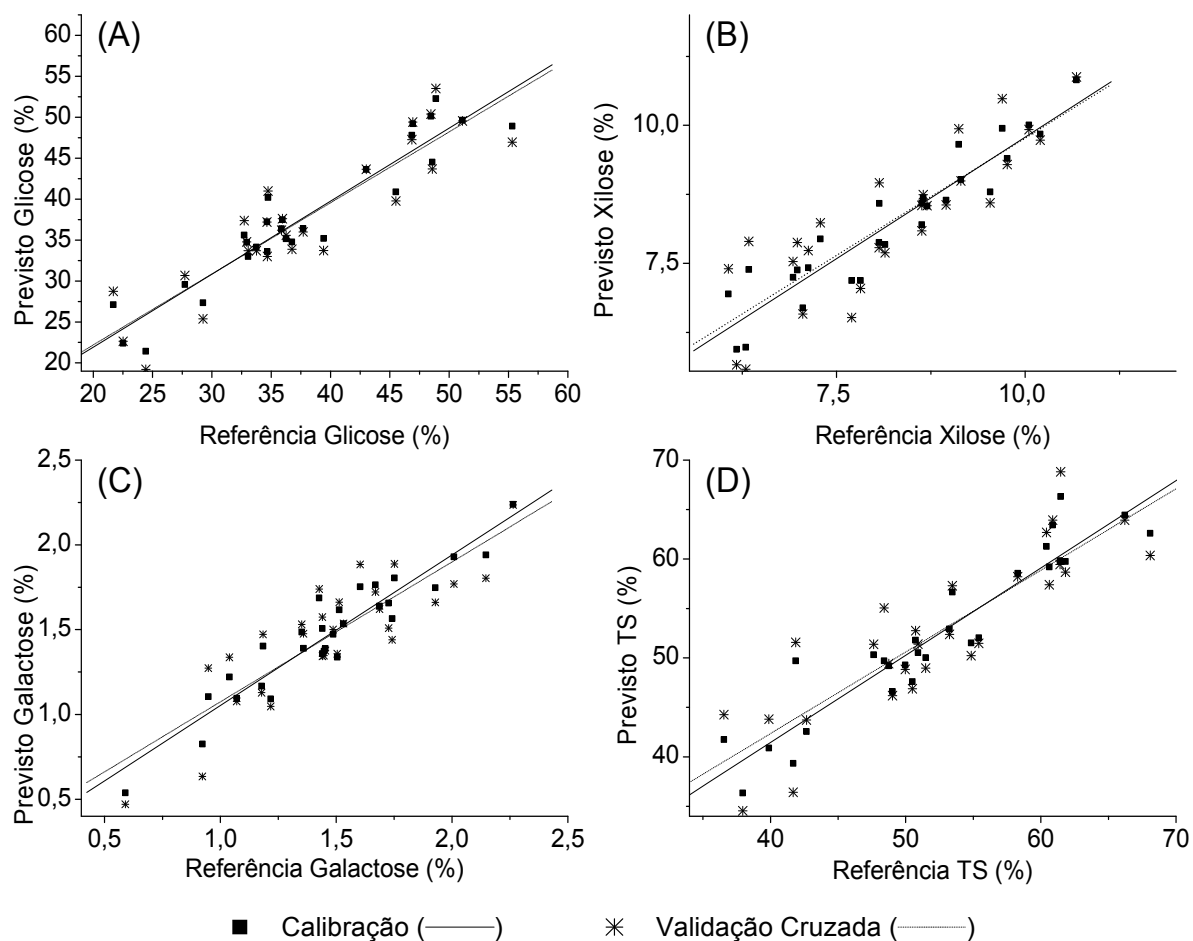


Figura 50. Gráficos de valores de referência versus valores previstos da calibração e validação cruzada para (A) glicose com 3 VL (B) açúcares totais (TS) com 4 VL (C) galactose com 4 VL e (D) xilose com 6 VL para os modelos de banana da fração DS.

O desempenho inferior de alguns componentes dos modelos DU em comparação com os modelos WU, provavelmente resulta do modo como as amostras se distribuem na cubeta durante a varredura. As amostras DU têm a mesma distribuição de tamanho de partícula das amostras WU, mas a ausência de umidade possibilita que as partículas menores não fiquem tão presas às frações maiores, tendo uma tendência para que essas se acumulem na janela da cubeta (segregação), resultando em espectros "seletivos" para as partículas menores quando a amostra está seca (Hayes, 2012).

Desta forma conclui-se que é possível determinar a composição química das biomassas lignocelulósicas, através de modelos de calibração multivariada. Além

desses componentes químicos, é possível ainda fazer a previsão de propriedades físicas de tais resíduos.

Assim uma propriedade de interesse em processos de biorrefinarias que pode ser desenvolvida para as amostras existentes, é o grau de cristalinidade. A cristalinidade da celulose é considerada um determinante importante das taxas e rendimentos dos processos de hidrólise (Yoshida *et al.*, 2008). A cristalinidade da celulose pode ser analisada por difração de raios-X (XRD). No entanto esse método apresenta algumas desvantagens, como operação complicada, altos custos e longos tempos de análise. Dessa forma a variação do índice de cristalinidade pode ser associada aos espectros NIR, os quais permitem revelar tais propriedades físicas. Trabalhos na literatura preveem a cristalinidade da celulose usando a espectroscopia NIR (Basch *et al.*, 1974; Jiang *et al.*, 2007; Qu e Wang, 2011; Zidan *et al.*, 2012).

Isso é possível, uma vez que a cristalinidade da celulose, que envolve as ligação de hidrogênio intermolecular e as forças de rede cristalina, são claramente evidentes nos espectros de infravermelho. Desde que os estiramentos C-O e O-H e modos de deformação C-H predominam na região NIR, é de esperar que esta região possa ser influenciada pela cristalinidade (Basch *et al.*, 1974).

Para os teores de cristalinidade da celulose (IC) das amostras de banana da fração DS, dois modelos de calibração PLS foram obtidos e comparados (Figura 51), com os dados apresentados na Tabela 10. O primeiro baseado no cálculo da cristalinidade pela intensidade e o segundo pelas áreas (Seção 3.2.5), ambos modelos utilizaram 7 VL, e apresentaram R^2_{cal} de 0,89 e 0,82 e R^2_{val} de 0,86 e 0,75, respectivamente, ao serem pré-tratados com 2D(31) na região do NIR. Os erros relativos foram de 6 e 10%, respectivamente, observando-se diferenças significativas entre os dois métodos. Os modelos de calibração para determinar o teor de cristalinidade pela intensidade e pela área, apresentaram respectivamente, valores de RER de 12,4 e 7%, sendo o primeiro aceito para análises de controle de qualidade, e o último apenas para análises de triagem (AACC, 1999).

No entanto avaliando as curvas de calibração de ambos os métodos (Figura 51), percebe-se que apesar do método B, apresentar valores menores de RER e maiores erros relativos, a sua curva de calibração (Figura 51 (B)) apresentou uma melhor

distribuição dos pontos ao longo da reta, o que pode proporcionar um melhor ajuste de novas amostras futuras, demonstrando assim, uma boa capacidade preditiva do índice de cristalinidade para os dois métodos propostos neste trabalho.

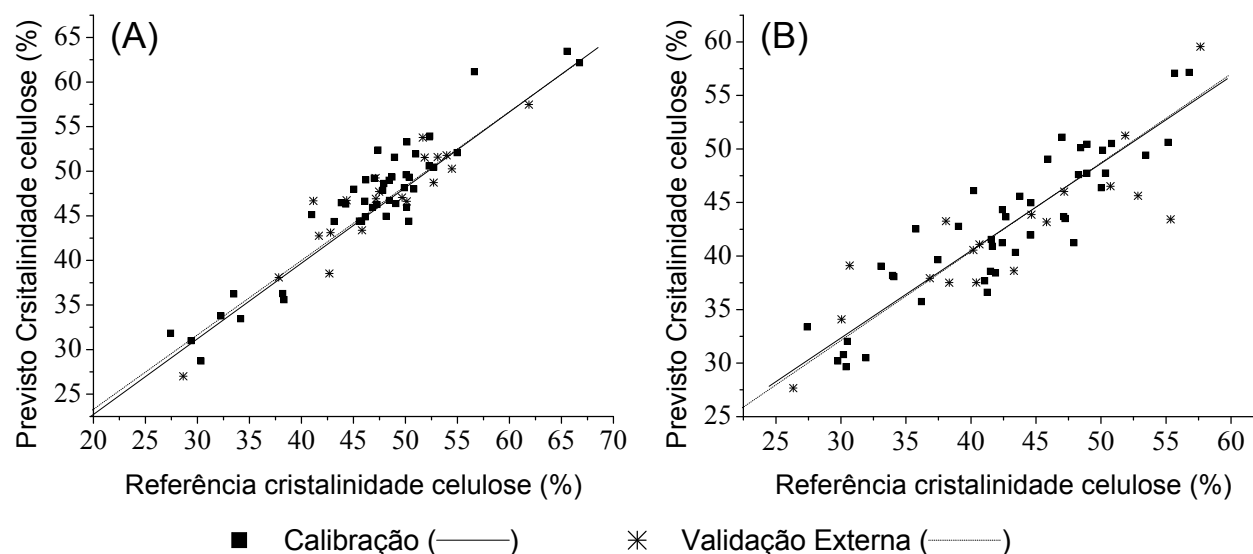


Figura 51. Gráficos de valores de referência versus valores previstos da calibração e validação externa para (A) teor de cristalinidade calculado pelo método da intensidade com 7 VL e (B) teor de cristalinidade calculado pela área com 7 VL, para os modelos de banana da fração DS.

Qu e Wang (2011) avaliaram o teor de cristalinidade de madeira de pinho usando a espectroscopia NIR. O teor de cristalinidade foi calculado utilizando as intensidades de acordo com o método 1 (Seção 3.2.5). Houve uma correlação significativa entre os espectros NIR e as análise por Raios-X, uma vez que o R^2_{cal} foi de 0,87 e o R^2_{val} de 0,83, para 7 VL. Dessa forma foi mostrado que a espectroscopia NIR pode prever rapidamente e com precisão a cristalinidade de madeira de pinhos. Esses resultados estão de acordo com os obtidos nesse trabalho (Tabela 10).

No entanto Kelley *et al.*, (2003) avaliando a cristalinidade da madeira, obtiveram pobre correlação entre a cristalinidade e os espectros NIR, com valores de $R^2_{cal} < 0,50$. Jiang *et al.*, (2007) avaliando a cristalinidade da madeira (pelo método 1) obtiveram excelentes resultados com o uso do espectro inteiro (Vis-NIR) com $R^2_{cal, val}$ de 0,95 e 0,86 com apenas 5 VL.

Apesar de a literatura relatar que o método pela intensidade (método 1) não fornece uma estimativa razoável de cristalinidade da celulose, uma vez que negligencia a variação na largura do pico, foi o método melhor modelado pelo NIR nesse trabalho e o mais usado na literatura para previsão do teor de cristalinidade pela espectroscopia NIR. O que acontece, é que a intensidade pode ser tomada como uma aproximação grosseira da contribuição da celulose amorfa ao espectro de difração, sendo útil apenas para comparar diferenças relativas entre as amostras (Park *et al.*, 2010). No entanto o cálculo da cristalinidade pela intensidade mostrou ser útil na previsão do teor de cristalinidade da celulose nas amostras de diversas frações botânicas da banana, com resultados satisfatórios.

Provavelmente o fator que tenha proporcionado um maior erro no cálculo pelas áreas seja o aumento da contribuição amorfa, uma vez que o principal contribuinte ao pico largo (veja Figura 22) seja a região amorfa, conhecida como celulose amorfa. A solução seria subtrair a contribuição amorfa do difratograma usando um padrão amorfo. O desafio é selecionar um padrão amorfo que seja semelhante ao componente amorfo das amostras. No entanto, como o objetivo do trabalho era apenas testar a viabilidade de modelar a cristalinidade pela espectroscopia NIR e modelos PLS, os resultados foram úteis como forma de demonstrar a possibilidade dessa técnica. Ou seja, o uso dessa metodologia se otimizada com valores de referência mais confiáveis, poderá levar a resultados ainda melhores.

Assim os principais modelos de regressão multivariada pelo método PLS da composição físico-química de biomassas lignocelulósicas e suas frações botânicas foram apresentados. Para a maior parte dos constituintes, resultados satisfatórios foram obtidos.

Não foram observadas diferenças tão significativas entre os modelos para os principais constituintes (lignina total e açúcares totais) entre os diferentes estados da biomassa (DS, DU e WU) com relação aos R^2_{cal} , R^2_{val} , RER e RSD. Em alguns modelos a fração DU apresentou resultados superiores, por exemplo (lignina total dos modelos de café e coco), em outros a fração WU (açúcares totais da banana). Ou seja, de forma geral, para esses dois principais componentes, as 3 frações apresentaram resultados satisfatórios, não observando grandes discrepâncias entre os valores. Observou-se

apenas que um número maior de fatores PLS foi escolhido para os modelos das frações WU e DU, principalmente na modelagem das ligninas. Para os açúcares essa diferença não foi tão notória. As diferenças físicas e químicas entre as frações de plantas distintas são mais significativas, explicando a necessidade de um número expressivo de fatores para a construção dos modelos por PLS, especialmente no estado WU e DU, que correspondem a frações brutas, não moídas.

Para os constituintes de maior importância, teores de açúcares totais e lignina total, os resultados foram todos superiores a $R^2_{cal, val} > 0,86$ e $0,83$; e $R^2_{cal, val} > 0,81$ e $0,74$, respectivamente, para todas as biomassas e frações, demonstrando assim o potencial dessas calibrações em tecnologias de biorrefinarias. O fato dos resultados das calibrações para o conjunto de dados WU serem comparáveis àqueles para os conjuntos DU e DS para os constituintes principais (em alguns casos sendo até mesmo superiores) é extremamente valioso, pois demonstra que a calibração pode ser realizada para as amostras úmidas, que são aquelas encontradas em sistemas *online*, reduzindo assim a necessidade de métodos de preparação demorados e laboriosos para a obtenção de amostras secas e com tamanho de partícula reduzido.

Esta é uma das grandes novidades e potencialidades deste trabalho, uma vez que poucas publicações são encontradas na literatura utilizando modelos baseados em biomassas minimamente processadas (frações DU e WU), adequadas para indústrias de biorrefinarias (Axrup *et al.* 2000; Cozzolino *et al.* 2006; Hayes, 2012). No entanto, as publicações que existem falham em vários parâmetros/sentidos, tais como excesso de VL e previsões pobres para os monossacarídeos individuais, o que significa que os diferentes açúcares não são distinguidos.

Quando o parâmetro RER possui valores maiores do que 10, as determinações são aceitáveis para controle de qualidade. Nesse trabalho, isso aconteceu para os constituintes majoritários, lignina total e açúcares totais da fração DS, e sendo para os constituintes minoritários, em geral, encontrados valores menores de R/SEP. Para as frações DU e WU, em geral, os valores de RER foram inferiores a 10, mas todos superiores a 4,00, o que justifica seu uso em processos de triagem. Para alguns parâmetros, tais como: lignina klason e resíduo ácido insolúvel (modelos de coco e banana da fração DS, respectivamente) lignina total (modelo do Coco, fração DU);

lignina ácida solúvel (modelo Banana, fração DU); extrativos (modelo do Coco da fração DS), manose e ramnose (modelos do Coco, fração WU) o valor de RER foi superior a 15,00%, indicando que os modelos são extremamente bons para uso em quantificações (AACC, 1999).

Alguns trabalhos da literatura relatam (Kelley *et al.*, 2004) que modelos de calibração envolvendo mais do que uma espécie, as diferenças entre elas poderiam sacrificar a precisão desses, devido a diferenças nas estruturas e nas propriedades físico-químicas. No entanto, esse trabalho demonstrou a possibilidade de combinar duas e até três biomassas num único modelo de calibração. Resultados satisfatórios, alcançando R^2_{cal} e R^2_{val} superiores a 0,90 e 0,85, respectivamente, para todos os modelos de ligninas foram obtidos. Da mesma forma, os teores de umidade e extrativos foram modelados satisfatoriamente, com erros relativos inferiores a 19%. O único ponto negativo foi o número maior de VL necessárias para a construção dos modelos combinando 3 biomassas quando comparados aos modelos simples, de uma única matéria-prima. Porém, esse ponto pode ser justificado pelo fato da maior variabilidade estrutural dos analitos, quando se tem uma diversidade tão grande de espécies vegetais.

4.5 Interpretação dos Coeficientes de Regressão

É muito comum na literatura quando se constrói modelos de calibração fazer uma modelagem “cega” sem uma interpretação crítica do que está sendo modelado; ou seja, sem uma interpretação própria da relação com as variáveis, procurando uma relação de causa-efeito desejável. Raros são os casos, pelo menos em artigos de caracterização de biomassas, que os autores apresentam e discutem os pesos (*loadings*) ou os chamados coeficientes de regressão (Kelley *et al.*, 2004; Lestander e Rhén, 2005; Everard *et al.*, 2012; Suchat *et al.*, 2013). Citando Kjeldahl e Bro (2010): “...nós olhamos para um grande número de erros comuns que ocorrem na quimiometria. Geralmente, os problemas parecem ser resultado de uma combinação de enganos na qual o operador apenas aperta o botão. O que geralmente acontece é que os softwares prontamente jogam os diagnósticos e os gráficos na face do usuário, e esse inexperiente é inclinado a aplicar esses resultados sem uma visão crítica. Usando

um conhecimento prévio, técnicas de diagnóstico amplamente utilizadas aparentam ser mais seguras e 'corretas' do que utilizar a razão, embora esse último seja frequentemente preferido. A única forma de seguir é usar da responsabilidade, decidindo o que é relevante baseado em conhecimentos biológicos/químicos e no raciocínio lógico, mantendo sempre o objetivo da modelagem em foco!”. Levando isso em conta, além das ferramentas convencionais de diagnóstico, uma interpretação cuidadosa do vetor dos coeficientes de regressão deve ser realizada, similar a abordagem proposta por Bro e Andersen (2003) para o vetor NAS, ou seja, focando na sua utilidade como uma ferramenta para caracterizar o sistema específico em análise. Isso é útil para identificar sinais espectrais atribuídos ao analito e também para detectar possíveis correlações espúrias, ao acaso. Essa análise assegura a obtenção de modelos confiáveis, visto que o sinal modelado tem relação direta com a concentração do analito de interesse.

Dessa forma os coeficientes de regressão para os modelos PLS desenvolvidos neste trabalho foram todos interpretados, como mostrado a seguir.

Os coeficientes de regressão para as ligninas (lignina total, lignina ácida solúvel, lignina klason e resíduo ácido insolúvel), dos modelos PLS combinados (Banana+Café) bem como o conjunto de banana, foram muito similares e apresentam sinais atribuíveis as estruturas da lignina (Figura 52), tais como: sobretons de estiramento Aril C-H em 676, 1134 e 1670 nm; primeiro estiramento de sobretom Aril O-H (1385 nm); e bandas de combinação de lignina na região de 2150 a 2400 nm (Wolkman, 1997; Ali *et al.*, 2001). As características bipolares na região de 660 nm confirmam a contribuição da clorofila (correlação negativa com teor de lignina) e quarto sobretom do estiramento aril C-H (correlação positiva) para essa região. Uma vez que os coeficientes apresentados foram obtidos a partir de espectros de segunda derivada, os coeficientes negativos correspondem a uma relação direta com o analito modelado.

A maior diferença entre os coeficientes de regressão para as ligninas foi observado para lignina ácida solúvel e seu correlograma é mostrado juntamente com um correlograma típico de lignina, especificamente a lignina Klason, na Figura 52.

A principal diferença é no segundo sobretom do grupo carboxila e estiramento de éster (1943 nm) e na região de combinação de C-Hn (2150 a 2400 nm). Todas as

frações de lignina, exceto a lignina ácida solúvel, apresentam um coeficiente de regressão positivo em 1943 nm, indicando que a lignina, em geral, apresenta uma correlação negativa com grupos carboxil. No entanto, a falta dessa correlação negativa com a fração de lignina ácida solúvel pode ser atribuída a inabilidade do método de determinação para diferenciar a lignina ácida solúvel de outros componentes ácidos solúveis que também absorvem na região do UV, assim como ácidos urônicos (que apresentam o grupo carboxila), açúcares solúveis, furfural e possivelmente ácidos oriundos da degradação de extrativos que não foram completamente removidos na etapa de extração (Sluiter *et al.*, 2010). Isso indica um possível teor de lignina ácida solúvel sobrestimado pelo método de referência, fato já reportado na literatura (Sluiter *et al.*, 2010).

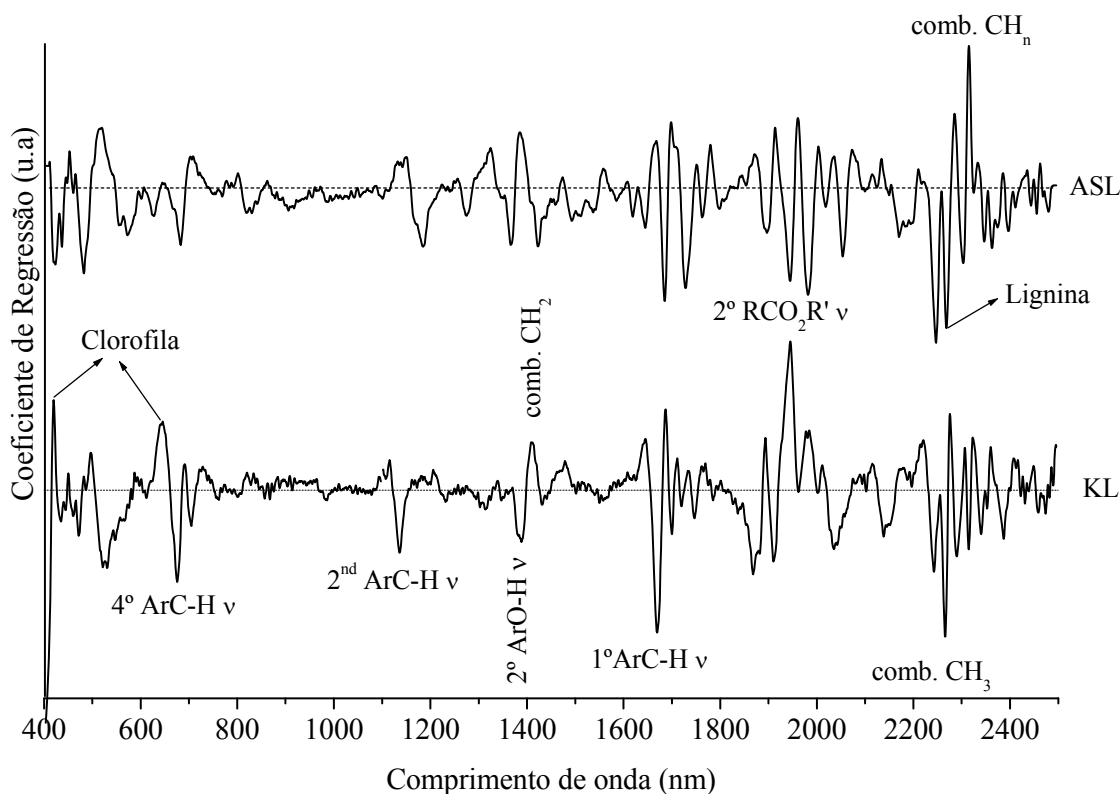


Figura 52. Coeficientes de regressão para os modelos PLS de ligninas (lignina Klason e lignina ácida solúvel). Os ordinais (1°; 2°; e 4°) indicam as bandas de sobretons; comb. se refere às bandas de combinação; e v são bandas de estiramento.

Para os extrativos (Figura 53), ambos os modelos, conjunto combinado de Banana+Café e conjunto de banana apenas, os coeficientes de regressão são muito

similares, exceto para a banda de combinação de Metil (2243 nm) que é mais proeminente no modelo combinado. Os sinais correlacionados com o teor de extrativos são atribuíveis a ceras (1938 nm), proteínas e/ou poliálcoois (2056 nm), componentes típicos presentes nos extrativos (Kuhad e Sigh, 1993). Por outro lado, os ácidos carboxílicos, segundo estiramento de sobretom na região (1910 nm) é positivamente correlacionado com os extrativos, provavelmente oriundo de longas cadeias de ácidos graxos solúveis em etanol, cujo conteúdo é correlacionado positivamente com a fração solúvel em etanol.

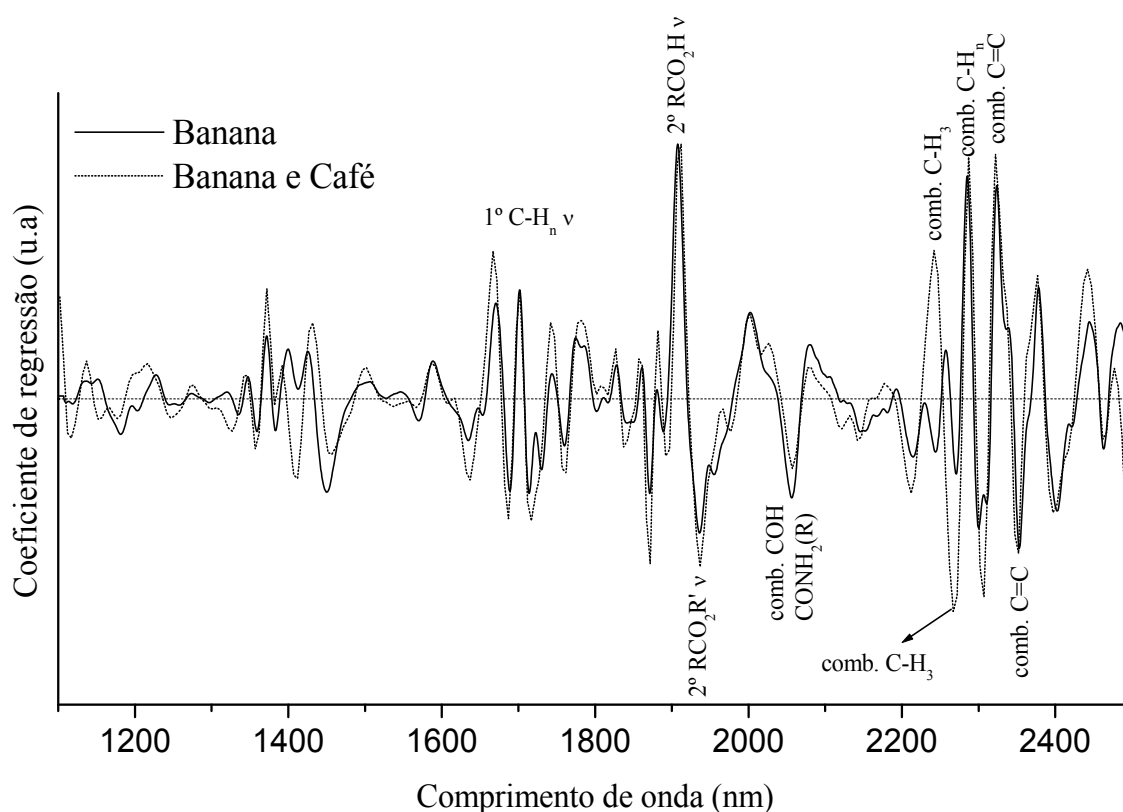


Figura 53. Coeficientes de regressão para os modelos PLS de extrativos (conjuntos combinados de B+C e conjunto apenas da banana). Os ordinais (1º; 2º; e 4º) indicam as bandas de sobretoms; comb. refere-se as bandas de combinação; e v são bandas de estiramento.

As regiões de alquil, de primeiro sobreton do estiramento (1625 - 1775 nm) e bandas de combinação (2200 - 2450 nm) são complexos, com coeficientes de regressão positivos e negativos. Isso é uma consequência da presença, nas amostras

analisadas, de compostos alquil solúveis e insolúveis em etanol, também correlacionados negativamente um com o outro.

Essa complexa estrutura do modelo PLS, indicada pelo aspecto dos coeficientes de regressão e alto número de VL; somado com a sua capacidade preditiva relativamente pobre, menor que a obtida para os modelos de lignina, é devido a mistura de multicompostos característicos da fração dos extrativos. Os melhores desempenhos dos modelos foram ajustados para os conjuntos de uma única biomassa do que para os modelos combinados, sendo um indicativo que nenhum solvente é igualmente aplicável a todas as biomassas (Hakkila, 1989).

Os coeficientes de regressão para os modelos PLS (Café e Coco) de açúcares totais são mostrados na Figura 54.

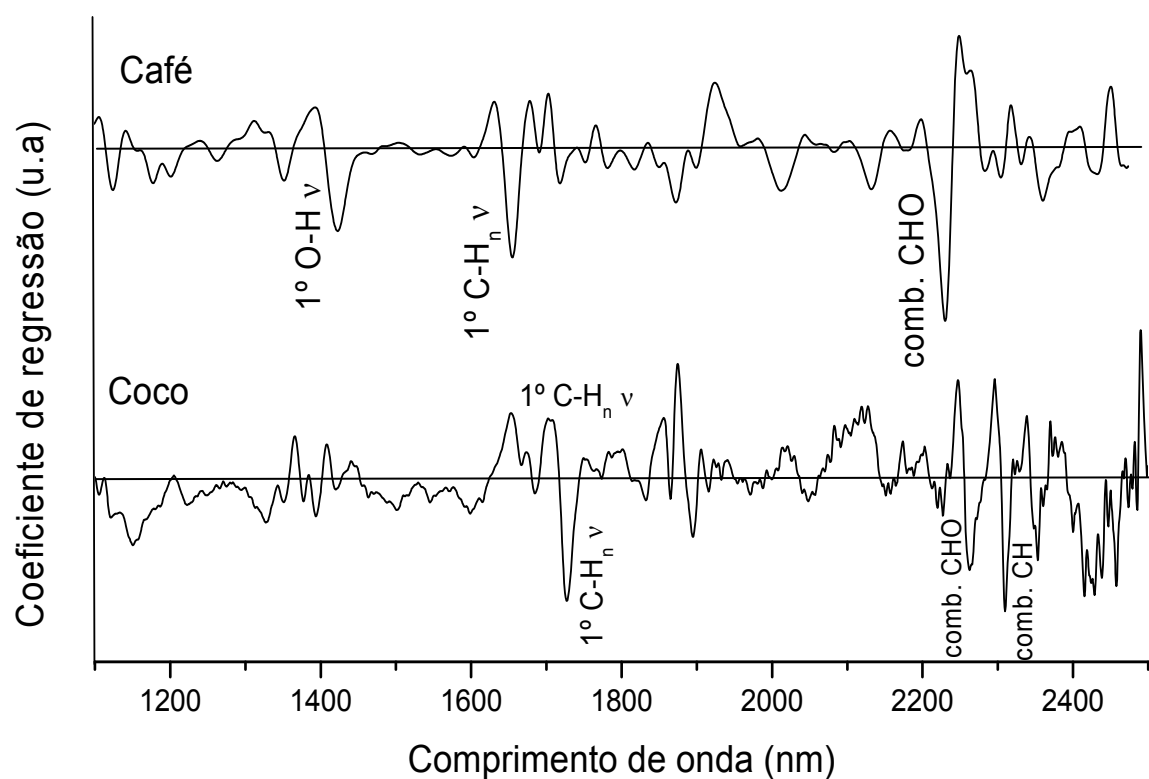


Figura 54. Coeficientes de regressão para o teor de açúcares totais (conjunto de coco e café) dos modelos PLS. Os ordinais (1° ; e 2°) indicam os sobretons; comb. refere se as bandas de combinação; e v são bandas de estiramento.

Uma vez que os coeficientes apresentados para os modelos de café e coco foram obtidos respectivamente, a partir de espectros de segunda e primeira derivada, os coeficientes negativos correspondem a uma relação direta com o analito modelado no primeiro caso.

Os coeficientes de regressão para o café apresentam bandas típicas de carboidratos. Em 1410 nm um sinal de primeiro sobretom do estiramento O-H e em 1669 a 1678 nm um sinal de primeiro sobretom do estiramento C-H. Esse último é atribuído ao estiramento dos grupos acetil da hemicelulose (Fackler *et al.*, 2007; Pires *et al.*, 2012). Ainda uma banda de combinação de CO em 2270 nm típica de polissacarídeos (Shenk *et al.*, 2008). Para a biomassa de coco bandas típicas de carboidratos também foram encontradas. Em 1724 nm um sinal de primeiro sobretom de estiramento C-H, provavelmente furanose ou piranose da hemicelulose (Mitsui *et al.*, 2008). Também uma banda de combinação de CO em 2270 nm e outra banda de combinação de CH em 2329 nm, ambas típicas de polissacarídeos (Shenk *et al.*, 2008).

Os modelos PLS para a umidade, independente do conjunto são similares e apresentam bandas típicas de água (livres e ligadas) (Malin *et al.*, 1999; Lenise e Mangrich 2000; Liu *et al.*, 2000). As regiões negativas correlacionadas com a umidade foram principalmente ácidos carboxílicos, que juntamente com o estiramento alquil de primeiro sobretom e bandas de combinação, podem ser atribuídas a ácidos graxos, componentes hidrofóbicos. Esses componentes provavelmente são ligações de H (primeiro sobretom de grupos de ligação de hidrogênio a 1380 nm). Proteínas e celulose (2000 nm) (Ali *et al.*, 2001) também apresentam uma correlação negativa com o teor de umidade.

Os coeficientes de regressão para a banana dos modelos PLS de cristalinidade da celulose apresentam bandas típicas de celulose cristalina em 1406 nm, 1560 nm e 1600 nm. Uma relação negativa foi encontrada nas regiões de 1428 nm, 1704 nm e 2064 nm, bandas essas, típicas de celulose amorfa (Tsuchikawa e Siesler, 2003).

Uma vez que as cinzas não interagem diretamente com o NIR, seu teor foi determinado de forma indireta com outras espécies que absorvem no NIR (Lestander and Rhén, 2005). Dessa forma a interpretação dos coeficientes de regressão não faz

sentido, uma vez que o modelo de cinzas provavelmente trabalha com correlação espúria, ou seja, devido ao acaso.

4.6 Figuras de Mérito em calibração multivariada

Os valores de RMSEP e RMSEC para os constituintes de cada modelo PLS foram inferiores a 4,50%. Os erros máximos encontrados entre os valores de referência e os valores previstos foram todos menores que 6%.

Visando verificar a robustez, o conjunto completo de dados foi aleatoriamente dividido, retirando 20%, 25% e 30% das amostras para validação de cada vez. Os testes foram repetidos 10 vezes (novas aleatorizações) para cada porcentagem de amostras excluídas, de modo a confirmar que, de fato, os modelos são robustos. Todos os modelos pré-processados com OSC apresentaram-se bastante robustos, com baixos desvios padrões e R^2_{val} similares, indiferente do conjunto utilizado para validação externa, assim como o número de amostras retiradas do conjunto de calibração (Figura 55-A e B).

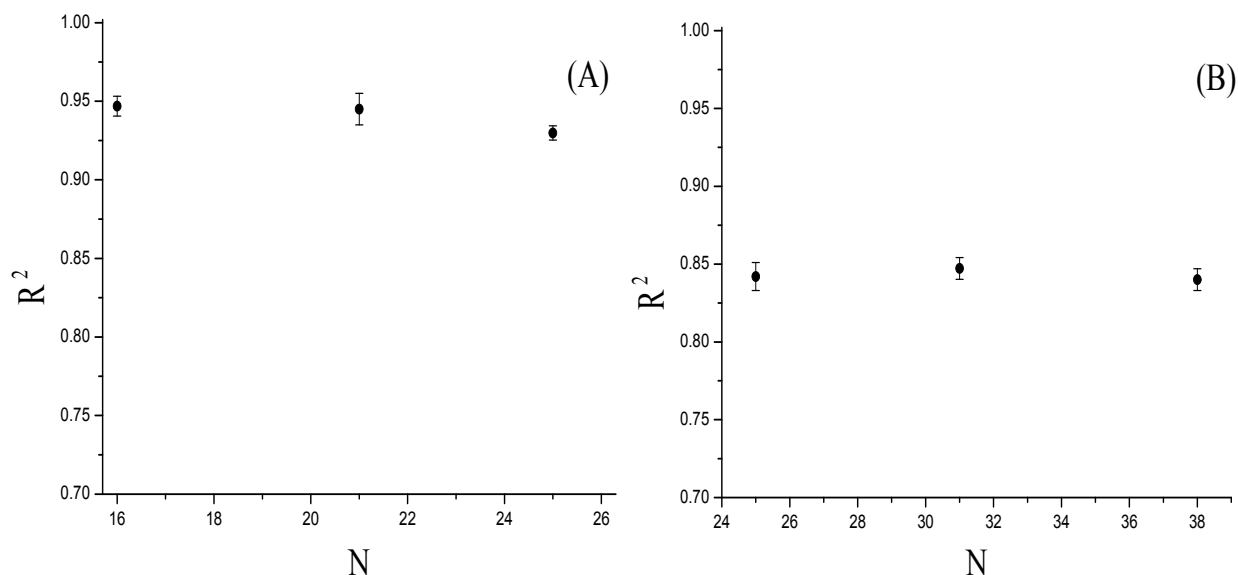


Figura 55. Gráfico do número de amostras no conjunto de validação versus R^2_{val} . (A) Valores médios e desvio padrão de R^2_{val} para o teor de resíduo ácido insolúvel para o modelo de banana. (B) Valores médios e desvio padrão de R^2_{val} para o teor de lignina total para o modelo de Banana + Café.

Avaliando o desvio padrão das duplicatas entre o método de referência e o método NIR, exceto para a umidade, o desvio padrão do NIR foi significativamente menor ($p < 0,05$) do que o método de referência, indicando que o método proposto é mais preciso. Porém nada pode ser dito sobre a exatidão, pois o valor real não é conhecido e o método de referência pode também ter baixa exatidão. A razão pela qual o desvio padrão da umidade pelo método NIR mostrou-se pior do que o método de referência, pode estar ligado ao fato deste não requerer vários passos analíticos na determinação da umidade. Por outro lado, a determinação da composição lignocelulósica pelo método de referência pode levar a erros aditivos, resultando em erros maiores na via úmida do que no método NIR.

Todos os modelos de calibração apresentaram bons valores de R^2_{cal} ($> 0,70$) entre o valor previsto e o medido, exceto para as cinzas e alguns açúcares minoritários. Frações relativamente baixas de *outliers* (no máximo 15%) foram excluídas dos modelos, praticamente mantendo a variabilidade do conjunto de dados original. Além disso, observou-se que todos os modelos seguem o princípio da parcimônia, com baixo número de variáveis latentes, baixas frações de *outliers* removidos e, portanto, os modelos obtidos são potencialmente mais robustos quando comparados com outros estudos (Lomborg *et al.*, 2010; Li *et al.*, 2010; Hayes, 2012).

O coeficiente de correlação dos modelos pseudo-univariados obtidos pelo NAS *versus* os valores de referência foram equivalentes aos obtidos nos modelos multivariados (Figura 56). Os gráficos da Figura 56 (C e D) mostram os ajustes dos modelos multivariados PLS construídos com base nos valores de referência *versus* os valores previstos. Também o ajuste dos valores escalares do NAS *versus* os valores de referência são mostrados nas Figuras 56 (A e B).

A linearidade dos modelos foi verificada pela análise dos gráficos de valores observados *versus* valores previstos, e também pelo gráfico de resíduos *versus* valores previstos. Os pontos devem ser distribuídos de forma linear em torno de uma linha diagonal no primeiro, ou distribuídos aleatoriamente em uma linha horizontal centrada na origem no último gráfico. Essa distribuição foi observada para a maioria dos modelos e a linearidade para a lignina do resíduo ácido insolúvel (AIR) e extrativos pode ser avaliada pela Figura 57.

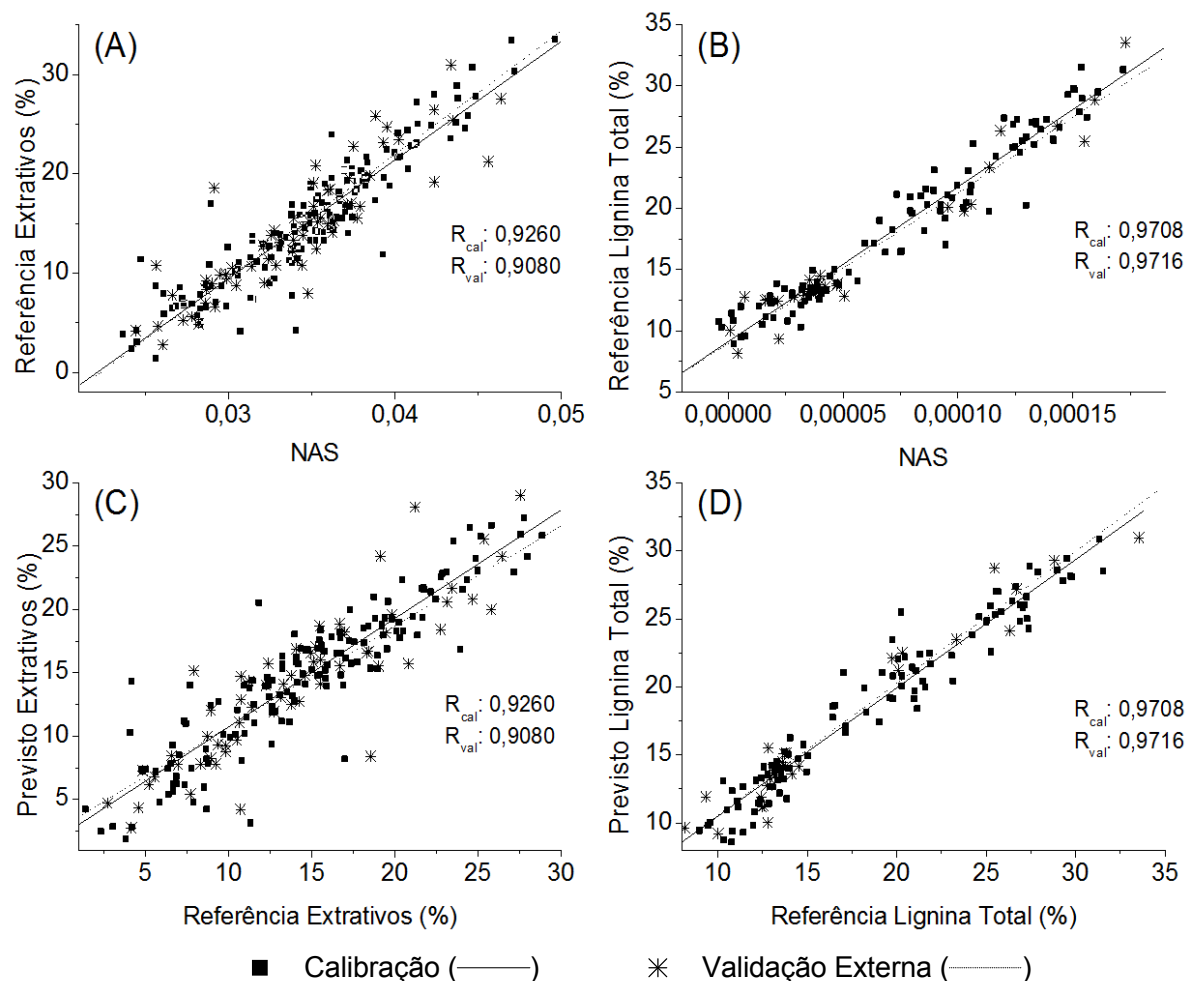


Figura 56. Representação dos modelos nas suas formas (A) pseudo-univariada dos valores de sinal analítico líquido (NAS) versus valores de referência da calibração e validação externa para o teor de extrativos do modelo de Banana+Café+Coco, (B) pseudo-univariada dos valores de sinal analítico líquido (NAS) versus valores de referência da calibração e validação externa para o teor de lignina total do modelo de Banana+Café+Coco (C) multivariada de valores de referência versus valores previstos da calibração e validação externa para o teor de extrativos do modelo de Banana+Café+Coco, (D) multivariada de valores de referência versus valores previstos da calibração e validação externa para o teor de lignina total do modelo de Banana+Café+Coco.

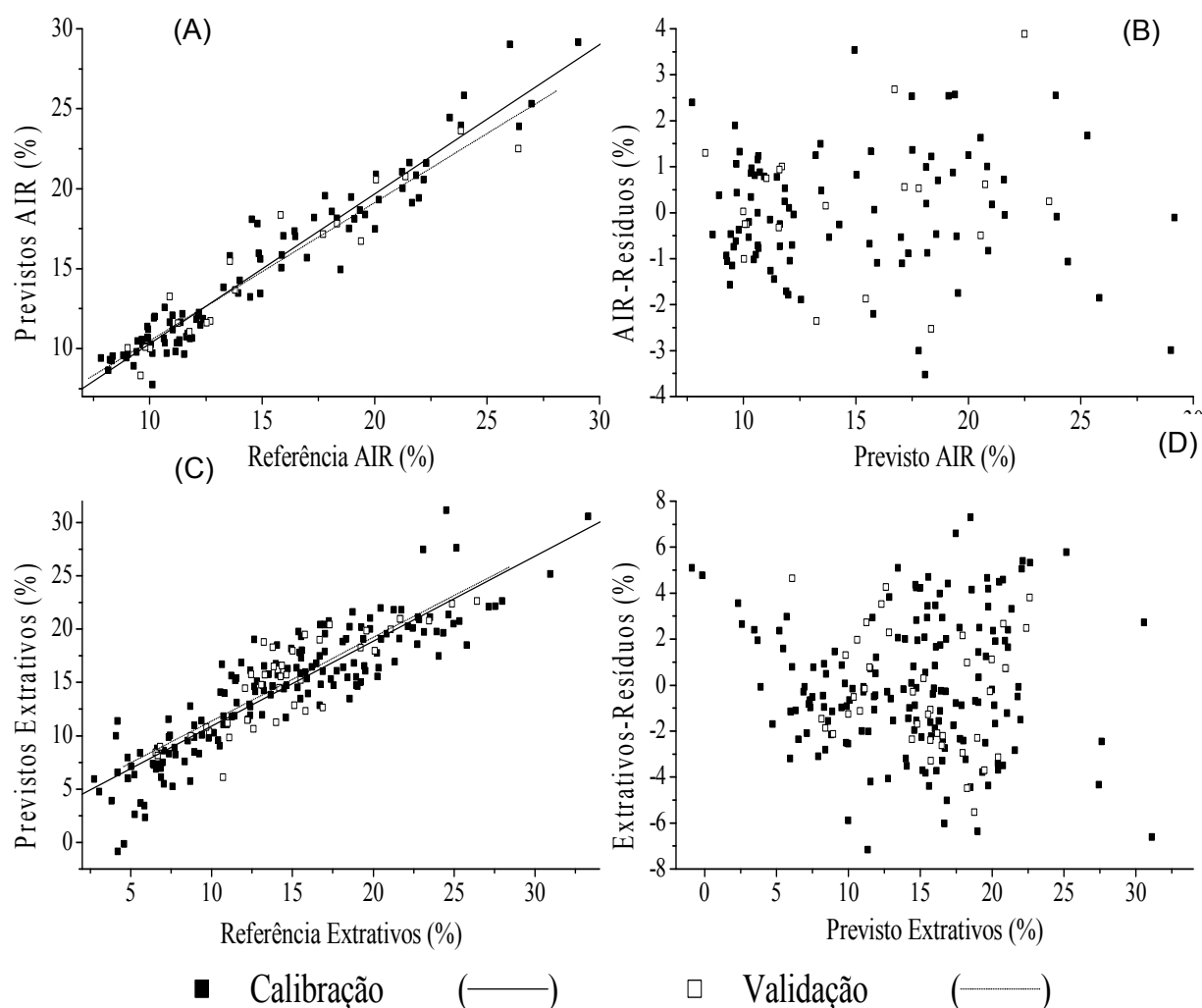


Figura 57. Gráfico dos valores previstos *versus* os de referência para (A) resíduo ácido insolúvel (AIR) e (C) extrativos da fração DS dos modelos de Banana+ Café; Gráfico dos resíduos *versus* os valores previstos para (B) resíduo ácido insolúvel (AIR) e (D) extrativos da fração DS dos modelos de Banana+ Café.

A linearidade das frações DU e WU, dos modelos de extrativos (Banana) e lignina ácida solúvel (ASL) do modelo de Café também foi verificada como mostra a Figura 58. Apesar do menor número de amostras quando comparado aos modelos da fração DS (Figura 57), os resíduos foram bem distribuídos, confirmado assim a linearidade dos modelos (Figura 58).

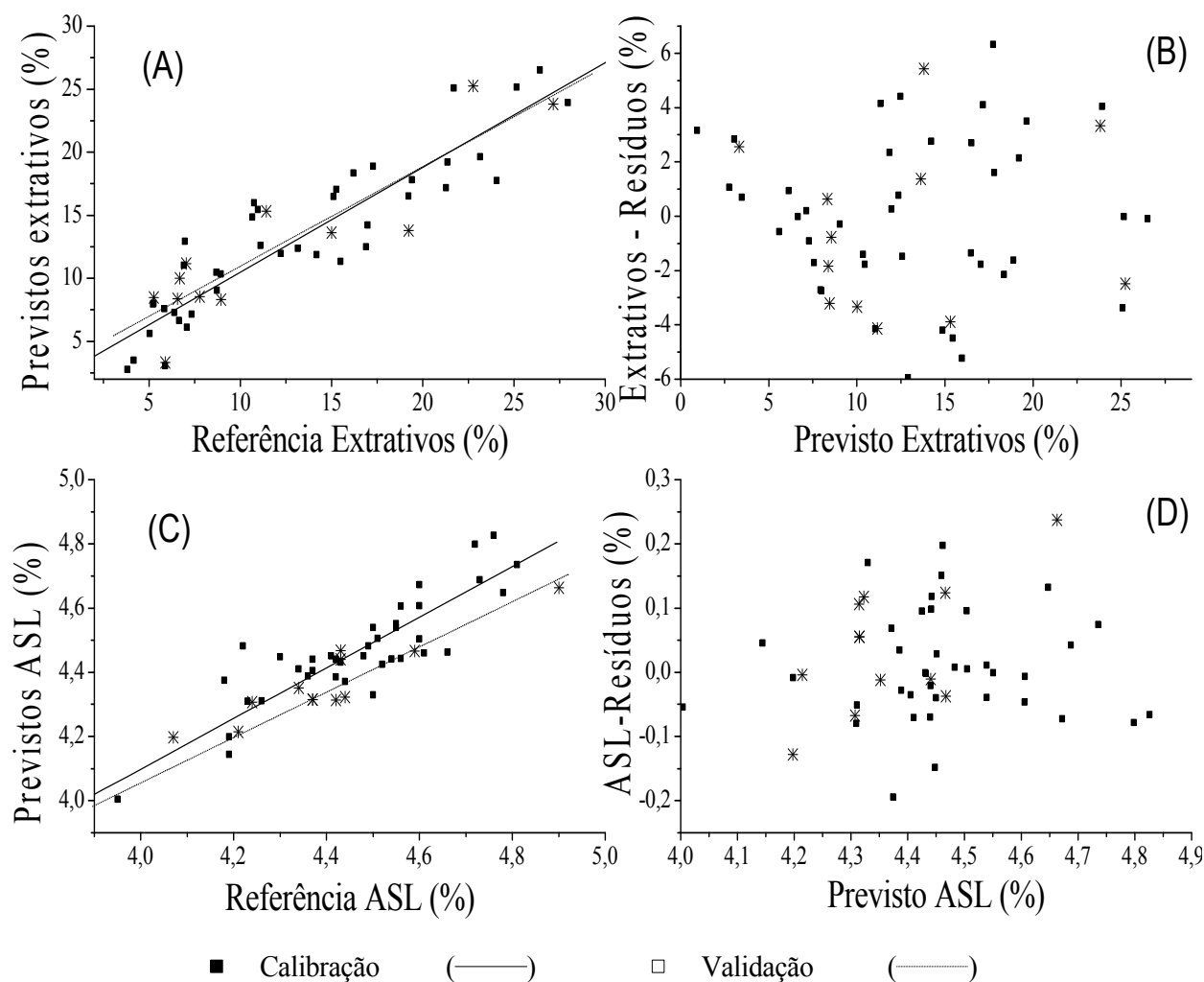


Figura 58. Gráfico dos valores previstos *versus* os de referência para (A) extrativos da fração WU do modelo de Banana e (C) lignina ácida solúvel (ASL) da fração DU do modelo de Café; Gráfico dos resíduos *versus* os valores previstos para (B) extrativos da fração WU do modelo de Banana e (D) lignina ácida solúvel (ASL) da fração DU do modelo de Café.

Finalmente, o viés não foi significativo ($p < 0,05$), quando o teste estatístico t foi aplicado, e, portanto ele pode ser negligenciado, com os modelos não apresentando erros sistemáticos.

Os modelos mostraram SEL superior para os casos em que os pré-tratamentos por OSC foram aplicados, variando de 0,12-0,25 (Tabela 13), em outras palavras, cerca de 12 a 25% do sinal original foi ortogonal para o espaço dos interferentes, transportando apenas a informação do analito modelado (Rocha *et al.*, 2012). No caso

do OSC uma alta SEL era esperada, uma vez que esse pré-processamento elimina os sinais que não têm qualquer relação (são ortogonais) com o analito de interesse. Alguns modelos empregando pré-tratamentos por 2D apresentaram valores de SEL acima de 0,60, indicando que mais de 60% do sinal original foi ortogonal aos interferentes. Por outro lado, para os modelos que empregaram outros pré-processamentos, principalmente a 1D e SNV+DT, a seletividade ficou na faixa de 0,01-0,08. Modelos de lignina (Café) da fração WU apresentam valores muito baixo, na faixa de 0,0002-0,0003, indicando que apenas 0,02 a 0,03 do sinal original foi devido ao analito de interesse.

Os valores de SEN estão no intervalo de $8,6774 \times 10^{-6}$ e 39,60 (Tabelas 13, 14 e 15). Os valores mais baixos de sensibilidade 10^{-6} - 10^{-4} foram obtidos para os modelos que empregaram derivadas, principalmente 2D, como pode ser verificado para todos os modelos de ligninas da fração DS (Tabela 8). Por outro lado, modelos utilizando apenas SNV, ou então derivadas seguidas de SNV (cinzas de Banana+Café da fração DS), umidade (Banana+Café da fração DS), arabinose e ramnose (Banana das frações DS e DU) apresentaram valores de sensibilidade, na faixa de 2,3 a 39,60. Isso, porque o espectro derivado apresenta menor intensidade, como já discutido na seção 4.4 e pode ser visto na Figura 42, e dessa forma, maiores coeficientes de regressão são necessários para manter a proporcionalidade com a concentração do analito. E, por conseguinte, o aumento dos coeficientes de regressão resultam numa diminuição da sensibilidade ($SEN=1/||b||$) (Faber, 1999). Uma vez que a SEN também pode ser obtida a partir dos vetores NAS normalizados para a concentração unitária (Ferré *et al.*, 2001), equação utilizada neste trabalho (Tabela 1), há uma relação direta entre a intensidade do NAS e a sensibilidade. Calculando-se dessa forma a sensibilidade é idêntica a obtida pelo inverso da inclinação da reta do modelo pseudounivariado (Ferré *et al.*, 2001).

No que diz respeito a γ^{-1} , que é a diferença de concentração mínima, estatisticamente perceptível pelo método, todos os valores foram inferiores a 0,6%, exceto para os modelos das ligninas (TL e AIR) e extrativos do café da fração DU, e ainda modelos de açúcares totais do café da fração DS (Tabelas 13, 14 e 15).

Os valores de LD indicaram que o método PLS-NIR foi útil na detecção de ligninas, extrativos, umidade, carboidratos, teor de cristalinidade da celulose e teor de cinzas superior a 0,0006%.

Com exceção das cinzas do modelo da Banana da fração WU (LQ =1,73; e **y** mínimo= 1,59%) extrativos do modelo do café da fração DU (LQ =9,11; e **y** mínimo= 7%) e xilose do modelo de café da fração DU (LQ =3,67; e **y** mínimo= 1,7%), os valores mínimos observados para todos os parâmetros foram maiores do que o LQ, confirmando o bom desempenho dos modelos.

Tabela 13. Resultados das figuras de mérito para os modelos PLS para a fração DS.

y	Conjunto de dados	SEL	SEN	γ	γ^{-1}	LD	LQ
TL	B+C	0,1927	$4,4799 \cdot 10^{-5}$	9,2059	0,1086	0,3259	1,0863
	B	0,2053	$2,8990 \cdot 10^{-5}$	5,9590	0,1678	0,5034	1,6782
	C	0,0777	$1,4893 \cdot 10^{-5}$	4,6698	0,2141	0,6424	2,1414
	COC	0,2499	$3,9459 \cdot 10^{-5}$	8,1084	0,1233	0,4069	1,2332
	B+C+COC	0,0715	$4,3627 \cdot 10^{-5}$	86,2345	0,0116	0,0348	0,1160
KL	B+C	0,1746	$4,8109 \cdot 10^{-5}$	9,8860	0,1011	0,3035	1,0115
	B	0,2453	$1,8137 \cdot 10^{-5}$	3,7269	0,2683	0,8049	2,6832
	C	0,0314	$4,8201 \cdot 10^{-4}$	952,7426	0,0010	0,0031	0,0105
	COC	0,0579	$3,4000 \cdot 10^{-3}$	77,33302	0,0129	0,0426	0,1293
	B+C+COC	0,0712	$4,9974 \cdot 10^{-5}$	98,7796	0,0101	0,0304	0,1012
ASL	B+C	0,2704	$1,5640 \cdot 10^{-4}$	32,1393	0,0311	0,0933	0,3111
	B	0,2700	$9,0355 \cdot 10^{-5}$	18,5671	0,0538	0,1616	0,5386
	C	0,0396	$7,3240 \cdot 10^{-5}$	22,9641	0,0434	0,1306	0,4355
	COC	0,1302	$5,0984 \cdot 10^{-5}$	10,4766	0,0954	0,3149	0,9545
	B+C+COC	0,0372	$2,5963 \cdot 10^{-4}$	513,1776	0,0019	0,0058	0,0195
AIR	B+C	0,1198	$4,3713 \cdot 10^{-5}$	8,9825	0,1113	0,3349	1,1130
	B	0,2586	$1,7167 \cdot 10^{-5}$	3,5277	0,2834	0,8504	2,8347
	C	0,0606	$9,1649 \cdot 10^{-4}$	1811,6	0,0005	0,0017	0,0055
	COC	0,2045	$7,6785 \cdot 10^{-4}$	17,46475	0,0572	0,1889	0,5725
	B+C+COC	0,0544	$3,9924 \cdot 10^{-5}$	78,9144	0,0126	0,0380	0,1267
EXTRAT	B+C	0,1563	0,1163	7,02946	0,1422	0,4694	1,4225
	B	0,0635	$9,9090 \cdot 10^{-3}$	328,974	0,0030	0,0100	0,0303
	C	0,1623	0,0181	10,4409	0,0957	0,2873	0,9578
	COC	0,0940	0,0043	2,4919	0,4013	1,2039	4,0130
	B+C+COC	0,1661	0,0425	24,5005	0,0408	0,1224	0,4082
CINZAS	B+C	0,1054	7,8521	274,6415	0,0036	0,0109	0,0364
	B	0,0915	$1,4150 \cdot 10^{-4}$	30,1025	0,0332	0,0997	0,3322
	C	0,0306	$1,9333 \cdot 10^{-4}$	60,6174	0,0165	0,0495	0,1650
	COC	0,0777	$5,6399 \cdot 10^{-4}$	119,9810	0,0083	0,0250	0,0833
	B+C+COC	0,0389	$3,2179 \cdot 10^{-4}$	636,0469	0,0015	0,0047	0,0157
UMIDA	B+C	0,0589	2,3170	8,663377	0,1154	0,3809	1,1542
	B	0,0473	$3,4924 \cdot 10^{-5}$	7,110569	0,1406	0,4640	1,4063
	C	0,0348	$5,7682 \cdot 10^{-5}$	18,0860	0,0553	0,1659	0,5529
	COC	0,2292	$7,7147 \cdot 10^{-5}$	16,4120	0,0609	0,1828	0,6093
	B+C+COC	0,0377	$7,5265 \cdot 10^{-5}$	148,7690	0,0067	0,0202	0,0672
GLIC	B	0,1392	0,5169	15,8706	0,0630	0,1890	0,6301
	C	0,1098	$2,6015 \cdot 10^{-6}$	8,1568	0,1225	0,3678	1,2260
	COC	0,1341	$2,0673 \cdot 10^{-5}$	6,4819	0,1542	0,4628	1,5428
XIL	B	0,0624	0,7479	22,9613	0,0435	0,1307	0,4355
	C	0,2388	3,3231	102,0232	0,0098	0,0294	0,0980
	COC	0,1763	$5,8968 \cdot 10^{-5}$	18,4893	0,0540	0,1623	0,5409

GALAC	B	0,0451	12,1875	374,1651	0,0026	0,0080	0,0267
	C	0,0348	0,2881	8,8446	0,1130	0,3392	1,1306
	COC	0,0740	0,0515	29,6631	0,0337	0,1011	0,3371
RAM	B	0,2182	28,0984	862,65	0,0011	0,0030	0,0120
	C	0,1363	0,0434	24,9730	0,0040	0,1201	0,4004
	COC	-	-	-	-	-	-
ARAB	B	0,2649	15,2674	468,7218	0,0021	0,0064	0,0213
	C	0,0319	0,1634	13,7147	0,0729	0,2187	0,7291
	COC	0,1196	0,0216	12,4411	0,0803	0,2411	0,8038
MAN	B	0,0117	0,3414	10,4820	0,0954	0,2862	0,9540
	C	0,1553	3,9141	120,1674	0,0083	0,0250	0,0832
	COC	-	-	-	-	-	-
TS	B	0,1981	0,5008	54,6529	0,0182	0,1287	0,8807
	C	0,0428	$2,0300 \cdot 10^{-6}$	0,6365	1,5710	4,7132	15,7106
	COC	0,2357	$1,5007 \cdot 10^{-5}$	4,7056	0,2125	0,6375	2,1251
IC	B ¹	0,0322	$4,1023 \cdot 10^{-6}$	8,1086	0,1233	0,3700	1,2333
	B ²	0,1390	$1,6751 \cdot 10^{-5}$	5,2522	0,1904	0,5712	1,9040

B:banana; C:café; Coc: coco; B+C: banana+café; B+C+Coc: Banana+Café+Coco; TL: lignina total; KL: lignina klason; ASL: lignina ácida solúvel; AIR: resíduo ácido insolúvel; Umid: umidade; Extr: extrativos; Glic: glicose; Xil: xilose; Gala:galactose; Ram: ramnose; Arab:arabinose; Man: manose; TS: açúcares totais; IC: Índice de cristalinidade; B¹: Índice de cristalinidade pela intensidade; B²: Índice de cristalinidade pela área; Sel:seletividade; Sen:sensibilidade; γ: Sensibilidade analítica; γ⁻¹: inverso da sensibilidade analítica; LQ: limite de quantificação; LD: limite de detecção.

Tabela 14. Resultados das figuras de mérito para os modelos PLS para a fração DU.

y	Conjunto de dados	SEL	SEN	γ	γ^{-1}	LD	LQ
TL	B	0,0803	0,9751	81,8422	0,0122	0,0367	0,1222
	C	0,0407	$4,6174 \cdot 10^{-5}$	1,0297	0,9711	2,9136	9,7120
	COC	0,0584	$1,0140 \cdot 10^{-5}$	2,1571	0,4635	1,3907	4,6358
KL	B	0,0486	0,7729	64,8771	0,0154	0,0462	0,1541
	C	0,0248	$3,6312 \cdot 10^{-5}$	8,0973	0,1234	0,3705	1,2350
	COC	0,0594	$1,0532 \cdot 10^{-5}$	2,2406	0,4463	1,3389	4,4631
ASL	B	0,0981	0,0529	30,4474	0,0328	0,0985	0,3284
	C	0,2126	0,0581	33,4808	0,0298	0,0896	0,2987
	COC	0,6578	0,0018	390,2017	0,0025	0,0077	0,0256
AIR	B	0,0301	0,0049	2,8353	0,3526	1,0581	3,5269
	C	0,0519	$7,3282 \cdot 10^{-5}$	1,6341	0,6119	1,8358	6,1194
	COC	0,1101	$1,9839 \cdot 10^{-5}$	4,2204	0,2369	0,7108	2,3695
EXTRAT	B	0,1150	0,0159	9,1384	0,1094	0,3283	1,0943
	C	0,0234	0,0019	1,0968	0,9117	2,7352	9,1174
	COC	0,0617	1,8474	155,0661	0,0064	0,0193	0,0645
CINZAS	B	0,0534	1,5460	167,9085	0,0059	0,0187	0,0483
	C	0,1180	0,0765	44,0593	0,0227	0,0681	0,2270
	COC	0,1695	0,0017	359,5214	0,0027	0,0083	0,0278
GLIC	B	0,0544	0,1281	3,9327	0,2542	0,7628	2,5428
	C	0,0328	$5,4336 \cdot 10^{-6}$	1,7037	0,5869	1,7609	5,8696
	COC	0,0747	0,0915	7,6832	0,1301	0,3905	1,3015
XIL	B	-	-	-	-	-	-
	C	0,0322	$8,6774 \cdot 10^{-6}$	2,7208	0,3675	1,1026	3,6754
	COC	0,0284	0,0572	4,8041	0,2081	0,6245	2,0815
GALAC	B	0,3590	16,7529	514,3258	0,0019	0,0058	0,0194
	C	0,3504	$6,6588 \cdot 10^{-5}$	20,8786	0,0478	0,1437	0,4790
	COC	0,0780	1,5408	129,3289	0,0077	0,0232	0,0773
RAM	B	-	-	-	-	-	-
	C	-	-	-	-	-	-
	COC	-	-	-	-	-	-
ARAB	B	0,6168	39,6108	1216,1	0,0008	0,0025	0,0082
	C	0,1045	$4,0615 \cdot 10^{-5}$	12,7348	0,0785	0,2356	0,7852
	COC	-	-	-	-	-	-
MAN	B	-	-	-	-	-	-
	C	-	-	-	-	-	-
	COC	-	-	-	-	-	-
TS	B	0,2279	0,1481	4,5465	0,2199	0,6598	2,1995
	C	0,0706	0,2314	7,1040	0,1407	0,4223	1,4077
	COC	0,1730	0,0462	3,8806	0,2576	0,7731	2,5769

Legenda idêntica da tabela 13.

Tabela 15. Resultados das figuras de mérito para os modelos PLS para a fração WU.

y	Conjunto de dados	SEL	SEN	γ	γ^{-1}	LD	LQ
TL	B	0,0390	0,0019	102,9440	0,0097	0,0291	0,0971
	C	0,0003	$2,9000 \cdot 10^{-5}$	6,1693	0,1620	0,4863	1,6209
	COC	0,0192	$5,9159 \cdot 10^{-5}$	116,9352	0,0008	0,0257	0,0855
KL	B	0,0417	0,0026	141,8053	0,0070	0,0212	0,0705
	C	0,0002	$2,6198 \cdot 10^{-5}$	5,5732	0,1794	0,5383	1,7943
	COC	0,0186	$5,7001 \cdot 10^{-6}$	11,2669	0,0887	0,2663	0,8876
ASL	B	0,0103	0,0023	123,7509	0,0080	0,0242	0,0808
	C	0,3975	0,1604	92,3812	0,0108	0,0325	0,1082
	COC	0,0651	$2,4425 \cdot 10^{-4}$	482,7779	0,0020	0,0062	0,0207
AIR	B	0,0415	0,0024	129,4168	0,0077	0,0232	0,0773
	C	0,0002	$2,6087 \cdot 10^{-5}$	5,5496	0,1801	0,5406	1,8019
	COC	0,0288	$6,1743 \cdot 10^{-6}$	12,2043	0,0819	0,2458	0,8194
EXTRAT	B	0,1609	0,0325	18,7376	0,0533	0,1601	0,5337
	C	0,1549	0,0140	8,0452	0,1242	0,3729	1,2430
	COC	0,1181	0,0189	10,8988	0,0917	0,2753	0,9175
CINZAS	B	0,1537	0,0026	5,7699	0,1733	0,5199	1,7331
	C	0,0236	$1,5656 \cdot 10^{-4}$	33,3050	0,0300	0,0901	0,3003
	COC	0,1963	0,0906	4979,7	0,0002	0,0006	0,0020
GLIC	B	0,0275	0,0682	1,6682	0,5994	1,7983	5,9945
	C	0,0772	$6,1726 \cdot 10^{-6}$	12,2001	0,0819	0,2459	0,8197
	COC	0,0673	$7,5437 \cdot 10^{-6}$	14,9110	0,0670	0,2012	0,6706
XIL	B	-	-	-	-	-	-
	C	0,1762	2,4913	87,1365	0,0114	0,0344	0,1148
	COC	0,2265	0,0079	4,5256	0,2209	0,6629	2,2097
GALAC	B	-	-	-	-	-	-
	C	0,0128	0,1055	8,8548	0,1129	0,3388	1,1293
	COC	0,6907	0,0018	563,6627	0,0017	0,0053	0,0177
RAM	B	-	-	-	-	-	-
	C	-	-	-	-	-	-
	COC	0,2970	0,0046	985,1950	0,0010	0,0030	0,0102
ARAB	B	-	-	-	-	-	-
	C	0,0634	0,8417	25,8415	0,0386	0,1161	0,3870
	COC	-	-	-	-	-	-
MAN	B	-	-	-	-	-	-
	C	-	-	-	-	-	-
	COC	0,2830	0,0018	376,5809	0,0026	0,0080	0,0266
TS	B	0,1004	0,0968	2,3681	0,4222	1,2668	4,2228
	C	0,1045	$5,3577 \cdot 10^{-6}$	10,5900	0,0944	0,2833	0,9443
	COC	0,0643	$5,0064 \cdot 10^{-6}$	9,8956	0,1010	0,3032	1,0105

Legenda idêntica da tabela 13.

5 CONCLUSÕES

A utilização de recursos naturais e renováveis é muito importante para uma sociedade economicamente viável e ambientalmente correta. As matérias-primas analisadas nesse trabalho são abundantes e promissoras no Brasil para oportunidades em biorefinarias. A caracterização físico-química das diferentes biomassas mostrou que, em geral, todos os resíduos são potenciais candidatos, pois apresentam parâmetros importantes para a produção de bioenergia, tais como, poder calorífico elevado, alto conteúdo de carboidratos, altos rendimentos em bioprodutos, baixo teores de extrativos e cinzas. Ainda quando se pensa na utilização da biomassa para obtenção de insumos aromáticos ou na produção de energia altos teores de lignina foram encontrados para algumas biomassas favorecendo os processos citados. Em alguns casos, como da casca de arroz, os altos teores de cinzas, não são prejudiciais, mas sim favorecem seu aproveitamento na obtenção de outros insumos importantes. Dessa forma as biomassas estudadas nesse trabalho, principalmente as cascas de café, fibras de coco e resíduos de banana, apresentam potencial para competir favoravelmente com outras fontes convencionais, como o bagaço e a palha da cana de açúcar, em tecnologias de biorrefinarias.

Uma vez que os principais constituintes (açúcares totais e lignina total) são os mais importantes na determinação dos rendimentos de biorrefinaria, as precisões dos modelos sugerem que a espectroscopia NIR tem potencial na análise de matérias-primas lignocelulósicas em processos rápidos, baratos e sem laboriosas etapas de preparação, o que indica seu potencial para desenvolvimento de sensores *online* em indústrias de biorrefinarias. Por outro lado, os carboidratos minoritários e constituintes majoritários das frações WU e DU podem ser previstos, ao menos, para fins de calibração de triagem (*screening calibration*).

Com exceção de alguns açúcares dos modelos de banana das frações DU e WU que não foram calibrados, todos os demais modelos, apresentaram resultados considerados razoáveis a bons.

Conclui-se que a espectroscopia NIR é uma técnica viável para a caracterização química, com vistas ao emprego em biorrefinarias, de resíduos de banana, coco e café separadamente, bem como combinando as biomassas em um único modelo, uma vez que %RMSEP relativos ficam em torno de 3%, e %ER menores que 20% (exceto para cinzas e a manose) e coeficientes de determinação superiores a 0,80 (exceto cinzas e alguns açúcares minoritários) o que é bastante satisfatório.

A grande variabilidade na composição química devido as diferentes frações botânicas foi muito importante para garantir a boa qualidade dos modelos.

Os melhores resultados foram obtidos quando os pré-tratamentos OSC, derivadas e derivadas+SNV foram aplicados. Com a exceção de alguns modelos (frações DU e WU, principalmente) os métodos foram considerados validados, como demonstram os resultados satisfatórios obtidos para todas as figuras de mérito, bem como para a validação externa.

As estatísticas de calibração e previsão demonstram que a espectroscopia NIR é adequada, e, quando acoplada a quimiometria é uma metodologia rápida, barata, não necessita de descarte de produtos nocivos ao ambiente, não destrói as amostras e de potencial para a análise da composição química das biomassas, sendo que calibrações precisas podem ser desenvolvidas para ligninas, açúcares majoritários, extrativos e umidade de ambas as amostras, secas e úmidas, independente do tamanho da partícula.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAF. Associação Brasileira de Produtores de Florestas Plantadas. (www.abraflor.org.br/). Acessado: Março/2013.
- AGUILAR, R. *et al.* Kinetic study of the acid hydrolysis of sugar cane bagasse. **J. Food Eng.** v. 55, p. 309-318, 2002.
- ALI, M. *et al.* Spectroscopic studies of the ageing of cellulosic paper. **Polym.** v.42, p.2893-2900, 2001.
- ANDRADE E CANELLAS. Disponível em <http://www.andradecanellas.com.br>. Capim-elefante ganha força na produção de eletricidade, 2012.
- ARREDONDO, H. I. V. *et al.* Ethanol production process from banana fruit and its lignocellulosic residues: Energy analysis. **Energy.** v.35, p.3081-3087, 2010.
- ASTM E 1655-00. Standard Practices for Infrared Multivariate Quantitative Analysis; 2005.
- ASTM D 3173-87. Standard method for determination of moisture content in biomass; 2003.
- ASTM D 3174-04. Standard method for ash in the analysis sample of coal and coke; 2004.
- ASTM D 3175-07. Standard method for volatile matter in the analysis sample of coal; 2007.
- AXRUP, L., MARKIDES, K., NILSSON, T. Using miniature diode array NIR spectrometers for analyzing wood chips and bark samples in motion. **J. Chemom.** v.14, p. 561–572, 2000.
- AZADI, R. *et al.* Liquid fuels, hydrogen and chemicals from lignin: A critical review. **Renew. Sus. Energ. Rev.** v. 21, p. 506–523, 2013.
- AZIZ, N. A. A *et al.* Chemical and functional properties of the native banana (*Musa acuminata* x *balbisiana* Colla cv. Awak) pseudo-stem and pseudo-stem tender core flours. **Food Chem.** v. 128, p.748–753, 2011.
- BANERJI, A.; BALAKRISHNAN, M.; KISHORE, V.V.N. Low severity dilute-acid hydrolysis of sweet sorghum bagasse. **Appl. Energ.** v.104, p.197–206, 2013.
- BARNES, R. J.; DHANOA, M. S.; LISTER, S.J. Standard Normal Variate Transformation and Detrending of Near-Infrared Diffuse Reflectance Spectra. **Appl. Spectrosc.** v.43, p.772-777, 1989.
- BASCH, A.; WASSERMAN, T.; LEWIN, M. Near-Infrared Spectrum of Cellulose: A New Method for Obtaining Crystallinity Ratios. **J. Polym. Sci.** v. 12, p. 1143-1150, 1974.
- BAUM, A. *et al.* Rapid near infrared spectroscopy for prediction of enzymatic hydrolysis of corn bran after various pretreatment. **New Biotechnol.** v. 29, p.293-301, 2012.
- BERENDT, R.T. *et al.* Solid-state NMR spectroscopy in pharmaceutical research and analysis. **Anal.Chem.** v.25, p.977–984, 2006.
- BEVILAQUA, D. *et al.* Cleaner Production: Levulinic Acid from Rice Husks. **J. Clean. Prod.** p.1-6, 2013. DOI: 10.1016/j.jclepro.2013.01.035

BIÈVRE, P. Measurement results without statements of reliability (uncertainty) should not be taken seriously. **Accred. Qual. Assur.** v.2, p.269, 1997.

BILBA, K.; ARSENE, M. A; OUENSANGA, A. Study of banana and coconut fibers. Botanical composition, thermal degradation and textural observations. **Bioresour. Technol.** v.98, p.58–68, 2007.

BIN, Y.; HONGZHANG, C. Effect of the ash on enzymatic hydrolysis of steam-exploded rice straw. **Bioresour. Technol.** v.101, p.9114–9119, 2010.

BINOD, P. *et al.* Bioethanol production from rice straw: An overview. **Bioresour. Technol.** v.101, p. 4767–4774, 2010.

BOZELL, J. J. *et al.* Production of levulinic acid and use as a platform chemical for derived products. **Resour. Conserv. Recycl.** v.28, p. 227-239, 2000.

BOYSWORTH, M. K.; BOOKSH, K. S. Aspects of multivariate calibration applied to near infrared spectroscopy. In: Burns, D. A. & Ciurczak, E. W. (2 ed.). Handbook of Near Infrared Analysis. New York, 2001.

BRERETON, R. G. Multivariate Pattern Recognition in Chemometrics, Illustrated by Case Studies, Elsevier, Amsterdam, 1992.

BRERETON R. G. Chemometrics: Applications of Mathematics and Statistics to Laboratory Systems, Ellis Horwood, Chichester, 1993.

BRERETON, R. G. Introduction to multivariate calibration in analytical Chemistry. **Analyst.** 2000, v.125, p. 2125-2154.

BRO, R.; ANDERSEN, C.M. Theory of net analyte signal vectors in inverse regression. **J.Chemometr.** v.17, p. 646–652, 2003.

BROSSE, N. *et al.* Miscanthus: a fast- growing crop for biofuels and chemicals production. **Biofuels, Bioprod. Bioref.** v.6, p. 580-598, 2012.

BRUUN, S. *et al.* Prediction of the degradability and ash content of wheat straw from different cultivars using near infrared spectroscopy. **Ind. Crop Prod.** v.31 321–326, 2010.

CASSALES, A. *et al.* Optimization of soybean hull acid hydrolysis and its characterization as a potential substrate for bioprocessing. **Biomass Bioenerg.** v. 35, p. 4675-4683, 2011.

CHANG, C.; CEN, P.; Ma, X. Levulinic acid production from wheat straw. **Bioresour. Technol.** v. 98, p.1448–1453, 2007.

CHARLTON, A. *et al.* The biorefining opportunities in Wales: Understanding the scope for building a sustainable, biorenewable economy using plant biomass. **Chem. Eng. Res. Des.** v. 87, p.1147–1161, 2009.

CHEN, H, *et al.* Qualitative and quantitative analysis of wood samples by Fourier transform infrared spectroscopy and multivariate analysis. **Carbohydr. Polym.** v. 82, p.772–778, 2010.

CHENG, H. *et al.* Geographical origin identification of propolis using GC-MS and electronic nose combined with principal components analysis. **Food Res. Int.** v.51, p.813-822, 2013.

CIRLINI, M. *et al.* HS-SPME/GC–MS and chemometrics for the classification of Balsamic Vinegars of Modena of different maturation and ageing. **Food Chem.** v.124, p.1678-1683, 2011.

CISSE, I. K. *et al.* Mechanical Characterisation of Foller Sandcretes with Rice Husk Ash Additions Study Applied to Senegal. **Cem. Concrete Res.** v. 30, p. 13-18, 2000.

CONAB. Companhia Nacional de Abastecimento. (www.conab.gov.br/). Acessado: Março/2013.

COZZOLINO, D. *et al.* Measurement of chemical composition in wet whole maize silage by visible and near infrared reflectance spectroscopy. **Anim. Feed Sci. Technol.** v.129, p. 329–336, 2006.

CRESPO, E. *et al.* Volatile organic compound emissions from elephant grass and bamboo cultivars used as potential bioethanol crop. **Atmospheric Environ.** v. 65, p. 61-68, 2013.

CRUZ, A.G. *et al.* Assessing the use of different chemometric techniques to discriminate low-fat and full-fat yogurts. **LWT - Food Sci. Technol.** v. 50, p. 210-214, 2013.

DAVIS, B. G.; FAIRBANKS, A. J. Carbohydrate Chemistry. In: Oxford University Press, Oxford, 2002.

DIBANET. Development of Integrated Biomass Approaches NETwork. Disponível em <http://www.dibanet.org>.

DHILLON, R.S; WUEHLISCH, G. Mitigation of global warming through renewable biomass. **Biomass Bioenerg.** v.4 8, p. 75-89, 2013.

DOMÍNGUEZ, M.T.G. Furfural production from Eucalyptus globulus: Optimizing by using neural fuzzy models. **Chem. Eng. J.** v.221, p.185–192, 2013.

EDWARDS, T. H.; WILLSON, P. D. Digital least squares smoothing of spectra. **Appl. Spectrosc.** v.28, p. 541–545, 1974.

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. Disponível em www.embrapa.br/

ENKE, C. G.; NIEMAN, T. A. Signal-to-noise ratio enhancement by least-squares polynomial smoothing. **Anal. Chem.** v.48, p. 705A-712A, 1976.

ESBENSEN, K. H. Multivariate data analysis— In practice (5th ed.), Oslo, 2004.

ESBENSEN, K. H.; GELADI, P. Principles of Proper Validation: use and abuse of re-sampling for validation. **J. Chemometr.** v. 24, p.168–187, 2010.

EVERARD, C. D. *et al.* Prediction of biomass gross calorific values using visible and near infrared spectroscopy. **Biomass Bioenerg.** v. 45, p. 203-211, 2012.

FABER, N. K. M. Mean centering and computation of scalar net analyte signal in multivariate calibration. **J. Chemometr.** v.12, p. 405-409, 1998.

FABER, N. K. M. Multivariate Sensitivity for the Interpretation of the Effect of Spectral Pretreatment Methods on Near-Infrared Calibration Model Predictions. **Anal. Chem.** v.71, p. 557-565, 1999.

FABER, N. K. M. Exact presentation of multivariate calibration model as univariate calibration graph. **Chemometr. Intell. Lab. Syst.** v. 50, p. 107-114, 2000.

FABER, N.K.M.; RAJKÓ, R. How to avoid over-fitting in multivariate calibration — The conventional validation approach and an alternative. **Anal. Chim. Acta.** v.595, p.98–106, 2007.

FACKLER, K. *et al.* Qualitative and quantitative changes of beech wood degraded by woodrotting basidiomycetes monitored by Fourier transform infrared spectroscopic methods and multivariate data analysis. **Microbiol. Lett.** v.271, p.162-169, 2007.

FANG, Q.; HANNA, M.A. Experimental studies for levulinic acid production from whole kerne grain sorghum. **Bioresour. Technol.** v.81, p. 187-192, 2002.

FAO. Food and Agriculture Organization (<http://www.fao.org>). Acessado: Jan/2013.

- FEARN, T. On orthogonal signal correction. **Chemometr. Intell. Lab. Syst.** v.50, p.47–52, 2000.
- FEARN, T. Are two pretreatments better than one? **NIR news**, v.14, p.9-11, 2003.
- FENGEL, D.; WEGENER, G. **Wood chemistry, ultrastructure, reactions**. Berlin, 1989.
- FERNANDES, D, R. *et al.* Levulinic acid esterification with ethanol to ethyl levulinate production over solid acid catalysts. **Appl. Catal. A: Gen.** v. 425– 426, p.199– 204, 2012.
- FERRÉ, J.; BROWN, S.D.; RIUS, F.X. Improved calculation of the net analyte signal in inverse multivariate calibration. **J. Chemometr.** v.15, p. 537–553, 2001.
- FERREIRA, M. M. C. *et al.* Quimiometria I: calibração multivariada, um tutorial. **Quim. Nova**, v.22, p.724-731, 1999.
- FERREIRA, M.H. *et al.* Development and validation of a chemometric method for direct determination of hydrochlorothiazide in pharmaceutical samples by diffuse reflectance near infrared spectroscopy. **Microchem. J.** v.109, p.158-164, 2013.
- FIGUEIREDO, A.; EVTUGUIN, D.; SARAIVA, J. Effect of high pressure treatment on structure and properties of cellulose in eucalypt pulps. **Cellul.** v.17, p.1193–1202, 2010.
- FONER, H.A.; ADAN, N. The Characterization of Papers by X-Ray Diffraction (XRD): Measurement of Cellulose Crystallinity and Determination of Mineral Composition. **J. Forensic Sci. Society.** v.23, p.313-321, 1983.
- GABRIELSSON, J.; TRYGG, J. Recent Developments in Multivariate Calibration. **Crit. Rev. Anal. Chem.** v.36, p.243–255, 2006.
- GARCÍA, S.G. *et al.* Environmental aspects of eucalyptus based ethanol production and use. **Sci. Total Environ.** v. 438, p.1–8, 2012.
- GELADI, P.; KOWALSKI, B.R. Partial least squares regression: A tutorial. **Anal. Chim. Acta.** v.185, p.1-17, 1986.
- GEORGES, F. Caracterização da palha da cana-de-açúcar do Rio Grande do Sul e de seus produtos de pirólise. Dissertação de Mestrado, Porto Alegre, 2011.
- GIRISUTA, B. Levulinic Acid from Lignocellulosic Biomass [Tese de Doutorado]. University of Groningen, Holanda, 2007.
- GIRISUTA, B. *et al.* A kinetic study of acid catalysed hydrolysis of sugar cane bagasse to levulinic acid. **Chem. Engin. J.** v.217, p. 61–70, 2013.
- GITELSON, A. A.; GRITZ, Y.; MERZLYAK, M.N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. **J. Plant Physiol.** v.160, p.271–282, 2003.
- GOLDEMBERG, J. The promise of clean energy. **Energ. Policy**, v.34, p.2185-2190, 2006.
- GOLDSTEIN, I. S. Overview of the chemical composition of wood. In: Lewin, M. & Goldstein, I. S. (eds.) **Wood Structure and Composition**. New York, 1991.
- GÓMEZ, E.O *et al.* Some simplified geometrical properties of elephant grass and sugarcane trash particles. **Fuel Process. Technol.** v.104, p.234-244, 2012.
- GÜTSCH, J.S.; NOUSIAINEN, T.; SIXTA, H. Comparative evaluation of autohydrolysis and acid-catalyzed hydrolysis of Eucalyptus globulus wood. **Bioresour. Technol.** v.109, p.77–85, 2012.
- GUIMARÃES, J. L. *et al.* Characterization of banana, sugarcane bagasse and sponge gourd fibers of Brazil. **Ind. Crop Prod.** v.30, p. 407–415, 2009.

- HASSUANI, S. J. Biomass power generation: sugar cane bagasse and trash. PNUD-CTC, Piracicaba, p.217, 2005.
- HAVERTY, D. et al. Autothermal, single-stage, performic acid pretreatment of *Miscanthus x giganteus* for the rapid fractionation of its biomass components into a lignin/hemicellulose-rich liquor and a cellulase-digestible pulp. **Bioresour. Technol.** v.109, p.173-177, 2012.
- HAYES, D. J. M. et al. The Biofine Process – Production of Levulinic Acid, Furfural, and Formic Acid from Lignocellulosic Feedstocks. **Biorefineries – Ind. Proc. Prod.** WILEY-VCH, Massachusetts, p.139, 2006.
- HAYES, D.J.M. Analysis of Lignocellulosic Feedstocks for Biorefineries with a Focus on The Development of Near Infrared Spectroscopy as a Primary Analytical Tool. Thesis of Doctor of Philosophy. Limerick, 2011.
- HAYES, D.J. M. Development of near infrared spectroscopy models for the quantitative prediction of the lignocellulosic components of wet *Miscanthus* samples. **Bioresour. Technol.** v.119, p.393-405, 2012.
- HILLIS, W.E.; BROWN, A.G. Eucalyptus for wood production. **Melbourn**, CSIRO, p.434, 1978.
- HONGZHANG, C.; BIN, Y.; SHENGYING, J. Production of levulinic acid from steam exploded rice straw via solid superacid, $S_2O_8^{2-}/ZrO_2-SiO_2-Sm_2O_3$. **Bioresour. Technol.** v.102, p. 3568–3570, 2011.
- IBGE. Instituto Brasileiro de Geografia e Estatística. Disponível em www.ibge.gov.br/. Acessado: Março/2013.
- ICDD. International Centre for Diffraction Data. Disponível em <http://www.icdd.com/>. Acessado: Setembro/2012.
- ICO. International Coffee Organization. Disponível em <http://www.ico.org/>. Acessado: Janeiro/2013.
- INMETRO. Instituto Nacional de Metrologia, Qualidade e Tecnologia. Orientações sobre validação de métodos de ensaios químicos, DQO-CGRE-008, 2003.
- JAIMES, H.C. et al. Plantain starch granules morphology, crystallinity, structure transition, and size evolution upon acid hydrolysis. **Carbohydr. Polym.** v.95, p. 207– 213, 2013.
- JIANG, Z.H. et al. Rapid prediction of wood crystallinity in *Pinus elliotii* plantation wood by near-infrared spectroscopy. **J. Wood Sci.** v. 53, p.449–453, 2007.
- JIN, S.; CHEN, H. Near-infrared analysis of the chemical composition of rice straw. **Ind. Crop. Prod.** v. 26, p.207-211, 2007.
- JOHAR, N.; AHMADA, I.; DUFRESNEC, A. Extraction, preparation and characterization of cellulose fibres and nanocrystals from rice husk. **Ind. Crop. Prod.** v. 37, p. 93– 99, 2012.
- JONES, P. D. et al. Nondestructive estimation of wood chemical composition of sections of radial wood strips by diffuse reflectance near infrared spectroscopy. **Wood Sci. Technol.** v.40, p.709-720, 2006.
- KAISER, H.F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika.** v. 23, p.187-200, 1958.
- KALAPATHY, U.; PROCTOR, A.; SHULTZ, J. An improved method for production of silica from rice hull ash. **Bioresour. Technol.**, v. 85, p. 285-289, 2002.
- KARAGOZ et al. Comparative studies of oil compositions produced from sawdust, rice husk, lignin and cellulose by hydrothermal treatment. **Fuel**, v. 84, 875–884, 2005.

- KELLEY, S. S. *et al.* Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. **Wood Sci. Technol.** v.38, p.257-276, 2004.
- KOWALSKI, B. R.; SEASHOLTZ, M. B. Recent developments in multivariate calibration. **J. Chemom.** v.5, p.129-145, 1991.
- KUHAD, R. C.; SINGH. Lignocellulose Biotechnology: Current and Future Prospects. **Critical Reviews Biotechnol.** v. 13, p. 151-73, 1993.
- KHALIL, H.P.S.A. *et al.* Thermal, spectroscopic, and flexural properties of anhydride modified cultivated Acacia spp. **Wood Sci. Technol.** v.45, p.597-606, 2011.
- KJELDAHLA, K.; BRO, R. Some common misunderstandings in chemometrics. **J. Chemom.** v. 24, p.558-564, 2010.
- KRASZNAI, D. J.; CHAMPAGNE, P.; CUNNINGHAM, M.F. Quantitative characterization of lignocellulosic biomass using surrogate mixtures and multivariate techniques. **Bioresour. Technol.** v.110, p.652-661, 2012.
- LEENAKUL, W.; TIPPAYAWONG, N. Dilute Acid Pretreatment of Bamboo for Fermentable Sugar Production. **J. Sustain. Energ. Environ.** v.1, p.117-120, 2010.
- LESTANDER, T. A.; RHEN, C. Multivariate NIR spectroscopy models for moisture, ash and calorific content in biofuels using bi-orthogonal partial least squares regression. **Analyst**, v.130, p.1182–1189, 2005.
- LESTANDER, T. A.; GELADI, P. NIR spectral information used to predict water content of pine seeds from multivariate calibration. **Can. J. For. Res.** v.35, p.1139–1148, 2005.
- LI, Y. *et al.* The Study of Enzyme Hydrolysis Saccharification Process of Stems and Leaves of Banana. **Energ. Procedia**, v.16, p. 223-228, 2012.
- LIGNIN INSTITUTE. Disponível em <http://www.lignin.org/about.html>. Acessado em: Setembro/ 2012.
- LIMA, K. M. G. *et al.* Sensores ópticos com detecção no infravermelho próximo e médio. **Quím. Nova**. vol. 32, p. 1635-1643, 2009.
- LIMAYEM, A.; RICKE, S.C. Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. **Progr. Energ. Combustion Scien.** v. 38, p.449-467, 2012.
- LIU, L. *et al.* Pretreatment of near infrared spectral data in fast biomass analysis. **J. Near Infrared Spectrosc.** v.18, p. 317-331, 2010.
- LIU, L. *et al.* Variability of biomass chemical composition and rapid analysis using FT-NIR techniques. **Carbohydr. Polym.** v.81, p.820–829, 2010.
- LOMBORG, C. J. *et al.* Power plant intake quantification of wheat straw composition for 2nd generation bioethanol optimization – A Near Infrared Spectroscopy (NIRS) feasibility study. **Bioresour. Technol.** v. 101, p.1199–1205, 2010.
- LORBER A. Error Propagation and Figures of Merit for Quantification by Solving Matrix Equations. **Anal. Chem.** v.58, p.1167–1172, 1986.
- MA, Y. *et al.* Comprehensive utilization of the hydrolyzed productions from rice hull. **Ind. Crop Prod.** v.33, p. 403–408, 2011.
- MADDEN, H.H. Comments on the Savitzky-Golay Convolution Method for Least-Squares Fit Smoothing and Differentiation of Digital Data. **Anal. Chem.** v. 50, p. 1383- 1386, 1978.

MARTENS H.; WOLD S, MARTENS M. A layman's guide to multivariate data analysis. In Food Research and Data Analysis. H, Martens H Russwurm (eds). Applied Science Publishers: London, p.473–492, 1983.

MARTENS, H.; NAES, T. Multivariate Calibration. John Wiley e Sons, New York, 1996.

MARTENS, H.; NIELSEN, J.P.; ENGELSEN, S.B. Light scattering and light absorbance separated by extended multiplicative signal correction. application to near-infrared transmission analysis of powder mixtures. **Anal. Chem.** v.75, p. 394-404, 2003.

MARTINS, A.F. *et al.* Caracterização dos produtos líquidos e do carvão da pirólise de serragem de eucalipto. **Quim. Nova** v.30, p.873-878, 2007.

MELLIGAN, F. *et al.* Characterisation of the products from pyrolysis of residues after acid hydrolysis of *Miscanthus*. **Bioresour. Technol.** v. 108, p. 258–263, 2012.

MICHEL, A.C.S. *et al.* Production of ethanol from soybean hull hydrolysate by osmotolerant *Candida guilliermondii* NRRL Y-2075. **Bioresour. Technol.** v.99, p. 2898–2904, 2008.

MOBLEY, P.R.; KOWALSKI, B.R.; WORKMAN J.J.; BRO, R. Review of Chemometrics Applied to Spectroscopy: Data Preprocessing. **Appl. Spectrosc. Rev.** v.34, p.347-368, 1996.

MONONO, E.M. *et al.* Variability in biomass yield, chemical composition, and ethanol potential of individual and mixed herbaceous biomass species grown in North Dakota. **Indus. Crops Prod.** v.41, p.331– 339, 2013.

MORGAN, D. R. Spectral Absorption Pattern Detection and Estimation. I. Analytical Techniques. **Appl. Spectrosc.** v.31, p. 404-415, 1977.

MURPHY J. D.; MCCARTHY, K. Ethanol production from energy crops and wastes for use as a transport fuel in Ireland. **Appl. Energ.** v. 82, p. 148-166, 2005.

MURRAY, I.; WILLIAMS, P.C. Chemical Principles of Near-Infrared Technology, In: Williams, P. and Norris, K. (eds.): Near-Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists, Minnesota, p. 29-31, 1987.

MUSSATTO, S. I. *et al.* Sugars metabolism and ethanol production by different yeast strains from coffee industry wastes hydrolysates. **Appl. Energ.** v. 92, p.763–768, 2012.

NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. Multivariate calibration and classification. Nir publications: Chinchester, p. 344, 2002.

NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. A User-Friendly Guide to Multivariate Calibration and Classification. NIR Publications, Chichester, 2007.

NAIK, S. *et al.* Characterization of Canadian biomass for alternative renewable biofuel. **Renew. Energ.** v.35, p.1624–1631, 2010.

NKANSAH, K. *et al.* Rapid characterization of biomass using near infrared spectroscopy coupled with multivariate data analysis: Part 1 yellow-poplar (*Liriodendron tulipifera* L.). **Bioresour. Technol.** v.101, p.4570-4576, 2010.

NREL. National Renewable Energy Laboratory. Determination of Structural Carbohydrates and Lignin in Biomass. Disponível em <http://www.nrel.gov/>. Acessado: Maio/2010.

OGEDA, T.L.; PETRI, D.F.S. Hidrólise enzimática de biomassa. **Quim. Nova**, v. 33, p.1549-1558, 2010.

OLIVIERI, A. C. *et al.* Uncertainty estimation and figures of merit for multivariate calibration. **Pure Appl. Chem.** v. 78, p. 633–661, 2006.

OTTO, M. Statistics and Computer Application in Analytical Chemistry. **Chemometrics**. Weinheim: Wiley, 1999.

PARAJO, J.C.; DOMÍNGUEZ, H.; DOMÍNGUEZ, J. M. Biotechnological production of xylitol. Part 2: operation in culture media made with commercial sugars. **Bioresour.Technol.** v.66, p.25–40, 1998.

PARK, S. *et al.* Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. **Biotechnol. Biofuels**, v.3, p.1-10, 2010.

PASQUINI, C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. **J. Braz. Chem. Soc.** v. 14, p.198-219, 2003.

PASTORE, T. C. M. *et al.* Efeito do intemperismo artificial em quatro madeiras tropicais monitorado por espectroscopia de infravermelho (DRIFT). **Quim. Nova**, v. 31, p. 2071-2075, 2008.

PEDRO, A. M. K.; FERREIRA, M. M. C. Simultaneously calibrating solids, sugars and acidity of tomato products using PLS2 and NIR spectroscopy. **Anal. Chim. Acta.** v. 595, p.221-227, 2007.

PENG, H. *et al.* Physicochemical characterization of hemicelluloses from bamboo (*Phyllostachys pubescens* Mazel) stem. **Indus. Crops Prod.** v.37, p.41– 50, 2012.

PEREIRA JR, N.; COUTO, M.A.P.G.; ANNA, L.M.M.S. Biomass of lignocellulosic composition for fuel ethanol production within the context of biorefinery. Series on Biotechnology. In: (1 ed.), Copiadora Amiga dos Estudantes Ltda; Rio de Janeiro, v.1, p. 1-47, 2008.

PEREIRA, P. H. F. *et al.* Preparação e Caracterização de Materiais Híbridos Celulose/ $\text{NbOPO}_4 \cdot n\text{H}_2\text{O}$ a partir de Celulose Branqueada de Bagaço de Cana-de-Açúcar. **Polímeros**, vol. 22, p. 88-95, 2012.

PIRES, E. N. *et al.* Efeito do Tratamento Alcalino de Fibras de Juta no Comportamento Mecânico de Compósitos de Matriz Epóxi. **Polímeros**, vol. 22, p. 339-344, 2012.

POKE, F. S. Predicting Extractives, Lignin, and Cellulose Contents Using Near Infrared Spectroscopy on Solid Wood in *Eucalyptus globulus*. **J. Wood Chem.Technol.** v. 26, p.187-199, 2006.

QU, Z.; WANG, L. Prediction of the crystallinity of white pine using near infrared spectroscopy. **Adv Mat Res.** v.183-185, p. 1215-1218, 2011.

RACKEMANN, D.W.; DOHERTY, W.O.S. The conversion of lignocellulosics to levulinic acid. **Biofuels, Bioprod. Bioref.** v.5, p.198–214, 2011.

RAFIQUL, I.S.M.; SAKINAH, A.M.M. Design of process parameters for the production of xylose from wood sawdust. **Chem. Eng. Sci.** v.71, p.431– 437, 2012.

RALPH, J. *et al.* Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids. **Phytochem. Rev.** v.3, p.29–60, 2004.

RAMBO, M.K.D. *et al.* Silica from rice husk ash as an additive for rice plant. **J. Agron.** v.10, p.99-104, 2011.

RAMBO, M.K.D. *et al.* Xylitol from rice husks by acid hydrolysis and *candida* yeast fermentation. **Quim. Nova**, p. 1- 6, 2013 (no prelo).

RAMBO, M.K.D; AMORIM, E.P; FERREIRA, M.M.C. Potential of visible-near infrared spectroscopy combined with chemometrics for analysis of some constituents of coffee and banana residues. **Anal. Chim. Acta**, v.775, p. 41-49, 2013.

REDDING, A.P. *et al.* High temperature dilute acid pretreatment of coastal Bermuda grass for enzymatic hydrolysis. **Bioresour. Technol.** v.102, p.1415-1424, 2011.

REZENDE *et al.* Chemical and morphological characterization of sugarcane bagasse submitted to a delignification process for enhanced enzymatic digestibility. **Biotechnol. Biofuel.** v.4, p. 3 -18, 2011.

RIBEIRO, P.R. *et al.* Furfural – da biomassa ao laboratório de química orgânica. **Quim. Nova**, v. 35, p.1046-1051, 2012.

ROBERTS, J. C. The Chemistry of Paper. Royal Society of Chemistry, Cambridge, 1996.

ROCHA, W. F.; NOGUEIRA, R.; VAZ, B .G. Validation of model of multivariate calibration: an application to the determination of biodiesel blend levels in diesel by near-infrared spectroscopy **J. Chemom.** v. 26, p.456-461, 2012.

RODRIGUES, J. A. R. Do engenho à biorrefinaria. A usina de açúcar como empreendimento industrial para a geração de produtos bioquímicos e biocombustíveis. **Quim. Nova**, v. 34, p.1242-1254, 2011.

RODRÍGUEZ, U.F. *et al.* Produção de complexos lignocelulíticos em substratos derivados de resíduos agroindustriais por fermentação semi-sólida. Workshop de Biotransformação e Biotransformação. Anais, São Paulo, 2008.

SALA, O. I₂- Uma molécula didática. **Quim. Nova**, v. 31, p.914-920, 2008.

SALIBA, E.O.S. *et al.* Ligninas – Métodos de obtenção e caracterização química. **Ciênc. Rural**, v.31, p.917-928, 2001.

SÁNCHEZ, C. *et al.* Furfural production from corn cobs autohydrolysis liquors by microwave technology. **Indus. Crops Prod.** v.42, p. 513– 519, 2013.

SANDERSON, M. A. *et al.* Compositional analysis of biomass feedstocks by near infrared reflectance spectroscopy. **Biomass Bioenerg.** v.11, p. 361-370,1996.

SANTOS, I.D. Influência dos teores de lignina, holocelulose e extrativos na densidade básica e contração da madeira e nos rendimentos e densidade do carvão vegetal de cinco espécies lenhosas do cerrado. Dissertação de mestrado em Ciências Florestais. Brasília, 2008.

SASMAL, S.; GOUD, V. V.; MOHANTY, K. Characterization of biomasses available in the region of North-East India for production of biofuels. **Biomass Bioenerg.** v.45, p.212-220, 2012.

SATYANARAYANA, K. G. *et al.* Studies on lignocellulosic fibers of Brazil. Part I: Source, production, morphology, properties and applications. **Compos. Part A- Appl. S.** v. 38, p.1694– 1709, 2007.

SAVITZKY, A.; GOLAY M.J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Anal. Chem.** v. 36, p. 1627-1638,1964.

SEVERO, E.T.D. *et al.* Composição Química da Madeira de *Eucalyptus citriodora* em Função das Direções Estruturais. **Silva Lusitana.** v. 14, p.113-126, 2006.

SEYE, O. *et al.* Estudo cinético da biomassa a partir de resultados termogravimétricos. **Enc. Energ. Meio Rural.** v.3, p.1-8, 2003.

SILVA, M.A.M. *et al.* Development and analytical validation of a multivariate calibration method for determination of amoxicillin in suspension formulations by near infrared spectroscopy. **Talanta**, v.89, p. 342– 351, 2012.

SHORT, S.M.; COGDILL, R.P.; ANDERSON, C.A. Determination of Figures of Merit for Near-Infrared and Raman Spectrometry by Net Analyte Signal Analysis for a 4-Component Solid Dosage System. **Pharm.Sci.Tech.** v.8, p.1-11, 2007.

SILVERSTEIN, R. M.; BASSLER, G. C.; MORRILL, T. C. Spectrometric Identification of Organic Compounds; In: (4 ed.). John Wiley e Sons; New York, 1981.

SJÖBLOM, J. *et al.* An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. **Chem. Intel. Lab. Syst.** v.44, p.229–244, 1998.

SOFFNER, M.L.A.P. Produção de polpa celulósica a partir de engaço de bananeira. Dissertação de mestrado, 2001, Piracicaba, São Paulo.

SUCHAT *et al.* Fast determination of the resin and rubber content in *Parthenium argentatum* biomass using near infrared spectroscopy. **Indus. Crops Prod.** v.45, p. 44– 51, 2013.

SUN, Z.Y. *et al.* Production of fuel ethanol from bamboo by concentrated sulfuric acid hydrolysis followed by continuous ethanol fermentation. **Bioresour. Technol.** v.102, p. 10929–10935, 2011.

SUN, Z.Y. *et al.* Reduction in environmental impact of sulfuric acid hydrolysis of bamboo for production of fuel ethanol. **Bioresour. Technol.** v.128, p. 87–93, 2013.

SHENK, J. S.; WORKMAN, J. J.; WESTERHAUS, M. O. Application of NIR Spectroscopy to Agricultural Products. In: (3 ed.). Burns, D. A. & Biurczak, E. W. Handbook of Near-Infrared Analysis, Florida, 2008.

SJOSTROM, E. Wood Chemistry: Fundamentals and Applications. Academic Press, London, UK, 1981.

SJOSTROM, E; WESTERMARK, U. Chemical composition of wood an pulps: basic constituents and their distribution. In: Eds. Sjöström E.; Alén, R. Analytical methods in wood chemistry , pulping and papermaking. Helsinki: Springer Series in Wood Science; 1999. p.1-20.

SLUITER, J. B. *et al.* Compositional Analysis of Lignocellulosic Feedstocks. Review and Description of Methods. **J. Agr. Food Chem.** v.58, p.9043–9053, 2010.

SZCZERBOWSKI, D. *et al.* Caracterização química da biomassa de cana-de-açúcar para avaliação de sua viabilidade à produção de bioetanol. **IX Simpósio de Hidrólise Enzimática de Biomassas.** 2009.

TAMAKI, Y.; MAZZA, G. Measurement of structural carbohydrates, lignins, and micro-components of straw and shives: Effects of extractives, particle size and crop species. **Indus. Crops Prod.** v.31, p. 534–541, 2010.

TIPPAYAWONG, N; CHANHOM, N. Conversion of Bamboo to Sugars by Dilute Acid and Enzymatic Hydrolysis. **Inter. J. Rene. Energ. Res.** v.1, p. 240-244, 2011.

TOMCZAK, F.; SYDENSTRICKER, T.H.D.; SATYANARAYANA, K.G. Studies on lignocellulosic fibers of Brazil. Part II: Morphology and properties of Brazilian coconut fibers. **Composites: Part A**, v.38, p.1710-1721, 2007.

THAKKER, C.; SAN, K.Y.; BENNETT, G.N. Production of succinic acid by engineered *E. coli* strains using soybean carbohydrates as feedstock under aerobic fermentation conditions. **Bioresour. Technol.** v.130, p.398-405, 2013.

TRUGILHO, P.F.; LIMA, J.T.; MORI, F.A. Correlação canônica das características químicas e físicas de clones de *Eucalyptus grandis* e *Eucalyptus saligna*. **Cerne.** v.9, p.66-80, 2003.

TRYGG, J.; WOLD, S. 02-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter. **J. Chemom.** v.17, p. 53-64, 2003.

TSAI, W.T. *et al.* Preparation and fuel properties of biochars from the pyrolysis of exhausted coffee residue. **J. Anal. Appl. Pyrol.** v.93, p.63–67, 2012.

TSUCHIKAWA, S.; SIESLER H. W. Near-Infrared Spectroscopic Monitoring of the Diffusion Process of Deuterium-Labeled Molecules in Wood. Part II: Hardwood. **Appl. Spectrosc.** v.57, p.675-681, 2003.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Estado da arte de figuras de mérito em calibração multivariada. **Quim. Nova**, v. 32, p. 1278-1287, 2009.

VÁVROVÁ, P. Wastewater Treatment, Plant Dynamics and Management in Constructed and Natural Wetlands. In: Vymazal, Jan (Ed.), New York, p. 65-87, 2008.

VERGNOUX, A. *et al.* Monitoring of the evolution of an industrial compost and prediction of some compost properties by NIR spectroscopy. **Sci. Total Environ.** v. 407, p. 2390-2403, 2009.

VERSARI, A. *et al.* Determination of grape quality at harvest using Fourier-transform mid-infrared spectroscopy and multivariate analysis. **Am. J. Enol. Viticult.** v.5, p.317–322, 2008.

XIAO, X. *et al.* Classification and analysis of corn steep liquor by UPLC/Q-TOF MS and HPLC. **Talanta**, v.107, p.344-348, 2013.

XIUJUAN, G. *et al.* Properties of Bio-oil from Fast Pyrolysis of Rice Husk. **Chinese J. Chem. Eng.** v.19, p. 116-121, 2011.

XU *et al.* Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review. **Appl. Energ.** v.104, p.801–809, 2013.

WESTERHUIS, J. A.; JONG, S.; SMILDE, A. K. Direct orthogonal signal correction. **Chem. Intel. Lab. Syst.** v.56, p.13–25, 2001.

WICKHOLM, K. Structural elements in native celluloses. Royal Institute of Technology, Sweden, 2001.

WIKLUND, S. *et al.* A randomization test for PLS component selection. **J. Chem.** v.21, p.427–439, 2007.

WOLD, S.; ESBENSEN, K.; GELADI, P. Principal Component Analysis. **Chem. Intell. Lab. Syst.** v.2, p.37-52, 1987.

WOLD, S. *et al.* Orthogonal signal correction of near-infrared spectra. **Chem. Intel. Lab. Syst.** v.44, p.175–185, 1998.

YANINI, N.; AMIN, N. A. S.; ASMADI, M. Optimization of levulinic acid from lignocellulosic biomass using a new hybrid catalyst. **Bioresour. Technol.** v. 116, p.58–65, 2012.

YANG *et al.* Dilute-acid conversion of cotton straw to sugars and levulinic acid via 2-stage hydrolysis. **Indus. Crops Prod.** v. 46, p. 205– 209, 2013.

YU, Y.; LOU, X.; WU, H. Some Recent Advances in Hydrolysis of Biomass in Hot-Compressed Water and Its Comparisons with Other Hydrolysis Methods. **Energ. Fuel.** v.22, p.46–60, 2008.

ZHANG L. H. *et al.* Effect of steam explosion on biodegradation of lignin in wheat straw. **Bioresour. Technol.** v.99, p.8512–8515, 2008.

ZHANG L. H. *et al.* Conversion of xylan, D-xylose and lignocellulosic biomass into furfural using AlCl_3 as catalyst in ionic liquid. **Bioresour. Technol.** v.130, p.110–116, 2013.

ZIDAN, A.S. *et al.* Crystallinity evaluation of tacrolimus solid dispersions by chemometric analysis. **Int. J. Pharm.** v.423, p.341– 350, 2012.