

**Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Departamento de Química**

**CRISTIANE DE CARVALHO GUIMARÃES**

**Desenvolvimento e validação de modelos quimiométricos  
baseados em espectroscopia no infravermelho próximo  
para a predição de parâmetros do caldo e da biomassa do  
sorgo sacarino visando à produção de etanol**

**Belo Horizonte  
2014**

UFMF/ICEx/DQ 1016<sup>a</sup>

D 554<sup>a</sup>

**CRISTIANE DE CARVALHO GUIMARÃES**

**Desenvolvimento e validação de modelos quimiométricos baseados em espectroscopia no infravermelho próximo para a predição de parâmetros do caldo e da biomassa do sorgo sacarino visando à produção de etanol**

Dissertação apresentada ao Departamento de Química do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do grau de Mestre em Química – Química Analítica.

**Belo Horizonte**

**2014**

Guimarães, Cristiane de Carvalho  
Desenvolvimento e validação de modelos  
quimiométricos baseados em espectroscopia no  
infravermelho próximo para a predição de parâmetros do  
caldo e da biomassa do sorgo sacarino visando à  
produção de etanol [manuscrito] / Cristiane de  
Carvalho Guimarães. 2014.

[xv], 101 f. : il.

Orientador: Marcelo Martins de Sena.

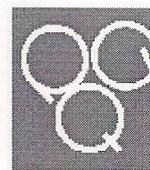
Dissertação (mestrado) - Universidade Federal de  
Minas Gerais. Departamento de Química.

Bibliografia: f. 84-90.

Anexos: f. 91-101.

1. Química analítica - Teses 2. Espectroscopia de infravermelho - Teses 3.  
Calibração - Teses 4. Sorgo sacarino - Teses I. Sena, Marcelo Martins de,  
Orientador II. Título.

CDU 043



**"Desenvolvimento e Validação de Modelos Quimiométricos Baseados em Espectroscopia no Infravermelho Próximo para a Predição de Parâmetros do Caldo e da Biomassa do Sorgo Sacarino Visando à Produção de Etanol"**

**Cristiane de Carvalho Guimarães**

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Prof. Marcelo Martins de Sena - Orientador  
UFMG

Prof. Adriana Ferreira Faria  
UFMG

Prof. Camila Nunes Costa Corgozinho  
UFMG

Belo Horizonte, 19 de março de 2014.

**“Se as coisas são inatingíveis...ora! Não é motivo para não querê-las...  
Que tristes os caminhos, se não fora a presença distante das estrelas!”**

**Mário Quintana**

## **AGRADECIMENTOS**

---

Esta dissertação de mestrado é fruto de um longo esforço coletivo. Em primeiro lugar, sou muito grata ao Prof. Marcelo Martins de Sena pela paciência, competência e amizade com que orientou o meu trabalho. Da mesma forma, agradeço aos Drs. Rafael Parrela e Maria Lúcia Simeone, pesquisadores da Embrapa Milho e Sorgo, por confiarem a mim este projeto. Também agradeço a colaboração do Carlos Henrique que dividiu comigo a tarefa de analisar tantas amostras! Aos meus queridos amigos Michelle e Vander Felipe pela ajuda, carinho e amizade. Ao Bruno, companheiro do grupo de pesquisa, meus sinceros agradecimentos por compartilhar comigo o seu conhecimento em Matlab. Ao Prof. Célio Pasquini e à sua aluna Lúvia Paulia, da UNICAMP, pelo empréstimo do polarímetro. Não poderia deixar de agradecer ao amigo Mário Ângelo por estar sempre ao meu lado nas dificuldades que encontrei pelo caminho e também aos professores, funcionários e colegas do Departamento de Química da UFMG por tudo que me ensinaram em seus cursos, seminários e conversas. Em especial, agradeço o incentivo e o amor dos meus pais, Paulo e Branca, do meu irmão Filipe e de todos os amigos que me ajudaram a seguir com alegria quando o cansaço teimava em aparecer. Finalmente, dou graças a Deus, meu melhor e fiel amigo, por me permitir alcançar essa conquista!

## RESUMO

---

O caldo do sorgo sacarino vem ganhando importância como matéria-prima para a produção de etanol de primeira geração na entressafra da cana-de-açúcar. Além disso, há uma grande expectativa na utilização do bagaço do sorgo para produção de etanol lignocelulósico, uma vez que esta cultura apresenta alto rendimento de biomassa por hectare e baixos custos de produção. Conseqüentemente, os programas de melhoramento genético estão buscando melhorar a qualidade desta matéria-prima com o objetivo de aumentar a produtividade. Tal prática gera um número excessivo de amostras para serem analisadas. Assim, o objetivo desse trabalho foi desenvolver métodos rápidos e baratos, baseados em calibração multivariada e espectroscopia no infravermelho próximo, para a determinação de parâmetros de qualidade da biomassa (celulose, hemicelulose e lignina) e do caldo (brix, leitura sacarimétrica, açúcares redutores, pol e pureza aparente) do sorgo sacarino, além da estimativa do rendimento teórico de etanol (RTE) baseada nos componentes da parede celular (celulose e hemicelulose). Os modelos para biomassa foram construídos com 957 amostras, obtidas a partir de mais de 100 genótipos diferentes, nas faixas de 21,4 a 49,1 % m/m, 18,4 a 34,8 % m/m, 221 a 412 L t<sup>-1</sup> e 1,8 a 11,5% m/m para celulose, hemicelulose, RTE e lignina, respectivamente. Os modelos para o caldo foram construídos com 500 amostras, parte delas proveniente de 275 linhagens contrastantes para qualidade e quantidades de açúcares no caldo, nas faixas de 5,5 a 18,1 °B para brix, 1,1 a 53,2 °Z para leitura sacarimétrica, 1,2 a 5,2 % m/v para açúcares redutores, 0,3 a 13,0 % m/v para pol e 9,8 a 83,0 % m/v para pureza aparente. Finalmente, os métodos foram submetidos a uma completa validação analítica multivariada, de acordo com os guias nacionais e internacionais, e considerados exatos, lineares, sensíveis e sem viés.

**Palavras-chave:** NIRS, calibração multivariada, biocombustíveis

## ABSTRACT

---

The sweet sorghum juice is gaining importance as a raw material for the first generation ethanol production in the period between harvests of sugar cane. Furthermore, there is a great expectation on the use of sorghum bagasse for the lignocellulosic ethanol production, since this culture has a high yield of biomass per hectare and low production costs. Consequently, breeding programs are seeking to improve this raw material quality in order to increase productivity. This practice has generated an excessive number of samples to be analyzed. Thus, the aim of this study was to develop rapid and low cost methods based on multivariate calibration and near infrared spectroscopy for the determination of quality parameters of sweet sorghum biomass (cellulose, hemicellulose and lignin) and juice (brix, saccharimetric reading, reducing sugars, pol and apparent purity), besides the estimate of the theoretical ethanol yield (TEY) based on cell wall components (cellulose and hemicellulose). The biomass models were built using 957 samples, obtained from over 100 different genotypes, in the ranges of 21.4 to 49.1% w/w, 18.4 to 34.8% w/w, 221 to 412 Lt<sup>-1</sup> and 1.8 to 11.5% w/w for cellulose, hemicellulose, lignin and TEY, respectively. The juice models were built with 500 samples, part of them from 275 contrasting strains for quality and for amount of sugars in the juice, in the ranges of 5.5 to 18.1 °B for brix, 1.1 to 53.2 °Z for saccharimetric reading, 1.2 to 5.2 % w/v for reducing sugars, 0.3 to 13.0 % w/v for pol and 9.8 to 83.0% w/v for apparent purity. Finally, a complete multivariate analytical validation was carried out in accordance with the Brazilian and international guidelines, and the methods were considered accurate, linear, sensitive and unbiased.

**Key words:** NIRS, multivariate calibration, biofuels

## SUMÁRIO

---

AGRADECIMENTOS.....	vi
RESUMO .....	vii
ABSTRACT .....	viii
SUMÁRIO.....	ix
LISTA DE FIGURAS .....	xi
LISTA DE TABELAS .....	xiii
LISTA DE ABREVIATURAS E SIGLAS .....	xiv
1 - INTRODUÇÃO.....	1
1.1 Bioenergia e Bioetanol .....	1
1.2 Sorgo sacarino.....	4
2 - MOTIVAÇÃO.....	7
2.1 Programas de melhoramento genético do sorgo sacarino .....	7
2.2 Parâmetros de qualidade e metodologias convencionais.....	8
2.2.1 Caracterização do caldo de sorgo sacarino .....	8
2.2.1.1 Brix .....	8
2.2.1.2 Pol e leitura sacarimétrica .....	9
2.2.1.3 Açúcares redutores.....	9
2.2.1.4 Pureza aparente do caldo.....	11
2.2.2 Caracterização da biomassa de sorgo sacarino .....	11
2.2.2.1 FDN.....	12
2.2.2.2 FDA.....	12
2.2.2.3 Lignina .....	13
2.2.2.4 Celulose.....	14
2.2.2.5 Hemicelulose .....	15
2.2.2.6 Rendimento Teórico de Etanol (RTE) .....	16
3 - REFERENCIAL TEÓRICO .....	18
3.1 Espectroscopia no infravermelho próximo – NIRS .....	18
3.2 Quimiometria e calibração multivariada .....	25
3.2.1 Definição dos termos .....	25
3.2.2 Métodos de calibração multivariada.....	27
3.2.3 Regressão por mínimos quadrados parciais .....	28
3.2.4 Detecção de amostras anômalas – <i>Outliers</i> .....	30
3.3 Validação de modelos multivariados e figuras de mérito.....	33
3.3.1 Veracidade .....	34
3.3.2 Precisão .....	34
3.3.3 Linearidade.....	35
3.3.4 Sensibilidade .....	38

3.3.5	Sensibilidade analítica .....	38
3.3.6	Seletividade .....	39
3.3.7	Faixa de trabalho .....	39
3.3.8	Viés (teste para erro sistemático – “ <i>Bias</i> ”) .....	39
3.3.9	Relação de Desempenho do Desvio, RPD ( <i>Residual Prediction Deviation</i> ).....	40
4 -	OBJETIVOS .....	42
5 -	PARTE EXPERIMENTAL.....	43
5.1	Amostras .....	43
5.2	Reagentes .....	44
5.3	Equipamentos.....	45
5.4	Softwares .....	46
5.5	Procedimentos.....	46
5.6	Métodos de referência .....	48
6 -	RESULTADOS E DISCUSSÃO.....	51
6.1	Resultados obtidos pelas metodologias de referência .....	51
6.2	Espectros NIR e desenvolvimento dos modelos PLS .....	54
6.3	Determinação do rendimento teórico de etanol .....	60
6.4	Otimização dos modelos através da detecção de amostras anômalas .....	63
6.5	Validação analítica multivariada.....	65
6.5.1	Linearidade.....	65
6.5.2	Veracidade, precisão e faixa de trabalho .....	69
6.5.3	Seletividade e sensibilidade analítica .....	77
6.5.4	Viés e RPD.....	78
6.5.5	Cartas de Controle .....	79
7 -	CONCLUSÃO.....	82
8 -	REFERÊNCIAS BIBLIOGRÁFICAS.....	84
	ANEXOS.....	91

## LISTA DE FIGURAS

---

<b>Figura 1</b> Emissões evitadas pela utilização de biocombustíveis em 2012 – Brasil.....	2
<b>Figura 2</b> Rotas para produção de bioetanol.....	3
<b>Figura 3</b> Planta de Sorgo sacarino .....	5
<b>Figura 4</b> Representação esquemática da produção de etanol a partir da biomassa lignocelulósica .....	6
<b>Figura 5</b> Equação da reação química envolvida no método de Lane-Eynon .....	10
<b>Figura 6</b> Representação esquemática da metodologia de extração sequencial baseada no método de van Soest.....	12
<b>Figura 7</b> Modelo estrutural da lignina proposto por Nimz e co-autores .....	13
<b>Figura 8</b> Celulose - Polímero linear de moléculas de D-glicose. ....	14
<b>Figura 9</b> Representação de um tipo de molécula de hemicelulose .....	15
<b>Figura 10</b> Algumas vibrações moleculares .....	20
<b>Figura 11</b> Representação esquemática dos modelos harmônico (A) e anarmônico (B) para a energia potencial de moléculas diatômicas <sup>33</sup> . de = distância de equilíbrio .....	23
<b>Figura 12</b> Alguns modos de medida empregados em espectroscopia NIR <sup>33</sup> : (a) reflectância difusa, (b) transreflectância e (c) transmitância. (Ip e Is = intensidade da radiação incidente e que emerge do material, respectivamente).....	24
<b>Figura 13</b> Decomposição em VL das matrizes <b>X</b> e <b>Y</b> para modelos PLS .....	29
<b>Figura 14</b> Distribuição residual normal. Dados se aproximam do ajuste de uma linha reta... 36	36
<b>Figura 15</b> Gráfico de correlação do teste de Durbin-Watson.....	37
<b>Figura 16</b> Acessório de transreflectância utilizado nas medidas dos espectros do caldo .....	48
<b>Figura 17</b> (a) Determinador de fibras Ankom e (b) bolsas de filtro utilizadas nas análises de FDA e FDN.....	49
<b>Figura 18</b> Polarímetro utilizado na determinação da leitura sacarimétrica.....	49
<b>Figura 19</b> Determinação do teor de açúcares redutores em Redutec. (a) antes e (b) depois da titulação .....	50
<b>Figura 20</b> Refratômetro digital para a determinação de brix .....	50
<b>Figura 21</b> Teores dos parâmetros da biomassa determinados pelas metodologias de referência para (a) celulose, (b) hemicelulose, (c) RTE e (d) lignina. ....	52

<b>Figura 22</b> Teores dos parâmetros do caldo determinados pelas metodologias de referência para (a) brix, (b) leitura sacarimétrica, (c) açúcares redutores, (d) pol e (e) pureza aparente.....	53
<b>Figura 23</b> Espectros NIR de 957 amostras da biomassa do sorgo sacarino .....	54
<b>Figura 24</b> Espectros NIR de 957 amostras da biomassa do sorgo sacarino após o pré-processamento SNV .....	55
<b>Figura 25</b> Coeficientes de regressão dos modelos PLS para (a) celulose, (b) hemicelulose, (c) RTE e (d) lignina.....	57
<b>Figura 26</b> Espectros NIR de 500 amostras do caldo do sorgo sacarino .....	58
<b>Figura 27</b> Espectros NIR de 500 amostras do caldo do sorgo sacarino após o pré-processamento 1ª DSG .....	59
<b>Figura 28</b> Coeficientes de regressão dos modelos PLS para (a) brix, (b) leitura sacarimétrica, (c) açúcares redutores, (d) pol e (e) pureza aparente.....	61
<b>Figura 29</b> Resíduos para as amostras da calibração (○) e validação (▼). Modelos PLS para determinação de (a) celulose, (b) hemicelulose, (c) RTE e (d) lignina em biomassa de sorgo .....	66
<b>Figura 30</b> Resíduos para as amostras da calibração (○) e validação (▼). Modelos PLS para determinação de (a) brix, (b) leitura sacarimétrica, (c) açúcares redutores, (d) Pol e (e) pureza aparente em caldo de sorgo .....	67
<b>Figura 31</b> Erros relativos das amostras de validação para os parâmetros da biomassa (a) celulose, (b) hemicelulose, (c) RTE e (d) lignina. Limites em vermelho (-20 a +10%)..	73
<b>Figura 32</b> Erros relativos das amostras de validação para os parâmetros do caldo (a) brix, (b) leitura sacarimétrica, (c) açúcares redutores, (d) pol e (e) pureza aparente. Limites em vermelho (-20 a +10 %).	74
<b>Figura 33</b> Cartas de controle para o monitoramento da estabilidade dos modelos NIRS da biomassa. (a) celulose, (b) hemicelulose, (c) lignina e (d) RTE.....	81

## LISTA DE TABELAS

---

<b>Tabela 1</b> Resultados para a otimização dos modelos PLS da biomassa (resultados finais em negrito) através das etapas de detecção de <i>outliers</i> .....	62
<b>Tabela 2</b> Resultados para a otimização dos modelos PLS do caldo (resultados finais em negrito) através das etapas de detecção de <i>outliers</i> .....	62
<b>Tabela 3</b> Parâmetros para a avaliação das principais FdM dos métodos NIRS desenvolvidos para a determinação de celulose, hemicelulose, RTE e lignina em biomassa de sorgo sacarino .....	75
<b>Tabela 4</b> Parâmetros para a avaliação das principais FdM dos métodos NIRS desenvolvidos para a determinação de brix, leitura sacarimétrica, açúcares redutores, pol e pureza aparente (Q) nas amostras de caldo de sorgo sacarino .....	76
<b>Tabela 5</b> Parâmetros dos testes estatísticos usados na avaliação das cartas de controle desenvolvidas para o monitoramento dos modelos NIRS .....	80

## LISTA DE ABREVIATURAS E SIGLAS

---

ASTM	<i>American Society for Testing and Materials</i> (Instituição normativa)
BAG	Banco Ativo de Germoplasma
BF	Brown Forsythe (Teste estatístico para avaliação da homocedasticidade dos resíduos)
DP	Desvio Padrão
DPR	Desvio Padrão Relativo
DSG	Derivada seguida de alisamento <i>Savitzky-Golay</i>
DW	Durbin Watson (Teste estatístico para avaliação da independência dos resíduos)
FDA	Fibras insolúveis em Detergente Ácido
FDN	Fibras insolúveis em Detergente Neutro
FdM	Figuras de Mérito
FT-NIR	Espectrofotometria de Infravermelho Próximo com Transformada de Fourier ( <i>Fourier Transform Near Infrared</i> )
INMETRO	Instituto Nacional de Metrologia, Normalização e Qualidade Industrial
IUPAC	União Internacional de Química Pura e Aplicada ( <i>International Union of Pure and Applied Chemistry</i> )
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MLR	Regressão Linear Múltipla ( <i>Multiple Linear Regression</i> )
MSC	Correção de Espalhamento Multiplicativo ( <i>Multiplicative Scatter Correction</i> )
NAS	Sinal Analítico Líquido ( <i>Net Analyte Signal</i> )
NIRS	Espectroscopia no Infravermelho Próximo ( <i>Near Infrared Spectroscopy</i> )
PCR	Regressão por Componentes Principais ( <i>Principal Component Regression</i> )
PLS	Mínimos Quadrados Parciais ( <i>Partial Least Squares</i> )
PMGS	Programa de Melhoramento Genético de Sorgo
Q	Pureza aparente do caldo
RIL	Linhagens Endogâmicas Recombinantes ( <i>Recombinant Inbred Lines</i> )
RJ	Ryan Joiner (Teste estatístico para avaliação da normalidade dos resíduos)
RMSEC	Raiz Quadradas dos Erros Quadráticos Médios de Calibração ( <i>Root Mean Square Error of Calibration</i> )

RMSECV	Raiz Quadrada do Erro Quadrático Médio de Validação Cruzada ( <i>Root Mean Square Error of Cross-Validation</i> )
RMSEP	Raiz Quadrada do Erro Quadrático Médio de Previsão ( <i>Root Mean Square Error of Prediction</i> )
RPD	Relação de Desempenho do Desvio ( <i>Residual Prediction Deviation</i> )
RTE	Rendimento Teórico de Etanol
S	Pol (Porcentagem de sacarose no caldo da cana/sorgo)
SDV	Desvio Padrão dos Erros de Validação ( <i>Standard Deviation of Validation Errors</i> )
SEL	Seletividade
SEN	Sensibilidade
SNV	Varição Normal Padrão ( <i>Standard Normal Variate</i> )
VL	Variáveis Latentes

# **1 - INTRODUÇÃO**

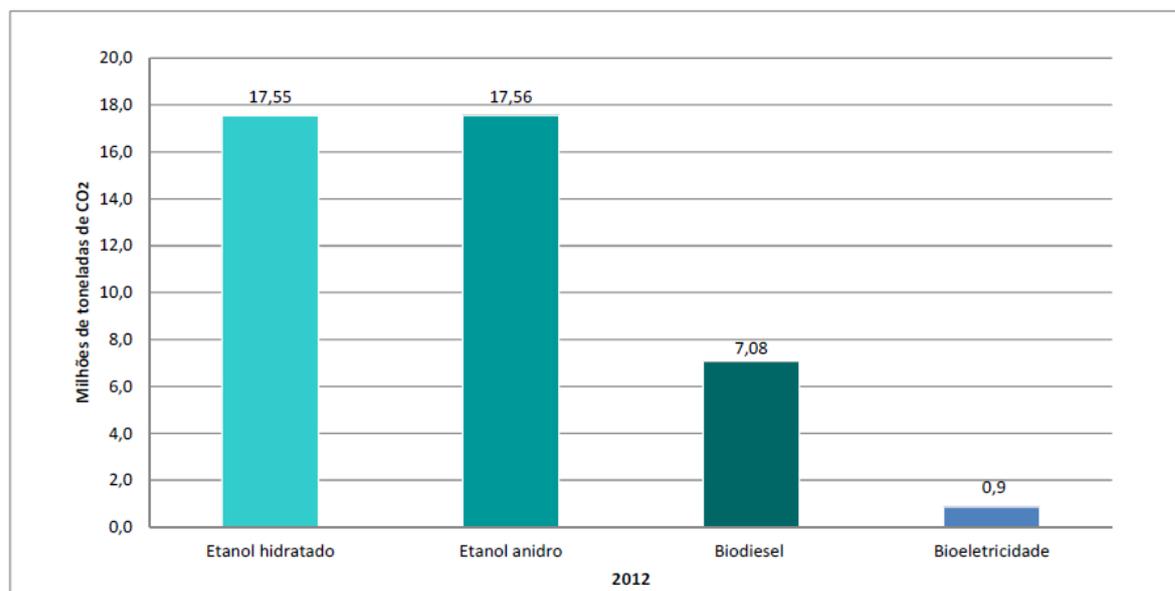
---

## **1.1 Bioenergia e Bioetanol**

O petróleo, principal fonte de energia do mundo moderno, alcançou posição de destaque durante décadas graças ao seu preço relativamente baixo e às facilidades de exportação e transporte do mesmo. Entretanto, a diminuição das reservas desse material fóssil e a elevação dos custos para sua obtenção, aliadas à crescente preocupação com a preservação ambiental, vêm exigindo soluções tecnológicas imediatas para a substituição dessa fonte de energia<sup>1</sup>. Tais soluções se baseiam na utilização de algumas fontes alternativas que permitam realizar em longo prazo e, com limitados impactos econômicos, a transição para uma nova era de energias mais limpas e renováveis<sup>2</sup>. Nesse sentido, é bastante intuitivo pensar em produzi-las com base em matéria orgânica renovável – a chamada bioenergia.

A bioenergia representa uma das melhores alternativas para captar e armazenar a energia solar sempre que haja terras livres para cultivo, clima adequado e conhecimento suficiente para aplicá-la. A energia solar na forma de biocombustível, produzido com eficiência e sustentabilidade, é capaz de atender às urgentes demandas para reduzir as emissões de gases de efeito estufa, melhorar a qualidade do ar nas metrópoles e competir em preço com as energias convencionais<sup>3</sup>. A principal diferença entre os biocombustíveis e os combustíveis fósseis, do ponto de vista de emissões, é que os combustíveis derivados de petróleo utilizam para a combustão o carbono das moléculas que estava estocado há milhões de anos nas jazidas petrolíferas, emitindo dióxido de carbono para a atmosfera, em um ciclo semiaberto, alterando sua composição química. No caso dos biocombustíveis, o ciclo de carbono é fechado, ou seja, todo o carbono emitido para atmosfera durante a combustão é reabsorvido no ciclo seguinte, para produção da mesma quantidade de biocombustível<sup>4</sup>. Além disso, o teor de oxigênio nos biocombustíveis é mais elevado, o que possibilita uma

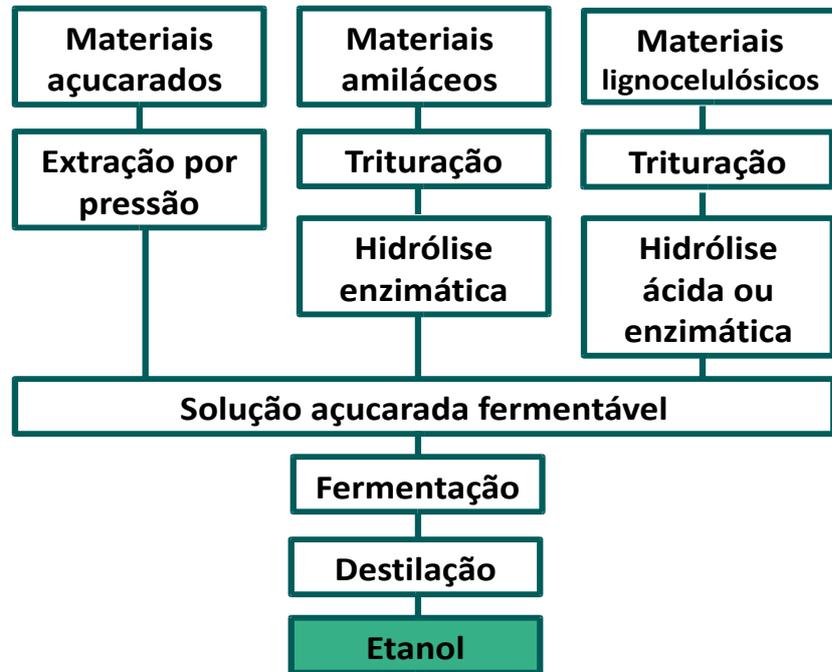
combustão mais limpa<sup>1,3</sup>. Na Fig. 1, observam-se as emissões evitadas no ano de 2012 no Brasil, medidas em toneladas de CO<sub>2</sub> equivalente, decorrentes do uso de biocombustíveis renováveis em detrimento de seus equivalentes fósseis<sup>5</sup>. Cerca de 35 milhões de toneladas de dióxido de carbono deixaram de ser lançadas na atmosfera com a substituição de gasolina por etanol anidro e hidratado.



**Figura 1** Emissões evitadas pela utilização de biocombustíveis em 2012 – Brasil

Dentre os biocombustíveis o que mais vem se destacando é o bioetanol, devido à sua efetiva inserção na matriz energética mundial e à maior maturidade do processo de produção do mesmo. Esse álcool é geralmente classificado em três tipos, dependendo da matéria-prima através da qual é obtido. O primeiro é derivado de materiais à base de açúcares simples, tais como cana-de-açúcar e beterraba sacarina. Por ser obtido a partir de fermentação direta desses açúcares, ele é denominado “etanol de primeira geração de tecnologias”. O segundo é derivado de materiais amiláceos como grãos de milho e trigo. O terceiro é produzido a partir de materiais lignocelulósicos, incluindo resíduos de culturas (por exemplo, palha de arroz, bagaço de cana-de-açúcar e palha de milho) e materiais lenhosos<sup>3</sup>. Esses dois últimos tipos são também chamados de “etanol de segunda geração de tecnologias”, pois, a matéria-prima

passa por um processo de hidrólise (ácida ou enzimática), antes da fermentação, para que os polissacarídeos sejam quebrados em açúcares fermentáveis. A Fig. 2 mostra as sínteses das rotas tecnológicas de obtenção de cada um deles<sup>3</sup>.



**Figura 2** Rotas para produção de bioetanol

Com relação ao bioetanol, o Brasil se encontra em uma posição privilegiada. O país é o segundo maior produtor de etanol de primeira geração (produzido a partir da fermentação do caldo da cana-de-açúcar) do mundo<sup>6</sup>. A produção e o uso de bioetanol como combustível veicular são praticados regularmente no Brasil desde 1931, com notável evolução durante as últimas décadas<sup>3</sup>. Embora a produção de etanol a partir do caldo da cana-de-açúcar seja um processo bem estabelecido no Brasil, com os menores custos, a maior produtividade e o melhor balanço energético do mundo, ainda há espaço para crescimento<sup>7</sup>. Existem diversas possibilidades de investimento em pesquisa, desenvolvimento e inovação para o aperfeiçoamento da produção de etanol, tanto pela busca de matérias-primas alternativas, que complementem a produção do álcool, quanto pelo desenvolvimento científico e tecnológico da obtenção do etanol lignocelulósico produzido a partir da celulose e hemicelulose.

A produção de bioetanol a partir de materiais celulósicos é assumidamente a que apresenta o maior potencial entre as alternativas possíveis para aumentar a produção do álcool no mundo. O bagaço produzido nas usinas de processamento de etanol de primeira geração, a partir da cana-de-açúcar, tem grande potencial para ser utilizado nesse processo. Entretanto, esta ainda não é uma realidade em escala industrial<sup>8</sup>. No Brasil, o desenvolvimento de combustíveis celulósicos segue com investimentos modestos. Apesar disto, no ano de 2012, foi anunciada a primeira usina comercial de etanol de segunda geração, localizada em Alagoas, cuja inauguração está prevista para 2014<sup>9</sup>.

Apesar da soberania da cana-de-açúcar como matéria-prima na obtenção de etanol tanto de primeira quanto de segunda geração, o setor sucroalcooleiro, por ser uma atividade agroindustrial, sofre a influência de uma característica intrínseca: a sazonalidade da produção da matéria-prima. O ciclo da cana-de-açúcar é de 12 a 18 meses o que dissocia o período de colheita das necessidades do consumo que se prolongam por todos os meses do ano. Assim, é preciso estocar uma grande proporção da produção de modo a permitir a manutenção da oferta de etanol no período de entressafra, quando as fábricas estão paralisadas<sup>10</sup>. Nesse contexto se insere a busca por matérias-primas capazes de complementar a oferta de bioetanol durante os períodos de entressafra da cana-de-açúcar.

## **1.2 Sorgo sacarino**

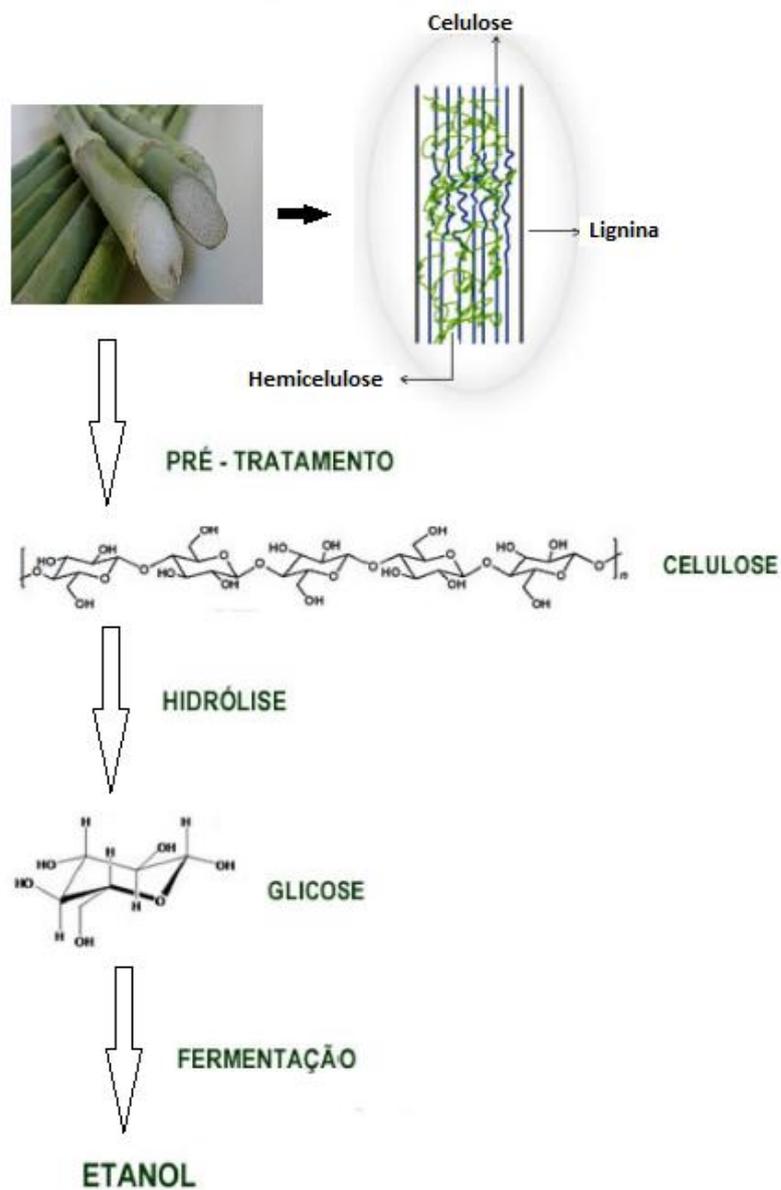
Uma das matérias-primas que se apresenta com grande potencial para produção de etanol, ao lado da cana-de-açúcar, é o sorgo sacarino (Fig. 3). O sorgo (*Sorghum bicolor* L. Moench) é uma planta de origem africana, da mesma família botânica do milho, conhecido por ser utilizado na alimentação animal, principalmente de bovinos<sup>11</sup>. Agronomicamente, os sorgos são classificados em cinco grupos: granífero, forrageiro para silagem e/ou sacarino, forrageiro para pastejo/corte, vassoura e biomassa.



**Figura 3** Planta de Sorgo sacarino

O sorgo do tipo sacarino apresenta colmos com caldo semelhante ao da cana-de-açúcar, é rico em açúcares fermentáveis, e pode servir para a produção de etanol de primeira geração na mesma instalação utilizada para cana<sup>12</sup>. O sorgo sacarino tem sido identificado como uma cultura de baixos custos de cultivo, pois, é uma planta com alta eficiência fotossintética, grande tolerância à seca e que requer o mínimo de fertilizantes para ser cultivada. Além disso, apresenta uma vantagem adicional em relação à cana-de-açúcar: seu período de maturação é de apenas 4 meses<sup>13</sup>.

Somado a isso, por ser uma matéria-prima que apresenta alto rendimento de biomassa por hectare e baixos custos de produção, o bagaço do sorgo sacarino é uma promessa futura na obtenção de etanol de segunda geração. Assim como outros materiais lignocelulósicos, ele contém quantidades consideráveis de carboidratos poliméricos (celulose e hemicelulose) que podem ser hidrolisados - por ácidos ou enzimas - em açúcares fermentáveis que serão posteriormente fermentados em etanol<sup>14</sup> (Fig. 4). Além dos carboidratos, a biomassa contém ainda quantidade apreciável de lignina, um complexo polímero fenólico, que torna o bagaço um excelente combustível para geração de eletricidade e vapor. Essa combustão da lignina contribui para a redução da energia fóssil utilizada nas usinas de produção de etanol<sup>15</sup>.



**Figura 4** Representação esquemática da produção de etanol a partir da biomassa lignocelulósica<sup>16</sup>

O cultivo do sorgo do tipo sacarino pode ser uma alternativa técnica e economicamente viável para fornecimento de matéria-prima às usinas alcooleiras, evitando o corte antecipado de cana-de-açúcar e a redução do rendimento industrial. Além disso, a inserção e expansão do sorgo sacarino em complemento à cana-de-açúcar para a produção de bioetanol podem estabilizar a oferta de álcool combustível no Brasil ao longo do ano<sup>17</sup>.

## **2 - MOTIVAÇÃO**

---

### **2.1 Programas de melhoramento genético do sorgo sacarino**

A utilização de sorgo sacarino para a produção de etanol de primeira e de segunda geração tem sido amplamente estudada pelo Programa de Melhoramento Genético de Sorgo (PMGS) da Embrapa. O desenvolvimento de novas variedades e híbridos dessa matéria-prima vem sendo realizado com o objetivo de aumentar a capacidade de produção do álcool<sup>18</sup>. Além disso, a redução dos custos de obtenção de etanol pode ser alcançada através da melhoria da qualidade da matéria-prima<sup>19</sup>.

Em um programa de melhoramento genético o ponto de partida é o banco de germoplasma, que reúne milhares de genótipos. Depois de serem obtidas por cruzamentos específicos definidos pelos pesquisadores, as sementes são enviadas às estações experimentais onde são plantadas. As plantas passam, então, por uma fase de seleção para a definição dos materiais com as características desejadas, por exemplo, maiores teores de açúcares no caldo e composição adequada da biomassa, um trabalho que é feito em conjunto com métodos biotecnológicos avançados. É nessa etapa do processo que as amostras devem ser caracterizadas em laboratório, pois, a matéria-prima precisa ser submetida a um controle de qualidade para garantir uma produção com maior rendimento e, portanto, com maior eficiência.

Devido á similaridade de composição, os métodos analíticos utilizados para avaliar os parâmetros de qualidade do caldo de cana-de-açúcar podem servir para a avaliação do caldo de sorgo<sup>18</sup>. Os parâmetros mais utilizados nessa avaliação são: teor de sólidos solúveis (brix), sacarose (leitura sacarimétrica, pol e pureza aparente), e açúcares redutores (glicose e frutose). Da mesma maneira, as metodologias e os parâmetros de qualidade empregados para avaliar a biomassa da cana-de-açúcar podem ser estendidos para o bagaço do sorgo sacarino utilizado

na produção de etanol de segunda geração<sup>18</sup>. Os métodos analíticos para avaliação de biomassa lignocelulósica devem possibilitar a quantificação dos polissacarídeos celulose e hemicelulose, bem como os teores de lignina. Além disso, é de interesse do PMGS estimar a quantidade de etanol celulósico que, teoricamente, pode ser obtido a partir de cada material desenvolvido.

É preciso levar em consideração que, para a caracterização dos materiais selecionados pelo PMGS, etapa imprescindível na otimização da produtividade, uma grande quantidade de amostras é gerada anualmente, tornando o processo impraticável com a utilização de metodologias de análise convencionais.

## **2.2 Parâmetros de qualidade e metodologias convencionais**

A caracterização da matéria-prima é realizada de acordo com a finalidade a que se destina. Para fins energéticos, nas rotas de produção de etanol de 1ª geração tecnológica faz-se a caracterização do caldo, e, para fins diferenciados (etanol de 2ª geração tecnológica) procede-se a caracterização da biomassa.

### **2.2.1 Caracterização do caldo de sorgo sacarino**

O caldo extraído do sorgo sacarino contém açúcares simples como glicose, frutose e sacarose<sup>20</sup>. Estes açúcares são diretamente fermentados para produzir o etanol, levando, dessa forma, ao processo mais barato. A seguir são apresentados os principais parâmetros a serem analisados na avaliação da qualidade do caldo, bem como uma breve descrição das metodologias de análises comumente utilizadas nessa avaliação.

#### **2.2.1.1 Brix**

O brix é o parâmetro mais utilizado na indústria do álcool. Por definição, ele expressa a porcentagem massa/massa dos sólidos solúveis contidos em uma solução pura de sacarose,

ou seja, mede o teor de sacarose em soluções puras<sup>21</sup>. A sua determinação quantitativa pode ser realizada através de refratômetros nos quais o índice de refração do caldo extraído é medido em escala brix (°B). Por definição, 1 °B corresponde a 1 grama de sacarose em 100 g de solução aquosa. No entanto, para o caldo de sorgo, assim como para o caldo de cana, ele representa a estimativa do teor de sacarose tendo em vista as demais impurezas presentes. Este parâmetro não fornece qualquer informação qualitativa acerca dos açúcares presentes.

### **2.2.1.2 Pol e leitura sacarimétrica**

Pol (S) é a medida da quantidade de sacarose na mistura de açúcares<sup>21,22</sup>. O cálculo do pol é feito através da determinação do grau de polarização da amostra. Essa medida, chamada de leitura sacarimétrica (°Z), é realizada em um polarímetro (“sacarímetro”) após a clarificação do caldo com uma mistura à base de alumínio.

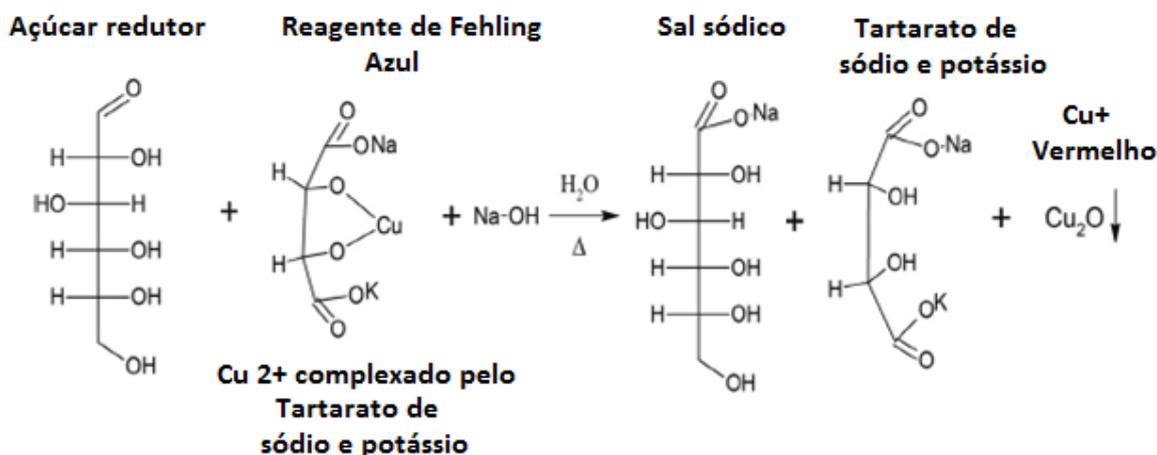
A equação completa para o cálculo do pol é a seguinte<sup>22</sup>:

$$S_{(\% \text{ do caldo})} = (1,00621 \times \text{LAI} + 0,05117) \times (0,2605 - 0,0009882 \times \text{B}) \quad (\text{Eq. 1})$$

na qual LAI é a leitura sacarimétrica obtida com a mistura clarificante à base de alumínio (°Z) e B é o brix do caldo (°B).

### **2.2.1.3 Açúcares redutores**

Os açúcares redutores (glicose e frutose) podem ser avaliados por uma titulação de oxirredução. O método mais comum, o de Lane-Eynon<sup>23</sup>, baseia-se no fato de que os sais cúpricos, em solução tartárica alcalina (solução de Fehling), podem ser reduzidos a quente por aldoses ou cetoses transformando-se em sais cuprosos, que se precipitam, conforme mostrado na Fig. 5.



**Figura 5** Equação da reação química envolvida no método de Lane-Eynon

Para alcançar maior exatidão dos resultados, a solução de Fehling deve ficar sob ebulição constante durante toda a titulação, pois o  $\text{Cu}_2\text{O}$  formado pode ser novamente oxidado pelo  $\text{O}_2$  do ar, mudando novamente a cor para azul. Além disso, a titulação deve levar no máximo três minutos, porque pode haver decomposição dos açúcares com o aquecimento prolongado. A relação entre o cobre reduzido e o açúcar redutor não é estequiométrica, e o resultado, obtido de tabelas ou padronizando-se a mistura de Fehling com uma solução de concentração conhecida de açúcar, é geralmente expresso em porcentagem de glicose.

A desvantagem deste tipo de método baseado na redução de metais e outros compostos é que os resultados dependem de um tempo de reação preciso, a temperatura da reação e a concentração dos reagentes devem ser cuidadosamente controladas e eles são bastante susceptíveis à interferência. A porcentagem de açúcares redutores presentes no caldo é dada pela equação 2:

$$\text{AR}_{(\%)} = C_p \times V_p / V_a \times \text{dil} \quad (\text{Eq. 2})$$

na qual  $C_p$  é a concentração da solução padrão de glicose em % m/v,  $V_p$  é o volume do padrão de glicose gasto para a padronização da solução de Fehling em mL,  $V_a$  é o volume de amostra gasto para a titulação da solução padronizada de Fehling em mL e dil é o fator de diluição da amostra (quando houver).

#### **2.2.1.4 Pureza aparente do caldo**

A pureza aparente do caldo (Q) é definida como a porcentagem de pol em relação ao brix, e é calculada pela equação<sup>22</sup>:

$$Q = 100 \times S \div B \quad (\text{Eq. 3})$$

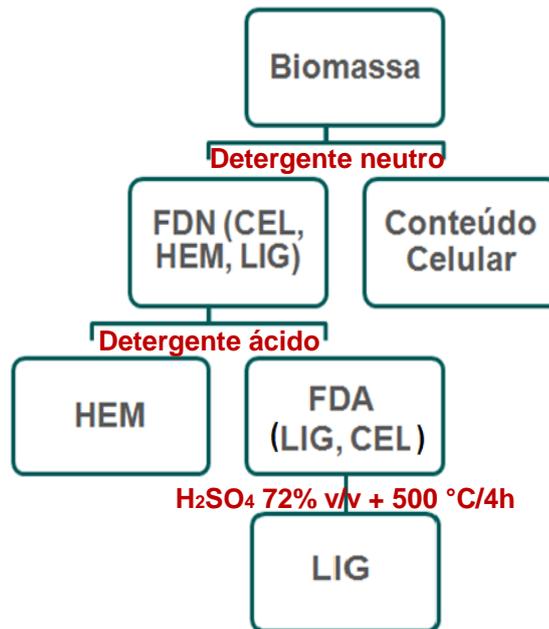
Caldos com pureza aparente abaixo de 75% são considerados de baixa pureza.

#### **2.2.2 Caracterização da biomassa de sorgo sacarino**

A biomassa lignocelulósica é basicamente composta de três polímeros: celulose, hemicelulose e lignina<sup>24</sup>. Para a produção de etanol de segunda geração é desejável que os teores desses compostos sejam determinados e, ainda, que seja feita uma estimativa do rendimento teórico de etanol com base nos teores dos polissacarídeos hidrolisáveis.

A seguir são apresentadas as principais metodologias para determinação desses parâmetros de qualidade da biomassa, bem como uma breve descrição das análises de referência.

A metodologia mais simples utilizada para as análises de celulose, hemicelulose e lignina é baseada no método de extração sequencial de van Soest<sup>25</sup>. Este método foi proposto em 1968 e considera que os constituintes das plantas podem ser divididos em conteúdo celular (lipídios, compostos nitrogenados, gorduras, amido e outros compostos solúveis em água) e parede celular (proteína insolúvel, hemicelulose, celulose e lignina). A metodologia é subdividida em FDA (fibras insolúveis em detergente ácido) e FDN (fibras insolúveis em detergente neutro)<sup>26</sup>, como mostrado na Fig. 6:



**Figura 6** Representação esquemática da metodologia de extração sequencial baseada no método de van Soest.

### 2.2.2.1 FDN

A célula vegetal é revestida por uma parede celular rígida composta basicamente por celulose, mas em células adultas esta parede sofre um espessamento que pode formar uma segunda parede composta por lignina e hemicelulose. O método de van Soest consiste, inicialmente, em separar o conteúdo celular da parede celular (Fig. 6). Isto é feito aquecendo-se parte da amostra em solução de detergente neutro. O conteúdo celular solubiliza-se no detergente, enquanto a parede celular não, podendo ser separada por filtração. As frações resultantes são denominadas fibras solúveis em detergente neutro, compostas por proteína, nitrogênio não proteico, lipídeos, pigmentos, açúcares, ácidos orgânicos e pectina, e FDN, constituída basicamente por celulose, hemicelulose e lignina.

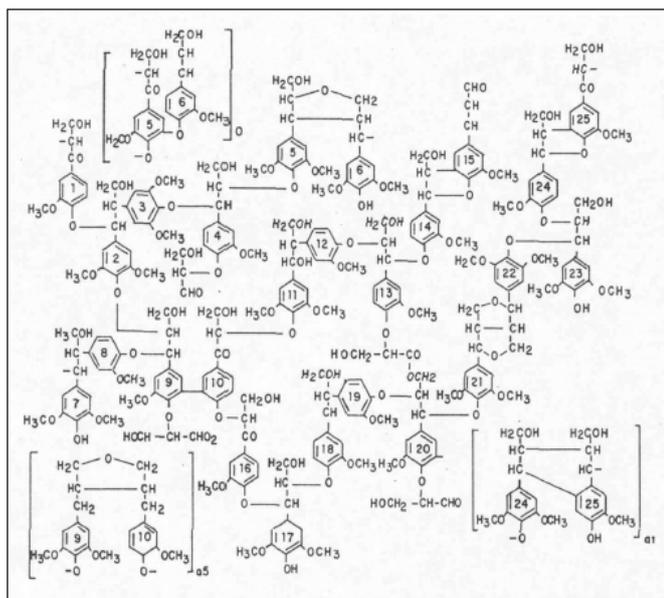
### 2.2.2.2 FDA

Quando se utiliza solução de detergente ácido, a hemicelulose solubiliza-se e a lignina ligada à celulose (lignocelulose) é separada por filtração (Fig. 6). As duas frações são

denominadas, respectivamente, fibras solúveis em detergente ácido e FDA (constituída de celulose e lignina).

### 2.2.2.3 Lignina

A lignina é um dos principais componentes da biomassa vegetal, possuindo natureza aromática e muito complexa. Ela está presente em muitas espécies vegetais e não possui a mesma estrutura química em todas elas. A presença de diferentes unidades precursoras e o elevado número de combinações possíveis entre essas unidades faz com que a estrutura da macromolécula de lignina seja bem mais complexa que as apresentadas pela celulose e pela hemicelulose. Na Fig. 7 é apresentado o modelo estrutural da lignina proposto por Nimz<sup>27</sup>:



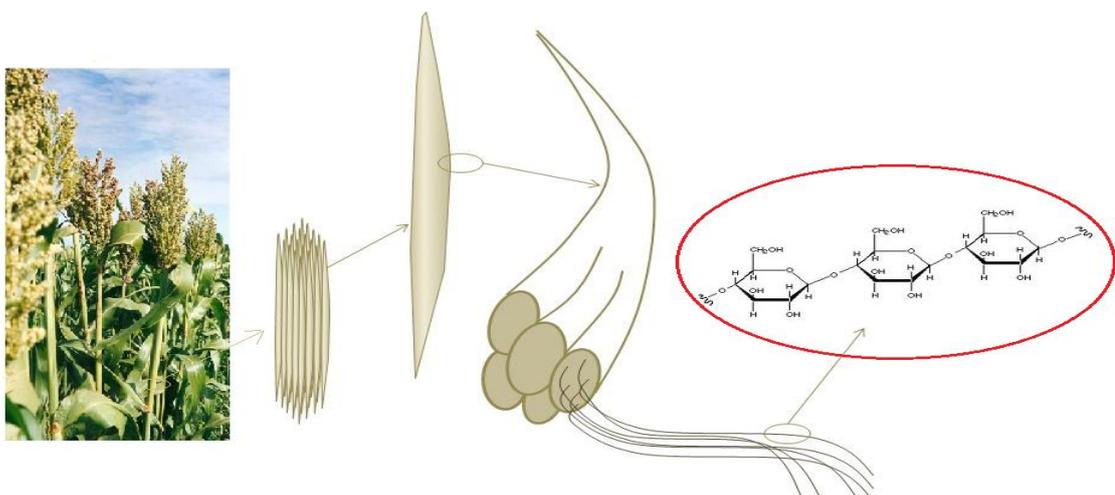
**Figura 7** Modelo estrutural da lignina proposto por Nimz e co-autores

Durante o desenvolvimento das células, a lignina é incorporada como o último componente na parede, interpenetrando as fibrilas e assim enrijecendo as paredes celulares. Ela tem um importante papel no transporte de água, nutrientes e metabólitos, sendo responsável pela resistência mecânica de vegetais, além de proteger os tecidos contra o ataque de microorganismos<sup>24</sup>. Os teores desejados de lignina na biomassa são controversos, pois, enquanto se espera maior produção de energia com o aumento do conteúdo de lignina nos

processos de cogeração, a eficiência da conversão dos polissacarídeos (celulose e hemicelulose) em etanol é prejudicada pela presença desse composto aromático. Dessa maneira, conhecer a quantidade e composição da lignina no sorgo sacarino é importante para as linhas de melhoramento genético que precisam adequar essa composição à finalidade que se pretende dar à biomassa. A determinação da lignina também segue a metodologia de van Soest e é feita sequencialmente à extração das fibras em detergente neutro e ácido, através da digestão do resíduo de FDN com solução de ácido sulfúrico e posterior calcinação (Fig. 6).

#### 2.2.2.4 Celulose

A celulose localiza-se principalmente na parede celular secundária e faz parte da estrutura fibrosa organizada. A celulose é um polímero orgânico, que é constituído unicamente por unidades de D-glicose, mantidas unidas numa cadeia molecular linear por ligações glicosídicas  $\beta$ -1,4<sup>28</sup> (Fig. 8).



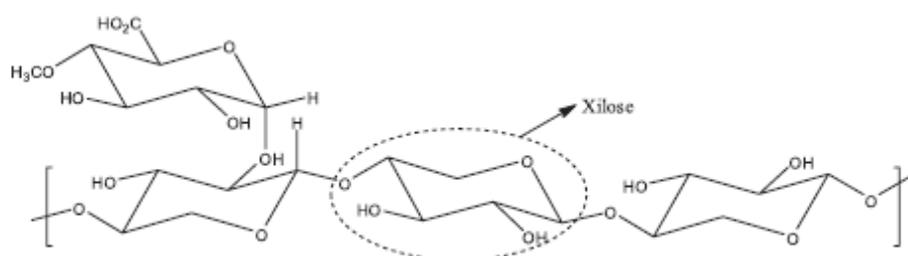
**Figura 8** Celulose - Polímero linear de moléculas de D-glicose.

As cadeias de celulose nas paredes celulares das plantas são arranjadas compactamente, de modo que suas fibras apresentam regiões nitidamente cristalinas, uma consequência do grande número de ligações de hidrogênio que resulta numa forte interação entre suas moléculas. A metodologia mais comum para a determinação de celulose se baseia

no teor de FDA. Esta fibra é composta basicamente de lignina, celulose e sílica. Uma vez determinados os teores de FDA e de lignina, efetua-se a incineração do resíduo da lignina. A celulose corresponde à parte do resíduo que se perde com a incineração. O que sobra, corresponde à cinza residual, que contém sílica. A celulose é então calculada por diferença de peso<sup>29</sup>.

### 2.2.2.5 Hemicelulose

A hemicelulose é um heteropolímero que possui muitas ramificações contendo diferentes carboidratos, como xilose, arabinose, manose, galactose, glicose, assim como ácidos urônicos<sup>24</sup>. Dependendo da predominância do tipo de açúcar as hemiceluloses podem ser chamadas de arabino-xilanas, mananas, glucanas ou galactanas. Esses açúcares contêm cinco (pentoses) ou seis (hexoses) carbonos em sua estrutura e são unidos por ligações glicosídicas quase sempre acetiladas, formando uma estrutura fraca e hidrofílica que serve como uma conexão entre a lignina e as fibras de celulose, além de conferir rigidez ao complexo celulose-hemicelulose-lignina. Comparadas com a celulose, as hemiceluloses apresentam maior susceptibilidade à hidrólise ácida, pois oferecem uma maior acessibilidade aos ácidos minerais comumente utilizados como catalisadores. Esta reatividade é usualmente atribuída ao caráter amorfo destes polissacarídeos<sup>24</sup>. Na Fig. 9 está representada uma molécula de hemicelulose. A proporção de hemicelulose é determinada através de cálculos pela diferença entre os teores de FDN e FDA.



**Figura 9** Representação de um tipo de molécula de hemicelulose

### 2.2.2.6 Rendimento Teórico de Etanol (RTE)

Para o PMGS é importante que sejam avaliados diversos cultivares de sorgo sacarino sob diferentes ciclos de cultivo, com o objetivo de aumentar o período de colheita para a produção de etanol em escala industrial. É interesse do Programa identificar a quantidade de etanol que poderia ser produzida por cada cultivar em diferentes períodos de maturação da planta (antes da maturação, no pico de maturidade e na maturação tardia) e sob diferentes tipos de manejo de cultivo e estresse (ataque de pragas, períodos de seca, deficiência de fertilidade, excesso de água). Além disso, esta estimativa também é importante na comparação do potencial de produção de etanol entre híbridos e linhagens do sorgo sacarino.

A forma mais usual de estimar o rendimento de etanol em escala laboratorial consiste na utilização de metodologias caras e demoradas. Tratam-se de etapas de hidrólise enzimática da biomassa e fermentação dos açúcares, seguidas de análises cromatográficas para a determinação do teor de álcool<sup>30</sup>. No entanto, alguns artigos têm correlacionado com sucesso a análise da composição de biomassa com a produção de etanol prevista para certas matérias-primas, incluindo o sorgo<sup>19,31</sup>.

Nesse contexto, baseado no trabalho de *Zhao et al*<sup>31</sup>, a equação a seguir pode ser usada para determinação do RTE a partir dos teores de celulose e hemicelulose presentes no bagaço do sorgo:

$$\text{RTE}_{\text{cel,hem}} = (\text{CEL} + \text{HEM}) \times (\text{BS}) \times 1,11 \times 0,85 \times 0,51 \times 0,85 \times 1000/0,79 \quad (\text{Eq. 4})$$

na qual  $\text{RTE}_{\text{cel,hem}}$  é o rendimento teórico de etanol de celulose e hemicelulose em  $\text{L ha}^{-1}$ , CEL + HEM são os teores de celulose e hemicelulose na matéria seca em %, BS é a biomassa seca em  $\text{t ha}^{-1}$ , 1,11 é o fator de conversão de açúcar de celulose e hemicelulose, 0,85 é a eficiência do processo de açúcar de celulose e hemicelulose, 0,51 é o fator de conversão de etanol a

partir de açúcar, 0,85 é a eficiência do processo de etanol a partir de açúcar,  $1000/0,79$  é a densidade do etanol  $\text{g mL}^{-1}$ .

A maioria das metodologias comumente usadas na caracterização do caldo e da biomassa, como foi exposto, é demorada, complexa, cara e demanda grande quantidade de reagentes, muitas vezes tóxicos, o que por sua vez leva à geração de um grande volume de resíduos químicos. Além disso, tendo em vista a quantidade de amostras geradas pelo PMGS, o uso dessas metodologias pode tornar as análises inviáveis de serem realizadas em tempo hábil para tomadas de decisão.

Dessa forma, a busca por metodologias mais rápidas e de baixo custo tem se tornado constante nos laboratórios que atendem a demanda do PMGS. Além disso, é preciso lembrar que a redução do uso de reagentes tóxicos e a diminuição da quantidade de resíduos gerados nas análises vão de encontro a uma questão bastante atual na pesquisa científica: a química verde<sup>32</sup>. A fim de atingir os objetivos da química analítica verde, métodos não destrutivos, ou pelo menos que evitem o tratamento químico das amostras, são desejáveis. Assim, cada vez mais, aumenta o interesse pelos métodos vibracionais de análise.

### **3 - REFERENCIAL TEÓRICO**

---

#### **3.1 Espectroscopia no infravermelho próximo – NIRS<sup>33-37</sup>**

Uma alternativa analítica para superar as desvantagens das metodologias tradicionais é o uso da espectroscopia no infravermelho próximo, NIRS (*Near Infrared Spectroscopy*) combinada com métodos de calibração multivariada. Esta técnica possibilita a determinação simultânea de várias propriedades (ou concentrações de vários analitos) de uma amostra usando uma única medida instrumental sem a necessidade de uma separação física dos componentes. Como consequência, ela apresenta baixo consumo de reagentes, pouca manipulação de amostras e alta velocidade de análise, o que pode ser crucial no mundo real, onde qualquer atraso na obtenção de resultados analíticos pode resultar em significantes perdas tanto de qualidade quanto financeiras. Com a utilização deste tipo de técnica, reagentes potencialmente perigosos são evitados e claros benefícios à saúde dos operadores e ao ambiente são alcançados. Além disso, há considerável redução de custos com tratamento de resíduos químicos.

A espectroscopia no infravermelho próximo é uma técnica vibracional que durante as últimas duas décadas tem feito progressos evidentes em suas aplicações em diversas áreas da ciência<sup>38-40</sup>. A radiação NIR emprega fótons de energia ( $h\nu$ ) na faixa de  $2,65 \times 10^{-19}$  a  $7,96 \times 10^{-20}$  J, o que corresponde à faixa de comprimento de onda de 750 a 2.500 nm (números de onda de  $13.330$  a  $4.000 \text{ cm}^{-1}$ ) do espectro eletromagnético. Esta faixa de energia é suficiente apenas para promover a vibração molecular. À temperatura ambiente a maioria das moléculas está em seus níveis de energia do estado vibracional fundamental. Átomos ou grupos de átomos participantes das ligações químicas estão se deslocando uns em relação aos outros em uma frequência que é definida pela força das ligações e pela massa dos átomos individuais ou grupos de átomos ligados. A amplitude dessas vibrações é de poucos nanômetros e aumenta

se alguma energia é transferida para a molécula. Esta energia (E) pode ser transferida na forma de um fóton de dado comprimento de onda ( $\lambda$ ) e pode ser dada por:

$$E = hv = hc/\lambda \quad \text{(Eq. 5)}$$

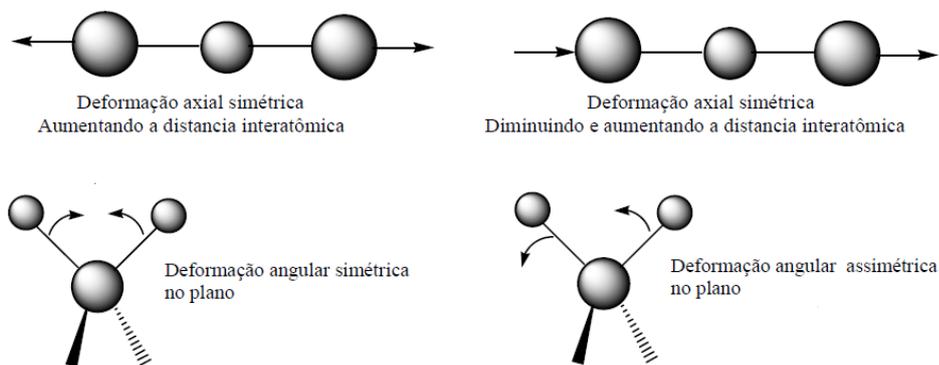
Na qual h é a constante de Planck e c é a velocidade da luz no vácuo.

A radiação de uma dada frequência, capaz de fornecer exatamente a energia entre dois níveis vibracionais, pode ser absorvida pela molécula e produzir excitação a níveis de energia vibracionais maiores. Entretanto, apenas essa energia não é suficiente para a absorção da radiação. Para que haja variações na amplitude dos movimentos das moléculas, é necessário que o campo de oscilação elétrica da onda eletromagnética possa interagir com a molécula. Isto só pode ocorrer se o deslocamento dos átomos no modo vibracional puder produzir uma mudança no momento de dipolo da molécula ou do grupo de átomos vibrantes.

O momento dipolar é determinado pela magnitude da diferença de carga e a distância entre os dois centros de massa. Quando uma molécula que possui essa variação do momento dipolar vibra, surge um campo que pode interagir com o campo elétrico associado à radiação. Dessa forma, a correspondência da energia da radiação com a diferença de energia entre dois níveis vibracionais provoca a resposta seletiva do sistema molecular à radiação incidente. Isto significa que em uma dada faixa de comprimento de onda, muitas frequências serão absorvidas, outras não, enquanto outras serão ainda apenas parcialmente absorvidas. Do mesmo modo, a rotação de moléculas assimétricas em torno dos seus centros de massa resulta em uma variação periódica do dipolo que pode interagir com a radiação.

As vibrações moleculares podem ser de dois tipos: deformações axiais (ou estiramentos) e deformações angulares. Quando a deformação ocorre na direção do eixo da molécula, a distância interatômica aumenta e diminui alternadamente e o modo de vibração é denominado deformação axial. As vibrações de deformação angular correspondem ao movimento de um grupo de átomos em relação ao resto da molécula, sem que as posições

relativas dos átomos do grupo se alterem. Essas deformações recebem a denominação de deformação angular simétrica e assimétrica no plano e deformação angular simétrica e assimétrica fora do plano (Fig. 10).



**Figura 10** Algumas vibrações moleculares

O modelo mecânico mais simples, conhecido como oscilador harmônico (Fig. 11a), pode ser usado, de forma mais didática, para representar as vibrações que acontecem de forma isolada em uma molécula. As características da vibração de estiramento podem se aproximar do modelo de duas massas ligadas por uma mola. As frequências fundamentais de quaisquer dois átomos ligados podem ser calculadas assumindo que a energia obedece à Lei de Hooke. 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, tendendo a restaurar as massas para sua posição original. A frequência ( $\nu$ ) da vibração é dada por:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \tag{Eq. 6}$$

em que  $k$  é a constante de força da ligação e  $\mu$  é a massa reduzida dos dois átomos de massas  $m_1$  e  $m_2$ , definida como:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad \text{(Eq. 7)}$$

A vibração molecular pode ser descrita por um modelo simplificado no qual a energia potencial (E) é dada como função da separação entre os átomos (y):

$$E = \frac{1}{2} k y^2 \quad \text{(Eq. 8)}$$

Este modelo é útil para o entendimento de energia vibracional e funciona bem para o cálculo de frequências fundamentais de moléculas diatômicas simples. Contudo, essa aproximação fornece apenas a média ou a frequência central de transições de estados vibracionais e rotacionais de ligações diatômicas. Em moléculas poliatômicas os elétrons sofrem influência de átomos ou grupos vizinhos e isso influencia o estiramento, o comprimento da ligação, o ângulo da ligação e, conseqüentemente, a frequência da vibração das ligações químicas. Estas diferenças específicas, que ocorrem devido a essas interações, proporcionam que cada substância apresente um espectro característico.

Para representar as diferenças citadas um modelo harmônico/quântico pode ser utilizado. Para a mecânica quântica as vibrações moleculares podem ter apenas determinados níveis de energia discretos definidos como:

$$E = \left( v + \frac{1}{2} \right) \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} \quad \text{(Eq. 9)}$$

em que h é a constante de Planck e v é o número quântico vibracional, que pode assumir valores positivos e inteiros incluindo o zero. Para moléculas poliatômicas, os níveis de energia se tornam numerosos e uma aproximação pode tratar essas moléculas como uma série de osciladores harmônicos diatômicos e independentes. Nesse caso, a equação para a energia potencial pode ser generalizada como:

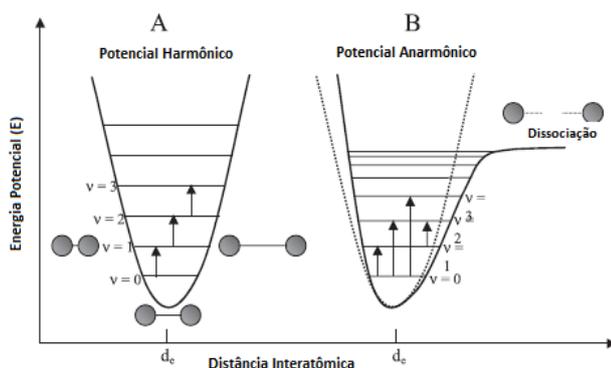
$$E_{(v_1, v_2, \dots)} = \sum_i^{i=3N-6} \left( v_i + \frac{1}{2} \right) \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} = \sum_i^{i=3N-6} \left( v_i + \frac{1}{2} \right) h \nu_m \quad \text{(Eq. 10)}$$

em que  $3N-6$  é o número de vibrações possíveis em uma molécula com  $N$  átomos (para moléculas lineares existem  $3N-5$  vibrações possíveis),  $\nu_m$  é a frequência vibracional do modelo clássico e  $\nu_1, \nu_2, \dots = 0, 1, 2, \dots$ .

Uma restrição imposta pelo modelo harmônico é que somente transições fundamentais poderiam existir, sendo esta restrição denominada de regra de seleção ( $\Delta\nu = \pm 1$ ). Uma vez que os níveis vibracionais são igualmente espaçados, apenas um único pico de absorção deveria ser observado para uma certa vibração molecular. Apesar de o modelo harmônico ajudar no entendimento da espectroscopia vibracional, ele apresenta algumas restrições por não permitir transições nas quais  $\Delta\nu$  é maior que 1. Transições com  $\Delta\nu \geq 2$  são proibidas pelo modelo harmônico e, neste caso, muitos dos fenômenos observados no espectro vibracional - bandas de sobretons - não poderiam existir. Além disso, as vibrações no modelo harmônico são independentes e suas combinações não existiriam sob as restrições impostas pelo modelo. No entanto, tanto bandas de sobretons quanto de combinação existem. Outra limitação apresentada por esse modelo é que à medida que dois átomos se aproximam, a repulsão coulombiana entre os dois núcleos produz uma força que age na mesma direção da força de restauração da ligação. Assim, espera-se que a energia potencial cresça mais rapidamente do que é previsto pelo modelo do oscilador harmônico. Por outro lado, quando a distância entre os átomos aumenta, um decréscimo na força de restauração e, portanto, na energia potencial, ocorre quando a distância interatômica se aproxima daquela em que ocorre a dissociação dos átomos.

Um modelo mecânico mais realístico para moléculas diatômicas pode ser visto na Fig. 11b. A molécula é ainda aproximada por duas massas ligadas por uma mola. No entanto, o modelo considera um comportamento não ideal do oscilador, que leva em conta a repulsão entre as nuvens eletrônicas quando os núcleos dos átomos se aproximam; e um comportamento variável da força da ligação quando os átomos se afastam uns dos outros. De

fato, em uma molécula real, o excesso de afastamento dos núcleos atômicos irá causar ruptura da ligação molecular com conseqüente dissociação dos átomos. O modelo anarmônico também impõe algumas restrições aos estados de energias possíveis das moléculas. Entretanto, ele prediz a ocorrência das transições com  $\Delta v \geq 2$  e a existência de bandas de combinação entre as vibrações. Sob as suposições do modelo anarmônico, as vibrações não são independentes e podem interagir umas com as outras.

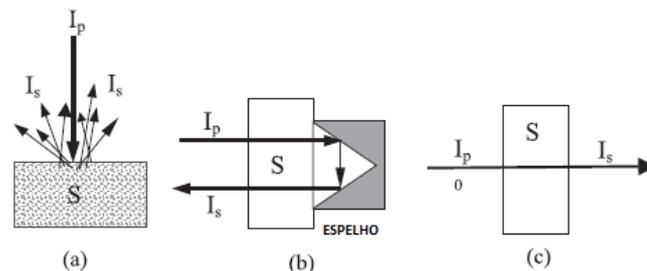


**Figura 11** Representação esquemática dos modelos harmônico (A) e anarmônico (B) para a energia potencial de moléculas diatômicas<sup>33</sup>.  $d_e$  = distância de equilíbrio

O espectro NIR pode ser dividido em região I (12500 – 8333  $\text{cm}^{-1}$ ), região II (8333-5555  $\text{cm}^{-1}$ ) e região III (5555-4000  $\text{cm}^{-1}$ ). Os limites dessas três regiões não são muito rigorosos. A região I é aquela na qual tanto bandas de combinação quanto de sobretons são observadas. Todas as bandas que aparecem nessa região são muito fracas, o que significa que ela mostra alta transparência. Um grande número de aplicações biomédicas e de produtos agrícolas é realizado nesta região. Na região II, podem ser observadas bandas resultantes do primeiro sobretom de estiramento vibracional de X-H (X=C, O, N) e vários tipos de modos de combinação de vibrações X-H. A região III é a dos modos de combinação cuja “permeabilidade” é mais pobre. Uma ampla gama de aplicações, tais como: estudos sobre ligações de hidrogênio e interações solvente-soluto usa as região II e região III.

Por se tratar de uma técnica vibracional, a maioria das vantagens da espectroscopia NIR advém da possibilidade de utilização de amostras intactas apresentadas diretamente ao

equipamento, sem a necessidade de pré-tratamentos. Este fato implica na promoção de interações não convencionais da radiação com a matéria por meio de diferentes modos de medidas. A Fig. 12 mostra alguns dos modos de medida empregados na obtenção do espectro NIR. A principal forma de obter a informação espectral é analisando a informação contida na porção da radiação difusamente refletida pelas amostras sólidas em vez do fraco sinal de transmitância. Na reflectância difusa (Fig. 12a), absorvância e espalhamento da radiação pelas amostras contribuem para mudança na intensidade de sinal. Na Fig. 12b está esquematizada outra possibilidade empregada na obtenção da informação, o modo de transflectância, que é uma forma especial de transmitância (Fig. 12c). A diferença em relação à transmitância simples é que o caminho óptico dobra, pois o feixe de radiação passa duas vezes através da amostra.



**Figura 12** Alguns modos de medida empregados em espectroscopia NIR<sup>33</sup>: (a) reflectância difusa, (b) transflectância e (c) transmitância. ( $I_p$  e  $I_s$  = intensidade da radiação incidente e que emerge do material, respectivamente)

Apesar da complexidade e da presença de bandas de absorção amplas e sobrepostas, o espectro NIR apresenta grande quantidade de informações, que têm sido intensamente explorada em análises químicas e físicas. Para fins quantitativos as informações presentes no espectro NIR podem ser empregadas na estimativa das concentrações ou propriedades físicas de uma dada substância em uma amostra, desde que elas reflitam mudanças significativas na intensidade de sinal ou no comprimento de onda do espectro produzido pela amostra. A aplicação de NIRS em análises de biomassa iniciou-se na última década<sup>41,42</sup>. Desde então,

muitos artigos têm sido publicados usando NIRS para caracterizar matérias-primas usadas para a produção de biocombustíveis, tais como palha de milho<sup>41</sup>, gramíneas<sup>43</sup> e grãos de cereais<sup>44</sup>. Apenas um artigo tratou do uso do NIRS para a determinação de três propriedades do bagaço de sorgo: poder calorífico, digestibilidade *in vitro* e proteína bruta<sup>45</sup>. Mais importante para os objetivos dos programas de melhoramento genético de plantas, focados em biocombustíveis, alguns artigos têm recentemente usado NIRS para prever o rendimento de etanol a partir de gramíneas (*Panicum virgatum*)<sup>43</sup>, de cereais como trigo, centeio e triticale<sup>44</sup>, beterraba<sup>46</sup> e milho<sup>47</sup>. Em todos eles o rendimento de etanol foi medido por HPLC após as amostras terem sido submetidas a um processo de fermentação.

Com relação às amostras líquidas, alguns trabalhos utilizando NIRS na predição de brix<sup>48</sup>, pH<sup>48</sup>, açúcares<sup>49,50</sup> e outros parâmetros<sup>51</sup> de qualidade em sucos de frutas já foram descritos. A utilização de NIRS na predição de alguns parâmetros de qualidade do caldo de cana-de-açúcar como pol<sup>52</sup>, brix<sup>52,53</sup> e açúcares redutores<sup>52</sup>, também já foi relatada.

É preciso destacar que os benefícios dessa técnica não podem ser considerados como totalmente explorados sem o uso da quimiometria, que permite a seleção de procedimentos de medidas ótimas e a extração da máxima informação química a partir dos dados espectrais.

## **3.2 Quimiometria e calibração multivariada**

### **3.2.1 Definição dos termos<sup>54-56</sup>**

A quimiometria pode ser definida como a aplicação de métodos matemáticos, estatísticos, gráficos ou simbólicos para maximizar a informação que pode ser extraída a partir de dados químicos e para planejar ou selecionar procedimentos e experimentos em condições ótimas. Esta disciplina foi introduzida por grupos de pesquisa em química, no final da década de 1960 e durante toda a década de 1970. Foram duas as razões para sua emergência: a necessidade de introdução de instrumentação que fornecesse respostas

multivariadas para cada amostra analisada e o aumento da disponibilidade dos computadores. O surgimento do pensamento quimiométrico veio da percepção de que a estatística univariada não era mais suficiente para descrever e modelar experimentos químicos. A quimiometria alcançou maturidade em meados da década de 1980, e hoje segue com algumas importantes linhas de aplicação: 1 – uso no planejamento e otimização experimental; 2 – reconhecimento de padrões e classificação de dados; 3 – processamento de sinais analíticos; 4 – calibração multivariada e 5 – métodos de inteligência artificial.

A calibração multivariada é a aplicação de maior sucesso da combinação da quimiometria com dados espectrais. Ela consiste na construção de uma relação entre certas propriedades químicas, físicas ou biológicas de uma amostra e seu espectro. Esta é uma extensão da lei de Beer aplicada a relações univariadas entre absorvância e concentração. Uma fórmula matemática é requerida com a finalidade de transformar os dados medidos (leituras espectrométricas) em informação relevante (concentrações químicas).

A equação de predição mais simples é a linear univariada:

$$C = b^0 + xb^1 \quad \text{(Eq. 11)}$$

na qual C é a estimativa da concentração, x é a variável medida e  $b^0$  e  $b^1$  são os coeficientes de regressão linear e angular, respectivamente. No entanto, contribuições de constituintes químicos interferentes com sinais instrumentais sobrepostos ou outros interferentes sistemáticos nos sinais medidos podem criar erros grosseiros nas concentrações preditas se esta calibração convencional univariada é utilizada. Dessa forma, muitos esforços têm sido direcionados para elaborar preparos de amostras prévios às medidas, purificando as amostras para obter medidas livres de interferências.

A calibração multivariada oferece uma solução alternativa para este problema específico: através da combinação de muitas medidas diferentes (ex.: absorvância da luz em

uma faixa de comprimentos de onda), interferentes podem ser modelados e eliminados. A equação de predição multivariada linear pode ser escrita como:

$$\mathbf{C} = \mathbf{b}^0 + \mathbf{x}\mathbf{b} \quad (\text{Eq. 12})$$

na qual  $\mathbf{x} = (x^1, x^2, \dots, x^p)$  representa um vetor linha, de  $p$  diferentes variáveis medidas, e  $\mathbf{b} = (b^1, b^2, \dots, b^p)$  é o correspondente vetor coluna de coeficientes de regressão. Aqui, iremos supor que a resposta instrumental lida,  $x$ , é linearizada. A determinação dos coeficientes de regressão  $b^0, b^1, \dots, b^p$  é a chamada calibração multivariada e pode ser feita de diversas formas. A principal delas é a calibração indireta para a qual é necessário um conjunto representativo de amostras de referência com valores medidos para ambos, espectros e concentrações dos constituintes a serem preditos. Este conjunto de dados de calibração é usado para a estimativa estatística dos parâmetros do modelo, que após a transformação matemática, rende os coeficientes de regressão desejados,  $b^0, b^1, b^2 \dots$

A técnica que mais utiliza a calibração multivariada indireta é a espectroscopia NIR. Ela era vista como uma técnica de pouca aplicação antes da introdução da quimiometria, pois as bandas de absorção no NIR são geralmente muito fracas e sobrepostas, dificultando análises qualitativas e tornando quase impossível a obtenção de um comprimento de onda seletivo que possibilite a calibração univariada.

### 3.2.2 Métodos de calibração multivariada<sup>57</sup>

Muitos métodos de regressão vêm sendo utilizados em química analítica para a construção de modelos de calibração multivariada, dentre esses os mais empregados têm sido a regressão linear múltipla (MLR), a regressão por componentes principais (PCR) e a regressão por mínimos quadrados parciais (PLS), que são métodos para ajuste linear entre as variáveis. A maioria dos métodos de calibração multivariada empregados em espectroscopia utiliza o ajuste linear, uma vez que, este representa o modelo de mais fácil elaboração e interpretação. Todas as técnicas multivariadas lineares buscam resolver o mesmo problema

genérico. Se  $n$  espectros de calibração são medidos a  $p$  comprimentos de onda, então  $\mathbf{X}$ , a matriz de dados espectrais, é definida como uma matriz  $n \times p$  contendo as variáveis espectrais como colunas. Similarmente,  $\mathbf{y}$  é o vetor de dimensão  $n \times 1$  contendo os valores de referência para as amostras de calibração (caso se deseje modelar mais de uma propriedade simultaneamente o vetor  $\mathbf{y}$  passa a ser uma matriz  $\mathbf{Y}$   $n \times k$ ). O objetivo do modelo multivariado linear é calcular um vetor predição  $\mathbf{p}$  de dimensão  $n \times 1$  que resolva a equação:

$$\mathbf{y} = \mathbf{X}^t \mathbf{p} + \mathbf{e} \quad (\text{Eq.13})$$

onde  $\mathbf{X}^t$  é a transposta da matriz  $\mathbf{X}$ . O vetor de erros,  $\mathbf{e}$ , é um vetor de dimensão  $n \times 1$ , que contém as diferenças entre os valores de referência  $\mathbf{y}$  e suas estimativas,  $\hat{\mathbf{y}}$ .

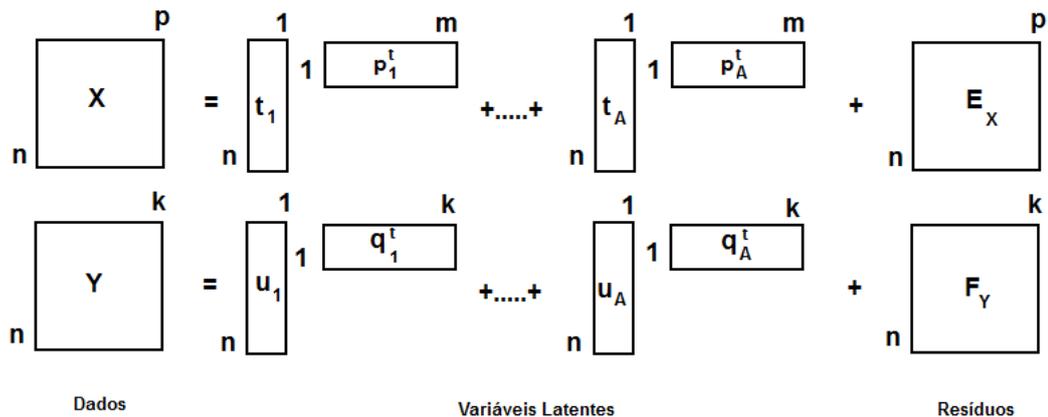
### 3.2.3 Regressão por mínimos quadrados parciais<sup>57-60</sup>

A modelagem mais simples de  $\mathbf{y}$  e  $\mathbf{X}$  é feita usando a MLR, que funciona bem desde que as variáveis em  $\mathbf{X}$  sejam poucas e não correlacionadas. No entanto, com as modernas medidas instrumentais, incluindo espectrômetros e cromatógrafos, as variáveis em  $\mathbf{X}$  tendem a ser numerosas, altamente correlacionadas e ruidosas. Uma forma de contornar esse problema é a utilização do método PLS, que nos permite avaliar esses sistemas mais complexos de maneira mais realística. O PLS envolve a decomposição dos dados espectrais da matriz  $\mathbf{X}$  em um produto de matrizes, levando-se em consideração as informações da matriz  $\mathbf{Y}$  (ou vetor  $\mathbf{y}$ ), de forma que haja a máxima covariância entre os dois conjuntos de dados. Assim, ele faz a decomposição em variáveis que não só modelam os sinais analíticos, mas também explicam a variância relacionada aos valores de referência. Cada componente do modelo recebe a denominação de Variável Latente (VL). A ideia principal aqui é que as matrizes de dados contêm informações redundantes e podem ser substancialmente reduzidas. Os termos reduzidos são mais fáceis de entender, têm mais estabilidade matemática e os resíduos devem conter somente ruído.

PLS é um método de calibração inversa, ou seja, a resposta analítica  $\mathbf{Y}$  (e.g. concentração) é uma função do sinal analítico  $\mathbf{X}$  (e.g. espectro). Assim, o modelo é calculado pela decomposição simultânea de  $\mathbf{X}$  e  $\mathbf{Y}$  em uma soma de  $A$  VL.  $\mathbf{X}$  é uma matriz  $n \times p$ , composta por  $n$  amostras e  $p$  variáveis, e  $\mathbf{y}$  é uma matriz  $n \times k$ , composta por  $n$  amostras e  $k$  propriedades ou concentrações a serem previstas. No caso mais simples e mais usual,  $k$  é igual a 1, sendo o modelo criado para prever apenas uma resposta analítica (PLS-1), mas a decomposição utilizando mais de uma propriedade simultaneamente também é utilizada (PLS-2). No método PLS a decomposição é feita segundo as equações 14 e 15 e a Fig. 13:

$$\mathbf{X} = \mathbf{TP}^t + \mathbf{E} \quad (\text{Eq. 14})$$

$$\mathbf{Y} = \mathbf{UQ}^t + \mathbf{F} \quad (\text{Eq. 15})$$



**Figura 13** Decomposição em VL das matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  para modelos PLS

Em que  $\mathbf{T}$  e  $\mathbf{U}$  são os escores,  $\mathbf{P}$  e  $\mathbf{Q}$  são os pesos (*loadings*) e  $\mathbf{E}$  e  $\mathbf{F}$  são os resíduos das matrizes  $\mathbf{X}$  e  $\mathbf{Y}$ , respectivamente. O vetor de coeficientes de regressão linear  $\mathbf{b}$  correlaciona os blocos  $\mathbf{X}$  e  $\mathbf{Y}$  de forma linear, para  $A$  VL, de acordo com a equação 16.

$$\mathbf{u}_A = \mathbf{b}_A \mathbf{t}_A \quad (\text{Eq. 16})$$

Os valores de  $\mathbf{b}_A$  são agrupados na matriz diagonal  $\mathbf{B}$ , que contém os coeficientes de regressão entre a matriz  $\mathbf{T}$ , dos escores de  $\mathbf{X}$ , e a matriz  $\mathbf{U}$  dos escores  $\mathbf{Y}$ . A melhor relação linear possível entre os escores desses dois blocos é obtida por meio de pequenas rotações das

VL dos blocos de  $\mathbf{X}$  e  $\mathbf{Y}$ . Os valores previstos para as novas amostras ( $\hat{\mathbf{Y}}$ ) podem ser então calculados de acordo com a equação 17, com base nos novos escores  $\mathbf{T}^*$ .

$$\hat{\mathbf{Y}} = \mathbf{T}^* \mathbf{B} \mathbf{q} \quad (\text{Eq.17})$$

Como as matrizes de escores e pesos dependem do número de VL  $A$ , a escolha desse parâmetro é de grande importância, para evitar sub ou sobre ajuste dos modelos. O primeiro caso implica a modelagem insuficiente dos dados para explicar toda informação sistemática, e o segundo, a inclusão de excesso de informação no modelo, que pode ser aleatória ou estar sistematicamente relacionada à presença de erros. De um modo geral, o risco de sobreajuste em calibração multivariada é muito maior do que o de subajustes. Para a escolha do número de VL, usa-se um método de validação cruzada, baseado num procedimento de reamostragem, e deve ser selecionado o número de VL correspondente ao modelo de menor raiz quadrada do erro quadrático médio de validação cruzada (RMSECV). Os principais métodos de validação cruzada são: 1) *leave-one-out*, que remove uma amostra de cada vez, sendo indicado apenas para modelos com pequeno número de amostras (aproximadamente 20); 2) blocos contíguos, que separa amostras em blocos de amostras sequenciais; 3) venezianas (*venetian blinds*), que separa amostras sistematicamente espaçadas; e 4) subconjuntos aleatórios, que separa, aleatoriamente, conjuntos de amostras.

### 3.2.4 Detecção de amostras anômalas – *Outliers*<sup>57,61</sup>

Desde que os parâmetros dos modelos tenham sido estimados, os coeficientes de regressão resultantes podem ser usados para predição de novas amostras. Entretanto, tais predições podem fornecer resultados de concentrações errôneas se obtidas a partir de amostras para as quais o modelo não se aplica. Tais amostras podem ser chamadas de “amostras anômalas” ou *outliers*. Testes estatísticos são aplicados para detectar *outliers* durante o desenvolvimento dos modelos de calibração e durante a validação dos modelos

construídos. *Outliers* incluem amostras com alta influência, amostras cujos valores de referência são inconsistentes com o modelo e/ou amostras cujos espectros apresentam alguma anomalia em relação ao conjunto de dados.

Erros e fenômenos inesperados são inevitáveis na construção de um modelo de calibração e podem ser causados por diversos motivos como, por exemplo, erros do operador, ruídos nos dados, problemas instrumentais e amostras com características muito distintas das demais. A presença de amostras anômalas no conjunto de calibração pode conduzir a modelos com baixa capacidade de previsão. Quando presentes no conjunto de validação, elas podem influenciar os resultados levando a estimativas que indicam que o modelo não é adequado ou que a sua capacidade é inferior à que poderia ser apresentada na ausência destas amostras anômalas. Assim, a identificação de anomalias é um passo importante para a otimização dos conjuntos de calibração e validação, sendo que, a exclusão dos *outliers* permite a construção de modelos mais eficientes, precisos e com melhor capacidade de previsão.

Os *outliers* na calibração são, usualmente, avaliados com base no *leverage* ( $h$ ) extremo, resíduos não modelados nos dados espectrais ( $X$ ) e resíduos não modelados nas variáveis dependentes ( $Y$ ). O *leverage* representa o quanto uma amostra está distante da média do conjunto de dados, ou seja, o “peso relativo” de uma amostra em relação às demais presentes em um mesmo conjunto. De acordo com a ASTM E1655-05<sup>57</sup>, amostras com  $h > (3A + 1)/n$ , onde  $n$  corresponde ao número de amostras da calibração, devem ser removidas do conjunto de calibração, e o modelo deve então ser reconstruído.

A identificação de anomalias em relação aos resíduos não modelados nos dados espectrais é obtida por comparação do desvio padrão residual total ( $s(\hat{\epsilon})$ ), definido como:

$$s(\hat{\epsilon})^2 = \frac{1}{nJ - J - A \max(n, J)} \sum_{i=1}^n \left( \sum_{j=1}^J (x_{ij} - \hat{x}_{ij})^2 \right) \quad (\text{Eq. 18})$$

em que  $J$  é o número de variáveis espectrais,  $A$  é o número de VL e  $n$  é o número de amostras da calibração, com o desvio padrão residual de uma amostra  $i$  ( $s(\hat{\epsilon}_i)$ ), que é calculado por:

$$s(\hat{\epsilon}_i)^2 = \frac{n}{nJ - J - A \max(n, J)} \sum_{j=1}^J (x_{i,j} - \hat{x}_{i,j})^2 \quad (\text{Eq. 19})$$

Se uma amostra apresentar  $s(\hat{\epsilon}_i) > 2s(\hat{\epsilon})$ , esta é removida do conjunto de calibração.

Com relação aos resíduos não modelados na variável dependente, os *outliers* são identificados através da comparação da raiz quadrada do erro médio da calibração (RMSEC) com o erro absoluto daquela amostra. Se a amostra apresentar erro absoluto maior do que (3xRMSEC), esta é definida como um *outlier*. O RMSEC é determinado como:

$$\text{RMSEC} = \frac{1}{n - A - 1} \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (\text{Eq. 20})$$

em que,  $y_i$  é o valor de referência e  $\hat{y}_i$  é o valor estimado.

É comum a contínua detecção de *outliers* após a reconstrução do modelo. Quando aplicações repetidas dos testes continuam a identificar *outliers*, um fenômeno conhecido como “*snowball*” (do inglês, efeito bola de neve) pode estar acontecendo. Este fenômeno indica que há algum problema com a estrutura dos dados espectrais. Para evitar este problema, não mais do que três rodadas de detecção de *outliers* (4 modelos) devem ser realizadas.

No conjunto de validação as amostras anômalas podem ser identificadas através do teste de resíduos Jacknife ( $J_{ei}$ )<sup>62</sup>. Através desse teste estatístico os *outliers* são detectados de acordo com a equação 21:

$$J_{ei} = r_i \sqrt{\frac{n - p - 1}{n - p - r_i^2}} \quad (\text{Eq. 21})$$

Na qual  $n$  é o número de pontos,  $p$  é o número de parâmetros do modelo,  $r_i = e_i/s_{ei}$  os resíduos padronizados. Este teste usa uma estimativa do desvio padrão independente do ponto. Então os resíduos são facilmente computados para cada ponto sem ter que ajustar as  $n$  regressões separadas, cada uma excluindo um ponto. Os resíduos Jackknife seguem a distribuição  $t$  com  $n-p-1$  graus de liberdade. Valores de  $J_{ei}$  maiores que valores de  $t_{\text{crítico}}$  são considerados *outliers* e eliminados até um total de  $2/9$  do total de dados<sup>62</sup>.

### 3.3 Validação de modelos multivariados e figuras de mérito

O objetivo da validação de um procedimento analítico é demonstrar que o mesmo é adequado para a sua finalidade. É por meio da validação que os laboratórios comprovam que os métodos de ensaio que executam conduzem a resultados confiáveis, rastreáveis, comparáveis e adequados à qualidade pretendida<sup>63</sup>. A validação de um método é um processo contínuo que começa no planejamento da estratégia analítica e continua ao longo de todo o seu desenvolvimento e transferência. Ela pode ser atestada através da determinação de parâmetros conhecidos como figuras de mérito (FdM), sendo os principais: veracidade, precisão, linearidade, sensibilidade, seletividade e faixa de trabalho<sup>64,65</sup>.

Estudos científicos que abordam a validação em calibração multivariada são recentes<sup>61,66-69</sup>. A grande maioria dos trabalhos e guias que tratam do tema ainda é referente à calibração univariada. Na área de alimentos e produtos agrícolas tanto os guias oficiais brasileiros<sup>70</sup> quanto os internacionais<sup>65,71-73</sup> ignoram completamente aspectos multivariados intrínsecos ao desenvolvimento de métodos NIRS. A necessidade de harmonização entre a legislação vigente e a demanda de métodos multivariados tem sido discutida recentemente<sup>61,67,69,74-76</sup>, levando à conclusão de que alguns conceitos como seletividade, sensibilidade e curvas de calibração precisam sofrer adaptações para se adequarem aos

modelos multivariados. A seguir são descritas as principais FdM e suas formas de aplicação nos modelos quimiométricos desenvolvidos.

### 3.3.1 Veracidade

A veracidade (termo que atualmente substitui o antigo significado de “exatidão”)<sup>77</sup> é avaliada através de um parâmetro específico não contemplado pelos guias tradicionais: RMSEP (estimado com as amostras de validação externa). O RMSEP, erro médio de predição da propriedade de interesse, é referido como uma estimativa da veracidade, entretanto, este é um parâmetro global que incorpora tanto erros sistemáticos quanto aleatórios. O RMSEP é calculado a partir da equação 22:

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

(Eq. 22)

em que,  $n_v$  é o número de amostras do conjunto de validação. É importante também, na avaliação da veracidade, que seja conhecida a faixa de erros relativos de predição para as amostras individuais<sup>76</sup>.

### 3.3.2 Precisão

A avaliação da precisão é realizada da mesma forma que nos modelos univariados. De acordo com a norma ASTM E 1655-05<sup>57</sup> a precisão dos valores estimados a partir de um modelo multivariado NIR deve ser calculada pela medida repetida dos espectros. Para a determinação da precisão em nível de repetitividade os espectros de, pelo menos, três amostras, com seis replicatas de cada, devem ser coletados no mesmo dia em três níveis de concentração, cobrindo a faixa útil do modelo. O cálculo é realizado utilizando-se a equação 23<sup>78</sup>. Para a determinação da precisão intermediária devem ser avaliadas as circunstâncias nas

quais o método será aplicado e assim proceder à variação de dia, número de analistas e laboratórios para a análise das replicatas.

$$\text{precisão} = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m (\hat{y}_{ij} - \hat{y}_i)^2}{n(m-1)}} \quad (\text{Eq. 23})$$

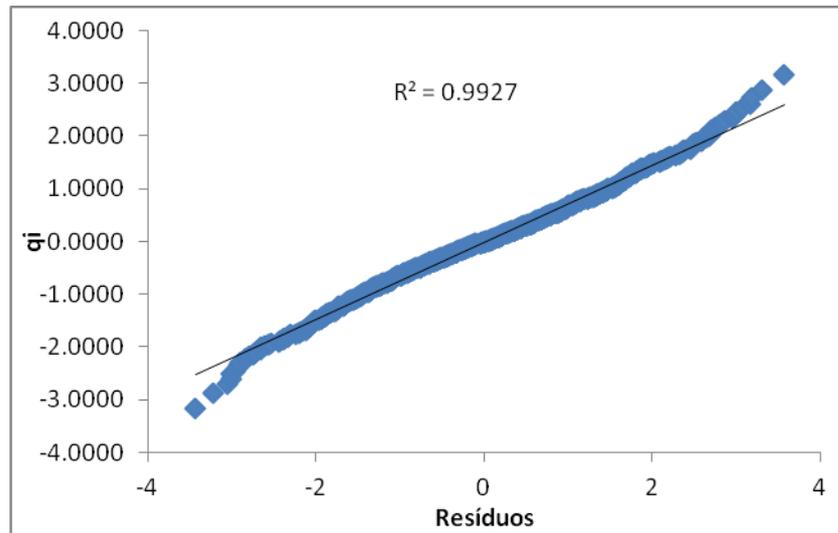
Em que  $n$  é o número de amostras utilizadas no cálculo da precisão,  $m$  é o número de replicatas de  $n$  e  $\hat{y}$  corresponde aos valores preditos pelo modelo.

### 3.3.3 Linearidade

De acordo com a definição do INMETRO<sup>77</sup>, linearidade é a habilidade de um método analítico em produzir resultados que sejam diretamente proporcionais à concentração do analito em amostras, em uma dada faixa de concentração. Normalmente, a linearidade pode ser inferida pela observação de parâmetros como o coeficiente de regressão “b”, o intercepto “a” e o coeficiente de correlação “r” de uma reta. No entanto, considera-se que um valor de  $r$  próximo de 1 não é suficiente para demonstrar a linearidade, pois deve-se também assegurar a aleatoriedade dos resíduos. Em calibração multivariada, deve-se analisar a curva dos valores de referência *versus* valores previstos (Anexos A e B) e/ou o gráfico dos resíduos para as amostras de calibração e validação, que podem indicar se a distribuição residual tem um comportamento aleatório<sup>78</sup>. No entanto, a observação da aleatoriedade dos resíduos através dos gráficos é subjetiva e deve ser confirmada através de testes estatísticos apropriados<sup>62</sup>.

A normalidade dos resíduos deve ser avaliada através do teste de Ryan-Joiner (RJ), no qual os resíduos ( $e_i$ ) são ordenados de forma crescente e plotados contra a porcentagem de pontos correspondentes ( $q_i$ ) a partir da distribuição normal padrão (gráfico quantil-quantil), conforme mostrado na Fig. 14. Caso os resíduos tenham comportamento normal, os dados devem se aproximar do ajuste de uma linha reta. Caso eles sigam uma distribuição alternativa,

o gráfico exibe algum grau de curvatura. O coeficiente de correlação ( $R_{eq}$ ) para os dados normalmente distribuídos é, então, comparado com os valores de  $R_{crítico}$  calculados, para a rejeição ou aceitação da hipótese nula de normalidade. Os resíduos podem ser considerados normalmente distribuídos se  $R_{eq} \geq R_{crítico}$ .



**Figura 14** Distribuição residual normal. Dados se aproximam do ajuste de uma linha reta

O teste Brown-Forsythe (BF), ou Levene modificado, deve ser usado na avaliação da homocedastidade dos resíduos, ou seja, a variação absoluta constante de todos os pontos da faixa de trabalho. Para este teste, os resíduos devem ser divididos em dois grupos de tamanho  $n_1$  e  $n_2$ , respectivamente, com relação aos valores da faixa de trabalho, ou seja, os resíduos correspondentes aos valores da metade inferior da faixa ficam no grupo 1 e os resíduos correspondentes à metade superior da faixa ficam no grupo 2. As medianas ( $\tilde{e}_1$  e  $\tilde{e}_2$ ) de cada grupo são, então, calculadas. Em seguida, os valores absolutos das diferenças entre os resíduos e as medianas de cada grupo  $d_{i1}=|e_i - \tilde{e}_1|$  e  $d_{i2}=|e_i - \tilde{e}_2|$  devem ser calculados. Por fim, a soma dos desvios quadrados de  $d_j$  ( $SSD_1$  e  $SSD_2$ ) de cada grupo devem ser estimadas. A estatística de  $t_L$  (t de Levene) é avaliada pela equação 24:

$$t_L = \frac{\bar{d}_1 - \bar{d}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) s_p^2}}$$

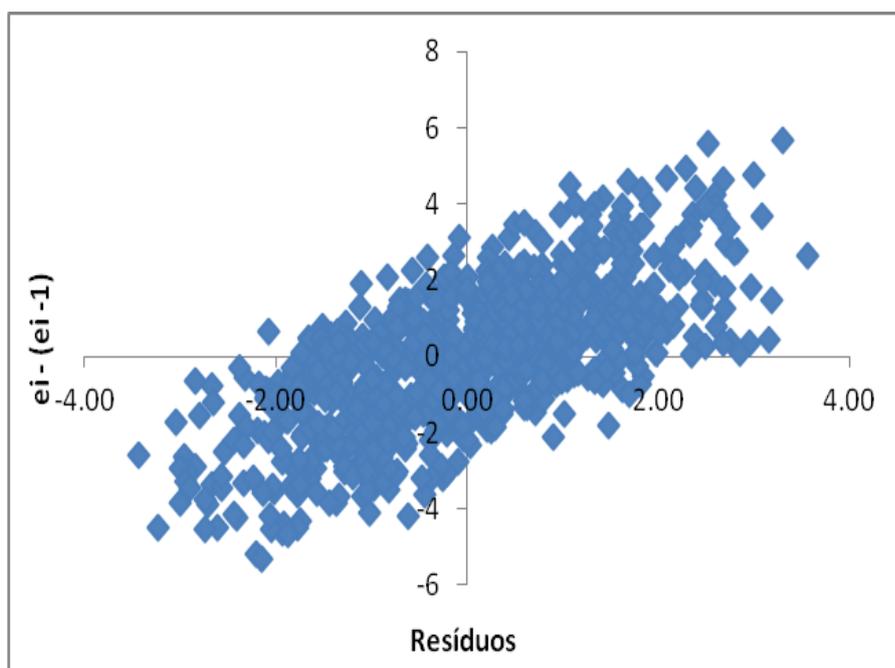
**Eq. 24**

na qual  $\bar{d}_j$  é a média de  $d_{ij}$  de cada grupo,  $n_j$  é o número de amostras de cada grupo e  $s_p^2 = \text{SSD}_1 + \text{SSD}_2 / (n_1 + n_2 - 2)$ . Os resíduos são considerados homocedásticos se  $t_L \leq t_{\text{critico}}$ .

Finalmente, a independência dos resíduos é avaliada com o teste de Durbin-Watson (DW). Este teste é definido pela equação 25:

$$d = \frac{\sum_{i=1}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad \text{Eq. 25}$$

A ausência de autocorrelação é verificada através da comparação entre os desvios padrão dos resíduos ordenados (denominador da Eq. 25) e das diferenças sucessivas desses mesmos resíduos (numerador dessa equação), conforme mostrado na Fig. 15. Para cada conjunto de dados, existem dois limites para  $d$  ( $d_L$  é o limite inferior e  $d_U$  é o limite superior). Se o valor de  $d$  estimado para um conjunto de dados estiver fora destes limites, o teste é inconclusivo ou existe autocorrelação (pode-se também calcular limites para estas duas situações). Os valores de DW variam de 0 a 4 com média 2. Se o valor calculado convergir para 2 significa que não há autocorrelação.



**Figura 15** Gráfico de correlação do teste de Durbin-Watson

### 3.3.4 Sensibilidade

A sensibilidade (SEN) é um parâmetro que demonstra a variação da resposta em função da concentração do analito<sup>77</sup>. A SEN pode ser considerada como a mudança na resposta líquida do instrumento dividida pela correspondente mudança no estímulo (a concentração do analito de interesse). Ela ainda é definida como a inclinação (coeficiente angular) em uma curva de calibração<sup>78</sup>. Suas dimensões são unidade de sinal x concentração<sup>-1</sup>. Para modelos de calibração multivariada inversa, como PLS, a SEN pode ser estimada por:

$$\hat{SEN} = \frac{1}{\|\mathbf{b}\|} \quad (\text{Eq. 26})$$

em que  $\mathbf{b}$  é o vetor dos coeficientes de regressão estimados pelo modelo PLS e " $\|\cdot\|$ " indica a norma euclidiana de um vetor.

### 3.3.5 Sensibilidade analítica

Como a SEN depende da técnica analítica empregada, uma FdM mais útil é a sensibilidade analítica ( $\gamma$ ). Este parâmetro, definido na equação 26, é a razão entre a SEN e o ruído instrumental ( $\delta x$ ), o qual é em NIRS estimado como o desvio padrão de várias medidas do acessório de medida vazio.

$$\gamma = \frac{\hat{SEN}}{\|\delta x\|} \quad (\text{Eq. 27})$$

O inverso desse parâmetro ( $\gamma^{-1}$ ) permite estabelecer a menor diferença de concentração entre amostras, a qual pode ser distinguida pelo método, considerando o ruído instrumental como única fonte significativa de erro. Este parâmetro também define o número de casas decimais e algarismos significativos que devem ser usados para expressar os resultados de predição.

### 3.3.6 Seletividade

A seletividade (SEL) é a capacidade de determinar uma espécie de interesse em amostras na presença de interferentes<sup>78</sup>, ou seja, é a medida do grau de sobreposição entre o sinal da espécie de interesse e os interferentes presentes na amostra. Logo, a SEL indica a parte do sinal que é perdida por essa sobreposição<sup>67</sup>. Para modelos de calibração multivariada esse parâmetro é definido como:

$$SEL_i = \frac{n\hat{s}_i}{\|x_i\|} \quad (\text{Eq. 28})$$

Em que  $n\hat{s}_i$  é o escalar NAS (Sinal Analítico Líquido - *Net Analyte Signal*)<sup>79</sup> estimado para a amostra 'i' e  $x_i$  o vetor de dados original.

Para métodos univariados o requisito de 100 % de SEL é usual, mas para métodos multivariados não existe interesse prático em estabelecer um valor mínimo. Diferentemente de métodos univariados, baixos valores de SEL podem ser obtidos mesmo para métodos multivariados exatos.

### 3.3.7 Faixa de trabalho

A faixa de trabalho corresponde ao intervalo de massas ou concentrações para o qual se pode afirmar que o método fornece um grau aceitável de exatidão e linearidade.

### 3.3.8 Viés (teste para erro sistemático – “*Bias*”)<sup>57,67</sup>

De acordo com a IUPAC, o viés é definido como a diferença entre a média populacional e o seu valor verdadeiro. Erros sistemáticos, ou seja, todos os componentes do erro que não são aleatórios correspondem à parcela fixa do viés em uma medida química. A existência desse tipo de erro afeta a precisão, a veracidade e a determinação dos intervalos de confiança. A norma E1655-05 da ASTM<sup>57</sup> sugere um teste *t* para avaliar se o viés incluso no

modelo é significativo, usando somente as amostras de validação. O viés médio pode ser calculado como:

$$\text{bias} = \frac{\sum_{i=1}^{n_v} (y_i - \hat{y}_i)}{n_v} \quad (\text{Eq. 29})$$

Em seguida, o desvio padrão dos erros de validação (SDV, “standard deviation of validation”) é determinado por:

$$\text{SDV} = \sqrt{\frac{\sum [(y - \hat{y}) - \text{bias}]^2}{n_v - 1}} \quad (\text{Eq. 30})$$

O valor de t é então obtido:

$$t = \frac{|\text{bias}| \sqrt{n_v}}{\text{SDV}} \quad (\text{Eq. 31})$$

Caso o valor de t calculado seja menor que o seu valor crítico para v graus de liberdade e 95% de confiança, o viés incluído no modelo pode ser considerado insignificante e desprezado.

### 3.3.9 Relação de Desempenho do Desvio, RPD (*Residual Prediction Deviation*)<sup>80</sup>

O RPD é uma FdM mais recente, utilizada para estimar a capacidade preditiva dos métodos e mais adequada para comparações em termos absolutos. A RPD é estimada para os conjuntos de calibração e validação de acordo com as equações 31 e 32, nas quais  $DP_{Cal}$  e  $DP_{Val}$  são os desvios padrões dos valores de referência dos conjuntos de calibração e validação, respectivamente.

$$RPD_{Cal} = \frac{DP_{Cal}}{RMSECV} \quad (\text{Eq. 32})$$

$$RPD_{Val} = \frac{DP_{Val}}{RMSEP} \quad (\text{Eq. 33})$$

Bons modelos de calibração devem possuir um valor de RPD acima de 2,4, enquanto valores entre 2,4 e 1,5 são considerados satisfatórios. Modelos com RPD menor que 1,5 não devem ser utilizados.

## **4 - OBJETIVOS**

---

O objetivo principal desse trabalho é o desenvolvimento de metodologias rápidas, limpas, confiáveis e de baixo custo para a caracterização da biomassa e do caldo do sorgo sacarino com o intuito de otimizar a pesquisa do PMGS da Embrapa Milho e Sorgo (Sete Lagoas-MG). Para alcançar tal objetivo foram seguidos os seguintes passos:

- A construção de modelos de calibração multivariada empregando o método de regressão por PLS para a determinação da composição (celulose, hemicelulose, lignina) da biomassa do sorgo sacarino utilizando espectroscopia no infravermelho próximo.
- A construção de modelos de calibração multivariada empregando o método de regressão por PLS para a determinação do RTE obtido a partir de cada cultivar de sorgo sacarino desenvolvido pelo PMGS da Embrapa utilizando espectroscopia no infravermelho próximo.
- A construção de modelos de calibração multivariada empregando o método de regressão por PLS para a determinação de parâmetros de qualidade (brix, leitura sacarimétrica, açúcares redutores, pol e pureza aparente) no caldo do sorgo sacarino utilizando espectroscopia no infravermelho próximo.
- A validação dos modelos de calibração multivariada pela determinação das FdM.

## **5 - PARTE EXPERIMENTAL**

---

### **5.1 Amostras**

Tanto as amostras de biomassa quanto as amostras de caldo analisadas neste estudo foram de diferentes genótipos de sorgo sacarino, dentre as quais algumas foram obtidas do PMGS e outras foram provenientes do Banco Ativo de Germoplasma (BAG), ambos pertencentes ao Centro Nacional de Pesquisas de Milho e Sorgo da Embrapa em Sete Lagoas, MG, Brasil.

#### **Biomassa**

Para a caracterização da biomassa foram conduzidos três ensaios nos campos experimentais da Embrapa em Sete Lagoas, entre 2010 e 2013, a partir dos quais três painéis de diversidade genética foram montados. Além disso, com o intuito de aumentar a variabilidade incorporada nos modelos de calibração, amostras oriundas de testes de maturação foram incluídas. Estas amostras foram coletadas desde o início do desenvolvimento da planta até o último estágio de maturação que ocorreu cerca de dez semanas depois.

Para todas as análises laboratoriais, amostras representativas dos colmos das plantas de sorgo, após a extração do caldo, foram pesadas frescas e secas em estufa de circulação forçada de ar por 72 h a 65 °C. As amostras secas foram moídas em moinho tipo Wiley até granulometria adequada (2 mm) e acondicionadas em potes de polietileno. Um total de 957 amostras, obtidas de mais de uma centena de híbridos e linhagens puras, foram usadas para construir os modelos.

## **Caldo**

Para a caracterização do caldo foi utilizado um total de 500 amostras. Parte delas são provenientes de 275 linhagens endogâmicas recombinantes (RIL) de sorgo sacarino, oriundas de duas linhagens contrastantes quanto à quantidade e qualidade de açúcares no caldo. Além disso, para assegurar uma alta variabilidade nos modelos de calibração, amostras derivadas de ensaios de maturação foram incluídas. Estas amostras foram coletadas desde o início do desenvolvimento da planta até o último estágio de maturação, que ocorreu cerca de dez semanas depois, garantindo teores de açúcares nos três níveis – baixo, moderado e alto.

Para todas as análises laboratoriais amostras representativas dos colmos das plantas foram desintegradas em desfibrador/homogeneizador. Em seguida, elas foram submetidas à prensa hidráulica, 250 kgf cm<sup>-2</sup> durante 1 minuto, para extração do caldo, segundo método descrito por Tanimoto<sup>81</sup>. O caldo extraído das amostras foi acondicionado em frascos de polietileno e enviado ao laboratório para análise. Uma vez que as amostras de caldo sofrem rápida fermentação e que não foi realizado qualquer teste de estabilidade, as mesmas foram analisadas no mesmo dia em que foram coletadas.

## **5.2 Reagentes**

Todos os reagentes utilizados neste trabalho foram de grau analítico e usados sem nenhuma purificação adicional.

### **Biomassa**

Para as análises da biomassa foram utilizados os seguintes reagentes: ácido sulfúrico (95-97% m/v), brometo de cetiltrimetilamônio – CTAB (99% m/m), dodecil sulfato de sódio (99% m/m), ácido etilenodiaminotetracético sal dissódico di-hidratado – EDTA (99-101%

m/m), tetraborato de sódio (borax) deca-hidratado (99-105% m/m), fosfato de sódio dibásico anidro (99% m/m), trietileno glicol e acetona (99,5% v/v).

## **Caldo**

Para as análises do caldo foram utilizados os seguintes reagentes: sacarose (99,9 % m/m), D-glicose anidra (99,7 % m/m), reagente clarificante à base de alumínio - Octapol®, solução de Fehling A (sulfato de cobre) e solução de Fehling B (tartarato de sódio e potássio e hidróxido de sódio).

## **5.3 Equipamentos**

Os espectros foram coletados em um espectrômetro Büchi NIRFlex N-500 FT-NIR (Flawil, Switzerland), equipado com acessório de reflectância difusa (biomassa) e de transreflectância (caldo).

## **Biomassa**

Os equipamentos usados para a avaliação da biomassa através das metodologias convencionais incluíram: um determinador de fibras Tecnal (Piracicaba, Brasil), modelo TE-149, com capacidade para 30 testes; um forno mufla Quimis (Diadema, Brasil), modelo Q318S, com rampas e patamares; uma balança analítica Mars Scales (North York, Canadá), modelo AUW220, com sensibilidade de  $\pm 0,1$  mg; uma estufa American Lab (San Francisco, EUA), modelo AL 100/150, com circulação e renovação de ar e um moinho tipo Wiley (Thomas Scientific, Swedesboro, EUA). Bolsas de filtro Ankom F57 (Ankom Technology, Macedon, EUA) e uma seladora de embalagem manual 40 cm modelo PCS400C também foram empregados para a determinação das fibras.

## **Caldo**

Os equipamentos usados para a avaliação da qualidade do caldo através das metodologias convencionais incluíram: refratômetro digital portátil Atago (Tóquio, Japão) com correção automática de temperatura modelo PAL-1 3810; polarímetro VIS/NIR com fonte de radiação tipo laser de diodo<sup>82</sup>; determinador de açúcares redutores (Redutec) Tecnal modelo TE-088 (Piracicaba, Brasil); prensa hidráulica para extração e esmagamento de cana Hidraseme modelo PHS 250 (Ribeirão Preto, SP); desfibrador/homogeneizador tipo betoneira IRBI (Araçatuba, SP).

### **5.4 Softwares**

O espectrômetro foi controlado e os dados foram adquiridos utilizando o *software* Büchi NIRWare Operator (Flawil, Switzerland). Os dados foram tratados utilizando-se o *software* MATLAB versão 7.13 (The MathWorks, Natick, MA, EUA). A rotina PLS utilizada é oriunda do PLS Toolbox versão 6.5 (Eigenvector Technologies, Manson, WA, EUA). Uma rotina própria, desenvolvida pelo grupo de pesquisa do professor Jez W. B. Braga, foi aplicada para a detecção de amostras anômalas<sup>75</sup>. Para a determinação da leitura sacarimétrica do caldo foi utilizado o *software* POL\_SAC<sup>82</sup>.

### **5.5 Procedimentos**

#### **Aquisição dos espectros NIR**

##### **Biomassa**

As amostras de biomassa de sorgo, secas e moídas, foram transferidas para placas de petri de vidro, de 100 mm de diâmetro interno, as quais foram utilizadas como célula de medida para a obtenção dos espectros NIR. Os espectros das amostras foram registrados em

triplicata na faixa de números de onda de 10.000 a 4.000  $\text{cm}^{-1}$  em passos de 4  $\text{cm}^{-1}$ , como a média de 32 varreduras, e à temperatura de  $25 \pm 2$  °C. Seis replicatas de amostras com três diferentes níveis de rendimento teórico de etanol - baixo, médio e alto - foram obtidas para a avaliação da repetibilidade. Estas replicatas também foram analisadas em três diferentes dias por um analista diferente para a estimativa da precisão intermediária. Dez replicatas espectrais da placa de petri vazia foram registradas, nas mesmas condições, com o objetivo de estimar o ruído instrumental.

### **Caldo**

As amostras de caldo de sorgo foram filtradas em algodão hidrófilo e transferidas para placas de petri de vidro de 100 mm de diâmetro interno, as quais foram utilizadas como célula de medida para a obtenção do espectro NIR. Um acessório metálico de transfectância foi então posicionado sobre a amostra para que o espectro fosse obtido (Fig. 16). Os espectros de transfectância das amostras foram registrados em triplicata na mesma faixa de número de onda e nas mesmas condições em que foram obtidos os espectros das amostras de biomassa. Seis replicatas das amostras com três diferentes níveis de açúcares, baixo, médio e alto, foram obtidas para a avaliação da repetibilidade. A precisão intermediária não foi avaliada neste caso, pois como as amostras não são estáveis, elas foram analisadas e descartadas no mesmo dia da coleta. Dez espectros de replicatas da placa de petri contendo apenas a placa de transfectância também foram obtidos para a estimativa do ruído experimental.

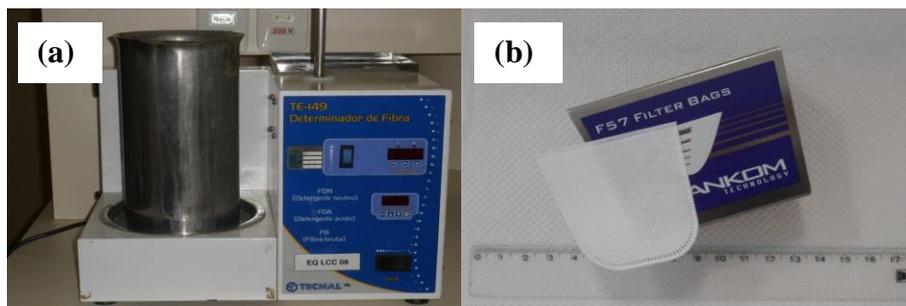


**Figura 16** Acessório de transfectância utilizado nas medidas dos espectros do caldo

## 5.6 Métodos de referência

### Biomassa

FDA, FDN e lignina em detergente ácido foram determinadas pelo método de extração sequencial desenvolvido por van Soest<sup>83</sup> e automatizado e simplificado pela Ankom Technology<sup>84</sup> (Fig. 17). Aproximadamente 500 mg de cada amostra foram colocados em bolsas de filtro, as quais foram seladas por aquecimento. Os filtros com amostras foram aquecidos em solução de detergente neutro. Nesse processo, o conteúdo celular foi solubilizado enquanto a parede celular (constituída de celulose, hemicelulose e lignina), insolúvel, foi separada por filtração. Os filtros com resíduos de FDN foram secos a 105°C e pesados para a determinação da porcentagem de FDN. Em seguida, eles foram submetidos a aquecimento com solução de detergente ácido. Nesse processo a hemicelulose foi solubilizada e o novo resíduo (consistindo quase que inteiramente de lignina e celulose) foi seco a 105°C e pesado para a determinação do teor de FDA. Para a estimativa da lignina, o resíduo de FDA foi submetido a uma solução de ácido sulfúrico 72% m/v durante 3h. Os filtros com os resíduos resultantes desse procedimento foram secos a 105 °C e pesados para a determinação da lignina. Em seguida, os resíduos foram submetidos a ignição à 500 °C por 4 h para a determinação das cinzas residuais. A quantidade de celulose foi determinada pela diferença entre FDA e lignina e o teor de hemicelulose foi encontrado como a diferença entre FDN e FDA. O rendimento teórico de etanol foi determinado através da equação 4.



**Figura 17** (a) Determinador de fibras Ankom e (b) bolsas de filtro utilizadas nas análises de FDA e FDN

## Caldo

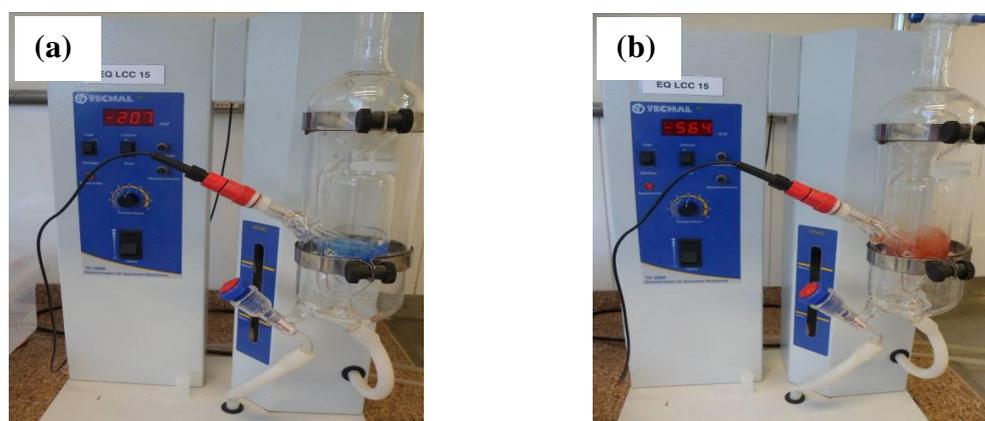
Para a determinação da leitura sacarimétrica ( $^{\circ}Z$ ) o caldo obtido a partir de cada amostra foi diluído com água deionizada na proporção de 1:1 (50 mL de caldo para 50 mL de água). Em seguida, essa solução foi misturada a aproximadamente 7,0 g de uma mistura clarificante à base de alumínio e filtrada em papel de filtro qualitativo. O filtrado foi então colocado em cubeta de caminho óptico conhecido e levado ao polarímetro (previamente calibrado com solução de sacarose a 15% m/v), para a determinação do grau de polarização em comprimento de onda de 589 nm (Fig. 18).



**Figura 18** Polarímetro utilizado na determinação da leitura sacarimétrica

Para a determinação dos teores de açúcares redutores (glicose e frutose) as amostras foram filtradas em algodão hidrófilo para a eliminação das partículas em suspensão. Em

seguida, elas foram diluídas com água deionizada na proporção de 5 mL do caldo para 50 mL total. Essa solução foi transferida para uma bureta e utilizada para titular um volume conhecido de reagente de Fehlling, previamente padronizado com solução de D-Glicose 0,25% m/v, mantido em ebulição constante no equipamento Redutec. O final da titulação foi acompanhado por diferença de potencial (Fig. 19) e o teor de açúcares redutores foi calculado através da equação 2.



**Figura 19** Determinação do teor de açúcares redutores em Redutec. (a) antes e (b) depois da titulação

O brix foi determinado automaticamente adicionando-se cerca de 0,5 mL de amostra na superfície do prisma do refratômetro digital (Fig. 20).



**Figura 20** Refratômetro digital para a determinação de brix

A determinação do pol e da pureza aparente do caldo foi feita com a utilização das equações 1 e 3, respectivamente.

## **6 - RESULTADOS E DISCUSSÃO**

---

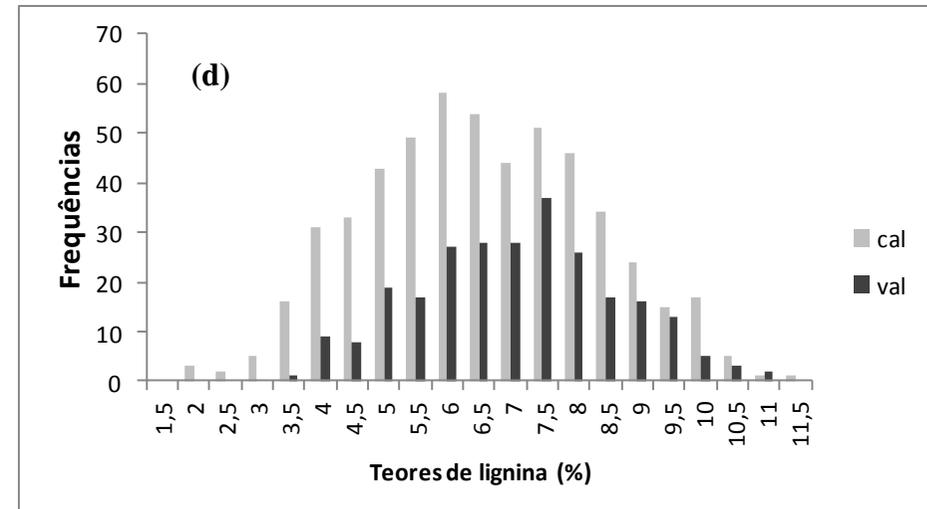
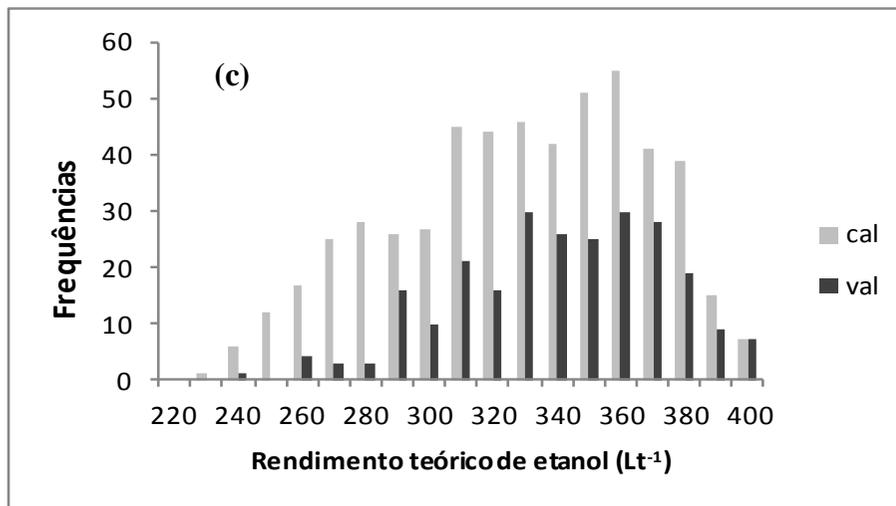
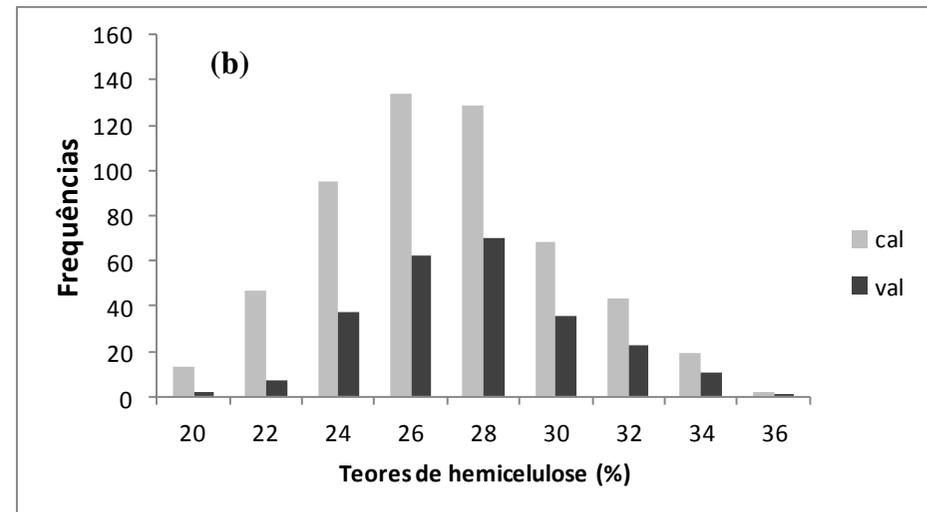
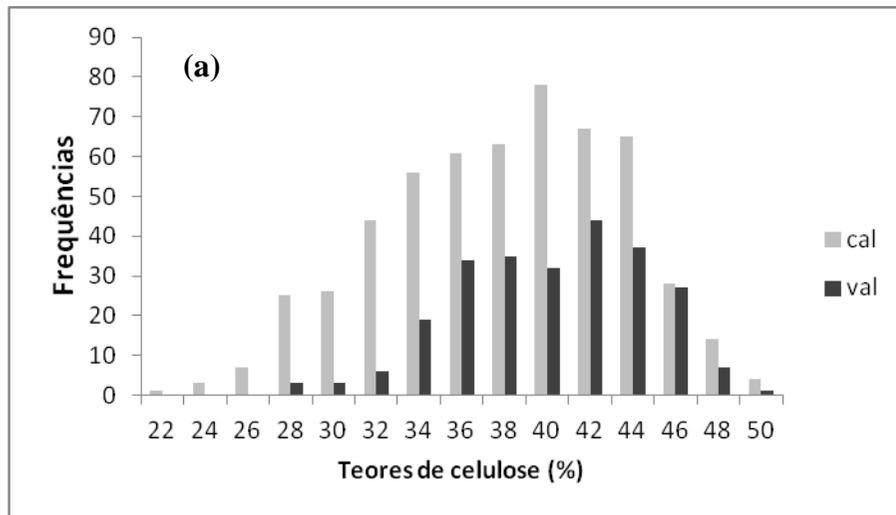
### **6.1 Resultados obtidos pelas metodologias de referência**

De acordo com a norma ASTM E1655<sup>57</sup> o conjunto de amostras utilizado para a construção dos modelos deve representar toda a variabilidade química e física normalmente encontrada na rotina de análises. Além disso, esta variabilidade deve estar distribuída uniformemente ao longo de ampla faixa de concentração, permitindo a construção de modelos com considerável faixa de trabalho.

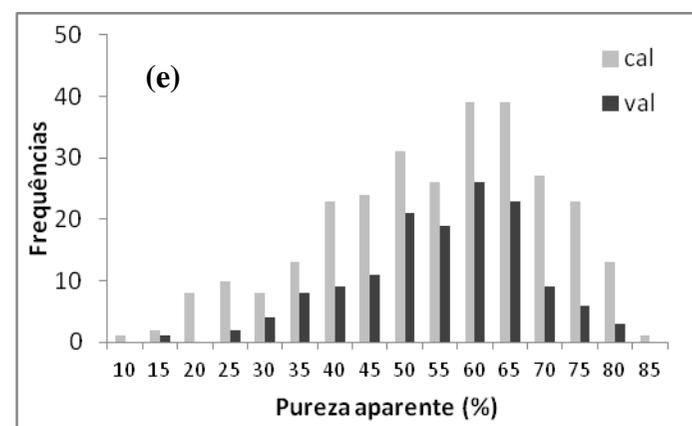
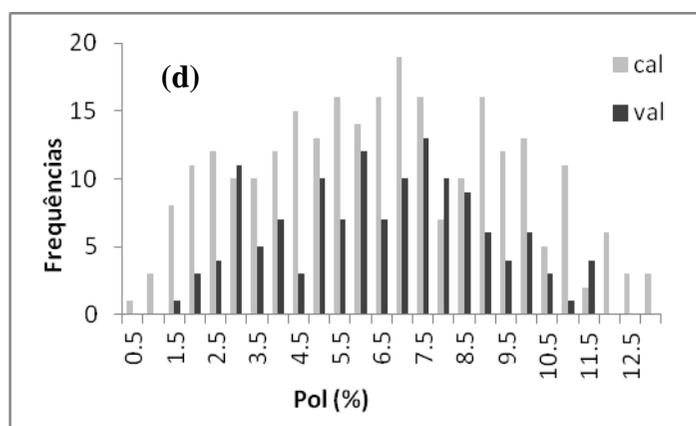
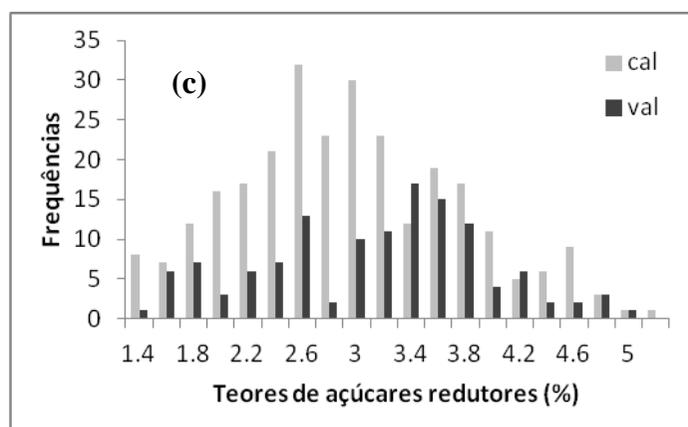
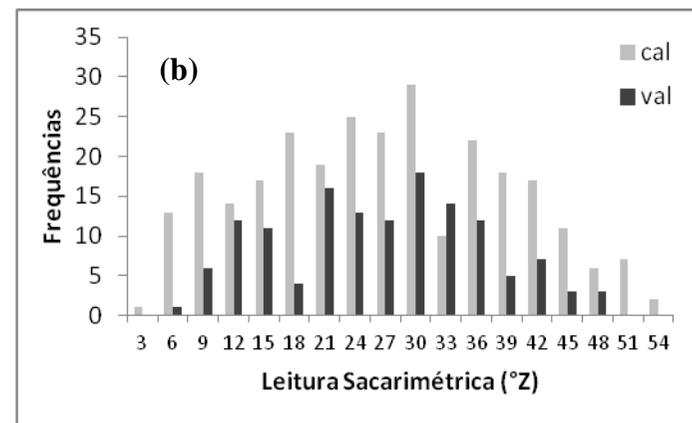
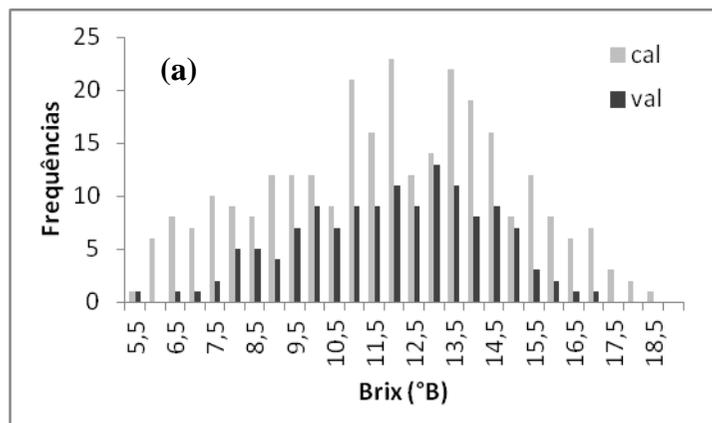
Nos histogramas das Fig. 21 e 22 estão representadas as faixas de concentração de cada parâmetro, obtidas pelas metodologias de referência, para as amostras utilizadas nos modelos bem como a distribuição dessas concentrações entre as amostras. Os resultados foram agrupados em dois conjuntos: calibração (cal) e validação (val). Pode-se inferir, a partir dos histogramas que os valores de todos os parâmetros estão distribuídos, homogeneamente, em ampla faixa de concentração e que o conjunto selecionado para validação tem faixa de concentração similar ao conjunto de calibração.

É preciso lembrar que nem todas essas amostras estão presentes nos modelos, uma vez que os mesmo passaram por etapas de eliminação de *outliers*.

As dimensões dos valores de referência também podem ser observadas pela avaliação das tabelas contidas nos anexos C a K.



**Figura 21** Teores dos parâmetros da biomassa determinados pelas metodologias de referência para (a) celulose, (b) hemicelulose, (c) RTE e (d) lignina

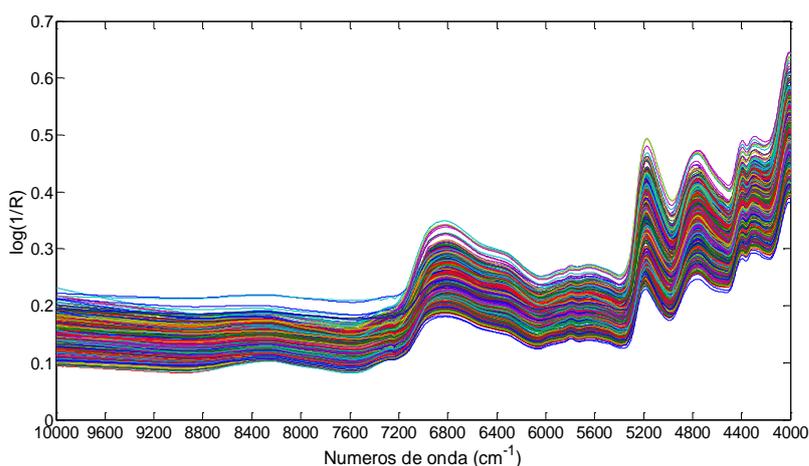


**Figura 22** Teores dos parâmetros do caldo determinados pelas metodologias de referência para (a) brix, (b) leitura sacarimétrica, (c) açúcares redutores, (d) pol e (e) pureza aparente

## 6.2 Espectros NIR e desenvolvimento dos modelos PLS

### Biomassa

Os parâmetros avaliados para a qualidade da biomassa foram celulose, hemicelulose, RTE e lignina. Celulose, polímero cristalino de glicose, hemicelulose, polímero amorfo de xilose e arabinose, e lignina, complexo polímero de alcoóis aromáticos, são os principais componentes da biomassa celulósica. Bandas de vibração associadas com estes componentes químicos<sup>42,85</sup> podem ser observadas na Fig. 23, que mostra o espectro das 957 amostras de biomassa de sorgo sacarino analisadas. Tais bandas serão especificamente atribuídas em seguida, na discussão sobre os coeficientes de regressão dos modelos PLS.

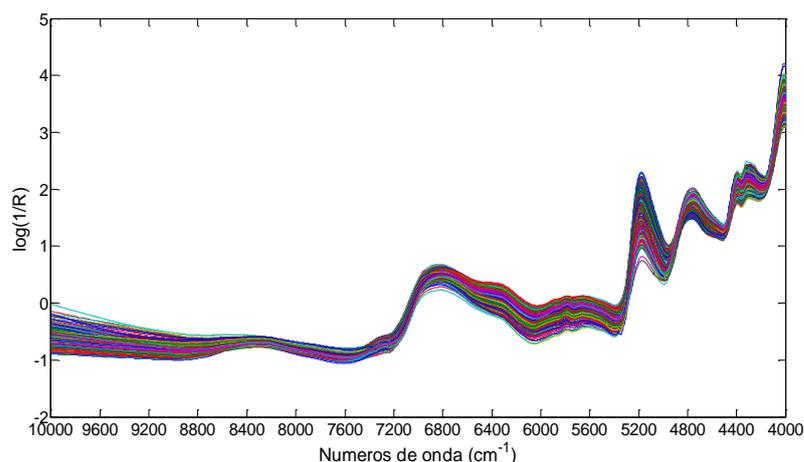


**Figura 23** Espectros NIR de 957 amostras da biomassa do sorgo sacarino

As regiões espectrais entre  $7.200-6.600\text{ cm}^{-1}$ ,  $6.000-5500\text{ cm}^{-1}$ ,  $5.400-4.600\text{ cm}^{-1}$  e  $4.600-4.000\text{ cm}^{-1}$  podem ser atribuídas ao primeiro sobretom de estiramento de O-H, primeiro sobretom de estiramento de C-H, bandas de combinação de O-H e bandas de combinação de C-H, respectivamente<sup>86</sup>.

As amostras foram divididas em 638 para o conjunto de calibração e 319 para o conjunto de validação usando o algoritmo Kennard-Stone<sup>87</sup> que garante, sistematicamente, a presença das amostras mais representativas no conjunto de calibração através da varredura

uniforme dos dados espectrais. Como a obtenção da informação espectral foi feita por reflectância difusa, foi necessária a utilização de um pré-processamento dos espectros para correção dos desvios de linha de base causados pelo espalhamento da radiação. Os desvios não lineares da linha de base foram corrigidos através do pré-processamento SNV (*standard normal variate*)<sup>88</sup>, como pode ser observado na Fig. 24. O método de pré-processamento alternativo, MSC (*multiple scatter correction*), também foi testado, mas forneceu resultados um pouco piores (valores de RMSEC e RMSEP mais elevados). Todos os modelos foram centrados na média e o número de VL foi escolhido por validação cruzada em blocos contínuos, considerando o menor RMSECV. Como as regiões espectrais entre 10.000 e 9.000  $\text{cm}^{-1}$  não apresentaram absorbâncias significantes, elas foram removidas dos modelos.



**Figura 24** Espectros NIR de 957 amostras da biomassa do sorgo sacarino após o pré-processamento SNV

Os coeficientes de regressão dos modelos PLS desenvolvidos para celulose, hemicelulose, lignina e RTE são mostrados na Fig. 25. Pela observação destas figuras é possível identificar os números de onda que apresentam os coeficientes mais positivos e, assim, mais contribuem para cada modelo.

Os coeficientes de regressão mais positivos e mais negativos para todos os quatro parâmetros são bastantes similares indicando que as regiões espectrais mais importantes para todos eles são praticamente as mesmas. Esse resultado era esperado, uma vez que tanto os

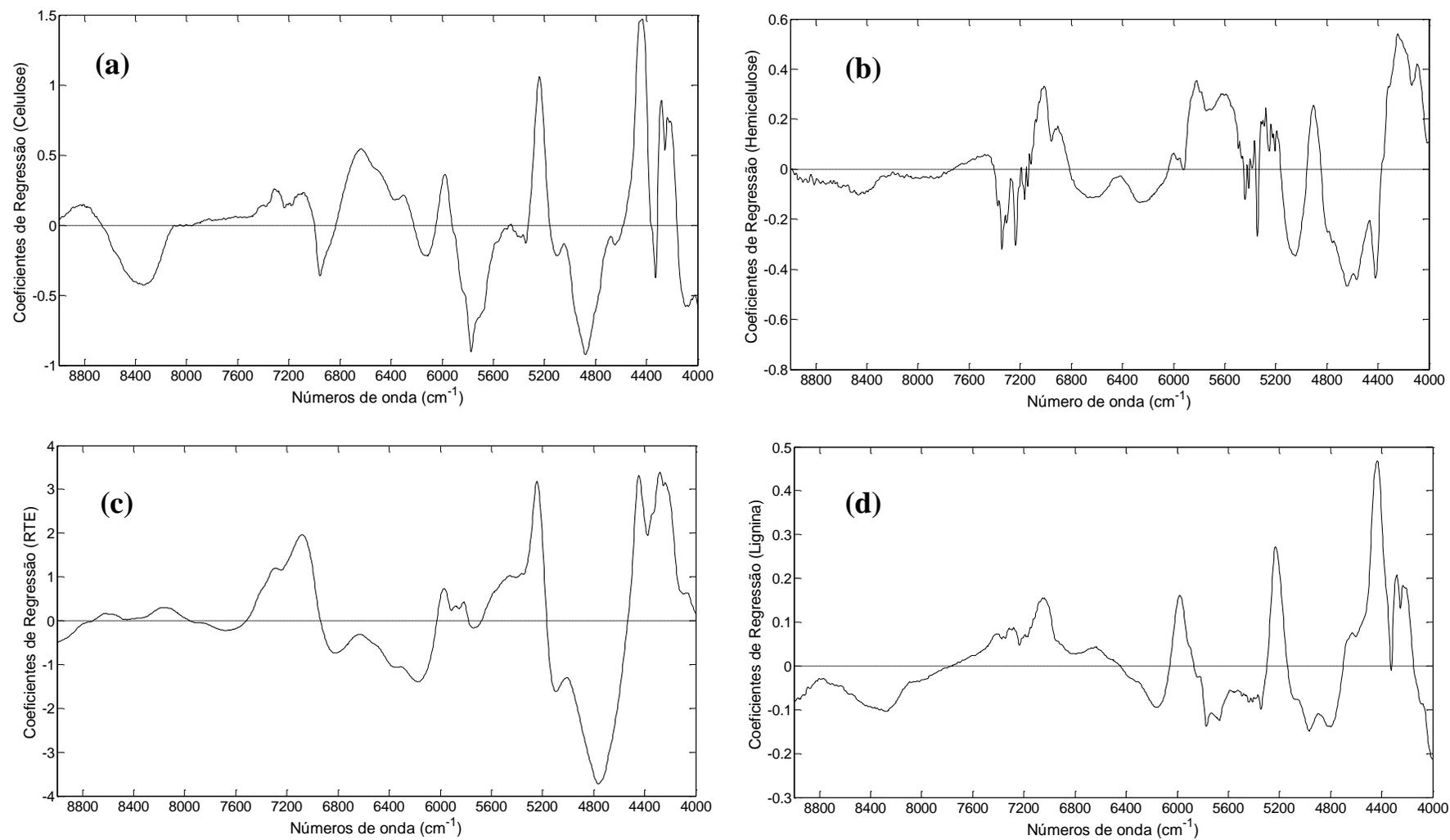
polissacarídeos (celulose e hemicelulose) quanto a lignina são moléculas orgânicas constituídas basicamente por ligações C-H e O-H, e o RTE é um parâmetro estimado a partir de celulose e hemicelulose.

Para a celulose (Fig. 25a) as variáveis espectrais mais importantes são uma banda larga entre 6.800-6.400  $\text{cm}^{-1}$  (relacionada ao primeiro sobretom do estiramento de O-H) e picos nas regiões de 5.970  $\text{cm}^{-1}$  (primeiro sobretom do estiramento de C-H), 5.210  $\text{cm}^{-1}$  (bandas de combinação do estiramento O-H e deformação de O-H), 4.415  $\text{cm}^{-1}$  (banda de combinação dos estiramentos de O-H e C-O), 4.290  $\text{cm}^{-1}$  (combinação do estiramento de C-H e deformação de  $\text{CH}_2$ ) e 4.230  $\text{cm}^{-1}$  (combinação de deformação de C-H e estiramento de C-H).

Para a hemicelulose (Fig.25b), os picos estão em cerca de 7.000  $\text{cm}^{-1}$  (primeiro sobretom do estiramento de O-H), 5.820  $\text{cm}^{-1}$  (primeiro sobretom do estiramento de C-H), 4.880 $\text{cm}^{-1}$ , 4.280 $\text{cm}^{-1}$  (combinação de estiramento de C-H e deformação de  $\text{CH}_2$ ) e 4.060 $\text{cm}^{-1}$ .

Para o RTE (Fig. 25c), os picos estão em cerca de 7.060  $\text{cm}^{-1}$ , 5.230  $\text{cm}^{-1}$  e 4.330  $\text{cm}^{-1}$ .

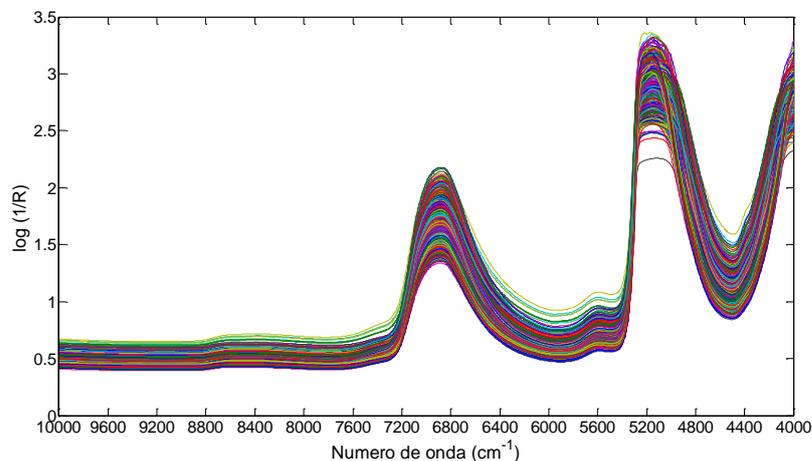
Para a lignina (Fig 25d), os picos estão em cerca de 7.040  $\text{cm}^{-1}$  (primeiro sobretom do estiramento de O-H), 5.940  $\text{cm}^{-1}$  (primeiro sobretom do estiramento de C-H de aromáticos), 5.230  $\text{cm}^{-1}$  (primeiro sobretom do estiramento de O-H de aromáticos), 4.415  $\text{cm}^{-1}$  (bandas de combinação de O-H e estiramento de C-O), 4290  $\text{cm}^{-1}$  e 4.210  $\text{cm}^{-1}$  <sup>42,85,86</sup>.



**Figura 25** Coeficientes de regressão dos modelos PLS para (a) celulose, (b) hemicelulose, (c) RTE e (d) lignina

## Caldo

Os parâmetros avaliados para a qualidade do caldo foram brix, leitura sacarimétrica, pol, pureza aparente e açúcares redutores. Os primeiros quatro parâmetros citados estão relacionados com os teores de sacarose, enquanto que o último se relaciona aos teores dos monossacarídeos glicose e frutose. Bandas de vibração associadas com estes componentes químicos do caldo podem ser observadas na Fig. 26, que mostra o espectro das 500 amostras de caldo de sorgo sacarino analisadas. Tais bandas serão especificamente atribuídas em seguida, na discussão sobre coeficientes de regressão dos modelos PLS.

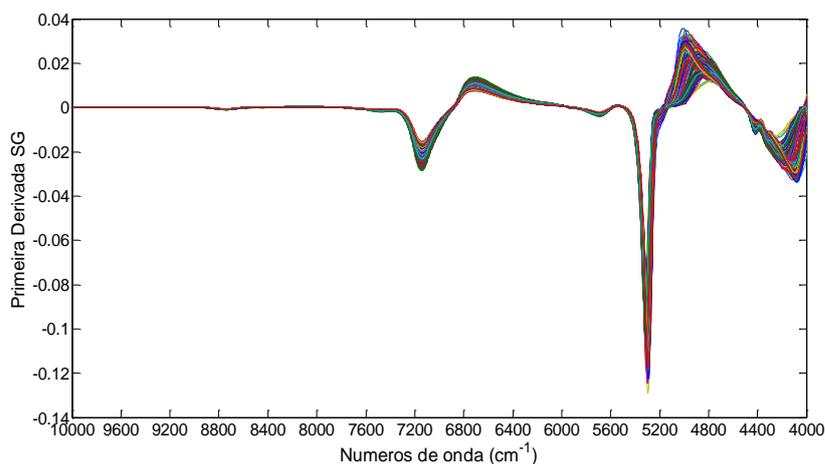


**Figura 26** Espectros NIR de 500 amostras do caldo do sorgo sacarino

Três regiões espectrais importantes ( $7.400-6.400\text{ cm}^{-1}$ ,  $5.800-4.600\text{ cm}^{-1}$  e  $4.600-4.000\text{ cm}^{-1}$ ) podem ser observadas nos espectros da Fig. 26. A região entre  $10.000$  e  $7.800\text{ cm}^{-1}$  é associada aos segundos e terceiros sobretons e é caracterizada por baixas intensidades de absorção e baixas razões sinal/ruído sendo, portanto, removida dos modelos.

As amostras foram divididas em 333 para o conjunto de calibração e 167 para o conjunto de validação usando o algoritmo Kennard-Stone<sup>87</sup>. A obtenção dos espectros foi feita por transfectância e, neste caso, a presença de desvios não lineares da linha de base foi corrigida através da aplicação de primeira derivada seguida de alisamento Savitzky-Golay (1ª DSG), com janela de 15 pontos<sup>89</sup> e polinômio de segunda ordem, como pode ser observado na

Fig. 27. O uso de derivadas reduz a sensibilidade das determinações analíticas, porém, elimina problemas relacionados à absorção de fundo devido à presença de pequenas partículas sólidas em amostras líquidas, que podem dispersar a luz provocando o deslocamento da linha de base. Como o uso de derivadas também aumenta o ruído, usa-se o alisamento simultaneamente.



**Figura 27** Espectros NIR de 500 amostras do caldo do sorgo sacarino após o pré-processamento 1ª DSG

Todos os modelos foram centrados na média e o número de VL foi escolhido por validação cruzada em blocos contínuos, considerando o menor RMSECV.

Os coeficientes de regressão dos modelos PLS desenvolvidos para brix, leitura sacarimétrica, açúcares redutores, pol e pureza aparente são mostrados na Fig. 28. Pela observação desta figura é possível notar que os números de onda que apresentam os coeficientes mais positivos e, assim, mais contribuem para cada modelo estão na região compreendida entre  $5.400\text{ cm}^{-1}$  e  $4.000\text{ cm}^{-1}$ .

Para brix (Fig. 28a), leitura sacarimétrica (Fig. 28b), pol (Fig. 28c) e pureza aparente (Fig. 28e), os coeficientes de regressão são bastante similares. Essa similaridade era esperada, uma vez que todos esses parâmetros estão relacionados aos teores de sacarose. Os números de onda mais positivos relacionados a esses parâmetros estão nas regiões próximas a  $4.350\text{ cm}^{-1}$  e entre  $4.500\text{--}4.700\text{ cm}^{-1}$ , que são o primeiro conjunto de bandas de combinação característico dos açúcares; a banda de absorção em  $4.405\text{ cm}^{-1}$  é associada à combinação de estiramento e

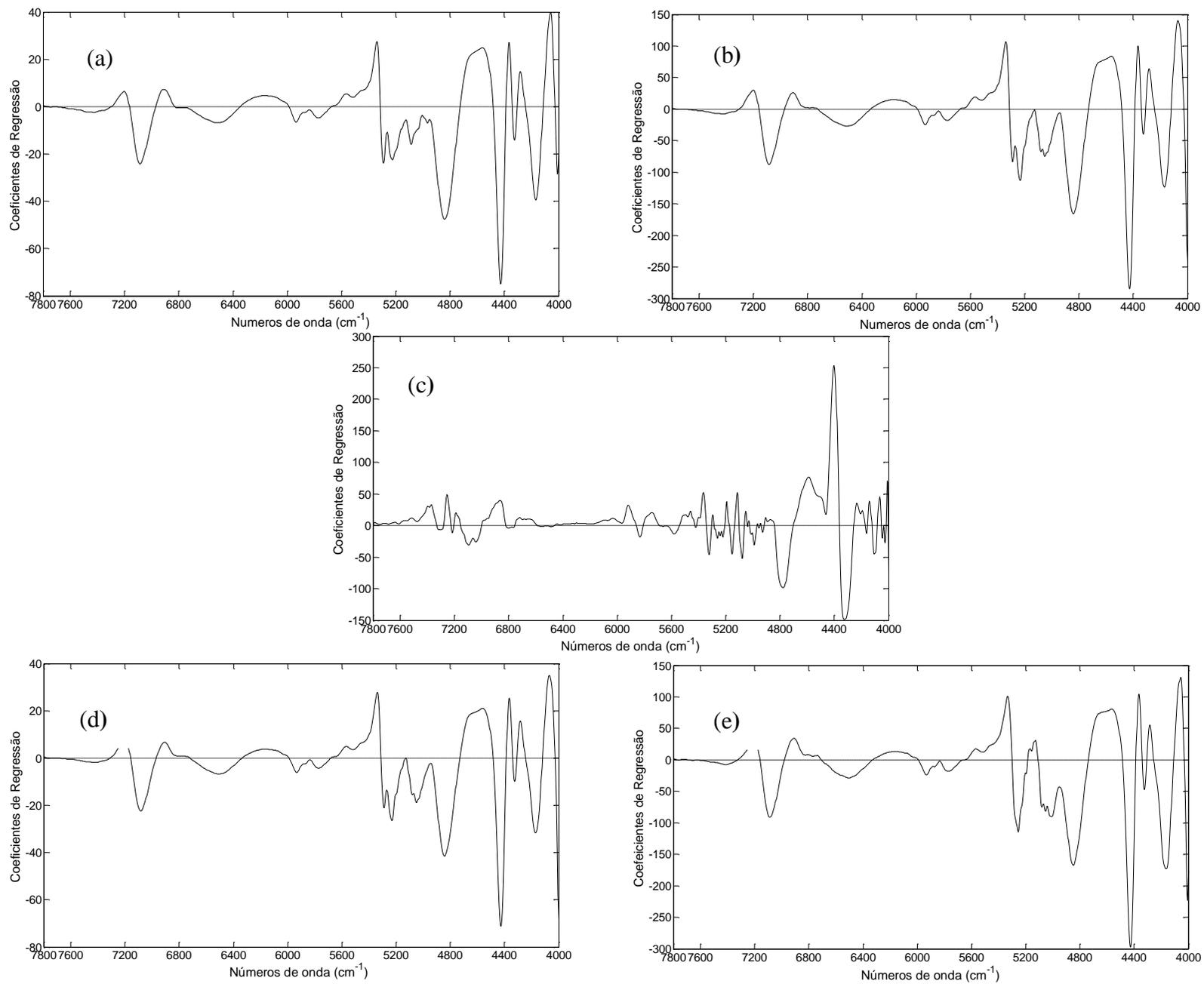
deformação de C-H. A região entre 6.024–5.682  $\text{cm}^{-1}$  é associada ao primeiro sobretom de modos de vibração de estiramento em grupos  $\text{CH}_3$  e  $\text{CH}_2$ . Bandas negativas podem ser observadas nas regiões de 7.140  $\text{cm}^{-1}$ , 5.300-4.890  $\text{cm}^{-1}$  e 4.450  $\text{cm}^{-1}$ . Estas bandas são associadas à água presente no caldo, mais especificamente a sobretons de estiramento de O-H<sup>50</sup>. Tais coeficientes negativos implicam em uma contribuição negativa para os modelos, ou seja, quanto maior a quantidade de água nas amostras, menor será a quantidade de açúcares.

A Fig. 28c mostra os coeficientes de regressão para os açúcares redutores. Os números de onda que contribuem positivamente para este modelo diferem um pouco dos demais. Destaca-se a região entre 4.300-4.700  $\text{cm}^{-1}$ , associada às bandas de combinações de açúcares.

### **6.3 Determinação do rendimento teórico de etanol**

Tendo em vista a necessidade de substituir a complexa metodologia de análise utilizada na determinação do RTE, baseada em hidrólise enzimática, fermentação e técnicas cromatográficas, nesse trabalho o modelo quimiométrico construído para a determinação desse parâmetro foi baseado na equação 4 desenvolvida por Zhao *et al*<sup>31</sup> com ligeira adaptação. Como não foram obtidas informações a respeito do rendimento de biomassa por hectare de cada cultivar utilizado no modelo, o termo BS (biomassa seca em  $\text{t há}^{-1}$ ) foi eliminado da equação 4 e os valores de RTE passaram, assim, a serem expressos em  $\text{L t}^{-1}$ .

A utilização dessa equação para a construção do modelo permitiu a determinação do RTE com base apenas nos teores dos polissacarídeos celulose e hemicelulose, reduzindo o tempo e o custo das análises.



**Figura 28** Coeficientes de regressão dos modelos PLS para (a) brix, (b) leitura sacarimétrica, (c) açúcares redutores, (d) pol e (e) pureza aparente

**Tabela 1** Resultados para a otimização dos modelos PLS da biomassa (resultados finais em negrito) através das etapas de detecção de *outliers*

	Celulose				Hemicelulose				RTE				Lignina			
	1°	2°	3°	4°	1°	2°	3°	4°	1°	2°	3°	4°	1°	2°	3°	4°
Ncal <sup>a</sup>	638	600	568	<b>542</b>	638	605	570	<b>550</b>	638	599	562	<b>534</b>	638	598	560	<b>532</b>
Nval <sup>b</sup>	319	319	319	<b>248</b>	319	319	319	<b>249</b>	319	319	319	<b>250</b>	319	319	319	<b>256</b>
NVL <sup>c</sup>	6	6	6	<b>6</b>	6	6	6	<b>6</b>	4	4	4	<b>4</b>	6	6	6	<b>6</b>
RMSEC <sup>d</sup>	2,7	1,7	1,6	<b>1,0</b>	3,5	2,0	1,8	<b>1,4</b>	18	14	12	<b>9</b>	1,4	1,0	0,9	<b>0,8</b>
RMSEP <sup>d</sup>	2,7	2,8	2,9	<b>1,5</b>	3,5	3,6	3,6	<b>1,7</b>	14	14	14	<b>12</b>	1,3	1,3	1,3	<b>0,8</b>

<sup>a</sup> Número de amostras da calibração, <sup>b</sup> Número de amostras da validação, <sup>c</sup> Número de variáveis latentes, <sup>d</sup> celulose, hemicelulose e lignina em %m/m. RTE em Lt<sup>-1</sup>

**Tabela 2** Resultados para a otimização dos modelos PLS do caldo (resultados finais em negrito) através das etapas de detecção de *outliers*

	Brix				Leitura Sacarimétrica				Açúcares redutores				Pol				Q			
	1°	2°	3°	4°	1°	2°	3°	4°	1°	2°	3°	4°	1°	2°	3°	4°	1°	2°	3°	4°
Ncal <sup>a</sup>	333	305	284	<b>284</b>	333	294	275	<b>275</b>	333	309	291	<b>276</b>	333	297	279	<b>264</b>	333	311	296	<b>288</b>
Nval <sup>b</sup>	167	167	167	<b>135</b>	167	167	167	<b>137</b>	167	167	167	<b>130</b>	167	167	167	<b>136</b>	167	167	167	<b>142</b>
NVL <sup>c</sup>	6	6	6	<b>6</b>	6	6	6	<b>6</b>	14	13	12	<b>12</b>	6	6	6	<b>6</b>	6	6	6	<b>6</b>
RMSEC <sup>d</sup>	0,7	0,6	0,5	<b>0,5</b>	4,7	3,6	3,3	<b>3,3</b>	0,6	0,4	0,4	<b>0,4</b>	1,2	0,9	0,8	<b>0,7</b>	8,8	7,1	6,6	<b>6,3</b>
RMSEP <sup>d</sup>	0,7	0,8	0,8	<b>0,4</b>	3,9	4,0	4,2	<b>3,2</b>	0,7	0,7	0,7	<b>0,4</b>	1,1	1,2	1,2	<b>0,8</b>	7,8	7,6	8,1	<b>5,8</b>

<sup>a</sup> Número de amostras da calibração, <sup>b</sup> Número de amostras da validação, <sup>c</sup> Número de variáveis latentes, <sup>d</sup> brix em °B, leitura sacarimétrica em °Z, açúcares redutores em %m/v, pol em %m/v e Q em %m/v

## 6.4 Otimização dos modelos através da detecção de amostras anômalas

Após os modelos PLS terem sido construídos, eles foram otimizados através da aplicação de procedimentos estatísticos para a detecção de *outliers*, baseados na literatura relevante<sup>57,61,62,75,76</sup>. Os *outliers* foram detectados com 95% de nível de confiança e os resultados para a biomassa e para o caldo estão sumarizados nas Tabelas 1 e 2, respectivamente. Amostras anômalas foram removidas até um limite de  $\frac{2}{9}$  (22%) do total do número de amostras, como estabelecido pelos guias brasileiros<sup>70</sup> e internacionais<sup>62,90</sup>. A otimização do conjunto de validação foi realizada após o final da otimização da calibração. Finalmente, não mais que três rodadas de detecção de *outliers* (quatro modelos) foram realizadas para que o efeito *snowball* fosse evitado.

Como pode ser visto nas Tabelas 1 e 2, o número de VL para os modelos da biomassa variaram de 4 a 6, enquanto quase todos os modelos do caldo usaram 6 VL, com exceção do modelo dos açúcares redutores para o qual foram usadas 12 VL. Esses números de VL podem ser considerados adequados tendo em vista a complexidade das duas matrizes (caldo e biomassa) e das metodologias utilizadas para a obtenção dos valores de referência. O número de VL mais alto utilizado para o modelo dos açúcares redutores pode ser atribuído à baixa concentração desses açúcares no caldo e à sobreposição das principais variáveis espectrais da sacarose àquelas relacionadas a esses açúcares. Além disso, a grande quantidade de água presentes nas amostras interfere consideravelmente nas medidas.

### Biomassa

Para a biomassa (Tabela 1), no modelo construído para a predição da celulose, 96 *outliers* foram detectados no conjunto de calibração (15%), dos quais 7, 26 e 63 foram baseados em alto *leverage*, altos resíduos espectrais (X) e altos resíduos na propriedade de interesse (Y), respectivamente; no conjunto de validação, 71 *outliers* (22%) foram removidos.

Para o modelo da hemicelulose, 88 *outliers* foram detectados no conjunto de calibração (14%), dos quais 6, 25 e 57 para alto *leverage*, altos resíduos em X e Y, respectivamente; no conjunto de validação 70 *outliers* (22%) foram removidos. Para o modelo do RTE, 104 *outliers* (16%) foram detectados no conjunto de calibração: 3, 28 e 73; e 69 (22%) no conjunto de validação. Para o modelo da lignina, 106 amostras anômalas (17%) foram detectadas no conjunto de calibração: 7, 23 e 76; e 63 (20%) no conjunto de validação.

### **Caldo**

Para o caldo (Tabela 2), no modelo construído para a predição do brix 49 *outliers* foram detectados no conjunto de calibração (15%), dos quais 8 por alto *leverage*, 12 por altos resíduos em X e 29 por altos resíduos em Y. No conjunto de validação, 32 (19%) amostras anômalas foram removidas. No modelo da leitura sacarimétrica 58 *outliers* foram eliminados do conjunto de calibração (17%), sendo 5, 15 e 38 por alto *leverage*, altos resíduos em X e altos resíduos em Y, respectivamente. No conjunto de validação foram detectados 30 *outliers* (18%). Para o modelo dos açúcares redutores 57 *outliers* foram identificados no conjunto de calibração (17%): 4, 5, 48; e 30 (22%) no conjunto de validação. Para o modelo do pol 69 amostras anômalas foram eliminadas do conjunto de calibração (21%), das quais 9 por alto *leverage*, 16 por altos resíduos em X e 44 por altos resíduos em Y; no conjunto de validação foram removidos 31 *outliers* (19%). Finalmente, para o modelo de pureza aparente 45 *outliers* foram detectados no conjunto de calibração (14%), entre eles 5 por alto *leverage*, 5 por altos resíduos em X e 35 por altos resíduos em Y. No conjunto de validação foram removidas 25 amostras anômalas (15%).

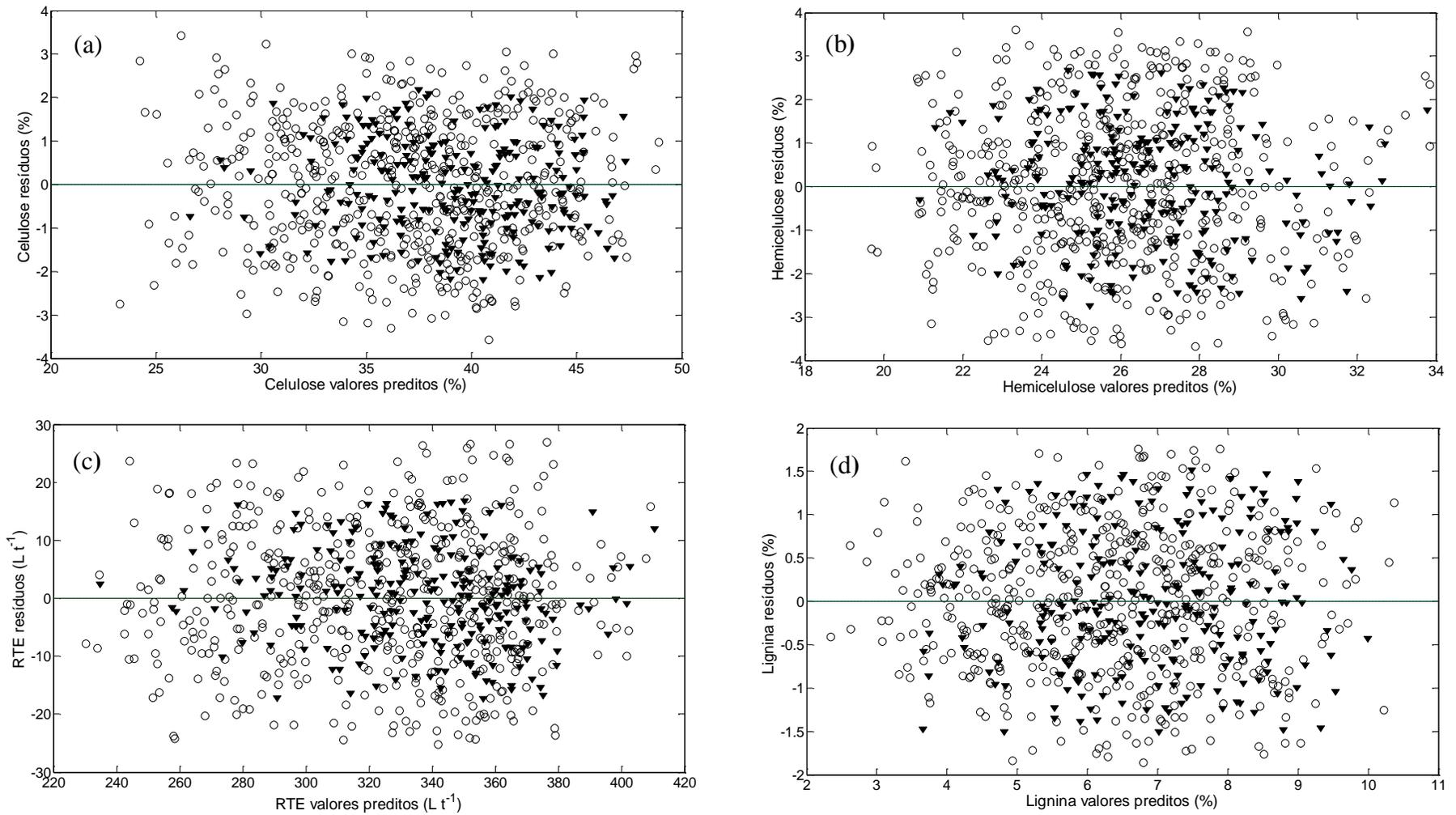
## **6.5 Validação analítica multivariada**

As Tabelas 3 e 4 resumem os parâmetros estimados na avaliação das FdM para os métodos propostos para a biomassa e para o caldo de sorgo sacarino, respectivamente. Nas subseções seguintes elas serão discutidas mais especificamente.

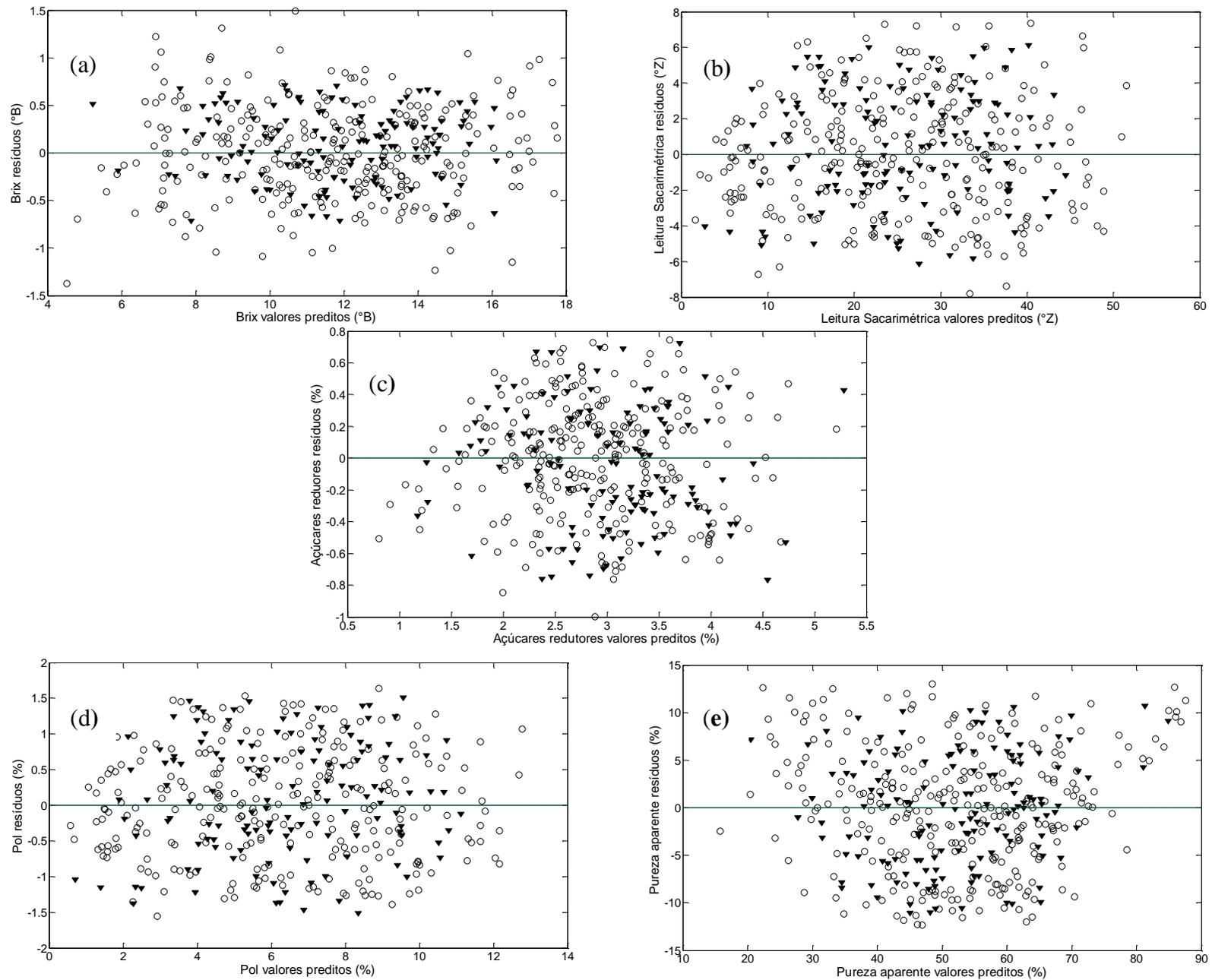
### **6.5.1 Linearidade**

A linearidade dos métodos desenvolvidos foi estimada através do ajuste dos valores preditos *versus* valores de referência (Anexos A e B). Qualitativamente, o gráfico dos resíduos deste ajuste para as amostras da calibração e da validação indica um comportamento linear se eles forem aleatoriamente distribuídos. Os resíduos para os modelos construídos para a biomassa e para o caldo de sorgo foram plotados em função dos valores preditos e são mostrados nas Fig. 29 e 30, respectivamente.

Como pode ser visto, nenhuma tendência sistemática foi observada nas distribuições residuais. Ainda assim, o comportamento aleatório desses resíduos foi confirmado através de testes estatísticos (Tabelas 3 e 4).



**Figura 29** Resíduos para as amostras da calibração (○) e validação (▼). Modelos PLS para determinação de (a) celulose, (b) hemicelulose, (c) RTE e (d) lignina em biomassa de sorgo



**Figura 30** Resíduos para as amostras da calibração (○) e validação (▼). Modelos PLS para determinação de (a) brix, (b) leitura sacarimétrica, (c) açúcares redutores, (d) Pol e (e) pureza aparente em caldo de sorgo

## **Biomassa**

Para a estimativa dos valores críticos dos testes estatísticos nos modelos da biomassa, foi considerado  $N=550$ , o maior número de amostras usado entre os métodos desenvolvidos, após a detecção de *outliers* (hemicelulose, Tabela 1).

Para o teste de RJ, as estimativas dos valores de  $R_{eq}$  estiveram todas abaixo do valor crítico de 0,9997 calculado a 95% de nível de confiança, mas elas estiveram todas acima de 0,9928, o valor crítico a 99%. A grande quantidade de amostras usadas (muito acima do normalmente utilizado em modelos univariados, para os quais este método foi inicialmente proposto) levou a elevados valores de  $R_{critico}$  e fez com que o teste de RJ se tornasse muito rigoroso. Assim, os resíduos para todos os modelos tiveram uma distribuição normal a 99% de nível de confiança.

No teste de BF os valores de  $t_L$  estimados estavam todos abaixo do  $t$  crítico de 1,96 (95% e infinitos graus de liberdade), confirmando então a homocedasticidade.

A independência dos resíduos (teste de DW) foi confirmada para todos os modelos, pois todos os valores estimados de  $d$  estavam dentro da faixa de aceitabilidade, entre  $dL = 1,81$  e  $dU = 2,19$  a 95% de nível de confiança.

## **Caldo**

Para a estimativa dos valores críticos para estes testes nos modelos do caldo, foi considerado  $N=288$ , o maior número de amostras usado entre os métodos desenvolvidos (pureza aparente, Tabela 2) após a etapa de detecção de *outliers*.

Para a normalidade dos resíduos, as estimativas dos valores de  $R_{eq}$  estavam todas abaixo do valor crítico de 0,9966, a 95% de nível de confiança, exceto para o modelo do brix que apresentou  $R_{eq}$  de 0,9988. No entanto, para todos os outros 4 modelos os valores de  $R_{eq}$  estavam acima de 0,9902, o valor crítico a 99%. Dessa forma, os resíduos para todos estes

modelos apresentaram uma distribuição normal, sendo que o brix a 95% de nível de confiança e os demais parâmetros a 99% de nível de confiança.

Na avaliação da homocedasticidade dos resíduos através do teste de BF os valores de  $t_L$  estimados estavam todos abaixo do  $t_{crítico}$  de 1,96 (95% e infinitos graus de liberdade), exceto para o modelo dos açúcares redutores, que apresentou valor de  $t_L$  de 2,03. Assim, os modelos de brix, leitura sacarimétrica, pol e pureza aparente foram considerados homocedásticos a 95% de confiança, e o modelo dos açúcares redutores a 98% ( $t_{crítico} = 2,33$  e infinitos graus de liberdade).

Por fim, a independência dos resíduos (teste de DW) foi confirmada, já que todos os valores de  $d$  estavam dentro da faixa de aceitabilidade, entre  $dL = 1,81$  e  $dU = 2,19$  a 95% de nível de confiança.

Uma discussão mais aprofundada sobre estes testes (cálculo dos valores críticos, por exemplo) pode ser encontrada no artigo de Souza & Junqueira<sup>62</sup> e nas referências nele citadas. Sendo que as premissas para a linearidade dos modelos foram verificadas, os parâmetros para os ajustes dos valores de referência *versus* valores preditos mostrados nas Tabelas 3 e 4 podem ser considerados válidos.

### **6.5.2 Veracidade, precisão e faixa de trabalho**

Os resultados de veracidade e precisão permitem atestar a exatidão dos métodos<sup>77</sup>.

A precisão expressa o grau de concordância entre os resultados de uma série de medidas realizadas para uma mesma amostra homogênea em condições determinadas<sup>78</sup> e é avaliada através dos desvios padrão relativos (DPR).

A veracidade pode ser avaliada através dos parâmetros absolutos RMSEC e RMSEP, mostrados nas Tabelas 1 e 2, para biomassa e caldo, respectivamente, e também pela observação dos erros relativos individuais (Tabelas 3 e 4 e Anexos C a K).

A avaliação dos erros relativos individuais foi realizada pela comparação dos mesmos com os limites estabelecidos pela normatização do MAPA<sup>91</sup>. De acordo com esta norma, prevista para calibração univariada, a veracidade é a concordância entre a média de um número suficientemente grande de resultados de um ensaio e o valor de referência aceito convencionalmente como verdadeiro através de ensaios de recuperação. No entanto, em calibração multivariada esta FdM pode ser avaliada pela comparação dos valores preditos pelo modelo com os valores obtidos pelos métodos de referência calculando-se os erros relativos individuais. Os limites de erros relativos utilizados para avaliação da veracidade foram de -20 a +10%.

## **Biomassa**

Os valores de RMSEP de 1,5%, 1,7%, 0,8% e 12 Lt<sup>-1</sup>; para celulose, hemicelulose, lignina e RTE, respectivamente (Tabela 1), indicam que os métodos fornecem resultados em bom acordo com os valores de referência.

Para celulose e RTE todos os valores de erros relativos estiveram dentro da faixa de  $\pm 10\%$ . Para hemicelulose, apenas 8 das 249 (3%) amostras da validação apresentaram erros fora da faixa de  $\pm 10\%$ , mas para lignina 69 das 256 (27%) das amostras apresentaram erros maiores do que os limites. A veracidade mais baixa desse último método é atribuída à faixa analítica menor para lignina. Os histogramas dos erros relativos das previsões dos parâmetros avaliados para a biomassa (amostras da validação) são mostrados na Fig. 31.

Para os parâmetros da biomassa, a precisão foi calculada nos níveis de repetibilidade e de precisão intermediária. Para todos os métodos, os valores de DPR variaram entre 0,5 e 1,5 % para repetibilidade e entre 0,5 e 1,6 % para precisão intermediária (Tabela 3). Todos esses resultados estão de acordo com os guias brasileiros que prescrevem aceitabilidade máxima para o DPR de 4% para repetibilidade e 10% para precisão intermediária<sup>91</sup>.

Considerando os estudos de linearidade e exatidão, as faixas de trabalho dos métodos desenvolvidos para análise da biomassa foram estabelecidas (Tabela 3): 21,4 a 49,1 % para celulose, 18,4 a 34,8 % para hemicelulose, 1,8 a 11,5 % para lignina e 221 a 412 L t<sup>-1</sup> para RTE.

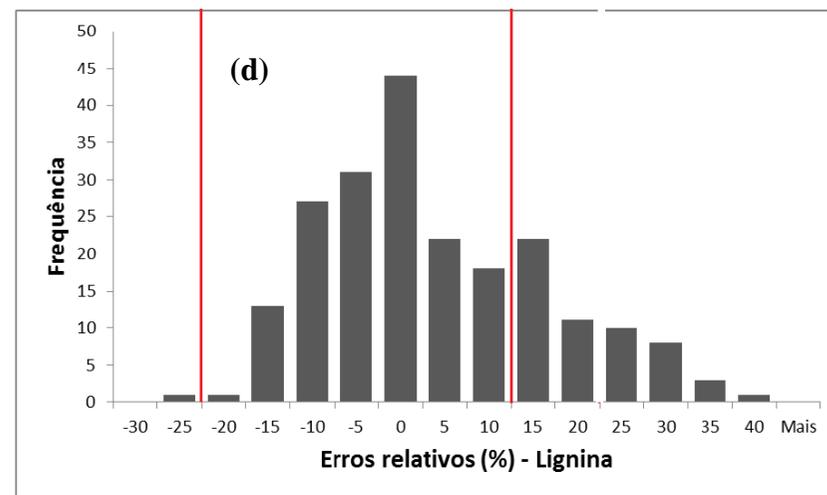
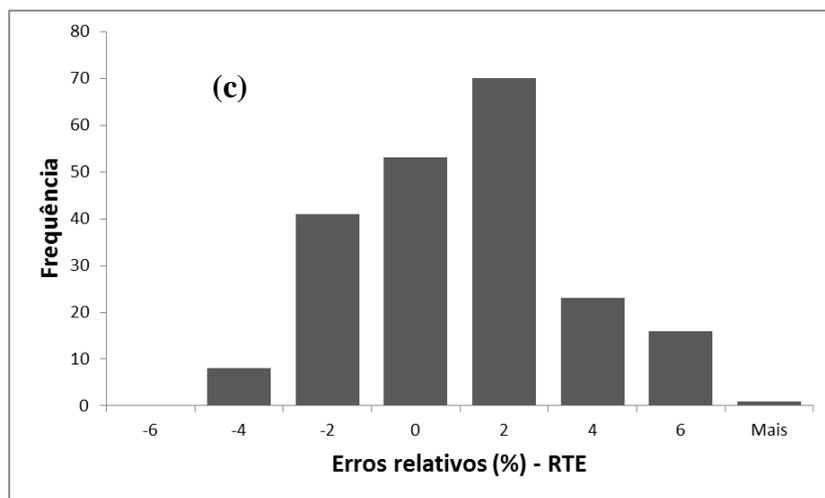
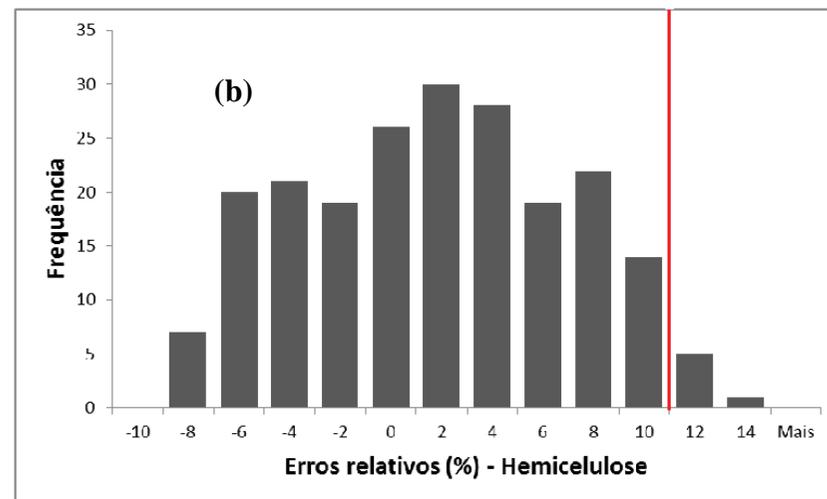
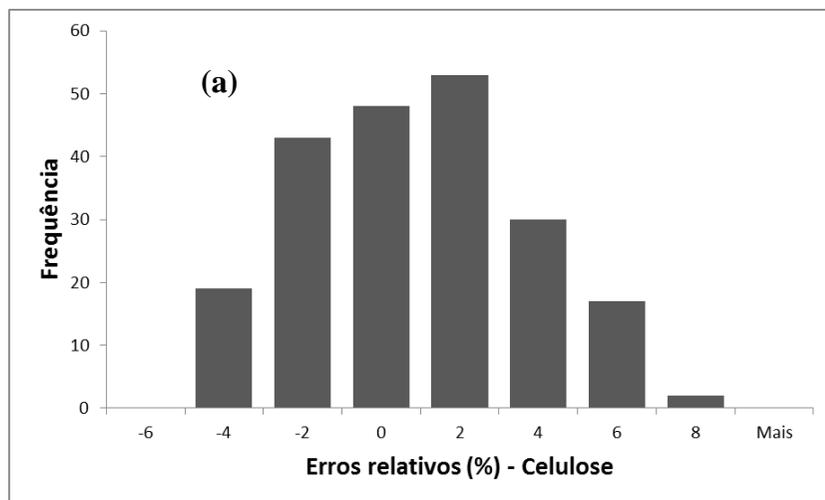
## **Caldo**

Os valores de RMSEP de 0,4 °B, 3,2 °Z, 0,4 %, 0,8 % e 5,8% para brix, leitura sacarimétrica, açúcares redutores, pol e pureza aparente, respectivamente (Tabela 2), indicam que os métodos fornecem resultados em bom acordo com os valores de referência. Com exceção do modelo do brix, cujos erros relativos (entre -8,3 e 11,0 %) estão mais próximos dos limites de aceitabilidade dos guias brasileiros (entre -20 e +10%), todos os demais parâmetros apresentaram faixas de erros próximas de ± 30%. Para leitura sacarimétrica, 49 das 137 (36%) amostras de validação apresentaram erros acima dos limites. Para açúcares redutores, 34 das 130 (26%) amostras de validação ultrapassaram a faixa de aceitabilidade. Para o modelo do pol 48 das 136 amostras (32%) apresentaram erros superiores aos limites. Finalmente, para pureza aparente 26 % (37 das 142) das amostras ficaram fora dos limites aceitáveis. A veracidade mais baixa para os métodos desenvolvidos para o caldo pode ser atribuída à maior complexidade das análises de referência e à baixa estabilidade das amostras. Os histogramas dos erros relativos das previsões dos parâmetros avaliados para o caldo (amostras da validação) são mostrados na Fig. 32.

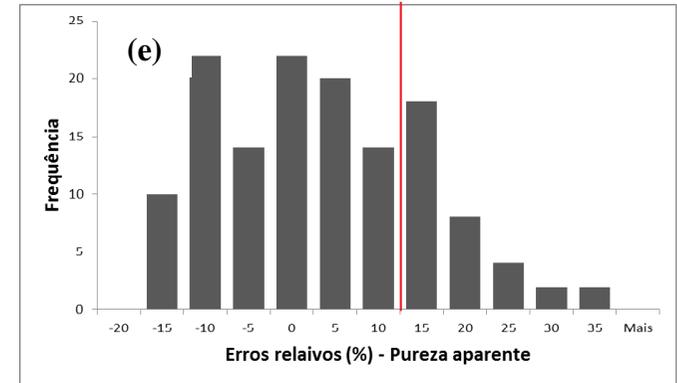
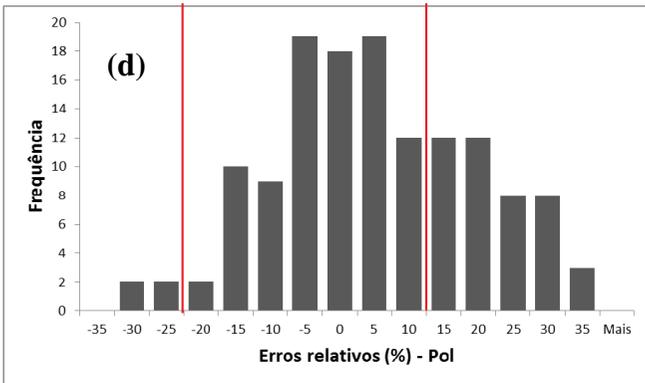
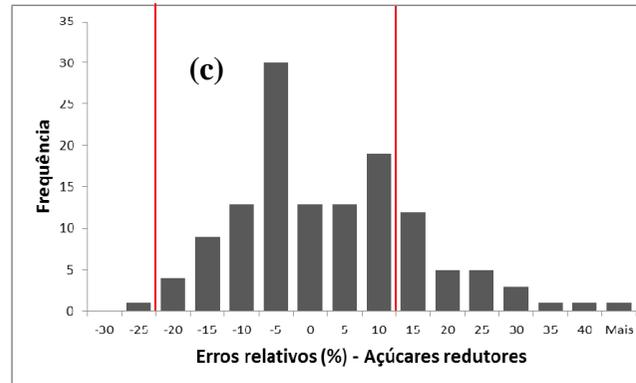
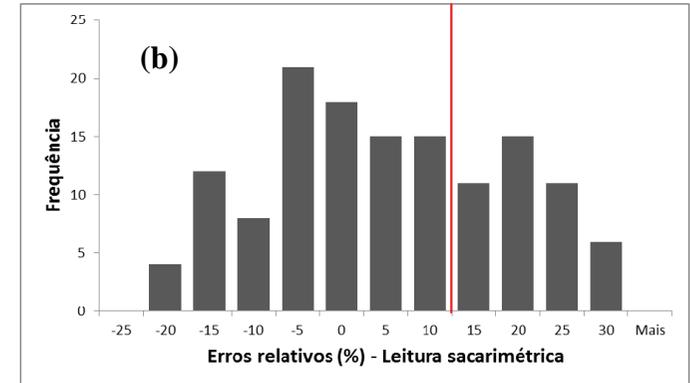
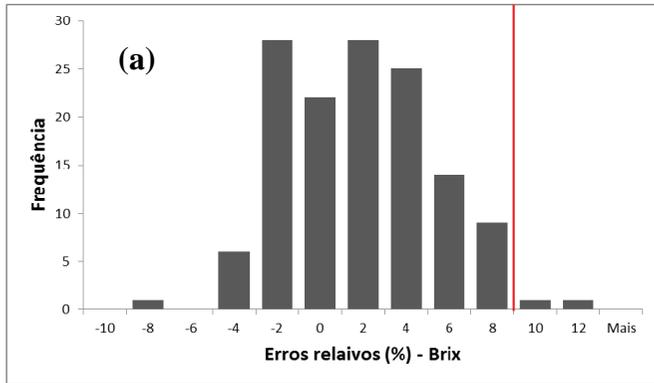
Para os parâmetros do caldo, a precisão foi calculada apenas no nível de repetibilidade. Os valores de DPR foram de 2,8 % para brix, 4,1 % para pureza aparente, 4,3% para leitura sacarimétrica, 4,7 % para pol e 14,3 % para açúcares redutores (Tabela 4). Apenas o resultado para o brix está de acordo com o limite estabelecido pelos guias brasileiros, que prescrevem aceitabilidade máxima para o DPR de repetibilidade de 4%<sup>91</sup>, no entanto os resultados para pureza aparente, leitura sacarimétrica e pol estão próximos a este

limite. O DPR mais elevado dos parâmetros avaliados para o caldo, possivelmente, se deve, em parte, ao uso alternativo do acessório de transflectância, em vez de cubetas, para avaliação das amostras. Esse acessório faz com que a sensibilidade da análise seja reduzida, o que se reflete também na precisão. Além disso, o caldo apresenta certa quantidade de amido que decanta rapidamente no fundo da placa, o que não permite grande reprodutibilidade na aquisição dos espectros. O valor ainda mais alto de DPR para os açúcares redutores pode ser atribuído à faixa analítica mais baixa deste parâmetro (1,2 a 5,2% - Tabela 4).

Considerando os estudos de linearidade e exatidão, as faixas de trabalho dos métodos desenvolvidos para análise do caldo foram estabelecidas (Tabela 4): 5,5 a 18,1 °B para brix, 1,1 a 53,2 °Z para leitura sacarimétrica, 1,2 a 5,2 % para açúcares redutores, 0,3 a 13,0 % para pol e 9,8 a 83,0 % para pureza aparente do caldo.



**Figura 31** Erros relativos das amostras de validação para os parâmetros da biomassa (a) celulose, (b) hemicelulose, (c) RTE e (d) lignina. Limites em vermelho (-20 a +10%)



**Figura 32** Erros relativos das amostras de validação para os parâmetros do caldo (a) brix, (b) leitura sacarimétrica, (c) açúcares redutores, (d) pol e (e) pureza aparente. Limites em vermelho (-20 a +10 %)

**Tabela 3** Parâmetros para a avaliação das principais FdM dos métodos NIRS desenvolvidos para a determinação de celulose, hemicelulose, RTE e lignina em biomassa de sorgo sacarino

FdM	Parâmetro	Celulose	Hemicelulose	RTE	Lignina
		Valor			
Veracidade	Erros relativos(min/max)	-5,3/6,5 %	-9,8/12,2%	-5,6/6,1 %	-28,8/37,6%
Precisão	DPR repetibilidade <sup>a</sup>	1,5%	0,5%	0,5 %	1,5%
	DPR precisão intermediária <sup>a</sup>	1,5%	0,7%	0,5 %	1,6%
Linearidade	R <sub>eq</sub> (teste RJ)	0,9964	0,9952	0,9980	0,9956
	t <sub>L</sub> (teste BF)	0,2400	1,9489	0,4736	1,1487
	d (teste DW)	1,95	1,93	1,96	1,87
	Inclinação <sup>b</sup>	0,93±0,01	0,70±0,02	0,91±0,01	0,79±0,02
	Intercepto <sup>b</sup>	2,69±0,43	7,67±0,61	28,37±4,19	1,29±0,13
	Coef. de correlação (r) <sup>b</sup>	0,9631	0,8390	0,9553	0,8920
Faixa de trabalho		21,4 a 49,1 %	18,4 a 34,8 %	221 a 412 L t <sup>-1</sup>	1,8 a 11,5%
Seletividade		0,13	0,11	0,26	0,14
Sensibilidade		0,07	0,15	0,02	0,27
Sensibilidade Analítica (γ)		1,7 % <sup>-1</sup>	3,4 % <sup>-1</sup>	0,5 t L <sup>-1</sup>	6,1 % <sup>-1</sup>
Viés	Viés ± SDV <sup>c</sup>	-0,031±1,079 %	0,106±1,354 %	0,049±8,182 Lt <sup>-1</sup>	0,362±0,778 %
	t estimado	0,46	1,23	0,09	0,36
RPD	RPD calibração	3,7	1,8	3,3	2,2
	RPD validação	2,7	1,7	2,7	2,0

<sup>a</sup> Resultados para três amostras em três níveis diferentes, <sup>b</sup> Valores para o ajuste linear das amostras de calibração, <sup>c</sup> Desvio Padrão dos erros de validação (*Standard Deviation of Validation Errors*)

**Tabela 4** Parâmetros para a avaliação das principais FdM dos métodos NIRS desenvolvidos para a determinação de brix, leitura sacarimétrica, açúcares redutores, pol e pureza aparente (Q) nas amostras de caldo de sorgo sacarino

FdM	Parâmetro	Brix	Leitura sacarimétrica	Açúcares redutores	Pol	Q
		Valor				
Veracidade	Erros relativos(min/max)	-8,3/11,0 %	-24,6/28,5 %	-26,6/37,0 %	-32,9/32,8 %	-19,7/34,3 %
Precisão	DPR repetibilidade <sup>a</sup>	2,8 %	4,3 %	14,3 %	4,7 %	4,1 %
Linearidade	R <sub>eq</sub> (teste RJ)	0,9988	0,9944	0,9952	0,9916	0,9905
	t <sub>L</sub> (teste BF)	1,2691	1,2976	2,0346	1,0385	1,7111
	d (teste DW)	2,14	2,11	1,94	2,16	2,11
	Inclinação <sup>b</sup>	0,97±0,01	0,93±0,02	0,82±0,03	0,94±0,02	0,84±0,02
	Intercepto <sup>b</sup>	0,32±0,12	1,83±0,46	0,52±0,08	0,36±0,11	8,54±1,30
	Coef, de correlação (r) <sup>b</sup>	0,9861	0,9631	0,9044	0,9683	0,9145
Faixa de trabalho		5,5 a 18,1 °B	1,1 a 53,2 °Z	1,2 a 5,2 %	0,3 a 13,0 %	9,8 a 83,0 %
Seletividade		0,08	0,08	0,01	0,08	0,08
Sensibilidade		0,02	0,01	0,01	0,02	0,01
Sensibilidade Analítica (γ)		2,7 °B <sup>-1</sup>	1,2 °Z <sup>-1</sup>	1,2% <sup>-1</sup>	2,7 % <sup>-1</sup>	1,2% <sup>-1</sup>
γ <sup>-1</sup>		0,4 °B	0,8 °Z	0,8 %	0,4 %	0,8 %
Viés	Viés ± SDV <sup>c</sup>	0,080±0,371 °B	0,418±3,170 °Z	-0,051±0,371 %	0,108±0,797 %	-0,086±5,811 %
	t estimado	2,50	1,54	1,58	1,58	0,18
RPD	RPD calibração	5,7	3,5	2,3	3,8	2,4
	RPD validação	6,2	3,1	2,3	3,0	2,1

<sup>a</sup> Resultados para três amostras em três níveis diferentes, <sup>b</sup> Valores para o ajuste linear das amostras de calibração, <sup>c</sup> Desvio Padrão dos erros de validação (*Standard Deviation of Validation Errors*)

### **6.5.3 Seletividade e sensibilidade analítica**

A SEL e a SEN foram estimadas com base no NAS. Em relação à SEN, os valores deste parâmetro não são apropriados para comparação com outros métodos. Dessa maneira, a sensibilidade analítica ( $\gamma$ ) também foi calculada a partir do ruído instrumental estimado. O inverso de  $\gamma$  fornece uma estimativa da diferença de concentração mínima que pode ser distinguida pelo método, independente da técnica analítica utilizada e considerando o ruído instrumental aleatório como a única fonte de erro.

#### **Biomassa**

No desenvolvimento dos modelos para biomassa, os valores estimados de SEL indicam que 13, 11, 14 e 26 % do sinal analítico foram usados para predição de celulose, hemicelulose, lignina e RTE, respectivamente.

O ruído instrumental de 0,042 foi estimado a partir dos desvios padrão dos espectros de dez replicatas da placa vazia e foi usado para o cálculo das sensibilidades analíticas ( $\gamma$ ) (Tabela 3).

#### **Caldo**

No desenvolvimento dos modelos para o caldo, os valores estimados de SEL indicam que 8 % do sinal analítico foi usado para predição de brix, leitura sacarimétrica, pol e pureza aparente, enquanto que apenas 1 % do sinal foi usado para a predição de açúcares redutores.

O ruído instrumental de 0,0083 foi estimado a partir dos desvios padrão dos espectros de dez replicatas da placa contendo apenas o acessório de transfectância e foi usado para o cálculo das sensibilidades analíticas ( $\gamma$ ) (Tabela 4).

#### 6.5.4 Viés e RPD

O viés foi calculado apenas com as amostras da validação<sup>57</sup> e as estimativas estão mostradas nas Tabelas 3 e 4.

De acordo com a literatura<sup>80</sup>, bons modelos de calibração devem ter valores de RPD maiores que 2,4, enquanto que valores entre 2,4 e 1,5 são considerados aceitáveis. Modelos com valores de RPD inferiores a 1,5 não são considerados adequados.

#### **Biomassa**

Para a biomassa, todos os valores de  $t$  calculados para a estimativa do viés foram menores que o valor de  $t_{\text{crítico}}$  (1,96, com infinitos graus de liberdade e 95% de nível de confiança), confirmando a ausência de erros sistemáticos nos modelos de predição.

Como pode ser observado na Tabela 3, os valores de RPD para celulose e RTE foram considerados bons, enquanto que para hemicelulose e lignina foram considerados satisfatórios.

#### **Caldo**

Para o caldo, o valor de  $t$  calculado (2,50) para o parâmetro brix (Tabela 4) foi menor que o valor de  $t_{\text{crítico}}$  apenas a 99 % de nível de confiança (2,58 com infinitos graus de liberdade), enquanto que os valores de  $t$  calculados para os demais parâmetros foram inferiores ao valor de  $t_{\text{crítico}}$  a 95% de nível de confiança (1,96 com infinitos graus de liberdade). Assim, a confirmação da ausência de erros sistemáticos para o modelo de brix se deu com 99% de nível de confiança, enquanto que para pol, leitura sacarimétrica, açúcares redutores e pureza aparente observa-se a ausência de erros sistemáticos com 95% de confiança.

Como mostrado na Tabela 4, os valores de RPD para brix, leitura sacarimétrica e pol foram considerados bons, enquanto que para açúcares redutores e pureza aparente foram considerados satisfatórios.

### **6.5.5 Cartas de Controle**

Após os métodos NIRS terem sido validados e implementados para a análise de biomassa, suas estabilidades foram monitoradas através de cartas de controle.

#### **Biomassa**

De abril a setembro de 2013, 50 amostras de biomassa de sorgo sacarino foram aleatoriamente selecionadas, duas por semana, e analisadas por NIRS e pelos métodos de referência. As cartas de controle construídas com estes resultados são mostradas na Fig. 33. Os limites indicados nessas cartas são baseados na normatização brasileira<sup>70</sup>, a qual estabelece que, no monitoramento de métodos quantitativos para análises agrícolas e de alimentos, dois terços das amostras devem ter erros relativos entre  $\pm 15\%$  (limites de alerta adotados nas cartas de controle), enquanto um terço pode apresentar erros máximos permissíveis de  $\pm 33\%$  (limites de ação). Como pode ser visto nesta figura, apenas duas amostras mostraram erros relativos entre os limites de alerta e de ação para os modelos da celulose e da hemicelulose (Fig. 33a-b), enquanto nenhuma amostra apresentou erro acima do limite de alerta para RTE (Fig. 33d). Os piores resultados observados foram para o modelo da lignina (Fig. 33c), no qual 15 das 50 amostras (30%) mostraram erros entre os limites de alerta e de ação. Apesar disso, este último resultado ainda está dentro dos limites de aceitação e todos os modelos foram considerados satisfatoriamente estáveis.

Para complementar a análise das cartas de controle desenvolvidas, a distribuição aleatória dos erros foi checada com os mesmo testes estatísticos previamente usados para a avaliação da linearidade. Uma vez que um comportamento aleatório dos resíduos das

previsões destas cartas indica a estabilidade dos modelos, justifica-se o emprego dos mesmos testes usados anteriormente para verificar a linearidade. Os parâmetros estimados são mostrados na Tabela 5.

**Tabela 5** Parâmetros dos testes estatísticos usados na avaliação das cartas de controle desenvolvidas para o monitoramento dos modelos NIRS

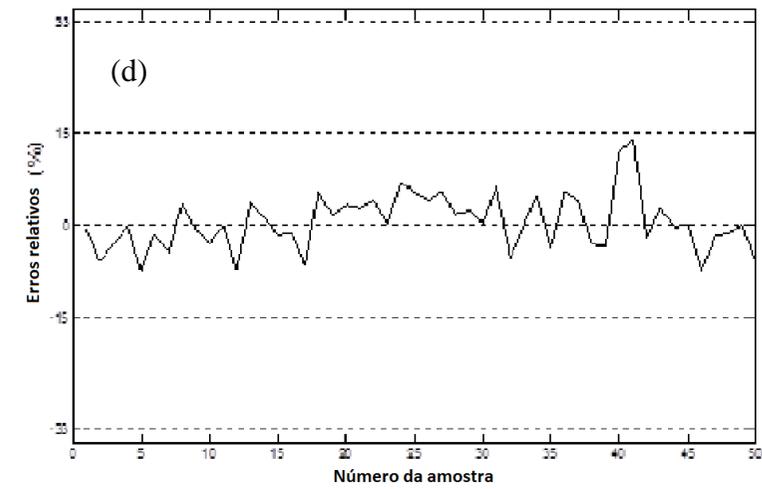
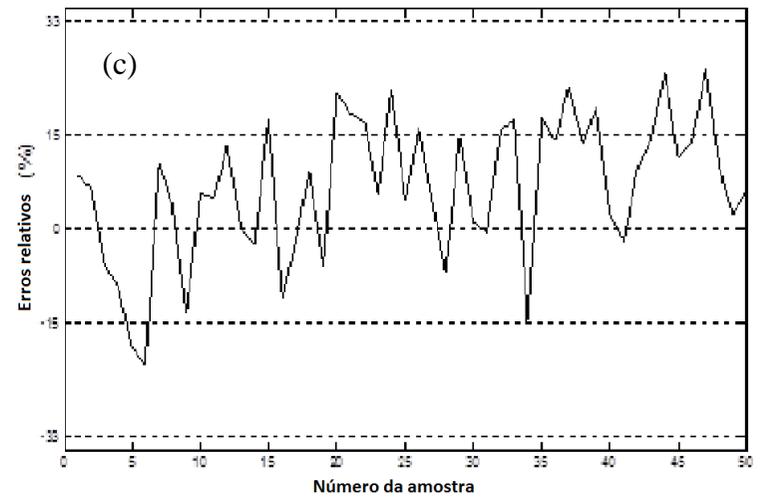
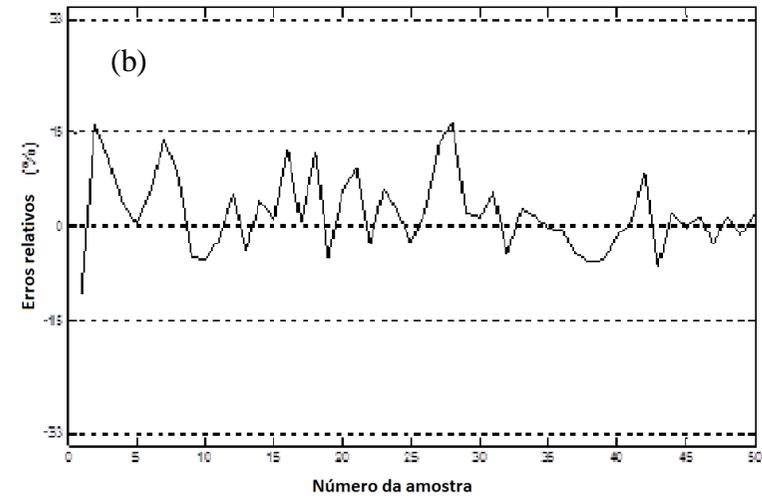
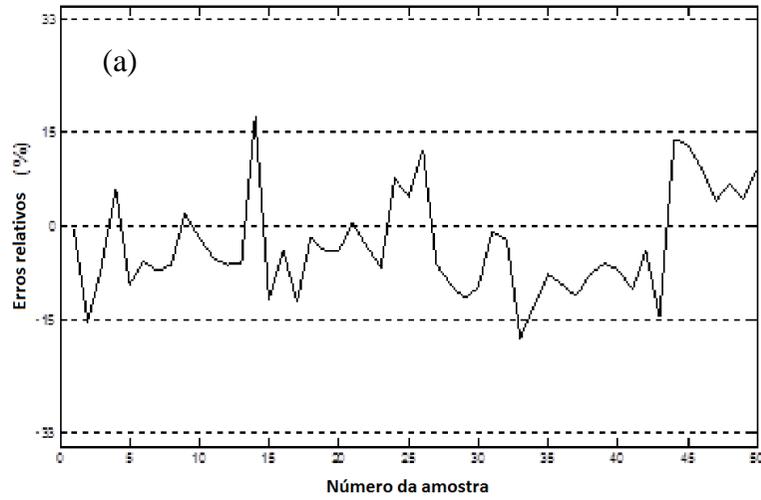
Teste/Parâmetro	Celulose	Hemicelulose	Lignina	RTE
RJ / $R_{eq}$	0,9763	0,9835	0,9861	0,9843
BF / $t_L$	1,53	2,36	0,23	0,30
DW / d	1,21	1,59	1,16	1,64

Para o teste de RJ, as cartas de controle para hemicelulose, lignina e RTE apresentaram valores de  $R_{eq}$  acima do  $R_{crítico}$  (0,9764, N=50) a 95% e a carta controle para celulose apresentou valores muito próximos deste limite. Assim, a 96% de nível de confiança, todos os erros foram considerados normalmente distribuídos. Para o teste de BF, celulose, lignina e RTE mostraram valores de  $t_L$  abaixo do valor crítico (2,01) a 95%, enquanto o  $t_L$  estimado para a hemicelulose foi aceitável apenas a 98% ( $t_{crítico} = 2,40$ ), garantindo a homocedasticidade dos erros neste nível. Finalmente, o teste de DW indicou que apenas a hemicelulose e o RTE apresentaram valores de d dentro da faixa de aceitabilidade (1,58-2,42, a 95%), enquanto celulose e lignina apresentaram erros com certo grau de dependência.

### **Caldo**

Como as amostras do caldo são instáveis e devem ser analisadas imediatamente após a coleta, não foi possível monitorar os modelos construídos para esta matriz. Além disso, não havia número de amostras suficiente para a construção das cartas de controle durante um longo período.

As cartas de controle são ferramentas de controle bastante úteis e devem ser utilizadas rotineiramente para a avaliação dos modelos construídos.



**Figura 33** Cartas de controle para o monitoramento da estabilidade dos modelos NIRS da biomassa. (a) celulose, (b) hemicelulose, (c) lignina e (d) RTE

## 7 - CONCLUSÃO

---

Considerando a necessidade de um grande número de experimentos para melhorar a qualidade do sorgo sacarino destinado à produção de etanol e a necessidade do PMGS da Embrapa de diferenciar híbridos de acordo com seu potencial de produção de álcool, métodos analíticos rápidos e de baixo custo são necessários para caracterização das amostras e para a estimativa do rendimento de etanol. Assim, este trabalho desenvolveu e validou métodos de calibração multivariada baseados em NIRS para determinação de parâmetros de qualidade em caldo e biomassa de sorgo sacarino visando à produção de etanol de primeira e de segunda geração, respectivamente.

Além disso, o rendimento teórico de etanol lignocelulósico foi determinado a partir dos componentes estruturais da parede celular, o que permitiu a eliminação das complexas metodologias de referência baseadas em hidrólise, fermentação e métodos cromatográficos. A estimativa do rendimento teórico de etanol a partir dos componentes estruturais é considerada mais confiável que a abordagem alternativa de quantificar os teores de monossacarídeos, uma vez que as características estruturais da planta, como a lignificação, afetam a disponibilidade dos monossacarídeos para a conversão em etanol<sup>92</sup>.

As metodologias desenvolvidas propiciaram uma economia significativa tanto em custo quanto em tempo de análises para o PMGS, uma vez que é necessário apenas um minuto para a obtenção dos espectros e predição simultânea de todos os parâmetros em cada amostra (biomassa ou caldo) e a utilização de reagentes é feita apenas nas etapas de calibração e manutenção dos modelos.

Todos os métodos desenvolvidos foram submetidos a uma validação analítica completa, de acordo com os guias nacionais e internacionais, e considerados exatos, lineares, sensíveis e sem viés. Finalmente, a estabilidade dos modelos da biomassa foi monitorada durante aproximadamente seis meses através do desenvolvimento de cartas de controle

apropriadas. Esta mesma estratégia não pôde ser aplicada às amostras de caldo, em virtude de sua indisponibilidade em quantidade suficiente.

Finalmente, o desenvolvimento deste trabalho numa parceria entre a universidade e uma empresa pública promoveu transferência de conhecimento e tecnologia no sentido de que as metodologias propostas serão efetivamente exploradas e utilizadas nas rotinas de análise da Embrapa, contribuindo para a agilidade e economia de recursos financeiros na pesquisa.

## **8 - REFERÊNCIAS BIBLIOGRÁFICAS**

---

1. R.C.C. Leite, M.R.L.V. Leal, O biocombustível no Brasil, *Novos Estudos*, 78 (2007) 15-21.
2. Companhia Nacional de Abastecimento (CONAB), Ministério da Agricultura, Pecuária e Abastecimento (MAPA). O etanol como combustível universal, Brasília, 2008.
3. Banco Nacional do Desenvolvimento (BNDES), Centro de Gestão e Estudos Estratégicos (CGEE), Organização das Nações Unidas para Alimentação e Agricultura (FAO), Comissão Econômica para a América Latina e o Caribe (CEPAL). Bioetanol de cana-de-açúcar – Energia para o desenvolvimento sustentável, Rio de Janeiro, 2008.
4. D.L. Gazzoni, Balanço de emissões de CO<sub>2</sub> por biocombustíveis no Brasil: histórico e perspectivas, *Documentos*, 334, Embrapa Soja, 2012, 102 p.
5. Empresa de Pesquisas Energéticas (EPE), Análise de Conjuntura dos Biocombustíveis, Brasília, 2012.
6. NOVA CANA <<http://www.novacana.com/n/etanol/mercado/futuro/eua-projecao-futuro-mercado-etanol-190213/#>> Acessado em outubro de 2013.
7. Federação das Indústrias do Estado de Minas Gerais (FIEMG), Núcleo de inteligência Competitiva- bioetanol, Etanol de segunda geração. <<http://www.fiemg.org.br>>, Acessado em setembro de 2013.
8. M.O.S. Dias, T.L. Junqueira, C.D.F. Jesus, C.E.V. Rossel, R. Maciel Filho, A. Bonomi, Improving second generation ethanol production through optimization of first generation production process from sugarcane, *Energy*, 43 (2012) 246-252.
9. Folha de São Paulo <<http://www1.folha.uol.com.br/mercado/1094693-1-usina-de-etanol-celulosico-do-pais-entra-em-operacao-em-2013.shtml>> Acessado em setembro de 2013.
10. Companhia Nacional de Abastecimento (CONAB). Os fundamentos da crise do setor sucroalcooleiro no Brasil. 2ª ed., Brasília, 2010.
11. P.M. Ribas, Sorgo: introdução e importância econômica, *Documentos*, 26, Embrapa Milho e Sorgo, 2003, 16 p.
12. F.O.M. Durães, Sorgo Sacarino: desenvolvimento de tecnologia agronômica. *Agroenergia em Revista*, 3ª ed., Brasília: Embrapa Agroenergia, 2011.
13. M. Calviño, J. Messing, Sweet sorghum as a model system for bioenergy crops, *Current Opinion in Biotechnology*, 23 (2012) 323-329.

14. A. Goshadrou, K. Karimi, M.J. Taherzadeh, Bioethanol production from sweet sorghum bagasse by *Mucor hiemalis*, *Industrial Crops and Products*, 34 (2011) 1219-1225.
15. T. Hattori, S. Morita, Energy Crops for Sustainable Bioethanol Production; Which, Where and How? *Plant Production Science*, 13 (2010) 221-234.
16. F.A. Santos, J.H. Queiróz, J.L. Colodette, S.A. Fernades, V.M, Guimarães, S.T. Rezende, Potencial da Palha da Cana-de-açúcar para a produção de etanol, *Química Nova*, 35 (2012) 1004-1010.
17. R.A. Miranda, Custo de produção e viabilidade econômica da produção de etanol a partir do sorgo sacarino plantado na entressafra da cana-de-açúcar, *Documentos*, 139, Embrapa Milho e Sorgo, 2012, 106-112.
18. A. Gomes, D. Rodrigues, P. Oliveira, Caracterização do sorgo para a produção de etanol, *Agroenergia em Revista*, 3ª ed., Brasília: Embrapa Agroenergia, 2011.
19. A.J. Lorenz, R.P. Anex, A. Isci, J.G. Coors, N. Leon, P.J. Weimer, Forage quality and composition measurements as predictors of ethanol yield from maize (*Zea mays* L.) stover, *Biotechnology for Biofuels*, 2 (2009) 1-8.
20. C.M.M. Machado, Produção de etanol de sorgo sacarino, *Agroenergia em Revista*, 3ª ed., Brasília: Embrapa Agroenergia, 2011.
21. Conselho dos produtores de cana-de-açúcar, açúcar e álcool do estado do Paraná (CONSECANA-PR), Normas operacionais de avaliação da qualidade da cana-de-açúcar. 1ª ed., Curitiba, 2000. < <http://www.faep.com.br/consecana/normasop.htm>> Acessado em outubro de 2013.
22. Conselho dos produtores de cana-de-açúcar, açúcar e álcool do estado de São Paulo (CONSECANA-SP), Manual de Instruções, 5ª ed., Piracicaba, 2006. <[http://www.orplana.com.br/manual\\_2006.pdf](http://www.orplana.com.br/manual_2006.pdf)> Acessado em outubro de 2012.
23. H. Lane, L. Eynon, Determination of reducing sugar by means of Fehling's solution with methylene blue as internal indicator, *Journal of the Society of Chemistry Industry*, 42 (1923) 32T-37T.
24. D. Fengel, G. Wegener, *Wood: Chemistry, Ultrastructure and Reactions*. Berlin: Walter de Gruyter, 1989. 613 p.
25. P.J. van Soest, R.H. Wine, Determination of lignin and cellulose in acid detergent fiber with permanganate, *Journal of the Association of Official Analytical Chemists*, 51 (1968) 780-785.

26. A.K.D. Salman, A.C.D. Ferreira, J.P.G. Soares, J.P. Souza, Metodologias para avaliação de alimentos para ruminantes domésticos, Documentos, 136, Centro de Pesquisa Agroflorestal de Rondônia, 2010, 21 p.
27. H.H. Nimz, D. Robert, O. Faix, M. Nembr, Carbon-13 NMR Spectra of lignins. 8. structural differences between lignins of hardwood, softwood, grasses and compression wood. *Holzforschung*, 35 (1981) 16-26.
28. H. Jorgensen, J.B. Kristensen, C. Felby, Enzymatic conversion of lignocellulosic into fermentable sugars: challenges and opportunities, *Biofuels, Bioproducts & Biorefining*, 1 (2007) 119-134.
29. D.J. Silva, A.C. Queiroz, *Análise de alimentos: métodos químicos e biológicos 3ª ed.*, Viçosa: UFV, 2006, 235 p.
30. T.L. Ogeda, D.F.S. Petri, Hidrólise enzimática de biomassa. *Química Nova*, 33 (2010) 1549-1558.
31. Y.L. Zhao, A. Dolat, Y. Steinberger, X. Wang, A. Osman, G.H. Xie, Biomass yield and changes in chemical composition of sweet sorghum cultivars grown for biofuel. *Field Crops Research*, 111 (2009) 55-64.
32. J. Moros, S. Garrigues, M. La Guardia, Vibrational spectroscopy provides a green tool for multi-component analysis. *Trends in Analytical Chemistry*, 29 (2010) 578-591.
33. C. Pasquini, Near Infrared Spectroscopy: fundamentals, practical, aspects and analytical applications. *Journal of the Brazilian Chemical Society*, 14 (2003) 198-219.
34. P. Valderrama, Avaliação de figuras de mérito em calibração multivariada na determinação de parâmetros de controle de qualidade em indústria alcooleira por espectroscopia no infravermelho próximo. Dissertação de mestrado. Universidade Estadual de Campinas, Campinas, 2005.
35. D.A. Skoog, F.J. Holler, T.A. Nieman, *Princípios da Análise Instrumental. 5ª ed.*, Porto Alegre: Bookman, 2002. 836p.
36. D.A. Burns, E. W. Ciurczak, *Handbook of Near-Infrared Analysis. 2<sup>nd</sup> ed.*, New York: Marcel Dekker, 1992, 814p.
37. Y. Ozaki, Near-Infrared Spectroscopy – Its versatility in Analytical Chemistry. *Analytical Sciences*, 28 (2012) 545-563.
38. E.W. Ciurczak, J.K. Drennen III, *Pharmaceutical and Medical applications of Near Infrared Spectroscopy*, New York: Marcel Dekker, 2002.
39. C.A. Roberts, J. Workman Jr., J.B. Reeves III, *Near infrared Spectroscopy in Agriculture*, Madison: American Society of Agronomy, 2004.

40. Y, Ozaki, W.F. McClure, A.A. Cristy, *Near Infrared Spectroscopy in food science and technology*, New York: Wiley Interscience, 2006.
41. B.R. Hames, S.R. Thomas, A.D. Sluiter, C.J. Roth, D.W. Templeton, Rapid biomass analysis - New tools for compositional analysis of corn stover feedstocks and process intermediates from ethanol production, *Applied Biochemistry and Biotechnology*, 105 (2003) 5-16.
42. F. Xu, J. Yu, T. Tesso, F. Dowell, D. Wang, Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review, *Applied Energy*, 104 (2013) 801-809.
43. K.P. Vogel, B.S. Dien, H.G. Jung, M.D. Casler, S.D. Masterson, R.B. Mitchell, Quantifying actual and theoretical ethanol yields for switchgrass strains using NIRS analyses, *Bioenergy Resource*, 4 (2011) 96-110.
44. F. Pohl, T. Senn, A rapid and sensitive method for the evaluation of cereal grains in bioethanol production using near infrared reflectance spectroscopy, *Bioresource Technology*, 102 (2011) 2834-2841.
45. C.A. Roberts, J.H. Houx, F.B. Fritschi, Near-infrared analysis of sweet sorghum bagasse, *Crop Science*, 51 (2011) 2284-2288.
46. C. Magaña, N. Núñez-Sánchez, V.M. Fernández-Cabanás, P. García, A. Serrano, D. Pérez-Marín, J.M. Permán, E. Alcade, Direct prediction of bioethanol yield in sugar beet pulp using near infrared spectroscopy, *Bioresource Technology*, 102 (2011) 9542-9549.
47. X. Hao, K. Thelen, J. Gao, Prediction of the ethanol yield of dry-grind maize grain using near infrared spectroscopy, *Biosystem Engergy*, 112 (2012) 161-170.
48. H. Cen, Y. He, M. Huang, Measurement of soluble solids contents and pH in orange juice using chemometrics and Vis-NIRS, *Journal of Agriculture and Food Chemistry*, 54 (2006) 7437-7443.
49. L. Xie, X. Ye, D. Liu, Y. Ying, Quantification of glucose, fructose and sucrose in bayberry juice by NIR and PLS, *Food Chemistry*, 114 (2009) 1135-1140.
50. F.J. Rambla, S. Garrigues, M. de la Guardia, PLS-NIR determination of total sugar, glucose, fructose and sucrose in aqueous solutions of fruit juice, *Analytica Chimica Acta*, 344 (1997) 41-53.
51. D. Cozzolino, W.U. Cynkar, N. Shah, P. Smith, Multivariate data analysis applied to spectroscopy: Potencial application to juice and fruit quality. *Food Research International*, 44 (2011) 1888-1896.

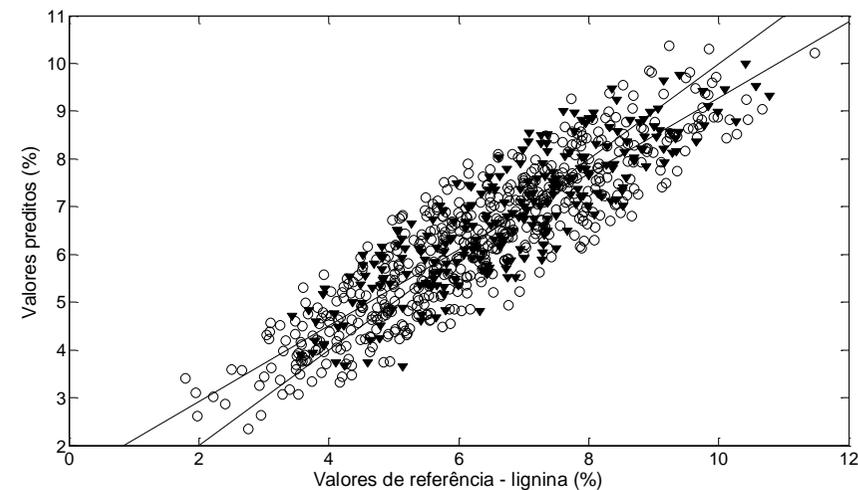
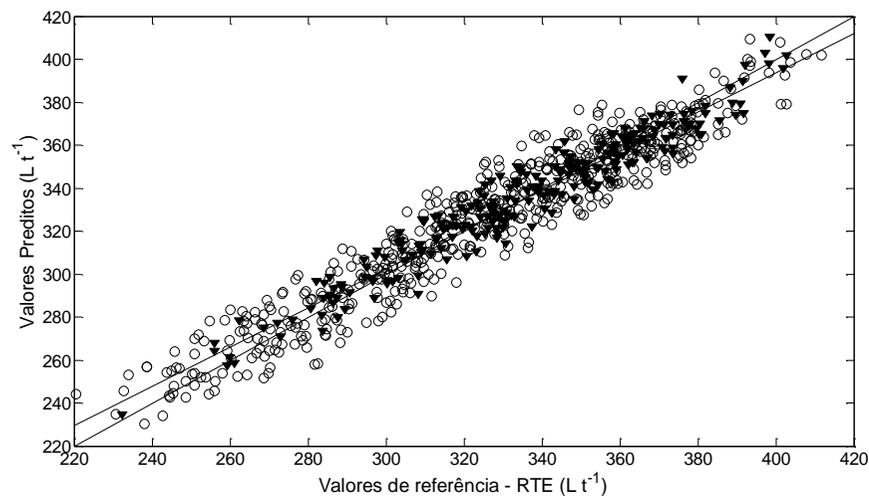
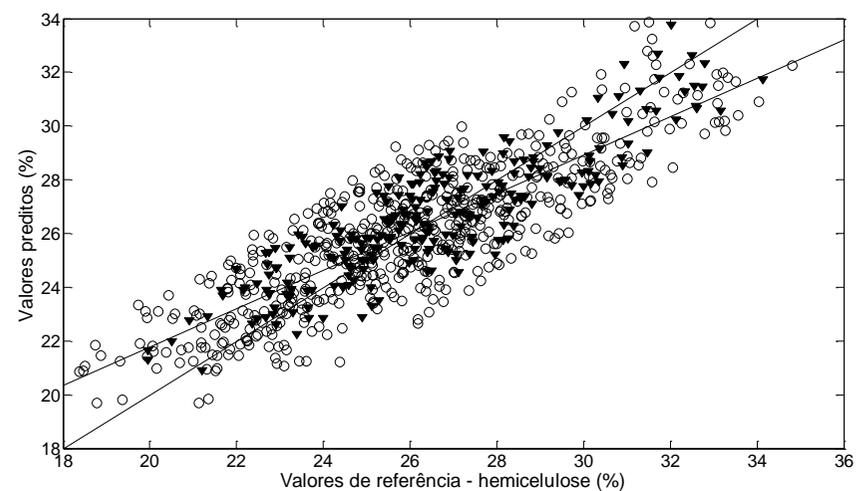
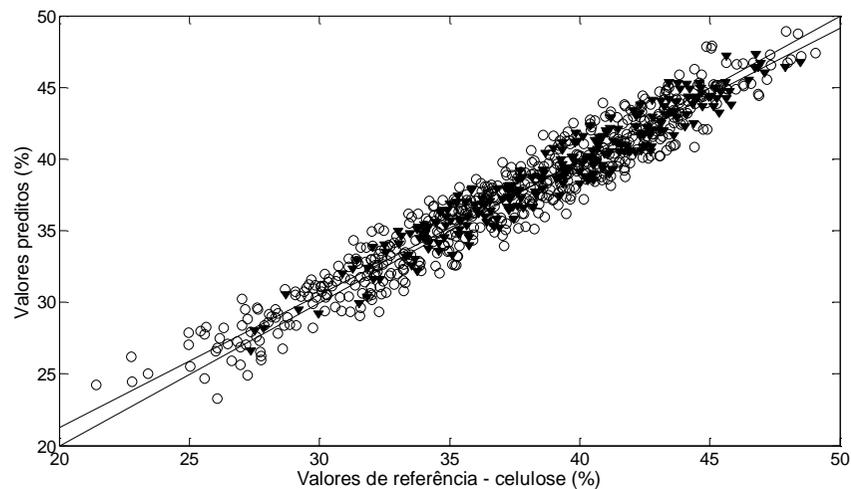
52. P. Valderrama, J.W.B. Braga, R.J. Poppi, Validation of Multivariate Calibration Models in the Determination of Sugar Cane Quality Parameters by Near Infrared Spectroscopy. *Journal of the Brazilian Chemical Society*, 18 (2007) 259-266.
53. N. Sorol, E. Arancibia, S.A. Bortolato, A.C. Olivieri, Visible/near infrared-partial least-squares analysis of brix in sugar cane juice. A test field for variable selection methods. *Chemometrics and Intelligent Laboratory Systems*, 102 (2010) 100-109.
54. P. Geladi, Chemometrics in spectroscopy. *Spectrochimica Acta Part B*, 58 (2003) 767-782.
55. M. Otto, *Chemometrics – Statistics and Computer Application in Analytical Chemistry*, Toronto: Wiley-VCH, 1999, 314 p.
56. T. Naes, H. Martens, *Multivariate Calibrate II – Chemometrics Methods*. *Trends in Analytical Chemistry*, 3 (1984) 266-271.
57. Standard practices for infrared multivariate quantitative analysis-E1655-05. ASTM International, West Conshohocken, 2012.
58. S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemometrics and Intelligent Laboratory Systems*, 58 (2001) 109-130.
59. P. Geladi, B. Kowalski, Partial Least Square Regression: a tutorial. *Analytica Chimica Acta*, 186 (1986) 1-17.
60. R. G. Brereton, *Applied Chemometrics for Scientists*, Chichester: Wiley, 2007.
61. P. Valderrama, J.W.B. Braga, R.J. Poppi, Variable selection, outlier detection and figures of merit estimation in a PLSR multivariate calibration model. A case study for the determination of quality parameters in alcohol industry by near infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, 52 (2007) 8331-8338.
62. S.V.C. Souza, R.G. Junqueira, A procedure to assess linearity by ordinary least squares method. *Analytica Chimica Acta*, 552 (2005) 25-35.
63. Instituto Nacional de Metrologia, *Normalização de Qualidade Industrial (INMETRO)*. DOQ-CGCRE-008, Brasília, 2007.
64. International Conference on Harmonization (ICH), *Tripartite Guideline-Q2A Text on validation analytical procedures*, Fed. Regist., 60 FR 11260, USA, 1995.
65. International Conference on Harmonization (ICH), *Tripartite Guideline-Q2B Validation of analytical procedures: Methodology*, Fed. Regist., 62 FR 27464, USA, 1997.
66. J.W.B. Braga, R.J. Poppi, Figures of merit for the determination of the polymorphic purity of carbamazepine by infrared spectroscopy and multivariate calibration. *Journal of Pharmaceutical Sciences*, 93 (2004) 2124-2134.

67. J.W.B. Braga, R.J. Poppi, Validação de modelos de calibração multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo. *Química Nova*, 27 (2004) 1004-1011.
68. J. Saurina, C. Leal, R. Compano, M. Granados, M. Dolors Prat, R. Tauler, Estimation of figures of merit using univariate statistics for quantitative second-order multivariate curve resolution. *Analytica Chimica Acta*, 432 (2001) 241-251.
69. A.C. Olivieri, N.M. Faber, J. Ferré, R. Boqué, J.H. Kalivas, H. Mark, Uncertainty Estimation and Figures of Merit for Multivariate Calibration. *Pure and Applied Chemistry*, 78 (2006) 633-661.
70. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária (MAPA). Instrução Normativa n° 24, Brasil, 2009.
71. Food and Agriculture Organization (FAO). Validation of analytical methods for food control. Report of a Joint FAO/IAEA Expert Consultation, FAO Food and Nutrition Paper N° 68, Roma, 1998.
72. European Commission (EC). Directive 657. Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results, Official Journal, L221, 8-36, Bruxelas, Bélgica, 2002.
73. M. Thompson, S.L.R. Ellison, R. Wood, Harmonized guidelines for single-laboratory validation of methods of analysis, *Pure Applied Chemistry*, 74 (2002) 835.
74. M.A.M. Silva, M.H. Ferreira, J.W.B. Braga, M.M. Sena, Development and analytical validation of a multivariate calibration method for determination of amoxicillin in suspension formulations by near infrared spectroscopy. *Talanta*, 89 (2012) 342-351.
75. M.H. Ferreira, J.W.B. Braga, M.M. Sena, Development and validation of a chemometric method for direct determination of hydrochlorothiazide in pharmaceutical samples by diffuse reflectance near infrared spectroscopy. *Microchemical Journal*, 89 (2012) 342-351.
76. B.G. Botelho, B.A.P. Mendes, M.M. Sena, Implementação de um método robusto para o controle fiscal de umidade em queijo minas artesanal. Abordagem metrológica multivariada. *Química Nova*, 36 (2013) 1416-1422.
77. Instituto de Metrologia, Normalização e Qualidade Industrial (INMETRO), Vocabulário Internacional de Metrologia – Conceitos fundamentais e gerais e termos associados (VIM 2012), 1ª ed. luso-brasileira, Portaria n° 232, Brasília, 2012.
78. P. Valderrama, J.W.B. Braga, R.J. Poppi, Estado da arte de figuras de mérito em calibração multivariada. *Química Nova*, 32 (2009) 1278-1287.

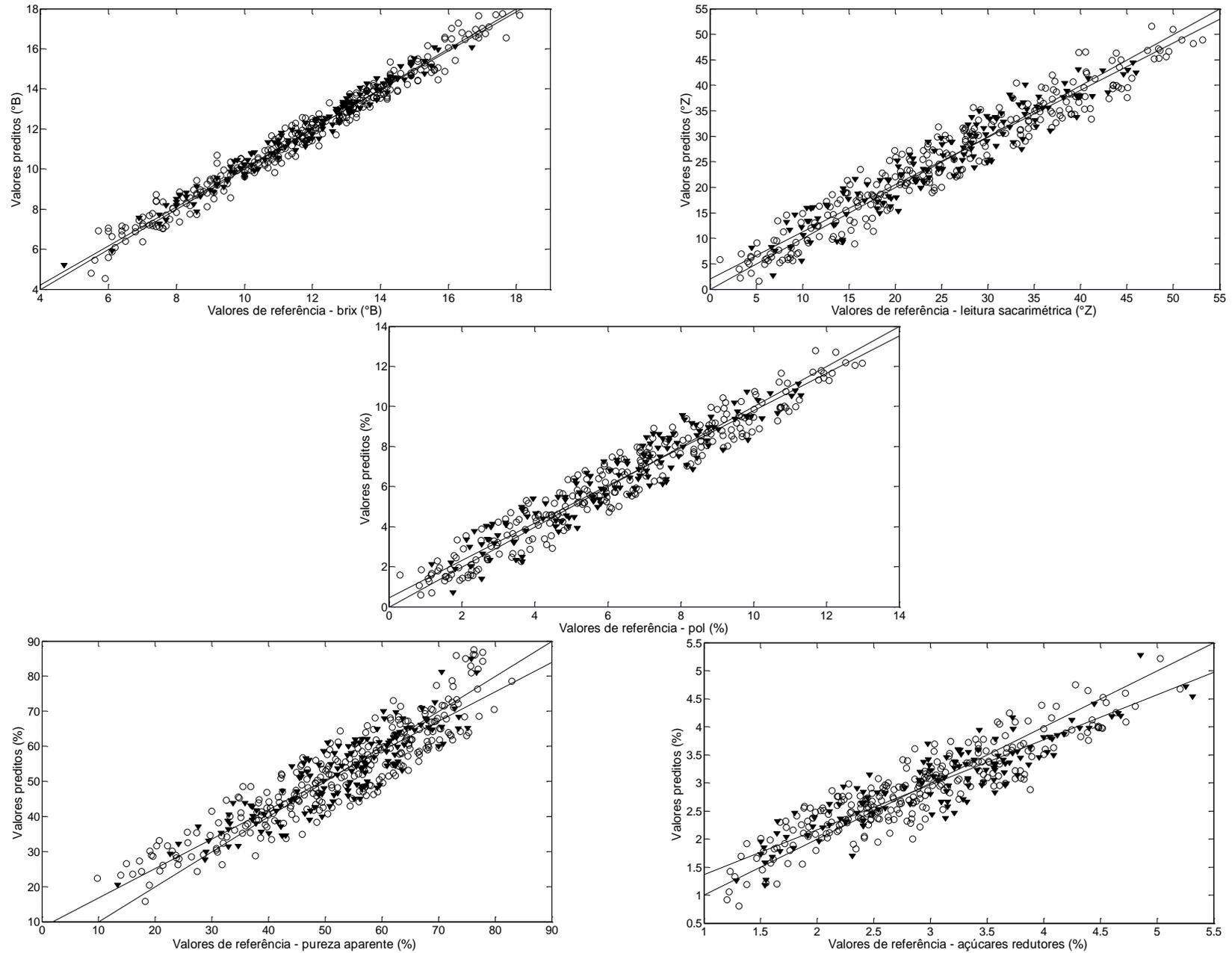
79. J. Ferré, S.D. Brown, F.X. Rius, Improved calculation of the net analyte signal in the inverse calibration, *Journal of Chemometrics*, 15 (2000) 537
80. P. Williams, *Near-infrared technology in the agricultural and food industries*, 2<sup>nd</sup> ed., Saint Paul: American Association of Cereal Chemists Inc., 2001.
81. T. Tanimoto, The press method of cane analysis, *Hawaiian Planters' Record*, 51 (1964) 133-150.
82. L.P.D. Ribeiro, *Espectropolarimetria e polarimetria baseadas em cristais birrefringentes para as regiões espectrais do visível e infravermelho próximo*, Tese de doutorado, Universidade Estadual de Campinas, Campinas, 2012.
83. P.J. van Soest, *Nutritional Ecology of Ruminant*, Ithaca: Cornell University Press, 1994.
84. Ankom Technology, <http://www.ankom.com/procedures.aspx>, acessado em novembro de 2013.
85. C.M. Popescu, M.C. Popescu, A near infrared spectroscopy study of the structural modifications of lime (*Tilia cordata* Mill.) wood during hydro-thermal treatment, *Spectrochimica Acta Part A*, 115 (2013) 227-233.
86. J.J. Workman Jr, L. Weyer, *Practical Guide to Interpretive Near-Infrared Spectroscopy*, Boca Raton: CRC – Press, USA, 2008.
87. R.W. Kennard, L.A. Stone, Computer aided design of experiments, *Technometrics*, 11 (1969) 137-148.
88. R.J. Barnes, M. S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, *Applied Spectroscopy*, 43 (1989) 772-777.
89. A. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, *Trends in Analytical Chemistry*, 28 (2009) 1201-1222.
90. W. Horwitz, Protocol for the design, conduct and interpretation of method-performance studies: Revised 1994, *Pure and Applied Chemistry*, 67 (1995) 331-343.
91. Ministério da Agricultura, Pecuária e Abastecimento (MAPA). Secretaria de Defesa Agropecuária. *Manual da Garantia da Qualidade*, Brasília, Brasil, 2011.
92. K.J. Han, W.D. Pitman, M. Kim, D.F. Day, M.W. Alison, M.E. McCormick, G. Aita, Ethanol production potential of sweet sorghum assessed using forage fiber analysis procedures, *GCB Bionergy* 5 (2013) 358-366.

## ANEXOS

**ANEXO A:** Valores referência *versus* valores preditos para os modelos da biomassa. (○) amostras de calibração (▼) amostras da validação.



**ANEXO B:** Valores referência *versus* valores preditos para os modelos do caldo. (○) amostras de calibração (▼) amostras da validação.



**ANEXO C:** Valores de referência, valores preditos e erros relativos de amostras de validação para o parâmetro celulose.

VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)
35,0	36,9	5,3	33,5	32,6	-2,8	42,6	41,0	-3,8	36,3	37,2	2,7
33,8	35,3	4,4	33,8	35,1	4,0	39,4	39,2	-0,4	34,0	34,8	2,5
40,9	39,6	-3,1	37,8	36,7	-3,1	41,1	40,4	-1,6	37,6	36,7	-2,4
34,1	34,4	1,0	40,7	39,9	-2,0	47,9	46,5	-3,1	42,3	41,6	-1,8
39,9	41,3	3,4	32,3	31,6	-2,2	40,7	41,0	0,8	40,0	38,3	-4,2
33,5	34,8	4,1	37,5	35,8	-4,3	40,6	40,0	-1,5	38,2	38,8	1,6
40,7	41,2	1,1	35,7	36,0	0,9	43,1	42,5	-1,5	39,3	40,7	3,3
37,5	37,7	0,7	35,4	35,7	0,9	35,0	35,7	2,0	40,9	39,4	-3,8
34,0	35,4	4,1	34,2	34,2	0,0	41,3	39,3	-4,7	44,2	44,9	1,6
34,6	33,6	-2,9	36,0	36,5	1,3	42,0	40,6	-3,2	34,1	34,8	2,2
37,3	37,8	1,5	44,7	44,7	-0,1	40,3	38,8	-3,8	47,0	46,7	-0,5
45,0	44,4	-1,3	33,8	32,2	-4,6	41,5	41,1	-0,7	45,1	44,4	-1,4
37,6	38,6	2,5	36,2	36,9	1,9	43,4	44,2	1,7	35,7	34,8	-2,6
35,2	37,0	5,0	35,7	34,0	-4,9	44,6	45,3	1,6	34,9	36,0	3,2
37,7	37,9	0,5	33,8	34,5	1,9	46,9	46,4	-0,9	41,6	41,0	-1,5
36,8	35,8	-2,6	36,3	36,0	-1,0	44,4	42,5	-4,2	45,1	43,7	-3,3
37,7	39,2	3,9	38,2	36,7	-4,1	40,1	39,7	-0,9	41,5	42,1	1,6
35,3	37,5	6,2	32,1	33,9	5,7	36,4	36,8	1,3	42,1	42,3	0,3
40,9	41,6	1,7	35,4	34,8	-1,6	40,5	39,6	-2,2	42,8	42,9	0,3
35,8	37,9	5,6	39,5	39,2	-0,8	40,3	38,5	-4,4	39,4	38,8	-1,5
39,3	40,0	1,8	36,7	35,4	-3,6	40,5	39,1	-3,6	45,3	44,3	-2,2
40,6	41,3	1,9	41,1	39,0	-5,3	41,2	42,2	2,4	40,0	39,2	-2,0
41,1	41,8	1,8	35,4	35,1	-0,7	44,6	45,3	1,7	44,7	45,0	0,8
33,1	34,1	3,2	37,3	36,4	-2,3	45,8	44,8	-2,2	40,7	41,4	1,6
35,9	38,0	5,9	42,7	41,7	-2,4	42,4	41,7	-1,7	39,3	39,7	0,9
31,9	32,4	1,7	34,0	34,6	1,8	42,4	41,9	-1,2	42,3	43,8	3,5
37,4	38,1	1,9	33,3	33,2	-0,4	34,2	33,8	-1,2	45,2	45,0	-0,5
37,2	36,7	-1,5	32,6	33,5	2,9	34,9	35,4	1,5	44,5	44,4	-0,4
43,1	44,1	2,2	36,9	35,2	-4,5	35,4	35,8	0,9	43,1	42,7	-1,0
44,4	44,2	-0,5	39,0	40,8	4,6	43,3	43,2	-0,4	37,6	37,8	0,4
39,5	39,4	-0,3	41,3	41,3	0,0	39,8	40,2	1,0	42,7	40,5	-5,1
31,3	32,3	3,4	42,8	42,1	-1,6	43,3	43,7	1,0	38,6	39,0	1,0
43,7	42,9	-2,0	46,5	45,5	-2,1	40,5	40,6	0,3	41,0	42,0	2,4
38,6	39,3	1,8	47,1	46,0	-2,3	40,4	40,2	-0,5	41,5	40,5	-2,4
35,8	34,5	-3,4	40,7	41,1	1,0	44,2	44,3	0,4	45,4	43,2	-4,7
42,6	42,2	-1,1	38,3	37,5	-2,1	45,4	44,7	-1,6	43,6	43,4	-0,7
29,2	29,5	1,1	40,4	42,3	4,8	40,5	38,5	-5,0	39,2	38,7	-1,4
36,1	36,2	0,3	34,1	35,0	2,8	44,1	42,3	-4,0	44,1	43,2	-2,1
37,2	37,4	0,5	33,0	35,0	6,0	41,0	40,4	-1,5	41,2	40,6	-1,5
28,7	30,6	6,5	34,9	36,4	4,3	42,5	40,9	-3,8	38,5	38,8	0,7
34,8	34,6	-0,7	37,3	38,2	2,5	41,7	40,5	-2,8	38,3	37,7	-1,5
40,2	38,8	-3,4	36,0	36,8	2,4	41,8	40,9	-2,1	35,6	37,0	3,9
35,1	33,3	-5,0	42,6	42,2	-0,9	45,7	44,2	-3,1	41,9	40,5	-3,4
38,6	38,6	0,1	34,7	36,5	5,0	37,2	38,0	2,2	36,4	35,6	-2,2
39,8	41,9	5,1	33,2	34,7	4,5	43,7	44,0	0,8	37,8	38,8	2,6
42,8	44,1	3,2	37,4	38,1	1,9	40,7	40,7	0,1	39,7	40,1	1,1
34,2	35,3	3,1	37,4	36,7	-2,0	45,6	45,4	-0,6	42,1	43,4	3,2
37,3	37,2	-0,3	34,5	35,6	3,4	34,7	36,3	4,7	44,7	43,7	-2,4
30,9	32,0	3,7	39,4	38,3	-2,7	40,1	41,6	3,6	46,8	47,3	1,1
32,5	34,1	4,8	39,2	37,5	-4,5	43,5	43,3	-0,4	27,4	26,6	-2,7
39,8	39,8	0,0	37,3	38,2	2,4	42,2	40,5	-3,9	31,8	30,4	-4,5
34,6	34,2	-1,3	36,5	37,1	1,5	42,4	40,6	-4,1	45,8	43,8	-4,4
27,9	28,2	1,4	43,2	42,0	-2,8	40,5	40,1	-0,9	45,6	47,2	3,4

**ANEXO D:** Valores de referência, valores preditos e erros relativos de amostras de validação para o parâmetro hemicelulose.

VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)
27,2	26,8	-1,6	25,7	26,1	1,4	27,9	28,3	1,6	24,5	25,2	3,0
26,8	26,0	-3,2	28,0	25,2	-9,8	27,1	25,0	-7,6	26,5	25,7	-3,1
22,9	23,3	1,6	23,7	23,9	0,7	26,1	27,0	3,3	24,6	26,0	5,3
28,4	26,3	-7,3	22,7	24,5	7,7	30,2	28,7	-4,8	27,9	27,4	-2,0
29,0	28,4	-1,9	22,7	25,3	11,4	25,4	26,5	4,4	28,8	27,1	-6,0
29,7	27,8	-6,5	25,8	25,6	-0,8	28,3	25,8	-8,7	26,4	25,4	-3,8
24,2	26,3	8,7	28,0	25,8	-8,1	22,0	24,7	12,2	27,4	25,7	-6,4
29,4	29,8	1,2	24,6	24,1	-2,0	25,5	25,9	1,7	27,7	27,3	-1,4
23,7	23,3	-1,7	24,7	24,3	-1,8	24,5	25,4	3,5	28,4	27,2	-4,1
23,3	23,9	2,9	23,2	23,7	1,9	22,9	25,5	11,2	30,0	27,8	-7,5
26,3	25,8	-1,9	26,0	27,8	7,0	23,5	25,9	10,1	31,5	29,0	-7,8
26,6	27,7	4,3	29,2	28,2	-3,4	27,1	27,4	1,0	30,9	28,6	-7,6
24,9	25,6	3,1	22,5	24,1	7,4	26,8	27,9	4,0	27,7	27,1	-2,2
27,2	27,7	1,6	26,1	25,0	-4,0	27,3	28,2	3,2	26,2	28,2	7,6
27,6	28,2	2,2	21,7	23,7	9,3	26,1	26,5	1,3	28,4	28,3	-0,5
26,1	26,7	2,1	24,9	22,9	-8,1	27,4	25,9	-5,7	30,1	28,3	-6,0
24,6	24,8	0,9	22,9	24,7	7,8	23,2	25,5	9,8	28,8	28,7	-0,5
29,6	28,0	-5,4	23,6	23,9	1,0	26,0	24,9	-4,1	29,5	27,9	-5,1
25,5	26,5	3,9	22,2	23,9	7,8	24,2	25,1	4,0	26,7	28,9	8,3
24,9	25,4	2,2	25,0	25,9	3,5	26,9	25,3	-5,8	30,0	28,3	-5,6
28,6	27,4	-4,4	24,3	25,9	6,6	25,8	26,3	2,0	25,7	26,7	4,0
27,1	24,6	-9,4	25,6	26,6	4,0	26,2	27,9	6,4	26,6	28,0	5,1
25,7	28,1	9,1	26,3	26,3	0,1	26,9	29,1	8,1	26,9	27,3	1,5
27,8	27,8	0,0	25,5	27,8	8,9	28,1	27,6	-1,8	28,8	28,8	0,0
27,4	26,7	-2,5	21,7	23,9	10,4	28,8	28,5	-1,1	28,7	28,1	-2,1
23,2	24,2	4,3	22,5	22,8	1,5	24,4	27,0	10,7	28,4	28,7	1,0
24,4	25,7	5,6	25,1	25,2	0,4	24,4	26,1	7,1	26,6	28,3	6,5
22,8	23,0	0,7	24,6	24,7	0,3	27,0	25,9	-3,9	28,1	27,8	-1,2
26,1	27,8	6,8	26,4	28,5	8,0	30,9	28,8	-6,7	28,2	26,3	-6,6
23,8	25,5	7,3	27,7	29,1	4,9	27,4	26,6	-2,8	25,3	25,5	0,7
26,0	27,4	5,4	26,3	28,6	8,6	28,6	27,4	-4,0	27,2	26,8	-1,3
28,3	27,5	-2,9	26,9	28,8	7,0	28,6	27,3	-4,5	26,3	25,9	-1,7
26,5	24,6	-7,1	25,0	25,0	0,2	26,5	27,5	3,8	22,2	24,0	8,0
30,4	29,2	-3,9	22,7	23,9	5,4	28,6	28,7	0,4	32,1	30,3	-5,8
28,3	26,9	-4,9	24,9	24,7	-1,0	26,1	27,3	4,7	32,6	30,7	-5,7
22,8	23,8	4,5	21,2	20,9	-1,4	27,0	26,7	-1,2	34,1	31,7	-7,0
25,8	26,9	4,2	25,5	25,8	1,0	25,6	25,9	1,3	33,2	30,6	-7,8
24,0	22,8	-4,8	26,4	26,1	-1,4	25,9	26,7	3,3	30,9	32,3	4,5
26,4	24,6	-6,9	23,2	25,1	8,0	24,6	25,8	4,7	31,7	32,7	3,1
27,2	26,3	-3,3	24,9	25,0	0,3	25,2	27,5	9,2	20,5	22,0	7,2
23,7	22,9	-3,4	27,1	25,3	-6,7	28,0	27,9	-0,5	23,3	23,1	-1,0
25,3	23,5	-7,0	26,4	25,3	-4,1	26,8	28,2	5,4	23,8	25,6	7,3
26,5	28,7	8,2	25,5	27,2	6,8	31,0	30,2	-2,7	22,7	24,9	9,5
24,8	25,8	4,2	24,6	24,4	-0,5	29,9	27,4	-8,2	20,9	22,8	8,9
21,3	22,9	7,4	24,2	25,1	3,5	28,2	29,6	5,1	32,6	30,7	-6,0
24,5	24,0	-1,7	25,1	25,8	3,1	30,4	27,9	-8,0	32,6	31,5	-3,2
25,7	27,4	6,6	25,9	26,7	3,0	30,1	28,9	-4,1	31,7	30,6	-3,5
23,6	25,7	8,7	25,3	26,0	2,4	28,3	29,4	4,0	32,0	33,8	5,5
22,9	22,6	-1,2	22,7	22,9	0,9	29,2	29,3	0,4	30,1	30,2	0,5
23,2	23,7	2,6	26,4	26,8	1,6	28,1	29,0	3,0	23,4	22,3	-4,8
25,8	26,3	2,1	26,8	25,4	-5,2	30,3	28,2	-6,9	31,4	30,6	-2,6
24,5	25,0	2,1	30,2	27,8	-7,7	26,2	27,3	3,9	30,7	30,5	-0,6
20,0	21,7	8,6	26,6	28,3	6,5	26,9	28,1	4,5	31,0	29,3	-5,4

**ANEXO E:** Valores de referência, valores preditos e erros relativos de amostras de validação para o parâmetro RTE.

VR (Lt <sup>-1</sup> )	VP (Lt <sup>-1</sup> )	ER (%)	VR (Lt <sup>-1</sup> )	VP (Lt <sup>-1</sup> )	ER (%)	VR (Lt <sup>-1</sup> )	VP (Lt <sup>-1</sup> )	ER (%)	VR (Lt <sup>-1</sup> )	VP (Lt <sup>-1</sup> )	ER (%)
327	328	0,2	313	327	4,6	357	356	-0,3	310	324	4,7
317	327	3	262	279	6,1	330	322	-2,6	348	355	2,1
290	292	0,2	319	331	3,8	328	331	1,1	331	329	-0,7
341	338	-1	256	264	3,2	318	323	1,5	332	341	2,6
358	359	0,2	300	297	-1,1	359	352	-2,1	324	334	3,1
357	356	-0,3	297	297	0,1	392	375	-4,3	370	370	-0,2
284	289	1,8	314	312	-0,6	341	331	-3	328	322	-2,1
284	296	4,2	308	291	-5,6	345	347	0,7	331	314	-5
309	311	0,6	302	297	-1,6	339	341	0,5	315	322	2,2
327	343	5	338	322	-4,5	353	340	-3,8	354	342	-3,5
305	306	0,4	298	311	4,4	314	324	2,9	374	359	-3,9
331	335	1,3	288	289	0,5	338	341	0,9	359	358	-0,1
372	364	-2,3	309	310	0,3	324	331	1,9	374	356	-4,6
327	330	0,8	309	313	1,6	333	344	3,1	364	352	-3,2
307	309	0,7	316	307	-2,7	362	363	0,4	347	350	0,9
345	337	-2,3	326	319	-2	366	371	1,3	364	366	0,5
324	327	0,8	302	297	-1,6	381	370	-2,8	355	358	1
309	314	1,7	312	310	-0,6	349	349	0,2	377	369	-2,3
357	344	-3,5	363	364	0,3	340	344	1,1	356	348	-2,3
323	311	-3,8	363	352	-3,1	327	325	-0,8	365	368	0,8
334	350	5	289	294	1,9	327	332	1,5	380	368	-3,1
351	346	-1,3	320	321	0,5	329	317	-3,5	358	365	2
351	345	-1,8	317	317	0,1	331	325	-1,7	347	347	-0,1
289	295	2,3	295	299	1,4	341	337	-1,1	351	349	-0,7
309	326	5,3	304	314	3,2	346	362	4,6	348	350	0,6
281	284	1,2	286	299	4,5	373	374	0,3	369	367	-0,5
325	342	5	295	304	2,8	362	361	-0,3	336	347	3,2
313	317	1,3	343	350	2	365	362	-0,9	359	366	2
346	357	3,1	344	339	-1,4	320	323	0,9	361	364	0,6
361	368	1,8	313	316	1,1	381	365	-4,1	371	370	-0,4
348	341	-2,1	343	329	-4,1	345	349	1,3	347	352	1,4
296	298	0,6	326	322	-1,4	369	364	-1,2	363	364	0,5
380	365	-3,9	350	348	-0,5	354	349	-1,4	356	354	-0,6
343	337	-1,7	297	289	-2,7	344	358	4,3	342	339	-0,9
300	296	-1,6	286	287	0,1	374	370	-1	326	327	0,3
351	343	-2,3	316	322	1,9	372	363	-2,3	322	332	3
273	271	-0,7	360	355	-1,3	339	338	-0,2	322	322	0,2
321	309	-3,8	378	370	-2,1	359	351	-2,1	284	281	-0,8
331	324	-1,8	374	365	-2,3	344	344	-0,2	359	353	-1,7
317	323	1,8	367	361	-1,5	335	348	3,9	371	359	-3,1
284	274	-3,6	370	375	1,1	364	361	-0,8	391	379	-2,9
330	330	-0,1	364	357	-2,1	335	350	4,5	388	387	-0,3
327	334	2	366	364	-0,6	362	362	0,2	403	402	-0,2
358	343	-4,3	303	315	3,8	362	351	-3,1	232	235	1,1
269	275	2,5	297	309	3,8	351	352	0,2	260	261	0,5
308	300	-2,8	256	268	4,7	379	376	-0,7	286	289	1,3
372	357	-3,8	327	331	1,4	334	342	2,5	272	277	1,9
333	334	0,4	304	316	4	361	364	1	259	258	-0,6
290	284	-2	314	328	4,5	368	374	1,5	399	411	3
323	318	-1,8	304	320	5,3	366	368	0,6	352	343	-2,6
356	361	1,5	349	352	0,8	362	360	-0,4	261	259	-0,9
352	352	0	304	311	2,2	367	354	-3,6	390	374	-3,9
348	335	-3,9	294	307	4,3	354	350	-1,2	382	378	-0,9

**ANEXO F:** Valores de referência, valores preditos e erros relativos de amostras de validação para o parâmetro lignina.

VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)
8,0	7,8	-2,1	6,3	6,1	-2,3	5,1	6,2	22,9	9,7	8,4	-13,2
5,0	6,5	29,0	3,9	4,1	4,5	7,5	6,1	-18,2	8,4	7,9	-5,9
5,6	5,3	-4,4	6,1	6,1	-0,1	6,6	6,4	-3,7	9,4	8,2	-12,6
7,5	7,4	-1,0	5,7	5,4	-5,5	5,8	6,9	20,1	7,3	8,5	17,1
7,0	6,8	-2,7	4,6	3,7	-18,7	8,9	9,0	0,5	6,0	7,5	25,4
6,6	6,2	-6,9	4,8	5,5	13,6	7,5	7,7	3,8	8,1	6,8	-15,5
7,2	7,6	5,2	5,6	4,7	-17,0	7,9	8,1	1,6	6,4	6,0	-5,2
6,7	6,3	-6,2	5,6	5,6	-0,3	6,4	6,8	5,1	4,8	5,9	22,9
7,7	7,6	-1,9	5,4	4,6	-15,0	7,4	7,4	-0,6	5,9	6,4	7,7
5,8	5,6	-3,0	4,8	4,2	-11,1	4,5	6,0	32,4	6,3	5,9	-6,7
7,3	5,9	-19,0	6,2	6,2	-0,8	7,9	7,2	-8,7	7,8	7,6	-1,8
6,5	6,6	1,9	7,4	7,1	-3,8	5,9	6,3	7,6	6,8	6,5	-4,9
7,0	7,9	13,9	6,7	6,3	-5,4	6,9	6,8	-2,1	6,7	6,8	0,9
6,0	5,8	-2,0	4,8	4,5	-5,7	8,4	7,8	-6,7	7,8	8,6	10,5
6,2	7,0	13,3	5,9	5,5	-7,7	8,4	9,2	9,6	4,6	5,6	22,0
8,9	8,5	-4,4	7,0	6,1	-12,2	5,6	6,7	20,3	7,8	9,0	15,3
6,2	6,1	-1,5	4,8	6,0	25,2	7,9	7,2	-8,5	5,4	6,1	11,9
6,6	6,7	1,6	6,0	5,4	-10,7	8,0	8,9	10,6	7,9	7,8	-0,8
5,4	6,1	13,9	5,8	6,1	4,9	6,7	7,1	5,8	8,4	8,6	1,6
7,2	7,1	-0,4	3,7	4,8	31,1	7,1	8,4	18,4	8,5	8,1	-4,8
6,1	6,7	9,7	7,9	8,0	2,0	7,3	6,2	-14,8	7,4	8,3	13,3
6,2	7,5	21,1	5,8	4,8	-16,8	6,2	5,9	-4,5	8,5	7,0	-17,6
5,8	6,2	7,5	4,1	4,5	8,0	9,4	8,6	-8,7	8,7	8,0	-8,5
8,3	7,9	-4,6	5,6	5,8	3,9	9,3	8,6	-7,4	9,0	8,7	-3,5
7,6	8,1	5,5	6,9	5,5	-19,6	9,1	9,1	-0,2	7,8	7,1	-9,7
7,4	7,5	1,4	5,1	4,9	-4,8	5,7	7,0	22,7	7,9	8,8	10,2
6,7	5,8	-13,3	4,5	5,0	10,3	7,4	6,5	-11,8	9,0	8,5	-6,3
7,5	6,8	-9,1	5,8	5,8	0,1	8,3	7,1	-14,7	9,2	8,2	-10,3
4,8	5,4	11,8	6,7	5,7	-15,1	6,4	5,8	-10,2	7,4	6,8	-8,3
7,0	6,9	-1,6	4,5	5,4	19,5	7,3	7,0	-3,4	8,1	7,3	-10,5
6,5	6,3	-3,4	5,8	5,2	-10,3	7,0	7,1	2,2	8,5	7,3	-13,8
9,8	8,7	-10,9	6,3	6,3	0,7	8,8	8,2	-6,8	9,3	8,1	-12,5
7,1	7,3	2,6	3,9	5,2	32,5	7,9	8,8	11,8	8,5	7,4	-13,0
5,5	5,4	-2,9	5,5	6,4	15,9	8,8	8,8	-0,2	9,3	8,4	-9,7
6,7	7,8	15,7	5,6	5,5	-2,3	7,3	7,9	8,6	7,7	7,3	-6,1
6,5	7,4	13,8	5,5	5,9	7,6	7,7	8,1	5,6	8,1	8,3	2,5
9,2	7,9	-13,5	6,5	5,6	-13,0	6,5	5,7	-11,4	8,0	7,7	-4,2
3,8	4,2	10,6	6,3	4,8	-23,8	6,4	5,6	-13,0	6,8	6,9	0,7
7,0	5,9	-15,6	4,7	5,3	13,7	6,1	6,0	-1,5	5,1	5,9	15,7
7,0	6,7	-3,6	4,5	5,8	27,6	7,1	7,3	2,2	5,4	4,7	-13,0
4,8	6,2	27,8	4,8	5,7	18,1	6,6	8,0	21,5	6,9	7,4	6,1
6,1	7,0	13,8	7,5	7,6	1,5	6,5	7,4	13,6	6,6	7,6	15,2
7,0	6,8	-3,5	8,8	8,8	-0,1	8,4	7,2	-15,1	7,3	6,6	-9,7
3,8	4,6	21,1	10,8	9,3	-13,5	8,1	8,9	10,0	8,6	8,8	2,1
5,4	5,4	-1,0	8,9	7,9	-11,6	7,6	9,0	18,2	4,2	3,7	-13,5
6,4	6,8	6,4	7,2	6,8	-5,7	7,1	8,6	20,8	5,8	5,3	-8,0
6,2	6,6	6,2	8,0	7,7	-3,4	5,9	6,7	13,2	5,1	3,7	-28,7
3,4	4,7	37,6	4,3	5,5	28,4	7,3	7,6	4,4	4,1	3,7	-8,9
5,1	6,5	28,2	4,4	5,0	14,7	7,5	7,4	-1,2	9,1	8,6	-5,3
8,7	8,0	-7,6	5,5	5,4	-1,7	8,6	7,6	-11,3	8,2	7,2	-11,8
8,8	8,2	-7,1	4,6	4,2	-9,0	8,6	8,4	-2,7	8,4	9,5	13,4
7,7	6,5	-15,3	4,8	5,5	13,6	7,1	6,6	-7,2	3,9	4,1	5,2
5,3	6,6	26,2	5,2	6,1	18,6	8,3	8,3	-0,2	7,3	8,3	14,7

**ANEXO G:** Valores de referência, valores preditos e erros relativos de amostras de validação para o parâmetro brix.

<b>VR</b> <b>(°B)</b>	<b>VP</b> <b>(°B)</b>	<b>ER</b> <b>(%)</b>									
14,3	14,4	0,7	13,9	14,0	0,4	8,3	8,8	6,3	13,9	14,2	2,0
15,3	14,8	-3,5	12,4	12,5	1,1	11,3	11,4	1,3	10,4	10,0	-3,7
13,0	13,5	4,1	12,1	12,5	3,2	10,3	10,8	5,2	14,0	14,0	-0,3
13,3	13,4	0,5	12,7	12,9	1,4	9,6	10,1	5,2	7,5	7,3	-3,0
12,0	11,7	-2,2	13,1	13,0	-0,5	11,0	11,4	4,0	14,3	14,3	-0,3
13,2	12,9	-2,5	11,8	12,3	4,3	9,6	9,9	2,8	6,9	7,6	9,8
12,5	12,6	1,0	12,9	13,2	2,6	8,6	9,1	6,2	11,5	11,5	-0,4
14,7	14,5	-1,5	15,3	15,4	0,6	7,7	8,2	6,4	10,7	11,3	6,0
11,6	11,9	2,2	14,5	14,5	0,0	8,7	8,6	-0,7	11,0	10,6	-4,0
16,7	16,1	-3,8	13,6	13,9	2,0	11,9	11,5	-3,5	11,8	11,3	-3,9
13,7	13,2	-3,5	12,9	13,3	3,4	13,4	13,0	-2,7	11,5	10,9	-4,8
14,2	13,8	-2,7	14,9	15,3	2,9	9,8	10,5	6,7	9,2	9,3	0,8
10,2	10,8	5,8	13,4	14,1	4,9	7,7	7,5	-3,2	11,0	11,2	2,1
14,5	14,5	0,3	12,3	12,4	1,0	10,1	10,4	3,4	8,0	8,2	2,5
14,3	14,6	2,3	14,9	15,2	2,2	9,2	9,2	0,1	9,9	10,1	1,6
13,9	14,5	4,6	13,4	13,7	2,3	10,8	10,8	-0,1	8,3	8,6	3,8
15,7	16,0	1,8	14,9	15,2	1,9	10,0	9,8	-2,2	11,3	11,3	0,2
15,6	16,1	3,0	12,1	12,4	2,7	9,5	9,5	0,1	9,2	9,6	3,9
14,6	15,1	3,7	13,0	13,3	2,4	11,1	11,8	6,4	8,6	7,9	-8,3
13,0	13,1	0,5	11,4	12,0	4,9	10,6	10,5	-1,1	9,4	9,1	-3,0
12,3	12,2	-0,6	12,8	13,0	1,4	10,2	10,2	0,0	12,0	11,5	-3,9
14,1	14,3	1,1	12,7	13,0	2,4	12,8	12,7	-0,6	12,6	12,2	-2,9
12,8	12,3	-3,7	13,1	13,4	2,1	11,9	12,0	0,5	11,9	11,5	-3,3
10,9	11,3	3,9	10,3	9,9	-3,7	13,9	13,5	-3,2	8,0	8,4	4,4
12,9	13,1	1,8	13,6	14,3	4,9	6,1	5,9	-3,1	9,1	8,9	-2,6
15,5	15,2	-2,2	14,9	14,6	-1,8	9,7	9,9	1,9	9,0	9,0	-0,3
15,0	15,5	3,6	11,8	11,1	-5,6	13,1	12,8	-2,2	14,2	14,5	1,8
14,4	14,6	1,3	10,9	10,8	-1,2	4,7	5,2	11,0	12,3	12,2	-1,0
12,0	11,7	-2,1	8,5	8,3	-2,8	10,0	10,2	2,4	10,0	9,6	-4,0
16,2	16,1	-0,5	12,2	11,5	-5,5	11,5	11,3	-1,6	10,2	10,8	5,6
11,1	11,7	5,3	8,1	8,7	7,8	9,7	10,4	7,3	12,6	11,9	-5,7
13,1	12,8	-1,9	8,0	8,5	6,5	7,5	7,7	3,1	11,0	10,9	-0,7
11,8	11,3	-3,9	9,5	9,4	-0,9	13,1	13,7	4,5	13,2	13,3	0,6
12,3	12,0	-2,4	11,2	11,1	-0,9	8,1	8,7	7,2			

**ANEXO H:** Valores de referência, valores preditos e erros relativos de amostras de validação para o parâmetro leitura sacarimétrica.

<b>VR</b> <b>(°Z)</b>	<b>VP</b> <b>(°Z)</b>	<b>ER</b> <b>(%)</b>									
35,3	35,9	1,7	34,1	40,2	18,0	20,7	22,6	9,1	29,6	31,6	7,0
39,8	37,7	-5,2	29,1	35,0	20,6	20,3	24,1	18,8	10,5	13,4	27,7
22,0	24,6	11,9	29,8	30,4	2,0	10,7	13,4	24,8	34,0	34,3	1,0
29,9	29,1	-2,7	35,5	37,7	6,3	23,5	23,4	-0,1	19,5	17,8	-8,7
23,6	29,0	23,0	32,3	38,2	18,2	18,4	21,4	16,6	35,1	33,2	-5,5
36,5	35,6	-2,4	36,8	31,1	-15,4	20,6	22,4	8,6	7,1	7,5	7,0
46,0	42,5	-7,5	28,2	33,8	19,5	11,7	13,9	19,6	31,5	34,5	9,5
35,2	31,9	-9,2	45,1	43,1	-4,6	15,7	18,7	19,3	4,5	5,8	28,1
30,9	33,5	8,2	43,0	38,6	-10,2	12,6	14,6	16,7	19,3	21,3	10,7
23,0	23,6	2,3	34,2	35,3	3,3	11,2	13,1	17,6	25,1	23,6	-6,2
41,3	37,9	-8,3	30,8	33,8	9,7	8,7	8,3	-4,6	26,9	23,5	-12,3
38,5	38,9	1,0	40,9	41,3	1,2	10,1	10,7	5,9	20,4	15,4	-24,6
38,2	37,7	-1,3	33,4	36,3	8,7	26,4	25,5	-3,3	24,7	23,6	-4,4
39,8	43,1	8,4	26,0	29,0	11,3	28,1	30,3	7,9	10,9	9,2	-15,7
33,8	36,6	8,3	38,3	40,4	5,5	14,4	16,7	16,4	13,7	10,4	-24,3
42,4	43,0	1,3	30,2	33,5	11,1	10,9	13,9	26,9	14,4	18,0	24,7
31,7	32,9	3,8	39,6	37,9	-4,2	9,9	7,6	-23,6	8,3	10,1	21,7
30,9	28,1	-9,0	25,2	26,3	4,3	17,0	20,7	21,8	22,8	19,9	-12,5
32,6	37,5	15,0	28,7	32,4	12,7	9,2	11,6	26,9	30,2	25,4	-16,0
26,1	28,9	10,6	25,1	28,4	13,4	15,9	18,7	17,6	18,7	16,9	-9,4
32,7	28,9	-11,9	10,5	12,9	23,5	20,1	18,0	-10,3	14,3	11,7	-18,3
20,2	25,3	24,8	28,2	28,7	1,8	21,9	25,7	16,9	18,4	14,9	-18,7
27,9	32,3	16,0	12,7	16,3	28,6	19,5	16,3	-16,7	28,5	23,9	-16,3
44,6	42,2	-5,5	29,9	34,0	13,6	16,7	17,4	4,4	28,4	26,0	-8,5
40,0	38,0	-5,0	23,0	21,3	-7,3	24,2	27,2	12,2	27,2	25,3	-7,0
27,4	26,4	-3,7	33,7	27,5	-18,3	21,6	26,3	22,0	9,9	12,2	23,0
45,6	44,5	-2,5	19,3	18,2	-5,6	6,8	5,8	-15,4	13,5	12,8	-5,1
30,6	31,7	3,5	14,4	11,3	-21,4	11,3	12,4	10,0	14,8	14,7	-0,1
29,9	24,9	-16,7	26,3	22,3	-15,5	18,1	17,2	-5,0	35,9	35,6	-0,9
27,4	27,8	1,6	23,7	22,0	-7,2	28,7	30,0	4,6	33,8	29,5	-12,6
31,6	30,4	-3,7	13,4	12,6	-5,7	22,2	21,1	-5,0	15,0	17,8	18,6
39,6	33,8	-14,7	8,6	10,7	23,4	24,7	23,8	-3,8	22,3	21,3	-4,6
28,2	27,6	-2,1	18,1	15,9	-11,9	23,1	21,0	-9,2	30,4	25,1	-17,3
25,0	29,7	19,0	22,2	21,6	-2,7	6,6	8,3	25,3	19,2	21,6	12,4

**ANEXO I:** Valores de referência, valores preditos e erros relativos de amostras de validação para o parâmetro pol.

VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)
8,8	9,0	2,4	7,3	8,7	19,4	5,1	6,2	20,1	2,7	3,4	25,7
9,8	9,5	-3,7	7,4	7,5	0,9	2,7	3,4	23,4	8,5	8,6	1,5
5,5	6,2	11,6	8,8	9,5	7,2	5,9	5,9	-0,7	4,9	4,5	-9,7
7,5	7,2	-3,3	8,1	9,6	18,7	4,6	5,5	18,0	8,7	8,3	-4,8
5,9	7,3	23,2	9,2	7,8	-14,6	5,2	5,7	9,6	1,8	1,9	4,6
9,0	8,9	-1,9	7,1	8,5	19,8	3,0	3,6	19,6	7,8	8,6	10,6
11,3	10,6	-6,5	11,2	10,8	-3,2	4,0	5,1	29,3	1,2	1,5	30,8
8,8	8,0	-8,2	10,7	9,7	-9,4	3,2	4,1	28,5	4,8	5,4	10,7
7,7	8,3	8,3	8,5	8,8	3,2	2,8	3,2	11,2	6,3	6,0	-5,8
5,8	5,9	1,3	7,7	8,4	9,8	2,2	2,0	-8,5	6,7	5,9	-12,0
10,2	9,4	-8,2	10,1	10,3	1,9	2,6	2,6	3,3	5,2	3,9	-23,6
9,6	9,7	1,9	8,3	9,2	10,4	6,6	6,4	-2,5	6,2	6,0	-3,9
9,5	9,4	-1,1	6,5	7,3	11,9	7,0	7,6	8,3	2,8	2,3	-15,4
9,8	10,7	9,3	9,5	10,2	7,5	3,6	4,6	26,0	3,5	2,4	-30,0
8,4	9,1	8,0	7,5	8,4	11,9	2,8	3,6	28,6	3,7	4,5	24,5
10,5	10,7	1,8	9,8	9,5	-3,0	2,5	1,9	-25,7	2,1	2,8	30,9
7,9	8,2	3,3	6,3	6,6	4,3	4,3	5,2	20,8	5,7	5,0	-12,9
7,7	7,0	-10,0	7,2	8,2	13,7	2,3	3,0	30,2	7,6	6,4	-16,1
8,1	9,4	15,4	6,3	7,2	14,8	4,0	4,7	16,1	4,7	4,3	-9,9
6,5	7,2	9,6	2,7	3,0	14,1	5,1	4,5	-11,5	3,7	2,6	-29,0
8,2	7,1	-13,3	7,1	7,2	2,6	5,5	6,5	18,4	4,7	3,8	-19,3
5,1	6,3	23,8	3,2	4,1	27,8	4,9	4,0	-18,4	7,2	6,1	-15,3
7,0	8,1	15,7	7,4	8,6	16,6	4,2	4,4	4,8	7,1	6,5	-8,4
11,1	10,5	-4,7	5,8	5,4	-6,9	6,1	6,8	12,4	6,8	6,4	-6,3
9,9	9,5	-4,2	8,3	6,9	-17,5	5,4	6,7	23,6	2,5	3,1	23,7
6,9	6,5	-4,9	4,9	4,6	-5,9	1,8	1,2	-30,9	3,4	3,3	-5,2
11,2	11,1	-1,1	3,7	2,8	-24,0	2,9	3,2	9,9	3,8	3,8	1,5
7,6	7,9	3,6	6,6	5,6	-15,1	4,6	4,3	-5,8	8,9	9,0	0,5
7,5	6,1	-18,3	5,9	5,6	-5,7	7,2	7,5	4,8	8,5	7,5	-11,7
6,9	6,9	0,2	3,4	3,2	-6,2	5,6	5,3	-6,0	3,8	4,5	19,4
8,0	7,5	-5,3	2,2	2,8	26,4	6,2	5,9	-4,5	5,6	5,3	-5,0
9,9	8,3	-15,3	4,6	4,0	-11,8	5,8	5,2	-10,8	7,6	6,2	-17,9
7,1	6,8	-3,7	5,6	5,5	-1,7	1,7	2,2	29,5	4,8	5,5	13,5
6,3	7,3	17,6	5,2	5,7	8,0	7,4	7,9	7,5	8,1	8,0	-1,0

**ANEXO J:** Valores de referência, valores preditos e erros relativos de amostras de validação para o parâmetro pureza aparente.

VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)
76,8	81,1	5,6	63,6	53,1	-16,5	54,1	48,8	-9,7	56,1	45,1	-19,7
70,6	81,3	15,1	55,7	55,1	-1,0	49,4	46,5	-6,0	54,0	47,5	-12,1
75,8	84,9	12,0	70,9	60,8	-14,2	42,2	37,1	-12,1	51,4	44,4	-13,6
45,7	51,4	12,4	56,9	55,4	-2,6	27,6	35,1	27,1	33,8	43,9	30,1
63,1	64,0	1,5	51,7	58,0	12,2	48,1	42,7	-11,2	22,7	29,3	29,4
61,2	61,1	-0,3	60,5	68,0	12,4	49,8	48,0	-3,7	56,3	53,7	-4,7
64,3	65,2	1,5	55,4	62,0	12,0	54,0	46,0	-14,8	33,2	38,0	14,4
45,9	52,5	14,3	58,6	56,9	-3,0	49,3	46,9	-4,9	60,8	58,5	-3,8
56,5	56,9	0,7	65,0	65,4	0,6	32,9	36,3	10,4	47,4	41,6	-12,3
47,2	56,1	19,0	59,2	65,0	9,7	52,2	46,7	-10,5	62,3	57,6	-7,6
61,5	61,7	0,3	70,0	60,0	-14,3	45,1	47,2	4,8	24,1	32,2	33,5
67,7	67,9	0,4	52,4	60,1	14,7	54,4	49,9	-8,2	54,6	58,1	6,4
63,9	60,5	-5,4	56,5	62,1	9,9	39,2	40,6	3,4	42,1	43,0	2,1
54,1	60,0	11,0	72,9	70,8	-2,9	35,9	43,8	22,2	59,2	48,2	-18,6
57,1	52,6	-7,8	73,5	65,1	-11,5	37,2	39,1	5,2	57,1	48,9	-14,4
70,6	65,6	-7,1	62,6	62,6	0,0	33,0	40,5	22,7	56,4	47,9	-15,1
66,9	66,1	-1,2	59,6	61,2	2,6	29,0	30,0	3,5	34,7	31,5	-9,0
68,3	65,4	-4,2	67,9	67,5	-0,6	29,5	33,0	12,2	38,7	35,7	-7,8
62,5	69,7	11,5	62,0	62,6	0,9	55,6	49,1	-11,7	36,9	43,1	17,0
62,4	63,5	1,8	53,0	54,5	2,8	52,2	53,7	2,9	51,6	46,1	-10,6
67,1	71,0	5,8	63,7	67,9	6,7	37,2	42,8	15,1	42,4	34,5	-18,6
60,3	70,0	16,1	56,1	61,0	8,8	32,8	38,8	18,2	49,5	40,0	-19,1
60,8	60,3	-0,9	62,1	54,9	-11,6	33,0	31,4	-4,7	59,6	49,0	-17,8
62,9	55,7	-11,4	65,8	64,6	-1,8	42,4	47,3	11,5	56,4	48,7	-13,6
57,5	62,1	8,1	52,1	52,0	-0,2	37,2	40,1	7,8	57,4	50,0	-12,8
51,0	55,0	7,7	55,2	60,7	10,0	50,8	43,0	-15,4	31,6	35,3	11,6
60,9	54,5	-10,4	48,9	54,9	12,2	49,7	50,1	0,8	37,8	38,0	0,5
46,8	52,8	12,8	55,5	54,4	-2,0	46,5	40,8	-12,1	41,6	38,5	-7,5
54,0	57,4	6,3	46,3	55,3	19,5	41,4	41,6	0,5	62,8	59,9	-4,6
75,1	65,2	-13,2	46,6	56,6	21,6	47,3	51,7	9,3	37,9	41,7	10,1
73,7	68,4	-7,2	50,5	61,1	21,0	45,5	51,0	12,1	55,2	45,1	-18,4
68,9	64,8	-6,0	49,7	58,7	18,2	28,8	27,8	-3,4	44,0	44,6	1,3
57,3	54,9	-4,2	48,9	46,3	-5,5	40,1	35,2	-12,2	61,1	55,9	-8,6
69,4	72,5	4,6	56,7	54,0	-4,8	47,2	39,1	-17,2	62,4	54,5	-12,6

**ANEXO K:** Valores de referência, valores preditos e erros relativos de amostras de validação para o parâmetro açúcares redutores.

VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)	VR (%)	VP (%)	ER (%)
1,5	1,3	-17,7	3,5	3,0	-13,0	3,3	2,7	-19,7	2,5	2,8	11,9
1,5	1,7	15,1	4,3	4,1	-3,1	2,0	2,4	20,7	4,2	3,9	-7,4
1,5	1,6	2,3	3,7	4,2	12,1	3,1	3,1	-1,0	2,3	2,3	2,4
1,3	1,3	-2,0	3,1	2,9	-9,4	2,6	2,8	8,3	2,2	2,6	19,2
1,5	1,2	-23,6	3,4	3,4	0,6	3,8	3,6	-5,6	2,6	2,8	8,5
1,7	1,8	6,8	3,7	3,4	-8,9	3,7	3,0	-19,0	2,1	2,2	6,6
4,7	4,2	-10,4	5,3	4,5	-14,4	3,3	3,6	6,6	2,4	2,7	10,3
3,7	3,0	-18,4	3,6	3,3	-8,3	3,8	3,1	-16,8	2,8	2,9	5,1
3,4	3,9	15,0	3,8	3,5	-9,1	3,1	3,5	10,0	3,2	3,4	4,8
3,1	3,0	-0,7	1,5	1,8	20,9	3,6	3,8	5,8	3,3	3,6	10,0
4,4	4,4	-0,7	1,7	2,3	40,5	3,7	3,5	-5,2	4,1	3,8	-5,5
5,3	4,7	-10,1	2,0	2,2	13,5	4,7	4,2	-8,9	4,1	3,9	-6,5
4,0	3,8	-4,6	2,4	2,2	-7,2	2,4	2,6	10,5	3,7	4,0	6,3
4,6	4,2	-9,0	4,0	3,6	-11,8	1,8	1,8	2,4	2,9	3,4	14,6
3,6	3,4	-6,2	3,9	3,6	-6,2	3,0	3,1	3,2	3,6	3,2	-9,4
3,3	3,3	0,8	3,7	3,2	-12,8	3,0	3,7	24,3	3,3	3,1	-7,4
4,9	5,3	8,8	2,5	3,2	28,0	1,5	2,0	29,8	3,3	3,3	1,6
2,4	2,3	-3,3	3,4	3,2	-5,3	1,6	2,1	27,7	3,3	2,7	-17,5
2,4	2,9	23,6	2,9	3,2	9,9	1,9	2,1	7,7	3,4	3,6	4,9
3,2	3,6	11,0	4,1	3,5	-14,6	2,2	2,4	7,4	3,6	3,3	-8,2
3,0	3,2	7,4	2,9	3,1	5,6	3,1	2,7	-14,0	3,2	3,4	5,3
3,6	3,3	-6,7	3,3	3,0	-11,3	1,6	1,7	4,9	3,0	3,4	14,7
4,4	4,0	-9,6	4,1	3,8	-7,1	1,7	2,0	17,8	2,8	2,6	-6,7
2,6	2,6	-1,9	4,3	4,0	-7,8	2,1	2,5	18,3	3,2	2,5	-23,2
3,3	3,6	10,2	3,3	3,1	-8,3	2,0	2,0	-2,6	2,5	2,5	-1,3
1,8	2,5	37,0	3,6	3,3	-6,1	3,3	3,5	7,6	3,6	3,1	-14,1
2,4	2,5	4,9	2,4	2,8	18,1	2,4	2,3	-3,3	3,0	2,8	-6,6
2,1	2,6	24,4	3,1	2,4	-24,2	3,2	2,7	-15,4	2,5	2,5	-0,2
3,6	3,5	-3,2	2,2	2,9	31,2	2,0	2,2	7,5	4,0	3,5	-12,6
3,6	2,8	-20,7	3,2	3,0	-6,5	3,5	3,2	-7,4	3,7	3,3	-8,7
2,4	2,2	-7,0	2,4	2,7	12,1	2,9	3,0	2,7	3,0	3,3	10,9
2,3	1,7	-26,6	3,0	2,4	-19,0	3,2	2,6	-18,1			
3,5	3,0	-14,2	3,9	3,3	-14,0	3,2	3,3	1,9			