

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA PROGRAMA
DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

ARIANE SILVEIRA SBRICE PINTO

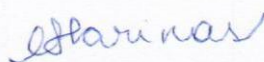
**MONITORAMENTO POR ESPECTROSCOPIA DOS COMPOSTOS
FENÓLICOS E FURALDEÍDOS GERADOS NO PROCESSAMENTO DE
BIOMASSA LIGNOCELULÓSICA**

São Carlos- SP

2017

MEMBROS DA BANCA EXAMINADORA DA DEFESA DE DISSERTAÇÃO DE ARIANE SILVEIRA SBRICE PINTO APRESENTADA AO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA DA UNIVERSIDADE FEDERAL DE SÃO CARLOS, EM 20 DE FEVEREIRO DE 2017.

BANCA EXAMINADORA:



Cristiane Sanchez Farinas
Orientadora, EMBRAPA/UFSCar



Marcelo Perencin de Arruda Ribeiro
UFSCar



Roberto de Campos Giordano
UFSCar



Marcel Otávio Cerri
UNESP

*Dedico este trabalho aos
meus pais, pelo apoio e
incentivo durante os anos de
dedicação.*

AGRADECIMENTOS

A Deus, pelas bênçãos concedidas nesses anos de graduação, pós-graduação, estudo, trabalho e em toda a minha vida;

À minha família que todos os dias me faz entender o sentido verdadeiro da palavra *amor*: em especial, aos meus pais Cinira e Rene, que sempre me incentivaram, apoiaram e acreditaram em meu sucesso;

Aos meus orientadores, Dra. Cristiane e Dr. Marcelo, pelo apoio, orientação, compreensão, dedicação e paciência durante a elaboração deste trabalho;

Aos amigos queridos pela paciência e companheirismo, especialmente ao Vitor, que contribuiu para a realização deste trabalho. Eles serão para sempre lembrados com muito carinho;

À UFSCar, ao Programa de Pós-Graduação em Engenharia Química e à Embrapa Instrumentação pelo auxílio e contribuição à minha formação;

Ao corpo técnico dos laboratórios da Embrapa Instrumentação e do Departamento de Engenharia Química, em especial para o Amadeus e para a Joana;

Aos professores que me proporcionaram ao longo desses dois anos todo o conhecimento e contribuíram de maneira indireta para que este trabalho fosse desenvolvido;

Ao órgão de fomento CNPq pelo apoio financeiro sem o qual a execução deste trabalho não seria possível;

Enfim, a todos que contribuíram para a realização desta etapa na minha vida...

Muito obrigada!

RESUMO

Um dos principais desafios na produção do etanol de segunda geração (2G) está relacionado à caracterização dos hidrolisados da biomassa lignocelulósica que, frequentemente, contêm quantidades elevadas de compostos fenólicos e furaldeídos. Estes componentes são responsáveis por inibir e desativar enzimas no processo de hidrólise, além de influenciarem negativamente na etapa de fermentação. A quantificação dos compostos fenólicos e furaldeídos poderia auxiliar na elucidação das limitações destes bioprocessos. Conseqüentemente, seria possível otimizar a produção de etanol 2G com estratégias aplicadas ao processo, tornando-o mais eficiente, robusto e tolerante a estes inibidores. Neste contexto, tecnologias rápidas, eficientes e de baixo custo para o monitoramento dos furaldeídos e compostos fenólicos torna-se crucial na otimização das etapas de produção do etanol 2G. A fim de alcançar este propósito, avaliou-se a viabilidade de monitorar os principais inibidores presentes no hidrolisado a partir de técnicas quimiométricas. Espectroscopias nas regiões do Ultravioleta-Visível bem como do Infravermelho Próximo associadas à regressão por Mínimos Quadrados Parciais (do inglês, *Partial Least Squares* - PLS) foram avaliadas no monitoramento de inibidores originados no processo de pré-tratamento do material lignocelulósico da cana-de-açúcar. Amostras de hidrolisado obtidas no pré-tratamento hidrotérmico da biomassa, bem como amostras sintéticas foram consideradas em diferentes conjuntos de calibração e teste dos modelos PLS. A escolha dos analitos monitorados foi baseada no efeito inibitório provocado nas etapas de hidrólise enzimática e fermentação. Assim, analisaram-se modelos de previsão para a concentração de vanilina, hidroximetilfurfural, furfural, bem como para os ácidos felúrico, gálico e p-cumárico. Os resultados garantiram o monitoramento destes compostos no hidrolisado do pré-tratamento hidrotérmico do bagaço da cana-de-açúcar em três condições de operação. As previsões de concentração foram satisfatoriamente precisas em modelagens PLS-UV-Vis, apresentando modelos PLS com potencial aplicação no monitoramento destes analitos durante o processo (desvio padrão de previsão na validação cruzada *leave-one-out* (RMSECV) em torno de 3,0 a 9,0% e desvio residual de previsão (RPD) de 2,0 até 5,0).

Palavras-chave: Compostos Fenólicos. Etanol de Segunda Geração. Espectroscopia no Infravermelho Próximo. Espectroscopia no Ultravioleta-Visível. Regressão por Mínimos Quadrados Parciais.

ABSTRACT

The major process challenge of the second-generation ethanol (2G) production is related with characterization of hydrolyzate from lignocellulosic biomass, which often contains high quantities of phenolic compounds and furan derivatives. These components of hydrolyzate are responsible for inhibit and deactivate enzymes during hydrolysis in addition to negatively influence the fermentation step. The phenolic compounds and furaldehydes quantification could help to highlight the bioprocesses limitations. As a result, it could allow the process improvement that may be characterized by more productive, robust and tolerant to these compounds. Concerning about this objective, rapid, efficient, and low-cost technologies for monitoring the phenolic compounds and furan derivatives are essential for better control of the pretreatment, hydrolysis and fermentation steps during 2G ethanol production. For achieving that goal it was verified the viability of monitoring phenolic compounds and furaldehydes by the use of chemometric techniques. The Ultraviolet Visible and Near Infrared spectral regions were analyzed in association with Partial Least Squares (PLS) regression for monitoring the inhibitors from pretreatment hydrolyzate of sugarcane bagasse. Hydrolyzate samples from liquid hot water pretreatment of biomass plus synthetic samples were evaluated on distinct calibration and test trials. The negative effect in both hydrolysis and fermentation process were considered for monitoring components from hydrolyzate and synthetic mixtures. Then, the concentration of vanillin, hydroxymethylfurfural, furfural, as well as ferulic, gallic and p-coumaric acids were analyzed. It was found that the most accurate PLS model could be used to monitor phenolic compounds and furaldehydes during the liquid hot water pretreatment of lignocellulosic material from three different operating conditions. The best predicting concentrations provided satisfactory accuracy for each analyte by presenting PLS-UV-Vis models with potential for process monitoring (standard deviation of prediction for cross-validation leave-one-out (RMSECV) around 3.0 to 9.0% and residual predictive deviation (RPD) was from 2.0 up to 5.0).

Keywords: Phenolic compounds. Second-Generation Ethanol. Near Infrared Spectroscopy. Ultraviolet-Visible Spectroscopy. Partial Least Squares Regression.

LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama esquemático dos principais compostos químicos liberados no pré-tratamento da biomassa que são responsáveis por prejudicar o processo fermentativo e a hidrólise enzimática da celulose.	34
Figura 2 - Fluxograma simplificado do processo fermentativo realizado pelo microrganismo <i>Saccharomyces cerevisiae</i>	36
Figura 3 - Regiões de espectroscopia e comprimentos de onda característicos.	45
Figura 4 - Representação do princípio básico da espectroscopia na região do NIR	48
Figura 5 - Representação esquemática do efeito de transição eletrônica entre orbitais (S_i) de um átomo com a incidência de comprimentos de onda na região do UV-Vis.	51
Figura 6 – Fórmulas moleculares dos compostos químicos quantificados nas amostras S e LHW.	61
Figura 7 – Gráficos dos espectros originais no UV-Vis (a) e NIR (b) das amostras sintéticas puras.	71
Figura 8 - Gráficos dos espectros das amostras de vanilina no UV-Vis (a) e NIR (b) do conjunto S pré-processados com a primeira e segunda derivadas.	73
Figura 9 – Gráficos de absorvidade molar no UV-Vis (a) e NIR (b) dos compostos químicos estudados nas amostras sintéticas.	79
Figura 10 - Gráficos de absorvidade molar no UV-Vis (a) e NIR (b) de amostras sintéticas para os compostos analisados nos modelos LHW.	87
Figura 11 - Gráficos dos espectros pré-processados com AMM no UV-Vis (a) e NIR (b) das amostras LHW.	90
Figura 12 – Gráficos de validação cruzada (CV-LOO) dos melhores modelos PLS, dado a bissetriz de referência. Validação (V) dos modelos com a barra do desvio padrão de previsão na CV (RMSECV). Posto que, para cada analito, tem-se: VAN, PLS-UV-Vis- M_2 (a); FEL, PLS-UV-Vis- M_1 (b); APC, PLS-UV-Vis- M_2 (c); HMF, PLS-UV-Vis- M_2 (d); FUR, PLS-UV-Vis- M_0 (e).	92
Figura 13- Resumo dos melhores modelos PLS-UV-Vis de cada analito para a validação cruzada (LOO) (a) e para a validação externa de cada modelo, apresentando o desvio padrão de previsão (b) (vide Tabela 13).	97

LISTA DE TABELAS

Tabela 1 - Concentração de fenólicos ao final da etapa de hidrólise do bagaço de cana-de-açúcar pré-tratado com explosão a vapor.....	37
Tabela 2 - Efeito no crescimento celular e na produção de etanol para a levedura <i>S. cerevisiae</i> de acordo com a presença de diferentes inibidores.....	41
Tabela 3 - Resumo do efeito inibitório e desativador de enzimas celulolíticas dos compostos fenólicos durante a hidrólise enzimática.....	43
Tabela 4 - Condições para aquisição de dados do hidrolisado advindo do pré-tratamento hidrotérmico (LHW) do bagaço de cana-de-açúcar (P_{sat}).....	60
Tabela 5 - Faixa de validade, média e desvio padrão dos dados considerados nos modelos PLS-UV-Vis para dado número de amostras (NA) utilizadas na calibração e validação dos modelos sintéticos.....	72
Tabela 6 - Faixa de validade, média e desvio padrão dos dados considerados nos modelos PLS-NIR para dado número de amostras (NA) utilizadas na calibração e validação dos modelos sintéticos.....	72
Tabela 7 - Modelos PLS-UV-Vis obtidos para o conjunto de amostras sintéticas.....	75
Tabela 8 - Modelos PLS-UV-Vis obtidos para o conjunto de amostras sintéticas pré-tratadas por MC e 1ª Derivada (PLS-UV-Vis- M_1).....	76
Tabela 9 - Modelos PLS-NIR obtidos para o conjunto de amostras sintéticas.....	77
Tabela 10 – Concentrações dos analitos no hidrolisado LHW aferidas por HPLC e o desvio padrão de amostragem ($DP^{ESTIMADO}$ - considerando o desvio padrão (DP) médio de 4 amostras distintas em cada hidrolisado) associado à quantificação concomitantemente às diluições das amostras.....	83
Tabela 11 - Faixa de validade, média, desvio padrão da média e a quantidade de amostras utilizadas na calibração e validação dos modelos PLS-UV-Vis- NIR com dados LHW.....	84
Tabela 12 - Modelos PLS-NIR obtidos para o conjunto de amostras LHW.....	85
Tabela 13 - Modelos PLS-UV-Vis obtidos para o conjunto de amostras LHW.....	89
Tabela 14- Resumo dos modelos PLS-UV-Vis aplicados no monitoramento <i>off-line</i> dos analitos presentes no hidrolisado advindo do tratamento hidrotérmico da biomassa de cana-de-açúcar.....	96

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

\bar{a}_j Média dos dados de absorvância no comprimento de onda j e i -ésima amostra

c Velocidade da luz (3×10^{10} m/s)

$\hat{c}_k^{(i)}$ Concentração predita na validação do modelo referente à amostra j e analito i

E_{v_i} Energia vibracional associada com o i -ésimo nível quântico de uma molécula

$E_{\text{eletrônica}}$ Energia eletrônica de uma molécula

$E_{\text{rotacional}}$ Energia rotacional de uma molécula

$E_{\text{vibracional}}$ Energia vibracional de uma molécula

h_{crit} Leverage crítico

ν_0 Frequência vibracional fundamental de determinada molécula

E_A Matriz $nd_{\text{cal.}} \times n_\lambda$ dos erros espectrais não ajustados pelo modelo de regressão linear

E_c Matriz $nd_{\text{cal.}} \times n_c$ dos erros de concentração não ajustados pelo modelo de regressão linear

E_0 Energia de ligação no nível fundamental

E_i i -ésimo nível de energia vibracional de determinada molécula

E_{total} Energia potencial total de uma molécula

I_0 Intensidade da radiação incidente na amostra

S_i i -ésimo orbital molecular

a_{ij} Absorvância no comprimento de onda j para a i -ésima amostra

e^- Elétron

n_c Quantidade de componentes químicos considerados na regressão linear

nd_{cal} . Número de amostras/dados utilizados na calibração do modelo de regressão linear

n_λ Quantidade de comprimentos de onda discretos

r_i Comprimento de ligação entre os átomos de determinada molécula

$c_k^{(i)}$ Concentração de referência na validação do modelo referente à amostra j e analito i

h_{ii} Vetor correspondente ao Leverage da amostra i (diagonal da matriz H)

1ª Deriv. Espectro com primeira derivada

2ª Deriv. Espectro com segunda derivada

2G Segunda geração

A Matriz de $nd_{cal} \times n_\lambda$, espectros da resposta do instrumento

Abs Absorbância da amostra

ANN *Artificial Neural Networks*

APC Ácido p-cumárico

C Concentração molar de determinado analito na solução sintética (cálculo da absorvidade)

C Matriz $nd_{cal} \times n_c$ de concentração de componentes

CLS *Classical Least Squares*

CV-LOO Validação cruzada *leave-one-out* dos modelos PLS (dados de calibração)

d Caminho óptico do espectrômetro (cálculo da absorvidade)

DP Desvio padrão dos dados

DP^{ESTIMADO} Desvio padrão (três amostras) médio dos quatro pré-tratamentos para cada composto químico

FEL Ácido felúrico

FUR Furfural

GAL Ácido gálico

h Constante de Planck ($6,62 \times 10^{-34}$ Js)

HMF Hidroximetilfurfural

HPLC *High Performance Liquid Chromatography*

ILS *Inverse Least Squares*

ILS *Inverse Least Squares Regression*

K Matriz $n_c \times n_\lambda$ de parâmetros ajustáveis

K_{opt} Matriz $n_c \times n_\lambda$ de parâmetros ajustáveis pela regressão linear

LD Conjunto de dados linearmente dependentes

LHW Pré-tratamento hidrotérmico (ou, do inglês, *Liquid Hot Water*)

LOO *Leave-one-out (cross-validation)*

LV *Latent Variables* (ou Número de Regressores)

MC Espectro com absorbâncias centradas na média (do inglês, *Mean Centered*)

m_i i -ésima massa esférica (em kg)

MLR *Multiple Linear Regression*

MLR *Multiple Linear Regression*

NA Número de amostras

NF *Neuro-Fuzzy*

NIR Infravermelho Próximo (do inglês, *Near Infrared*)

NL Região espectral não considerada na análise

NP Nenhum pré-processamento espectral aplicado

P Loadings (ou autovalores)

PC *Principal Components*

PCA *Principal Component Analysis*

PCR *Principal Component Regression*

PLS *Partial Least Squares*

PLS-NIR Técnica de análise multivariada (PLS) associada à espectroscopia na região do Infravermelho Próximo

PLS-NIR/ M₀ Pré-processamento de centragem na média aplicado nos espectros de amostras LHW para a região do NIR

PLS-NIR/ M₁ Pré-processamento de centragem na média e seleção de comprimentos de onda (NL [1.389 – 1.538] -[1.887 – 2105] nm) aplicados em espectros de amostras LHW para a região do NIR

PLS-NIR/ M₂ Pré-processamento de centragem na média, derivada de primeira ordem e seleção de comprimentos de onda (NL [1.389 – 1.538] -[1.887 – 2105] nm) aplicados em espectros de amostras LHW para a região do NIR

PLS-NIR/ M₃ Pré-processamento de centragem na média, derivada de segunda ordem, e seleção de comprimentos de onda (NL [1.389 – 1.538] -[1.887 – 2105] nm) aplicados em espectros de amostras LHW para a região do NIR

PLS-NIR/ MS₀ Nenhum pré-processamento aplicado nos espectros de amostras sintéticas para a região do NIR

PLS-NIR/ MS₁ Pré-processamento de seleção de comprimentos de onda (NL [1.389 – 1.538] - [1.887 – 2105] nm) aplicado em espectros de amostras sintéticas para a região do NIR

PLS-NIR/ MS₂ Pré-processamento de centragem na média, derivada de primeira ordem e seleção de comprimentos de onda (NL [1.389 – 1.538] - [1.887 – 2105] nm) aplicados em espectros de amostras sintéticas para a região do NIR

PLS-NIR/ MS₃ Pré-processamento de centragem na média, derivada de segunda ordem, e seleção de comprimentos de onda (NL [1.389 – 1.538] - [1.887 – 2105] nm) aplicados em espectros de amostras sintéticas para a região do NIR.

PLS-UV-Vis Técnica de análise multivariada (PLS) associada à espectroscopia na região do Ultravioleta-Visível

PLS-UV-Vis/ M₀ Pré-processamento de centragem na média e seleção de comprimentos de onda (NL [325 – 600] nm) aplicados nos espectros de amostras LHW para a região do UV-Vis

PLS-UV-Vis/ M₁ Pré-processamento de centragem na média, derivada de primeira ordem e seleção de comprimentos de onda (NL [325 – 600] nm) aplicados em espectros de amostras LHW para a região do UV-Vis

PLS-UV-Vis/ M₂ Pré-processamento de centragem na média e derivada de segunda ordem e seleção de comprimentos de onda (NL [325 – 600] nm) aplicados em espectros de amostras LHW para a região do UV-Vis

PLS-UV-Vis/ MS₀ Nenhum pré-processamento aplicado nos espectros de amostras sintéticas para a região do UV-Vis

PLS-UV-Vis/ MS₁ Pré-processamento de centragem na média e derivada de primeira ordem aplicados em espectros de amostras sintéticas para a região do UV-Vis

PLS-UV-Vis/ MS₂ Pré-processamento de centragem na média e derivada de segunda ordem aplicados em espectros de amostras sintéticas para a região do UV-Vis

P_{opt} Matriz $n_\lambda \times n_c$ de parâmetros ajustáveis pela regressão linear

R^2 Coeficiente de determinação (Razão entre a variância explicada do modelo de previsão e a variância dos dados)

RER *Range Error Ratio*

RMSECV *Root Mean Square Error of Cross Validation*

RMSEP *Root Mean Square Error of Prediction*

RPD *Residual Predictive Deviation*

RS Resíduo de Student

S Conjunto de amostras sintéticas

SPA *Successive Projection Algorithm*

T Scores (ou autovetores)

UV-Vis Ultravioleta–Visível (do inglês, *Ultraviolet -Visible*)

V Validação externa dos modelos PLS (dados de teste)

VAN Vanilina

u número quântico vibracional de determinada molécula

E Energia de uma molécula

I Intensidade da radiação transmitida através da amostra

L_{resc_i} Resíduo de Student da amostra i corrigido pela *Leverage*

b Caminho óptico (em cm)

c Velocidade da luz (3×10^{10} m/s)

e Energia da radiação eletromagnética ($e = e_{fóton}$, em Joules)

k constante de elasticidade da mola considerada a lei de *Hooke* (em N/m)

ϵ Absortividade molar de um analito puro em dado solvente

κ Constante de proporcionalidade da lei de *Lambert-Beer* (em L/g cm)

λ Comprimento de onda (distância entre dois picos sucessivos da onda eletromagnética, medida em m)

μ massa reduzida considerada a lei de *Hooke*

ν Frequência da radiação eletromagnética (em 1/s)

ς Concentração do analito presente na amostra (em g/L)

SUMÁRIO

1. INTRODUÇÃO E JUSTIFICATIVA	27
2. OBJETIVO	31
2.1. Objetivo Geral	31
2.2. Objetivos Específicos	31
2.2.1. Aquisição de Amostras	31
2.2.2. Caracterização do Hidrolisado.....	31
2.2.3. Protocolo de Monitoramento dos Inibidores	31
3. REVISÃO BIBLIOGRÁFICA.....	33
3.1. Etapas de Produção do Etanol.....	33
3.2. Compostos Fenólicos na Produção de Etanol 2G.....	37
3.3. Furaldeídos na Produção de Etanol 2G	38
3.4. Inibidores Presentes no Hidrolisado do Pré-tratamento Hidrotérmico e as Principais Consequências Durante a Produção de Etanol 2G.....	39
3.5. Fundamentos e Aplicações das Técnicas de Espectroscopia	44
3.5.1. Espectroscopia no Infravermelho Próximo (NIR).....	46
3.5.2. Espectroscopia no Ultravioleta Visível (UV-Vis)	50
3.6. Análise Quimiométrica em Espectroscopia	52
3.7. Calibração Multivariada dos Espectros.....	53
4.MATERIAIS E MÉTODOS	59
4.1. Preparação de Amostras.....	59
4.2. Método Analítico de Referência para as Amostras do Hidrolisado	60
4.3. Reagentes	61
4.4. Varredura de Espectros.....	61
4.5. Tratamentos Matemáticos dos Dados Espectroscópicos	62
4.5.1. Pré-processamento dos Espectros.....	62

4.5.2. Calibração Multivariada	63
4.5.3. Validação cruzada e Teste dos Modelos.....	63
4.5.4. Outliers: Teste Estatístico de Leverage e Resíduo de Student	65
5.RESULTADOS E DISCUSSÃO	69
5.1. Conjunto de Dados Sintéticos (S)	69
5.2. Conjunto de Dados Reais do Pré-tratamento Hidrotérmico (LHW)	82
6.CONSIDERAÇÕES FINAIS.....	99
7.REFERÊNCIAS	101

1. INTRODUÇÃO E JUSTIFICATIVA

Na busca de inovações sustentáveis visando otimizar a produção de etanol, tecnologias que possibilitem a ampliação da produção deste biocombustível a partir de coprodutos da cana-de-açúcar, tais como a palha e o bagaço, vem sendo desenvolvidas, viabilizando, assim, a produção do etanol de segunda geração (2G) ou, como também é conhecido, etanol celulósico. O processo de obtenção do etanol 2G apresenta vantagens ambientais significativas, pois proporciona o aumento da produção deste biocombustível sem expandir áreas de plantio. Além disso, também contribui com a redução na emissão de gases do efeito estufa (Pippo, 2011). Apesar destes benefícios, a produção em escala comercial do etanol 2G ainda está em fase de aperfeiçoamento e apresenta desafios a serem vencidos. O alto custo de processamento, a elevada quantidade de sólidos nos reatores, bem como a presença de contaminantes e inibidores nas reações bioquímicas ainda são obstáculos que devem ser superados (Martins *et al.*, 2015, Schneiderman *et al.*, 2015). Neste contexto, é industrialmente desejável a aplicação de técnicas de monitoramento de processo que viabilizem desenvolver tecnologias de controle e otimização para a redução dos custos de processamento deste biocombustível.

Na produção de etanol 2G, durante a etapa de pré-tratamento, quantidades elevadas de ácidos alifáticos, furaldeídos e compostos fenólicos são liberados da lignina ou gerados devido à desintegração da parede celular vegetal em proporções variáveis de acordo com a severidade do pré-tratamento (Guo *et al.*, 2014; Zeng *et al.*, 2014; Jönsson e Martín, 2016; Sun *et al.*, 2016). Dentre estes contaminantes, os fenólicos e furaldeídos destacam-se por afetarem negativamente as etapas subsequentes ao pré-tratamento durante a produção do etanol 2G. Os compostos fenólicos são responsáveis, principalmente, minimizar a eficiência da etapa de hidrólise, enquanto que os furaldeídos reduzem o desempenho da levedura *Saccharomyces cerevisiae* no processo fermentativo (Jönsson e Martín, 2016, Zeng *et al.*, 2014, Mosier *et al.*, 2005). Na etapa de hidrólise enzimática, os compostos fenólicos desativam e/ou inibem os catalisadores biológicos (Qin *et al.*, 2016, Ximenes *et al.*, 2011b, Michelin *et al.*, 2016, Tejirian e Xu, 2011, Sun *et al.*, 2016), o que leva à grande quantidade de enzimas requeridas ao longo do processo e, conseqüentemente, demanda de alto investimento econômico para aplicações em escala industrial (Li *et al.*, 2011; Guo *et al.*, 2014; Jönsson e Martín, 2016). Enquanto que, durante a fermentação, tanto os compostos fenólicos quanto os furaldeídos são responsáveis por prejudicar o crescimento e a

produtividade de etanol pela levedura (Delgenes *et al.*, 1996; Larsson *et al.*, 2000). A fim de atuar na otimização destes bioprocessos para a redução dos custos atrelados à reposição de enzimas e à produtividade de etanol, torna-se essencial o monitoramento dos principais compostos fenólicos e furaldeídos liberados do material lignocelulósico na etapa de pré-tratamento durante a produção do etanol 2G.

A caracterização do ambiente químico, usualmente, requer a amostragem manual associada à realização de procedimentos indispensáveis para a preparação de amostras utilizadas em análises subsequentes. A quantificação dos furaldeídos e compostos fenólicos extraídos da biomassa pode ser efetuada com precisão a partir da Cromatografia Líquida de Alta Eficiência (do inglês, *High Performance Liquid Chromatography* - HPLC) (Acosta-Estrada *et al.*, 2014). No entanto, este procedimento apresenta risco de contaminação, além de não viabilizar a otimização e o controle do bioprocessos industrialmente devido ao tempo despendido na análise das amostras (Khoddami *et al.*, 2013). Considerando estes aspectos, a aplicação de técnicas que permitam o monitoramento do processo a partir de informações diretas, adquiridas de maneira rápida, precisa e reproduzível torna-se crítica.

A espectroscopia associada às técnicas de calibração multivariada representa uma alternativa às análises químicas convencionais. Esta abordagem possibilita a quantificação de multiconstituintes com o mínimo, ou nenhuma, preparação da amostra, de forma rápida e não destrutiva (Cozzolino *et al.*, 2004; Hongqiang e Hongzhang, 2008). Diferentes técnicas espectroscópicas têm sido investigadas em aplicações que abrangem diferentes regiões espectrais, dentre elas o ultravioleta-visível (UV-Vis) e o infravermelho próximo (NIR), que possuem potencial aplicação no monitoramento de compostos fenólicos e furaldeídos (Cozzolino, 2015; Frizon *et al.*, 2015; Khani *et al.*, 2016; Monroy *et al.*, 2015).

A regressão por mínimos quadrados parciais (do inglês, *Partial Least Squares* - PLS) é umas das principais técnicas de análise multivariada encontradas na literatura e vem se mostrando eficiente no monitoramento de diversos componentes em meios reacionais complexos (Belguidoum *et al.*, 2014; Frizon *et al.*, 2015; Milligan *et al.*, 2014; Pinto *et al.*, 2014a; Ribeiro *et al.*, 2008; Rodríguez-Zúñiga *et al.*, 2014; Sampaio *et al.*, 2014; Sills e Gossett, 2012). O uso de espectros de NIR e UV-Vis associados à regressão PLS vem sendo aplicados de forma eficaz na quantificação de compostos presentes em misturas multicomponentes, tais como: etanol e glicose (Pinto *et al.*, 2016c); glicerol e ácidos acéticos

(Blanco *et al.*, 2004); polifenóis (Belguidoum *et al.*, 2014; Martelo-Vidal e Vázquez, 2014). Entretanto, quando a relação entre a concentração dos componentes e o espectro deixa de ser linear, técnicas como Redes Neurais Artificiais (do inglês, *Artificial Neural Networks* - ANN) e Sistemas *Neuro-Fuzzy* (NF) devem ser aplicados (Kadlec *et al.*, 2009).

Apesar da vasta aplicação de técnicas de espectroscopia associadas à análise multivariada na quantificação de analitos distintos, ainda não há estudos sobre a caracterização individual dos compostos fenólicos e furaldeídos presentes no hidrolisado da biomassa lignocelulósica. O protocolo de monitoramento dos compostos fenólicos e furaldeídos em questão auxiliaria na redução do impacto negativo dos componentes presentes no hidrolisado, visto que permitiria a implementação de estratégias de processo, tais como, a otimização das condições de operação do pré-tratamento da biomassa, ou ainda a utilização de proteínas de sacrifício ou outros aditivos na etapa de hidrólise enzimática (Florencio *et al.*, 2016; Qin *et al.*, 2016). Portanto, o escopo desta dissertação foi investigar diferentes regiões espectrais associadas à análise multivariada de dados no monitoramento *off-line* dos principais compostos fenólicos e furaldeídos liberados no pré-tratamento hidrotérmico da biomassa de cana-de-açúcar.

2. OBJETIVO

2.1. Objetivo Geral

O presente estudo teve como objetivo o desenvolvimento de uma metodologia de monitoramento *off-line* dos principais compostos fenólicos e furaldeídos presentes nos hidrolisados da biomassa lignocelulósica de cana-de-açúcar. O protocolo de monitoramento fundamenta-se em técnicas de espectroscopia associadas à quimiometria, sendo aplicáveis em protocolos de otimização e controle de processos durante a produção de etanol 2G.

2.2. Objetivos Específicos

2.2.1. Aquisição de Amostras

Obter amostras sintéticas (solução aquosa contendo apenas os principais inibidores dos bioprocessos envolvidos na produção de etanol 2G) e reais de hidrolisado do pré-tratamento hidrotérmico da biomassa de cana-de-açúcar.

2.2.2. Caracterização do Hidrolisado

Desenvolver a melhor metodologia de caracterização dos compostos fenólicos e furaldeídos presentes no hidrolisado de cana-de-açúcar a partir da cromatografia líquida de alta eficiência.

2.2.3. Protocolo de Monitoramento dos Inibidores

Levantamento e análise dos espectros nas regiões do UV-Vis e NIR. Calibrar/testar modelos para a previsão da concentração dos principais compostos fenólicos e furaldeídos presentes no hidrolisado da biomassa lignocelulósica a partir destas técnicas de espectroscopia associadas à análise multivariada de dados. Monitorar *off-line* estes componentes do hidrolisado em amostras sintéticas e reais do pré-tratamento.

3. REVISÃO BIBLIOGRÁFICA

A motivação deste trabalho foi contribuir para a viabilização de uma metodologia de monitoramento dos compostos fenólicos e furaldeídos presentes no hidrolisado de biomassa lignocelulósica da cana-de-açúcar. Assim, nas seções subsequentes foi detalhado o processo de produção de etanol, expondo o porquê e em quais etapas há a liberação de compostos fenólicos e furaldeídos, além de elucidar seu efeito negativo em cada etapa do bioprocessamento. A seguir, foram descritos os principais embasamentos das técnicas quimiométricas utilizadas na estratégia de monitoramento destes inibidores de processo, as quais se fundamentam na espectroscopia e na regressão multivariada de dados.

3.1. Etapas de Produção do Etanol

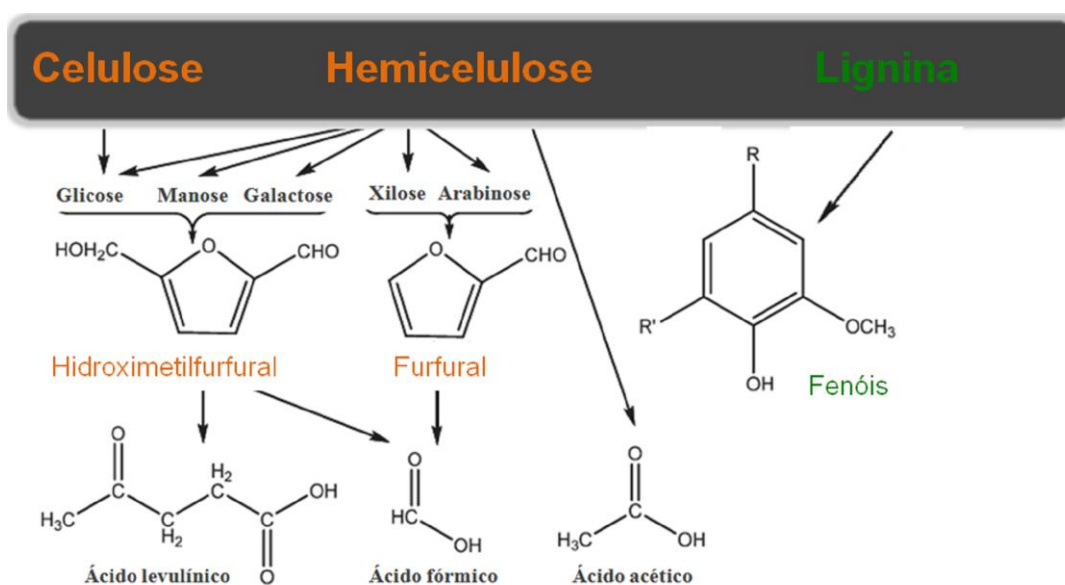
O etanol é um composto químico cuja fórmula molecular é C_2H_6O . Trata-se de um álcool orgânico produzido industrialmente através da fermentação alcoólica realizada pela levedura *Saccharomyces cerevisiae*. Esse processo ocorre a partir do consumo de açúcares simples, que podem ser do caldo de cana-de-açúcar (conhecida como fermentação de primeira geração, 1G) ou obtidos a partir da hidrólise de açúcares superiores presentes no material celulósico das plantas (conhecida como fermentação de segunda geração, 2G).

O processo de produção do etanol 2G pela rota química, ou enzimática, pode ser dividido em três etapas distintas: pré-tratamento da biomassa lignocelulósica, hidrólise química ou enzimática, e fermentação alcoólica.

O primeiro passo do processo de produção do etanol celulósico constitui no pré-tratamento da biomassa lignocelulósica. A finalidade principal desta etapa é aumentar a susceptibilidade da fase sólida ao processo de hidrólise enzimática. O hidrolisado obtido nesta etapa apresenta características particulares que dependem do tipo de pré-tratamento e da biomassa (Martin *et al.*, 2002; Mosier *et al.*, 2005; Jönsson e Martín, 2016; Sun *et al.*, 2016). Além disso, o hidrolisado obtido pode conter quantidades elevadas de ácidos alifáticos, furaldeídos e compostos fenólicos que são liberados da lignina ou durante a degradação da biomassa (Martin *et al.*, 2002; Mosier *et al.*, 2005; Jönsson *et al.*, 2013; Zeng *et al.*, 2014; Ko, J. *et al.*, 2015; Ko, J. K. *et al.*, 2015; Jönsson e Martín, 2016; Sun *et al.*, 2016). Os furaldeídos são responsáveis por reduzir tanto a produção de etanol, quanto o crescimento celular da levedura durante a fermentação alcoólica (Delgenes *et al.*, 1996; Larsson *et al.*, 2000). Já os

compostos fenólicos merecem destaque, pois são os principais responsáveis por inibir e desativar enzimas no processo de hidrólise, além de influenciarem negativamente o crescimento de microrganismos nos processos fermentativos (Michelin *et al.*, 2016; Qin *et al.*, 2016; Zha *et al.*, 2014). A Figura 1 apresenta o diagrama que esquematiza a origem dos principais compostos químicos liberados no pré-tratamento da biomassa, responsáveis por prejudicar o processo fermentativo e a hidrólise enzimática do material lignocelulósico.

Figura 1 – Diagrama esquemático dos principais compostos químicos liberados no pré-tratamento da biomassa que são responsáveis por prejudicar o processo fermentativo e a hidrólise enzimática da celulose.



Fonte: Imagem adaptada de Jönsson *et al.* (2013).

Com o objetivo de viabilizar a utilização da biomassa como matriz energética no processo de produção de etanol 2G, esta deve ser hidrolisada em açúcares redutores. A conversão da celulose em açúcares de cadeias carbônicas pequenas pode ser realizada a partir da hidrólise química ou enzimática. A conversão enzimática do material lignocelulósico tem sido apontada como rota de grande interesse industrial para o aumento da produtividade do etanol (Knauf e Moniruzzaman, 2004; Wyman, 1999).

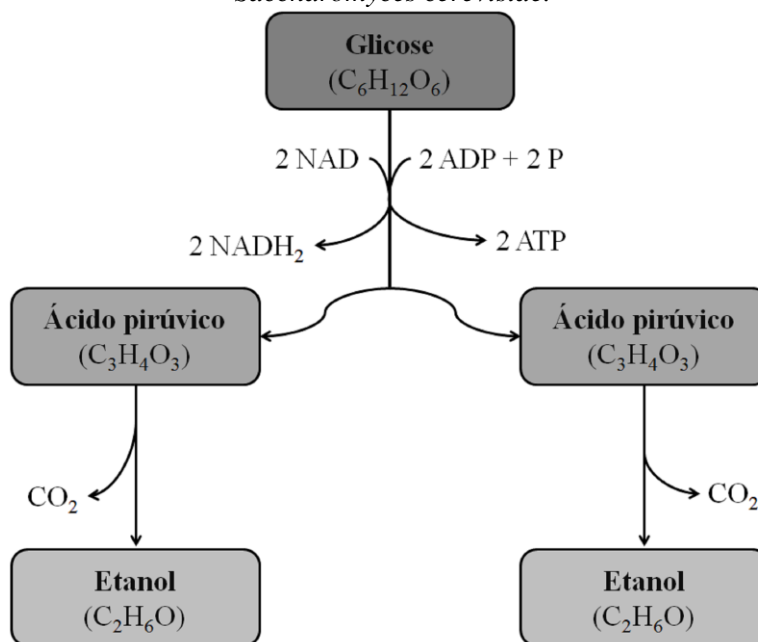
Na natureza, o grau de ordenação da celulose requer que os microrganismos celulolíticos produzam uma complexa mistura de enzimas, as celulasas e hemicelulasas, para efetuar a degradação da celulose cristalina e amorfa. Esse complexo de enzimas é necessário para a solubilização completa e efetiva da biomassa, produzindo um efeito sinérgico no processo de

hidrólise (Himmel *et al.*, 1999). Atualmente, o mecanismo de hidrólise enzimática mais aceito descreve a ação sinérgica de pelo menos três classes de enzimas: as endoglucanases, as exoglucanases e as β -glicosidases (ou celobiasas) (Zhang e Lynd, 2004).

O rendimento e a taxa de hidrólise enzimática têm grande impacto na conversão global da biomassa em etanol, o que representa um gargalo na produção de etanol 2G. A eficiência desta etapa está diretamente relacionada com a transferência de massa do processo devido ao alto teor de sólidos presentes, além de ser limitada pela inibição/desativação das enzimas celulolíticas por compostos químicos liberados da lignina e oriundos da degradação da celulose ou hemicelulose (Jönsson e Martín, 2016; Schneiderman *et al.*, 2015). Conseqüentemente, nesta etapa do processo, há alto custo atrelado à elevada carga de enzimas celulolíticas requerida, o que é um dos principais obstáculos a serem superados no processo de produção do etanol 2G. Considerando a inibição/desativação das enzimas celulolíticas, os compostos fenólicos devem ser destacados (Michelin *et al.*, 2016; Qin *et al.*, 2016; Ximenes *et al.*, 2011). Interferir nas etapas de produção do etanol 2G pode minimizar o efeito negativo dos inibidores presentes no hidrolisado durante a hidrólise enzimática. A adição de proteínas de sacrifício no hidrolisado do pré-tratamento, por exemplo, são alternativas para otimizar a eficiência da produção de etanol 2G (Florescio *et al.*, 2016; Jönsson e Martín, 2016).

Após o processo de hidrólise segue a fermentação. Nas reações bioquímicas intracelulares envolvidas na fermentação alcoólica, as moléculas de substrato são convertidas pela levedura em ácido pirúvico, o que acarreta na produção de etanol com a liberação do dióxido de carbono e a formação de adenosina trifosfato (ATP) (Lehninger, 1986). A Figura 2 ilustra este processo de forma simplificada.

Figura 2 - Fluxograma simplificado do processo fermentativo realizado pelo microrganismo *Saccharomyces cerevisiae*.



Fonte: *Lehninger (1986)*.

Os compostos químicos liberados no pré-tratamento da biomassa lignocelulósica também apresentam efeito negativo sobre a eficiência das reações bioquímicas que ocorrem durante o processo fermentativo. O crescimento celular, bem como a produtividade de etanol da levedura são reduzidos na presença de alguns inibidores advindos da lignina, como os compostos fenólicos, e da degradação de celulose e hemicelulose, tais como os furaldeídos (Jönsson e Martín, 2016a).

Salienta-se que a redução da eficiência das etapas de hidrólise enzimática e fermentação alcoólica está, dentre outros fatores, diretamente atrelada à quantidade dos compostos químicos inibidores liberados e/ou formados durante o pré-tratamento do material lignocelulósico. A Tabela 1 exibe, por exemplo, a concentração observada após o processo de hidrólise enzimática de alguns dos principais interferentes do processo de produção do etanol 2G.

Tabela 1 - Concentração de fenólicos ao final da etapa de hidrólise do bagaço de cana-de-açúcar pré-tratado com explosão a vapor.

Compostos Inibidores	Concentração (g/L)
Vanilina	0,05
Ácido Felúrico	0,21
Ácido p-Cumárico	0,48
Hidroximetilfurfural	0,40
Furfural	1,10
Fenólicos Totais	3,20

Fonte: *Martin et al.* (2002).

Neste contexto, estudos que viabilizem o monitoramento dos principais inibidores dos bioprocessos envolvidos na produção de etanol 2G, tais como os compostos fenólicos e furaldeídos, são de interesse, uma vez que, o monitoramento do hidrolisado proporcionaria atuar nas etapas de pré-tratamento, hidrólise ou fermentação a fim de otimizá-las.

3.2. Compostos Fenólicos na Produção de Etanol 2G

Os compostos fenólicos são intrínsecos das plantas, podendo ser solúveis ou insolúveis. Os compostos fenólicos insolúveis estão ligados aos componentes estruturais das plantas, tais como a celulose, hemicelulose, lignina, pectina e proteínas estruturais (Wong, 2006). A liberação destes compostos químicos é realizada na etapa de pré-tratamento e também durante o processo de hidrólise, no qual as enzimas (celulases, hemicelulases, glucanases, xilanases) desempenham a desintegração da matriz da parede celular vegetal facilitando a liberação dos fenólicos insolúveis (Zheng *et al.*, 2009). A solubilização gradual dos fenólicos no meio reacional faz com que a desativação/inibição das celulases e hemicelulases sejam acentuadas ao longo do processo, prejudicando, assim, a sua eficiência (Michelin *et al.*, 2016; Qin *et al.*, 2016; Sun *et al.*, 2016; Tejirian and Xu, 2011; Ximenes *et al.*, 2011).

A concentração de cada composto fenólico no hidrolisado depende não só da biomassa vegetal, mas também das condições de operação do processo, ou seja, da austeridade do pré-tratamento (Ko *et al.*, 2015a; Martin *et al.*, 2002; Michelin *et al.*, 2016; Sun *et al.*, 2016). De acordo com a literatura, a severidade do pré-tratamento influencia diretamente na concentração e no tipo de compostos fenólicos presentes no hidrolisado (Sun *et al.*, 2016). A variação na quantidade de certos inibidores relaciona-se diretamente com a proporção de desativação ou inibição das enzimas durante a etapa de hidrólise e também com a intensidade

na redução da eficiência da etapa fermentativa (Kim *et al.*, 2015; Larsson *et al.*, 2000; Sun *et al.*, 2016).

Além do mais, a intensidade do efeito de inibição/ desativação das enzimas celulolíticas pode variar dependendo do composto fenólico, do tipo de enzima utilizada e de qual micro-organismo a enzima foi produzida (Michelin *et al.*, 2016; Qin *et al.*, 2016; Ximenes *et al.*, 2011). Na etapa fermentativa, o efeito negativo destes compostos químicos está atrelada à minimização do crescimento celular da levedura em até, aproximadamente, 8%, além de reduzir a produção de etanol por volta de 78%, dependendo da concentração e de qual inibidor está presente no meio reacional (Delgenes *et al.*, 1996; Larsson *et al.*, 2000; Qin *et al.*, 2016).

Os compostos fenólicos monoméricos que mais influenciam negativamente a fermentação alcoólica, bem como a atividade das enzimas por inibição e/ou desativação são: os ácidos gálico (GAL), felúrico (FEL), p-cumárico (APC) e a vanilina (VAN) (Larsson *et al.*, 2000; Michelin *et al.*, 2016.; Palmqvist e Hahn-Hägerdal, 2000; Ximenes *et al.*, 2010, 2011).

3.3. Furaldeídos na Produção de Etanol 2G

Além dos compostos fenólicos, destaca-se que a elevada severidade do pré-tratamento da biomassa vegetal também acarreta na formação de furaldeídos. As condições do processo de pré-tratamento favorecem a geração destes compostos químicos, visto que os furaldeídos originam da degradação térmica de monossacarídeos (vide Figura 1) sob pH ácido (Kabel *et al.*, 2007). Os furaldeídos são tóxicos ao processo de produção do etanol 2G devido à susceptibilidade da levedura *Saccharomyces cerevisiae*. Os furanos são responsáveis por reduzir o crescimento celular de leveduras, bem como sua produtividade de etanol (Banerjee *et al.*, 1981; Carrasco *et al.*, 2010).

Os principais furaldeídos presentes no hidrolisado do pré-tratamento do bagaço de cana-de-açúcar são o hidroximetilfurfural (HMF) e o furfural (FUR). O mecanismo de inibição provocado pelo HMF e pelo FUR no metabolismo celular é similar, visto que estudos cinéticos mostraram que este processo está relacionado com a atuação de algumas enzimas específicas da levedura (Carrasco *et al.*, 2010).

Segundo Delgenes *et al.* (1996), o furfural na concentração de 1,0 g/L reduz o crescimento celular em 20% e a produção de etanol em 19% para a levedura *Saccharomyces cerevisiae*.

No mesmo estudo, o efeito inibidor da fermentação causado pelo HMF foi mais representativo, pois, na concentração de 1,0 g/L, este composto químico reduziu o crescimento celular em 35% e a produção de etanol em 29%.

3.4. Inibidores Presentes no Hidrolisado do Pré-tratamento Hidrotérmico e as Principais Consequências Durante a Produção de Etanol 2G

A redução na eficiência dos bioprocessos envolvidos na produção de etanol 2G provém principalmente do efeito inibitório provocado nas reações bioquímicas e/ou enzimáticas por determinados compostos químicos gerados durante o pré-tratamento da biomassa. Estes inibidores podem ser liberados da lignina, tais como os compostos fenólicos, ou advindos da degradação de açúcares como, por exemplo, os furaldeídos (vide Figura 1), sendo que ambos são responsáveis por prejudicar as etapas de hidrólise e fermentação alcoólica durante a produção do etanol 2G (Jönsson e Martín, 2016). Salienta-se que o grau de inibição depende diretamente da severidade do pré-tratamento e do material lignocelulósico, visto que estes fatores definem a presença e quantidade de determinados compostos químicos no hidrolisado (Delgenes *et al.*, 1996; Larsson *et al.*, 2000; Sun *et al.*, 2016).

Meios de cultivo que contenham furaldeídos e compostos fenólicos apresentam crescimento celular da levedura *S. cerevisiae* prejudicado concomitantemente com a diminuição na produtividade de etanol (Banerjee *et al.*, 1981; Delgenes *et al.*, 1996; Larsson *et al.*, 2000). Na concentração de 0,20 g/l, por exemplo, destaca-se o caráter inibidor da vanilina e do ácido p-cumárico, visto que estes compostos levaram à redução de, aproximadamente, 70% na produção de etanol, além de diminuir o crescimento de células em torno de 8%. O ácido felúrico não prejudicou o processo tanto quanto os analitos citados ao fixar esta concentração, apresentando diminuição de 31% na formação do produto e de 6%, na multiplicação celular (Larsson *et al.*, 2000). Na etapa fermentativa da produção de etanol 2G, tanto o HMF quanto o furfural diminuíram a produção deste biocombustível tão significativamente quanto os demais analitos. A redução na produtividade de etanol foi de, no máximo, 57% em 0,50 g/L de furfural e de 81% em 2,0 g/L de HMF (Delgenes *et al.*, 1996; Taherzadeh *et al.*, 2000). Além da consequência isolada de cada composto químico na fermentação, devido a possíveis efeitos sinérgicos, a presença de mais de um inibidor pode maximizar a retração do crescimento celular da *S. cerevisiae* (Zha *et al.*, 2014). O aparecimento ou avanço da fase *lag* durante o cultivo da levedura *S. cerevisiae* também foi

observado na presença de inibidores como o FUR, HMF, VAN e FEL, seguido ou não de redução no crescimento celular (Zha *et al.*, 2014). A Tabela 2 mostra o resumo do efeito negativo sobre o processo fermentativo da *S. cerevisiae* para diferentes concentrações de inibidores.

Tabela 2 - Efeito no crescimento celular e na produção de etanol para a levedura *S. cerevisiae* de acordo com a presença de diferentes inibidores.

Composto Químico	Concentração (g/l)	Efeito na fermentação ¹		Referências
		Y_x ² (g/g)	Y_{EtOH} ³ (g/g)	
VAN	0,02	0,079	0,73	Larsson <i>et al.</i> (2000)
	0,20	0,075	0,73	
	0,50	0,490	0,78	Delgenes <i>et al.</i> (1996)
	1,00	0,054	0,58	Larsson <i>et al.</i> (2000)
FEL	0,02	0,072	0,68	Larsson <i>et al.</i> (2000)
	0,20	0,063	0,31	
	1,00	0,008	0,14	
APC	0,02	0,078	0,70	Larsson <i>et al.</i> (2000)
	0,20	0,078	0,68	
	1,00	0,073	0,42	
FUR	0,50	0,530	0,57	Delgenes <i>et al.</i> (1996)
	1,00	0,190	0,20	
	2,00	0,100	0,11	Taharzadeh <i>et al.</i> (2000)
	4,00	0,110	0,44	
HMF	1,00	0,350	0,29	Delgenes <i>et al.</i> (1996)
	3,00	0,170	0,17	
	5,00	0,110	0,05	

¹ As condições do processo variam de acordo com a referência

² O rendimento de biomassa (Y_x) foi baseado no aumento de peso seco por gramas de glicose consumida.

³ O rendimento em etanol (Y_{EtOH}) foi baseado em gramas de etanol formados por gramas de glicose consumida.

Além de influenciarem o processo fermentativo, os compostos fenólicos devem ser destacados principalmente por afetarem negativamente a atividade de enzimas celulolíticas e hemicelulolíticas durante a hidrólise enzimática (Michelin *et al.*, 2016; Qin *et al.*, 2016; Sun *et al.*, 2016; Tejirian e Xu, 2011; Ximenes *et al.*, 2011). Durante o processo de hidrólise, as ligações não específicas ou não produtivas entre os derivados da lignina e a enzima podem ser mediadas por interações hidrofóbicas, eletrostáticas ou de pontes de hidrogênio (Saini *et al.*, 2016). Tejirian e Xu (2011) observaram que a cinética de inibição das enzimas poderia ser mista ou não competitiva, dependendo dos compostos fenólicos analisados no processo de hidrólise. No entanto, devido à complexidade do hidrolisado advindo do pré-tratamento da biomassa, é possível que mais de um tipo interação química esteja envolvida na ligação não produtiva das enzimas (Saini *et al.*, 2016). Ademais, o efeito de inibição/desativação nas celulasas/xilanases por fenólicos mono- ou oligoméricos é vasto e depende do componente inibidor além de sua concentração, bem como da fonte microbiana e do tipo das enzimas envolvidas no processo (exo-/endocelulasas, β -glicosidase, xilanase ou β -xilosidase) (Michelin *et al.*, 2016; Ximenes *et al.*, 2010; Ximenes *et al.*, 2011).

Segundo estudos realizados por Ximenes *et al.* (2011), a desativação da enzima β -glicosidase (obtidas a partir do *T. reesei*, considerando 100% de atividade enzimática de 37 IU/mL), por exemplo, pode levar à redução da atividade enzimática em, aproximadamente, 70% para o GAL, 80% para o APC, 38% para o FEL e 25% para a VAN (fixando a concentração de fenólico em 4,5 g/L e medindo a nova atividade após 24h de incubação do analito com a enzima, seguindo de 24h de hidrólise). Além da desativação, pode ocorrer também a inibição enzimática. O GAL apresentou redução na atividade de celulasas em até, ~30%, dada a referência de atividade enzimática máxima (37 IU/mL) de β -glicosidases produzidas pelo *T. reesei*, fixando-se ainda a concentração de 1,9 g/L de inibidor em ensaios de hidrólise de 15-60 min. (Ximenes *et al.*, 2011). Qin *et al.* (2016) verificaram que, para a concentração de 5,0 g/L de fenólico, houve inibição aparente das celulasas, de até, ~15% ao considerar o FEL e, aproximadamente, 38% de inibição aparente ao utilizar a VAN ou APC no processo de hidrólise (porcentagens baseadas na inibição aparente ($I_{app} = 1 - \%_{conversão}$ da celulose) em 24h de hidrólise, dado a conversão de referência sem inibidor). A Tabela 3 mostra, resumidamente, o efeito dos compostos fenólicos na hidrólise enzimática.

Tabela 3 - Resumo do efeito inibitório e desativador de enzimas celulolíticas dos compostos fenólicos durante a hidrólise enzimática.

Composto Químico	Concentração (g/L)	Perda de atividade enzimática ⁴		Enzimas	Referência
		Desativação ⁵ (%)	Inibição (%)		
GAL	1,92	-	30	β -glicosidase (<i>T. reesei</i>)	Ximenes <i>et al.</i> (2011)
	4,50	65	-		
VAN	4,50	65	-	endo-, exo-, β -glicosidase (<i>T. reesei</i>)	Ximenes <i>et al.</i> (2011)
	5,00	-	38		
FEL	4,50	38	-	endo-, exo-, β -glicosidase (<i>T. reesei</i>)	Ximenes <i>et al.</i> (2011)
	5,00	-	15		
APC	4,50	80	-	endo-, exo-, β -glicosidase (<i>T. reesei</i>)	Ximenes <i>et al.</i> (2011)
	5,00	-	38		

⁴ $I_{app} \% (24h \text{ de hidrólise}) = 100 \cdot (1 - \text{conversão de celulose com fenólico} / \text{concentração de celulose sem fenólico})$.

⁵ A análise da hidrólise enzimática ocorre depois da pre-incubação das enzimas com o fenólico.

Em suma, estes compostos químicos afetam substancialmente o processo de produção do etanol 2G, logo a aplicação de tecnologias de baixo custo, rápidas além de eficientes no monitoramento das concentrações de compostos fenólicos e furaldeídos possibilitaria maior controle sobre as etapas do processo bioquímico. Nas seções 3.5 a 3.7 foram apresentados os fundamentos das técnicas de quantificação propostas no presente trabalho.

3.5. Fundamentos e Aplicações das Técnicas de Espectroscopia

À temperatura ambiente, as moléculas estão, na maior parte, em seus níveis de energia de vibração fundamental. Este nível energia pode ser alterado através da incidência de radiação eletromagnética (e), definida pela equação 1.

$$e = h\nu \quad (1)$$

Onde e é a energia da radiação eletromagnética (em Joules), h é a constante de Planck ($6,62 \times 10^{-34}$ Js) e ν é a frequência (em 1/s).

A radiação eletromagnética é explicada pela combinação alternada entre os campos elétrico e magnético. A radiação apresenta o comportamento de onda, devido à forma com que ocorre sua movimentação através do espaço. Assim, pode ser definida por seu comprimento de onda (λ) ou frequência (ν). A relação entre estas grandezas está apresentada na equação 2.

$$\nu = \frac{c}{\lambda} \quad (2)$$

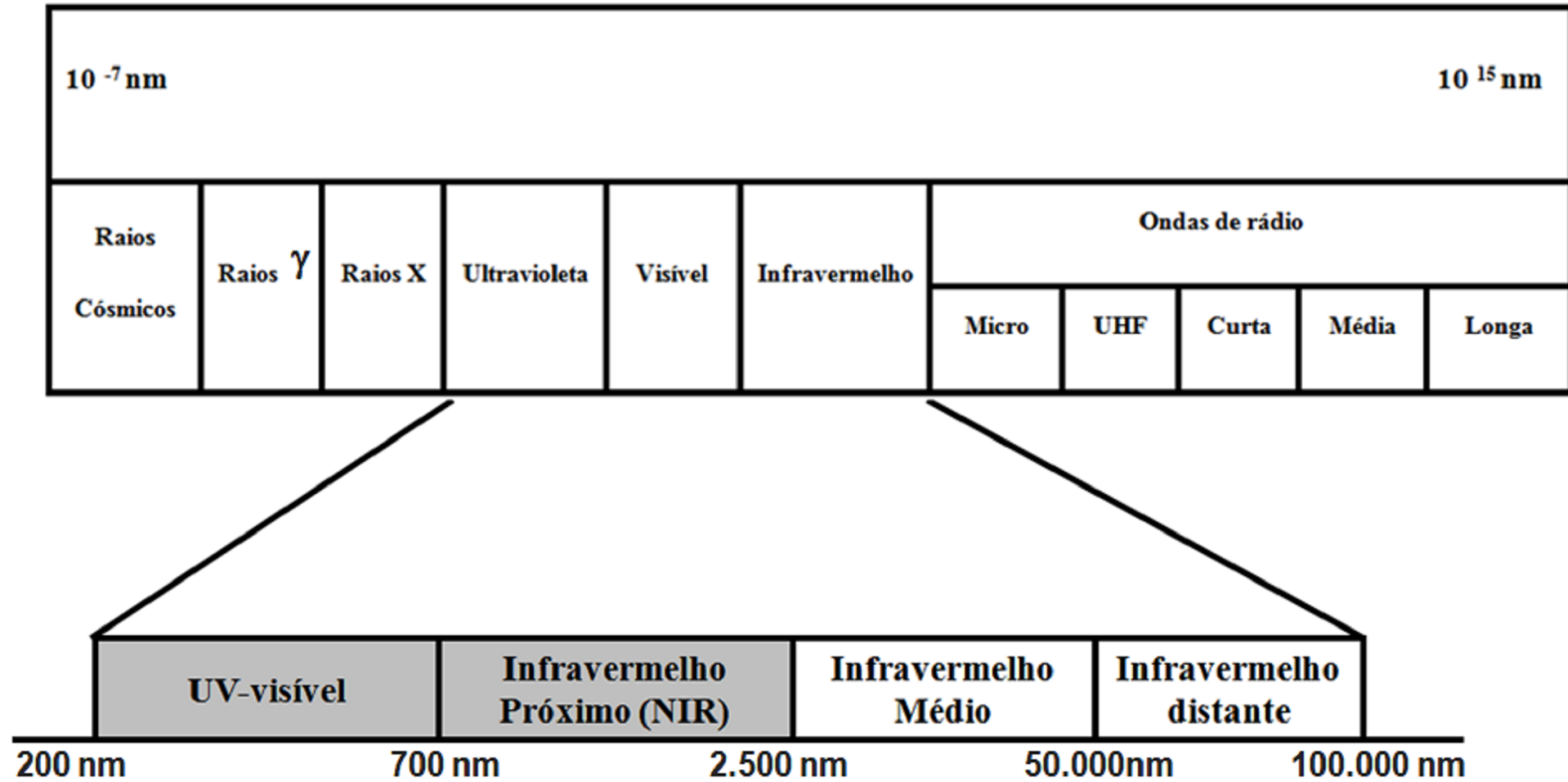
Na qual, c é a velocidade da luz (3×10^{10} m/s) e λ é o comprimento de onda (distância entre dois picos sucessivos da onda eletromagnética, medida em metros).

Portando, reescrevendo a equação 2 em 1, tem-se a equação 3.

$$e = e_{fóton} = \frac{hc}{\lambda} \quad (3)$$

A consequência direta destas equações é que a radiação eletromagnética com comprimentos de onda curtos apresenta maior energia. A seguir, a Figura 3 ilustra as regiões espectrais e os comprimentos de onda em nanômetros.

Figura 3 - Regiões de espectroscopia e comprimentos de onda característicos.



Fonte: Acervo pessoal.

Observando-se a Figura 3 e a equação 3, verifica-se, portanto, que a radiação na região do UV-Vis apresenta maior energia do que na do NIR. Salienta-se que a radiação pode ser explicada também como uma corrente de partículas definidas como fótons ao invés de onda. Os fótons de luz incidentes na matéria podem levar a diferentes interações, tais como absorção, espalhamento, reflexão, fluorescência ou mesmo reações fotoquímicas. No caso de absorção dos fótons pela matéria, a energia presente na radiação eletromagnética provoca o aumento da energia fundamental dos átomos ou moléculas. Por definição, geralmente representa-se a energia potencial total de uma molécula (E_{total}) pela soma da energia eletrônica ($E_{eletrônica}$), vibracional ($E_{vibracional}$) e rotacional ($E_{rotacional}$), como segue na equação 4 (Owen, 2000).

$$E_{total} = E_{eletrônica} + E_{vibracional} + E_{rotacional} \quad (4)$$

Os níveis, ou estados, de energia fundamental da molécula são discretos, apresentados quantidades de energia distintas, sendo que:

$$E_{eletrônica} > E_{vibracional} > E_{rotacional}$$

Ao expor a matéria a um fóton de luz, altera-se o estado fundamental de energia das moléculas/átomos de acordo com a quantidade de energia associada ao fóton incidente (Owen, 2000). Conforme a relação anterior, necessita-se de mais energia para alterar o nível eletrônico da molécula do que seu nível rotacional. Partindo-se deste princípio, as diferenças básicas entre os fundamentos de cada uma das técnicas espectroscópicas estudadas neste trabalho foram explicados, visto que as regiões do UV-Vis e NIR compreendem aos fótons com energias distintas.

Nas seções 3.5.1 e 3.5.2, os fundamentos de cada espectroscopia foram delineados.

3.5.1. Espectroscopia no Infravermelho Próximo (NIR)

Átomos ou grupos de átomos que participam da ligação química são deslocados uns em relação aos outros com uma frequência que é definida pela força da ligação e pela massa dos átomos ligados. As amplitudes destas ligações são de poucos nanômetros e pode aumentar se certa quantidade de energia foi transferida para a molécula (Siesler *et al.*, 2002).

A região referente à espectroscopia NIR corresponde ao intervalo de 700 - 2.500 nm (Figura 3). O principal fundamento da espectroscopia no Infravermelho Próximo (NIR) está relacionado às ligações químicas e seus níveis de vibração, sendo assim uma espectroscopia vibracional.

A vibração das ligações químicas nas moléculas é resultado da combinação entre força de ligação e massa dos átomos envolvidos. O estado de vibração fundamental pode ser alterado com o aumento da amplitude da ligação atômica, que ocorre através de transferência de energia. Como o descrito anteriormente, esta energia pode ser fornecida por um fóton de certo comprimento de onda.

O modelo clássico mais simples para explicar a interação da radiação com a matéria na região espectral do infravermelho próximo considera uma molécula diatômica como duas massas esféricas (m_1 e m_2) conectadas por uma mola com constante de elasticidade (k). Neste caso, assume-se que a vibração entre as ligações químicas é apenas harmônica e pode ser descrita pela lei de *Hooke*.

A lei de *Hooke* estabelece que a energia (E) deste sistema pode ser representada pela equação 6.

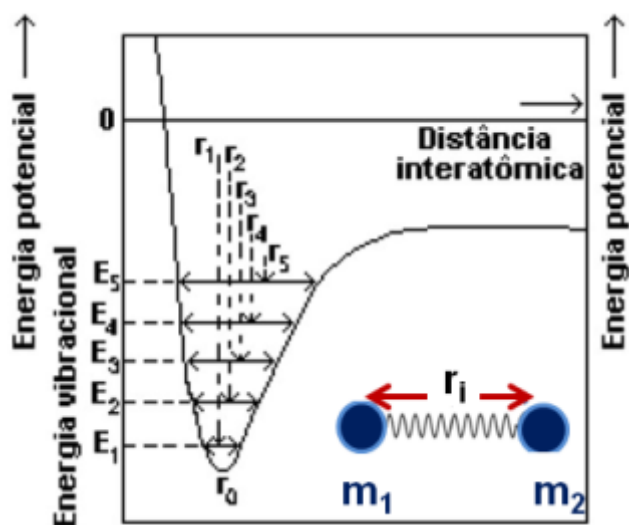
$$E = \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} \quad (6)$$

Em que h é a constante de Planck (em J.s), k é a constante de elasticidade (em N/m) e μ é a massa reduzida (em kg) calculada pela equação 7.

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

A Figura 4 ilustra o princípio da lei de *Hooke*, na qual foi apresentada a energia de ligação no nível fundamental (E_0), além de identificar variações do comprimento de ligação (r_i) de acordo com o aumento da energia vibracional da molécula (E_i).

Figura 4 - Representação do princípio básico da espectroscopia na região do NIR (lei de Hooke).



Fonte: *Acervo pessoal.*

Entretanto, sistemas moleculares possuem estados de energia quantizados e não podem assumir os perfis de energia contínuos preditos pelo modelo clássico mola-bola proposto por *Hooke*.

O sistema molecular pode admitir somente alguns níveis de energia discretos ($E_{vibracional}$) definidos pela mecânica quântica, de acordo com a equação 8 (Siesler *et al.*, 2002).

$$E_{vibracional} = \left(v + \frac{1}{2} \right) h\nu_0 \quad (8)$$

Onde v é o número quântico vibracional, $E_{vibracional}$ é a energia associada com o v -ésimo nível quântico e ν_0 é a frequência vibracional fundamental, obtida através da equação 9.

$$\nu_0 = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (9)$$

Uma restrição adicional imposta pelo modelo quântico harmônico é que as transições podem somente ocorrer entre níveis de energia adjacentes, ou seja, quando $\Delta v = \pm 1$. Assim, a energia da radiação eletromagnética que é absorvida para promover a molécula para um nível excitado deve ser igual à diferença entre dois níveis de energia adjacentes. Desta forma, a energia do fóton deve ser dada pela equação 10.

$$e_{\text{fóton}} = E_{v_2} - E_{v_1} = \Delta v h \nu_0 \quad (10)$$

A combinação da energia da onda eletromagnética com a diferença de energia entre dois níveis vibracionais causa uma resposta seletiva do sistema molecular à radiação incidente. Significa dizer que em uma dada faixa de comprimento de onda, algumas frequências serão absorvidas, outras não serão absorvidas, enquanto que outras podem ser parcialmente absorvidas. Entretanto, somente a energia combinada entre fótons e níveis vibracionais não é suficiente para a absorção da radiação. Para uma vibração ser ativa, é necessário que o campo elétrico da onda eletromagnética possa interagir com a molécula. Isto pode ocorrer somente se o deslocamento dos átomos em um modo vibracional puder produzir uma mudança no momento de dipolo da molécula ou no grupo local de átomos que estão vibrando. Se as ligações moleculares seguissem o modelo vibracional ideal de oscilador harmônico, o momento dipolo aconteceria somente se $\Delta v = \pm 1$. Conseqüentemente, a diferença de energia entre dois níveis adjacentes ($e_{\text{fóton}} = E_{v_2} - E_{v_1} = h\nu$) será sempre a mesma e nunca haverá uma banda de absorção no espectro vibracional. Em sistemas moleculares reais, a mudança do momento dipolo elétrico corresponde às transições de energia entre níveis energéticos variáveis e as transições possíveis são aquelas de $\Delta v = \pm 1, \pm 2, \pm 3, \pm 4$ etc. As ocorrências espectrais na frequência do infravermelho próximo são dominadas por *overtones* e combinações de bandas de absorção cujas transições de momento dipolo elétrico das moléculas assumem $\Delta v = \pm 2$ (do estado de energia $v = 0$ até $v = 2$ – transições pouco prováveis).

Os dados espectrais da região do infravermelho próximo são comumente associados às técnicas quimiométricas para quantificar multiconstituintes de amostras sem a necessidade de pré-processamentos laboratoriais, de forma rápida e não destrutiva (Cozzolino *et al.*, 2004; Hongqiang e Hongzhang, 2008). Salienta-se que a utilização da calibração multivariada com a espectroscopia na região do NIR apresentou resultados promissores em aplicações distintas de quantificação como, por exemplo: etanol e glicose (Liebmann *et al.*, 2009; Pinto *et al.*, 2016; Pinto *et al.*, 2014a); glicerol e ácidos acéticos (Blanco *et al.*, 2004); glicose e xilose (Morita *et al.*, 2014); celulose, hemicelulose e lignina (Rodríguez-Zúñiga *et al.*, 2014); compostos fenólicos (Cozzolino *et al.*, 2004). Com estes resultados, a regressão multivariada de dados juntamente com o NIR mostrou-se uma técnica analítica com potencial para reduzir drasticamente o tempo e o custo de análise no monitoramento de compostos químicos em

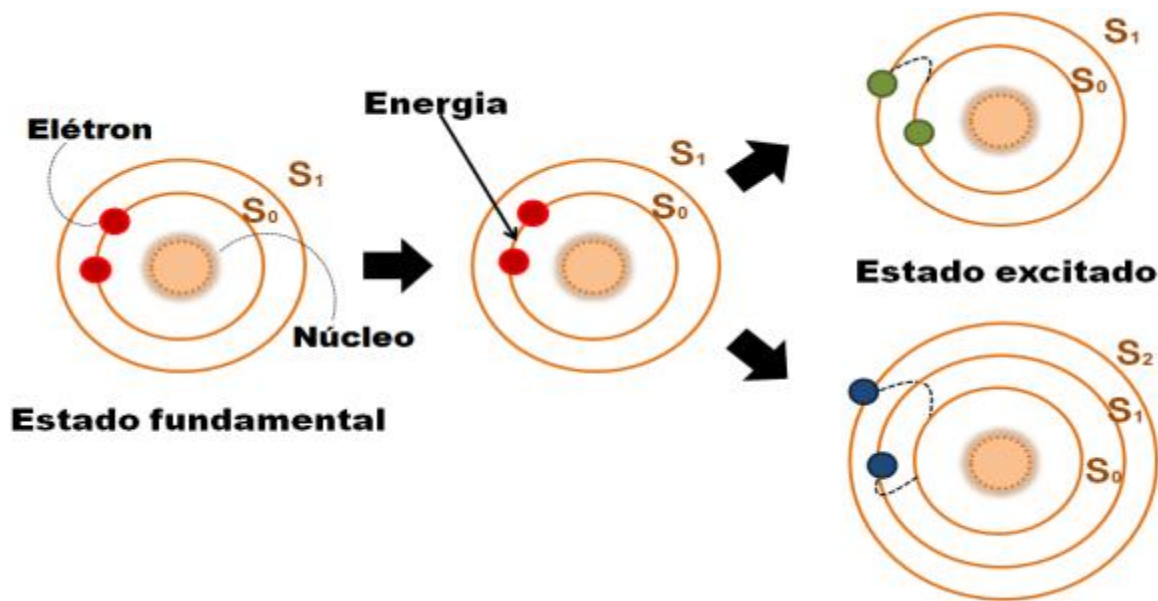
relação à quantificação realizada por HPLC, tanto para aplicações comerciais de produção quanto para pesquisa (Frizon *et al.*, 2015; Monrroy *et al.*, 2015).

Ressalva-se que, embora haja vasta aplicabilidade da espectroscopia na região do NIR na quantificação de analitos em misturas complexas, bandas de absorção específicas podem afetar negativamente o processo de modelagem. A fim evitar esta interferência, a literatura propõe a seleção de diferentes regiões espectrais para calibrar modelos de quantificação específicos na região do infravermelho próximo (Finn *et al.*, 2006; Morita *et al.*, 2011; Nishii *et al.*, 2012; Pinto *et al.*, 2016c). Além disso, os principais grupos funcionais envolvidos neste tipo de espectroscopia são aqueles que possuem átomos de hidrogênio, tais como $C-H$, $N-H$, $O-H$, e $S-H$ (Siesler *et al.*, 2002). Por conseguinte, a principal desvantagem de considerar esta região espectral é que somente átomos ligados ao hidrogênio produzem bandas intensas nesta região (harmônicas e regiões de combinação) . Desta forma, uma vasta gama de compostos não pode ser analisada através da varredura de amostras no NIR (Lourenço *et al.*, 2012).

3.5.2. Espectroscopia no Ultravioleta Visível (UV-Vis)

Espectroscopia UV-Vis corresponde à radiação no comprimento de onda do intervalo de 200 - 700 nm, o que proporciona fótons de maior energia do que na região do NIR. Nesta técnica, para algumas moléculas e átomos, a radiação possui energia suficiente para proporcionar transições de elétrons entre os orbitais eletrônicos. Ou seja, o comprimento de onda absorvido possui energia suficiente para deslocar um elétron (e^-) de um baixo nível de energia para um alto nível de energia (Owen, 2000). A Figura 5 ilustra esquematicamente este efeito em um átomo de orbitais S_i .

Figura 5 - Representação esquemática do efeito de transição eletrônica entre orbitais (S_i) de um átomo com a incidência de comprimentos de onda na região do UV-Vis.



Fonte: *Acervo pessoal*.

Nas moléculas, os níveis vibracionais e rotacionais dos orbitais moleculares são sobrepostos nos níveis eletrônicos. Devido à grande variabilidade de transições em diferentes níveis de energia que podem ocorrer, sucede o alargamento das bandas de absorção (Lourenço *et al.*, 2012).

A técnica de espectroscopia no ultravioleta visível é adequada para a detecção de alterações nos níveis eletrônicos de baixa energia, associada com grupos funcionais mais reativos (encontrados, por exemplo, em: compostos aromáticos e ligações múltiplas). O custo de operação relativamente baixo do UV-Vis, bem como a possibilidade de detectar transferências individuais de elétrons entre orbitais ou bandas atômica, iônica e molecular são as principais vantagens associadas à técnica espectroscópica. Além disso, a quimiometria com base no UV-Vis mostrou-se eficiente para quantificação de compostos fenólicos individuais e totais durante o processo de produção de vinho, bem como aqueles presentes no café (Belguidoum *et al.*, 2014; Cozzolino, 2015; Khani *et al.*, 2016; Martelo-Vidal e Vázquez, 2014).

3.6. Análise Quimiométrica em Espectroscopia

A simples inspeção visual dos dados espectroscópicos obtidos pelas técnicas propostas, geralmente, não é suficiente para extrair as informações relevantes sobre a composição das amostras. Assim, a fim de que as informações mais significativas dos espectros sejam extraídas, estes devem ser estatisticamente tratados a partir da análise multivariada de dados.

A calibração multivariada associada a diferentes espectroscopias tem sido investigada na quantificação de compostos químicos a partir de diferentes comprimentos de onda, incluindo o ultravioleta-visível (UV-Vis), infravermelho próximo (NIR), infravermelho médio (MIR), Raman e fluorescência (Lourenço *et al.*, 2012).

As técnicas espectroscópicas na região do ultravioleta-visível (UV-Vis) e infravermelho próximo (NIR), quando associadas à calibração multivariada, destacam-se por apresentarem potencial aplicação no monitoramento de compostos fenólicos e furaldeídos durante a produção de etanol 2G, visto que estas técnicas espectroscópicas foram eficientes na quantificação de fenólicos em misturas multiconstituintes (Belguidoum *et al.*, 2014; Cozzolino, 2015; Frizon *et al.*, 2015; Khani *et al.*, 2016; Martelo-Vidal e Vázquez, 2014; Monroy *et al.*, 2015; Siesler *et al.*, 2002).

A análise de componentes principais (PCA), a regressão por componentes principais (PCR), e a regressão por mínimos quadrados parciais (PLS) são exemplos de técnicas quimiométricas mais utilizadas na calibração multivariada. Nestas regressões a relação linear entre a absorbância e a concentração do analito para determinado comprimento de onda é essencial (Li *et al.*, 2004; Swierenga *et al.*, 2000).

Atualmente, a regressão por mínimos quadrados parciais (do inglês, *Partial Least Squares* - PLS) é uma das principais técnicas aplicadas na quantificação de analitos, visto que esta recupera as informações desejadas dos espectros, fornecendo modelos que apresentam fácil interpretação (Lourenço *et al.*, 2012). A técnica de regressão por PLS é baseada principalmente no rearranjo da matriz de absorbâncias com o intuito de representar as variações presentes em muitas variáveis, através de poucos fatores. Ou seja, há a construção de um novo sistema de eixos (denominado variáveis latentes) que representará a natureza multivariada das absorbâncias e das concentrações em poucas dimensões (ou em determinado número de variáveis latentes (do inglês, *Latent Variables* - LV)) (Brereton 2007; Ferreira *et al.*, 1999). A determinação das LV que serão utilizadas nos modelos de previsão deve ser realizada a partir

da validação cruzada (vide seção 4.5.3). Nesta metodologia a escolha do número de LV se baseia na avaliação da magnitude dos erros de previsão de dado modelo PLS, o que definirá a acurácia da modelagem (Brereton, 2007). As LV devem ser escolhidas cuidadosamente a fim de evitar o sobreajuste do modelo de regressão linear (Drennen JK, 1991).

O PLS associado a dados espectroscópicos vem se mostrando eficiente no monitoramento de compostos químicos presentes em meio reacional complexo (Bao *et al.*, 2014; Frizon *et al.*, 2015; Khani *et al.*, 2016; Wulandari *et al.*, 2016). A aplicação desta metodologia de quantificação destaca-se pela capacidade de monitorar multiconstituintes em misturas, com o mínimo, ou nenhuma, preparação da amostra, de forma rápida e não destrutiva (Blanco e Villarroya, 2002). Neste contexto, destaca-se a vasta aplicabilidade de modelos PLS-NIR-UV-Vis na quantificação de analitos em diferentes bioprocessos (Milligan *et al.*, 2014; Pinto *et al.*, 2016 ; Ribeiro *et al.*, 2008; Rodríguez-Zúñiga *et al.*, 2014; Sampaio *et al.*, 2014).

3.7. Calibração Multivariada dos Espectros

A análise espectrofotométrica para a quantificação de compostos químicos presentes em misturas multiconstituintes são baseadas, na maioria das vezes, na lei de *Lambert-Beer*. Esta lei define que a quantidade de luz absorvida é proporcional ao número de moléculas que a absorvem. Com base na lei de *Lambert-Beer* (Owen, 2000), tem-se a definição de transmitância pela equação 13.

$$\text{Transmitância} = -\left(\frac{I}{I_0}\right) = e^{-\kappa d \varsigma} \quad (13)$$

Na qual I_0 é a intensidade da radiação incidente, I corresponde à intensidade da radiação transmitida, κ é uma constante de proporcionalidade, d refere-se ao caminho óptico (em cm) e ς equivale à concentração do analito (geralmente expressada em g/L).

Ao linearizar a equação 13, tem-se a equação 14.

$$\text{Abs} = -\log_{10} T = \log_{10} \left(\frac{I_0}{I}\right) = \varepsilon d \varsigma \quad (14)$$

Em que *Abs* é a absorbância e ε é, por definição, a absorvidade molar de um analito específico em determinado solvente definida por $\varepsilon = \text{Abs}/C.d$, sendo C a concentração molar do analito na amostra (em mol/L) e d o caminho óptico (em centímetros).

A equação 14 é conhecida por lei de Beer. A dependência linear entre concentração e absorvância de um analito para um determinado comprimento de onda é a base para métodos quantitativos.

O modelo de regressão por mínimos quadrados clássico (do inglês, *Classical Least Squares* - CLS) deriva diretamente desta lei. Considerando $nd_{cal.}$ amostras de calibração, n_λ comprimentos de onda discretos e n_c componentes químicos, a representação matricial desta regressão pode ser escrita pela equação 15 (Haaland e Thomas, 1988).

$$\mathbf{A} = \mathbf{CK} + \mathbf{E}_A \quad (15)$$

Onde \mathbf{A} é a matriz de $nd_{cal.} \times n_\lambda$ da resposta do instrumento (espectros de calibração), \mathbf{C} é a matriz $nd_{cal.} \times n_c$ de concentração de componentes, \mathbf{K} é a matriz $n_c \times n_\lambda$ de parâmetros que representam o produto entre a absorvância específica de vezes o comprimento do percurso óptico de dado componente puro, e \mathbf{E}_A é a matriz $nd_{cal.} \times n_\lambda$ dos erros espectrais não ajustados pelo modelo. Então, \mathbf{K} é uma matriz de parâmetros ajustáveis, que pode ser calculado utilizando um método de mínimos quadrados clássico para minimizar $\|\mathbf{E}_A^T \mathbf{E}_A\|$ (Nelles, 2001).

$$\mathbf{K}_{opt} = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{A} \quad (16)$$

A regressão CLS se fundamenta no cálculo do \mathbf{K}_{opt} que minimiza a variância do erro na absorvância, \mathbf{E}_A . Assim, este método torna-se ideal para prever absorvâncias, e não concentrações.

A previsão de concentração pode ser melhorada consideravelmente usando a *Lei de Lambert-Beer Inversa*, na qual, tem-se:

$$\mathbf{C} = \mathbf{A} \mathbf{P}_{opt} + \mathbf{E}_C \quad (17)$$

Em que \mathbf{c} é a matriz de concentrações ($nd_{cal.} \times n_c$) de apenas um analito, \mathbf{E}_C é a matriz de resíduos ($nd_{cal.} \times n_c$) e \mathbf{A} é a matriz absorvâncias ($nd_{cal.} \times n_\lambda$). A matriz \mathbf{P}_{opt} minimiza o erro $\|\mathbf{e}_C^T \mathbf{e}_C\|$ na previsão de \mathbf{c} , assim:

$$\mathbf{P}_{opt} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{C} \quad (18)$$

Este método é denominado *Inverse Least Squares* (ILS). Destaca-se que a matriz resultante $\mathbf{A}^T\mathbf{A}$, a ser invertida, tem dimensão $n_\lambda \times n_\lambda$, então, neste tipo de regressão linear o número discreto de comprimentos de onda utilizadas no modelo não pode exceder o número de amostras usadas na calibração ($n_\lambda \leq nd_{cal.}$) para que \mathbf{P}_{opt} seja ajustável. Entretanto, mesmo que n_λ de \mathbf{A} não excedesse $nd_{cal.}$, poderiam surgir problemas de multicolinearidade quando n_λ aumentasse. A multicolinearidade faz com que $\mathbf{A}^T\mathbf{A}$ seja singular ($rank(\mathbf{A}) \leq nd_{cal.}, n_\lambda$), ou seja, os componentes da matriz resultante podem formar um conjunto de dados linearmente dependentes (LD). Conseqüentemente, $\mathbf{A}^T\mathbf{A}$ não admite inversa e, portanto, o parâmetro \mathbf{P}_{opt} não pode ser ajustado pela regressão ILS.

A fim de sanar os problemas de singularidade da matriz $\mathbf{A}^T\mathbf{A}$, outras metodologias de regressão multivariada foram propostas. A maior parte delas divide-se em duas grandes categorias (Araújo *et al.*, 2001). O primeiro grupo inclui os métodos que reduzem a singularidade da $\mathbf{A}^T\mathbf{A}$, escolhendo n'_λ comprimentos de onda específicos (colunas de \mathbf{A}) em vez de usar o espectro completo. A vantagem desta abordagem é que a matriz resultante ($nd_{cal.} \times n'_\lambda$) é também uma matriz de absorbâncias. Assim, não há modificação do significado físico da equação 11. O segundo grupo inclui os métodos que tratam o espectro completo em um novo sistema de coordenadas e, em seguida, reduzem o número de entradas do modelo (normalmente chamado de variáveis latentes, ou, ainda, componentes principais). A abordagem deste grupo tem a vantagem de utilizar apenas parte da informação completa, ou seja, consideram-se a combinação linear das bandas de absorção que contém a maior variância dos dados. Desta forma, o significado físico do modelo previsto pela *Lei de Lambert-Beer Inversa* é perdido ao considerar um novo sistema de coordenadas para representar \mathbf{A} (Brereton, 2007).

O algoritmo de projeção sucessiva (do inglês *Successive Projection Algorithm*, SPA) é um exemplo de métodos pertencentes ao primeiro grupo. Este método se baseia na seleção progressiva, que consiste em recolher n variáveis minimamente redundantes da matriz \mathbf{A} . A partir de uma coluna inicial (\mathbf{a}_0) de \mathbf{A} , o algoritmo determina qual coluna, \mathbf{a}_1 , tem a maior projeção no subespaço, \mathbf{S}_0 , ortogonal à \mathbf{a}_0 . Na próxima iteração, SPA é restrito ao subespaço \mathbf{S}_0 (de dimensão $nd_{cal.} - 1$). A projeção do \mathbf{a}_1 sobre \mathbf{S}_0 é a nova coluna inicial (\mathbf{a}_0'). As iterações são realizadas até n variáveis de \mathbf{A} (\mathbf{a}_{ns}) sejam escolhidas. Então, MLR é aplicado usando uma nova matriz $\hat{\mathbf{A}}$, construída com as n colunas escolhidas de \mathbf{A} .

Há ainda o método *Principal Component Regression* (PCR), que pertence ao segundo grupo. Nesta metodologia, aplica-se a análise de componente principal (do inglês, *Principal Component Analysis*, PCA) na matriz \mathbf{A} , em seguida utiliza-se as componentes de \mathbf{A} como variáveis independentes no modelo MLR para a previsão da matriz \mathbf{C} .

No PCA, a matriz \mathbf{A} é projetada para um novo conjunto de eixos de coordenadas ortonormal ($\boldsymbol{\alpha}$, autovetores de $\mathbf{A}^T\mathbf{A}$), que são calculados a fim de maximizar a variância dos dados ($\mathbf{A}^T\mathbf{A}$) ao longo dos eixos (Abdi, 2010). Então, encontra-se uma função linear que maximize a variância em \mathbf{A} , expressada pela combinação linear de cada comprimento de onda (n_λ). A equação 19 expressa matematicamente a combinação linear entre n_λ variáveis independentes.

$$\mathbf{A}' = \mathbf{TP} \quad (19)$$

Em que, as colunas da matriz \mathbf{T} são chamadas de componentes principais (do inglês, *Principal Components* - PC) ou de *scores* (autovetores). A matriz \mathbf{P} é denominada *loadings* (autovalores). Salienta-se que cada coluna de \mathbf{T} maximiza a variância de \mathbf{A}' , sendo que cada uma é linearmente independente e ortonormal à outra. É possível determinar n PC, mas é desejável que haja a redução das variáveis independentes. Assim, a maior variância de \mathbf{A}' deve ser representada por p componentes principais ($p \leq n$) (Jolliffe, 1986).

Algebricamente, a obtenção das componentes principais é baseada na técnica dos Multiplicadores de Lagrange. Jolliffe (1986) apresenta todo detalhamento matemático, bem como a metodologia para escolha do número de PCs ótimo (p). Então, utiliza-se a matriz \mathbf{A}' para prever a matriz \mathbf{C} a partir do ILS.

O PCR proporciona modelos com previsões razoáveis de \mathbf{C} , visto que a ortogonalidade dos vetores singulares (PC) elimina o problema de multicolinearidade dos dados. Entretanto, as componentes principais são escolhidas para a melhor representação de \mathbf{A} , não considerando a estimativa ótima da matriz \mathbf{C} . Ou seja, não há garantia de que as PC escolhidas para representar a absorvância (\mathbf{A}) serão significativos para a previsão da concentração (\mathbf{C}) (Abdi, 2010). O SPA também apresenta este inconveniente, no qual o critério relevante de cada entrada baseia-se apenas nos dados de entrada (\mathbf{A}). Portanto, em alguns casos, tanto o SPA quanto o PCR podem descartar entradas que são altamente relevantes para o desempenho do modelo de estimativa para as concentrações.

A solução deste problema está na utilização da regressão por *Partial Least Squares* (PLS). Na regressão PLS, determina-se o conjunto de scores (chamados de variáveis latentes-LV) que permite a decomposição de \mathbf{A} e \mathbf{C} simultaneamente. Desta forma, ambas as matrizes \mathbf{A} e \mathbf{C} são projetadas em um novo conjunto comum as duas, a fim de maximizar a covariância $\mathbf{A}^T\mathbf{C}$ (Haaland e Thomas, 1988). Neste caso, tanto as informações de entrada como as informações de saída são levadas em consideração na transformação das coordenadas do sistema, podendo gerar modelos de previsão melhores.

Em decorrência desta peculiaridade, a regressão PLS frente às demais técnicas é vantajosa. Assim, atualmente, o PLS é método mais reportado na literatura para a calibração multivariada espectrométrica (Bao *et al.*, 2014; Cozzolino, 2015; Khani *et al.*, 2016; Pinto *et al.*, 2016c; Sarraguça *et al.*, 2009). Abdi (2010) apresenta detalhadamente o desenvolvimento matemático para esta técnica e Ferreira *et al.* (1999) reportou um tutorial sobre calibração multivarida na linguagem MATLAB.

4. MATERIAIS E MÉTODOS

4.1. Preparação de Amostras

O conjunto de dados considerados nos modelos de calibração deste trabalho contém tanto amostras (S), quanto amostras de hidrolisado do pré-tratamento hidrotérmico (LHW, do inglês, *Liquid Hot Water*) do bagaço de cana-de-açúcar.

As amostras sintéticas foram preparadas a partir da solubilização em água/etanol dos principais compostos químicos prejudiciais ao processo de hidrólise enzimática e de fermentação alcoólica. Assim, prepararam-se soluções-padrão de vanilina, ácidos gálico e p-cumárico. A solubilização dos analitos foi padronizada utilizando soluções de água:etanol em proporções que viabilizaram a solubilização completa destes compostos na concentração de 5 g/L (vanilina: 0,17 kg de etanol por 1kg de solução; ácido gálico: 0,17 kg de etanol por 1kg de solução; ácido p-cumárico: 0,36 kg de etanol por 1kg de solução). A partir das soluções-padrão, realizaram-se diversas diluições e misturas entre os componentes estudados (proporções definidas em laboratório). O número total de amostras sintéticas utilizadas na modelagem do PLS-NIR foi de 78 e do PLS-UV-Vis foi de 112.

As amostras reais foram obtidas a partir do pré-tratamento LHW do material lignocelulósico cordialmente cedido pela Destilaria Santa Clara (Jaboticabal, SP). A biomassa *in natura* foi previamente moída no moinho de facas e, em seguida, peneirada, resultando em partículas com diâmetro inferior a 2 mm. O pré-tratamento LHW foi realizado de acordo com a metodologia descrita por Vasconcellos *et al.*(2015) no reator de 5L da *Parr Instruments* (Modelo 4580). A fim de obter modelos de monitoramento aplicáveis em diferentes condições de operação do processo estudado, foram avaliadas três condições de temperatura e tempo de digestão da biomassa para uma carga de sólidos fixada em, aproximadamente, 10% ($m_{biomas. seca}/m_{água}$). No total, 4 experimentos foram realizados. A Tabela 4 ilustra as condições de pré-tratamento hidrotérmico avaliadas neste trabalho. Após esta operação, o hidrolisado foi separado da biomassa pré-tratada através do processo de filtração simples.

Tabela 4 - Condições para aquisição de dados do hidrolisado advindo do pré-tratamento hidrotérmico (LHW) do bagaço de cana-de-açúcar (Psat).

Ensaio	Carga de sólidos (% m _{biomas. seca} /m _{água})	Temperatura (°C)	Tempo (minutos)	Fator de Severidade ⁶
LHW 01	10	195	10	4,10
LHW 01'	9,3	195	10	4,10
LHW 02	10	170	15	3,39
LHW 03	10	220	5	4,70

Cerca de 40 amostras foram obtidas a partir da diluição do hidrolisado advindo de cada pré-tratamento LHW, totalizando 160 amostras. Além de diluídas, as amostras de hidrolisado foram modificadas com a adição de pequenos volumes das soluções sintéticas citadas anteriormente, como o sugerido em trabalhos anteriores (Martelo-Vidal e Vázquez, 2014b; Milligan *et al.*, 2014).

Com o intuito de prevenir a degradação dos compostos químicos presentes nas amostras estudadas, as soluções padrão, bem como as amostras do hidrolisado foram armazenadas sob refrigeração e na ausência de luz até que fossem utilizadas.

4.2. Método Analítico de Referência para as Amostras do Hidrolisado

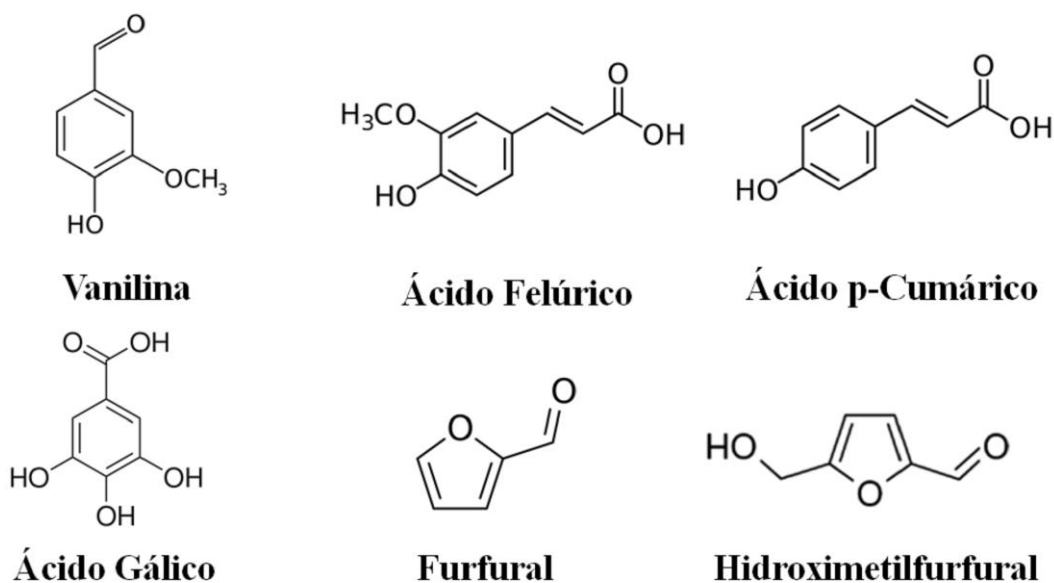
Os inibidores presentes na água do pré-tratamento LHW foram quantificados por HPLC-UV. Utilizou-se a coluna C-18 (Microsorb-MV 100 com 250 cm x 4,6 mm e membrana de 5 μ m) com detector UV-Vis em 277 nm e sob refrigeração (T ~ 20 °C). Metanol foi utilizado como solvente e uma solução aquosa de ácido trifluor acético (0,025% v/v) como fase móvel (5:95%, v/v) para a caracterização dos analitos de interesse. O tempo de retenção considerado foi de 35 minutos (100% de ácido trifluor acético) em rampa, partindo de 95% de fase móvel e 5% de metanol. A vazão da coluna foi mantida constante em 0,8 ml/min. Após a aquisição dos dados, consideraram-se o equilíbrio de 10 min. até 5% de metanol e 5 min. de estabilização. As amostras foram filtradas para remover partículas suspensas com diâmetro superior a 0,45 μ m. Nesta metodologia, foi possível quantificar a concentração dos principais componentes do hidrolisado: vanilina (VAN, 4-hidroximetil-2-furaldeído), ácido gálico (GAL, ácido 3,4,5-trihidróxibenzóico), ácido p-cumárico (APC, ácido 4-hidroxicinâmico), ácido felúrico (FEL –

⁶ Definido por OVEREND, CHORNET (1987)

ácido 4-hidroxi-3-metoxicinâmico), hidroximetilfurfural (HMF, 5-hidroximetil-2-furaldeído) e furfural (FUR, 2-furaldeído).

A Figura 6 apresenta as fórmulas moleculares destes compostos químicos.

Figura 6 – Fórmulas moleculares dos compostos químicos quantificados nas amostras S e LHW.



Fonte: Acervo pessoal.

4.3. Reagentes

Todos os reagentes utilizados possuíam pureza analítica. Água *mili-Q* foi utilizada em todos os padrões, bem como nas diluições realizadas. Vanilina, ácido gálico, ácido p-cumárico, ácido felúrico e hidroximetil-furfural foram comprados da Sigma Aldrich (Brasil), enquanto que o furfural foi comprado da Impex (LABIMPEX Indústria e Comércio de Produtos para Laboratórios Ltda.). Todas as soluções obtidas foram utilizadas sem purificação adicional.

4.4. Varredura de Espectros

Os espectrômetros FT-NIR da Perkin Elmer (*Spectrum 100N*- cubeta de quartzo) e UV-Vis da Shimadzu (UV-1600, *software* UVProbe 2.31- cubeta de quartzo) foram usados na varredura *off-line* de todas as amostras avaliadas neste estudo. O intervalo de absorvâncias obtido no FT-NIR correspondeu ao intervalo de 700 a 2.500 nm, enquanto que no UV-Vis este foi de 200 a 600 nm. Os espectros no NIR foram obtidos a uma temperatura de 19 °C e para o UV-Vis na temperatura média de 27 °C.

4.5. Tratamentos Matemáticos dos Dados Espectroscópicos

Os procedimentos de refinamento e tratamento dos espectros obtidos, bem como as análises estatísticas necessárias, foram implementados no Matlab[®] (*PLS tool box* - 2011a software, *Mathworks*, USA).

4.5.1. Pré-processamento dos Espectros

Os espectros originais foram pré-processados sendo que, para este estudo, foram considerados absorbâncias centradas na média, bem como as derivadas de primeira e segunda ordem dos espectros alisados baseadas no método Savitzky-Golay (Rinnan *et al.*, 2009; Savitzky e Golay, 1964).

O objetivo dos pré-processamentos de dados foi remover matematicamente variações aleatórias ou sistemáticas indesejáveis que não seriam removidas naturalmente durante a análise quimiométrica dos dados e que poderiam interferir nos resultados finais, ressaltando ou mascarando as informações relevantes do espectro (Brereton, 2007; Rinnan *et al.*, 2009).

4.5.1.1. Absorbâncias Centradas na Média

A absorbância centrada na média dos dados é apenas uma translação de eixos dada pela equação 20.

$$a_{ij} = a_{ij} - \bar{a}_j \quad (20)$$

Dado a i -ésima amostras, \bar{a}_j é a média dos dados de absorbância no comprimento de onda j . Geometricamente equivale a fazer uma translação do sistema de eixos ao longo do vetor de absorbâncias para o centro do conjunto de dados, o qual equivale ao valor médio das absorbâncias. Neste pré-processamento há apenas a translação da origem do sistema de eixos para o centroide, conservando-se a variação das bandas de absorção (Ferreira *et al.*, 1999).

4.5.1.2. Derivadas

O principal objetivo de derivar os espectros obtidos pelo NIR é remover o deslocamento e inclinação na linha de base dos espectros, pois estas imperfeições produzem variações indesejáveis nas medidas de absorvância (Brereton 2007).

O deslocamento do espectro inteiro de uma quantidade constante na linha de base é denominado *offset* e seus efeitos podem ser contornados tomando-se a primeira derivada do conjunto de dados (Brereton 2007). Assim, a translação da linha de base não afeta as tarefas de calibração ou de previsão.

Neste trabalho, utilizou-se o algoritmo de Savitsky-Golay no cálculo das derivadas e alisamento prévio dos espectros (Brereton, 2007; Savitzky and Golay, 1964).

Ressalta-se que cada ordem de derivada adicionada pode produzir um erro extra nos dados, relacionado aos erros de cálculo (Ferreira *et al.*, 1999).

4.5.2. Calibração Multivariada

Neste trabalho, a regressão PLS foi realizada a partir do *PLS tool box* do Matlab[®]. A escolha do número ótimo de variáveis latentes dos modelos PLS foi realizada a partir da validação cruzada (LOO), os conceitos básicos desta técnica são descritos na próxima seção.

4.5.3. Validação cruzada e Teste dos Modelos

A validação cruzada é uma metodologia para sistematizar o aumento da complexidade de um modelo empírico evitando, ao mesmo tempo, o sobreajuste do modelo (Nelles, 2001). A validação cruzada é comumente utilizada na escolha do número de componentes principais do modelo PLS. Esta escolha deve ser baseada na avaliação cuidadosa da magnitude dos erros de previsão de um dado modelo de calibração, visto que definir um número de regressores inadequado podem acarretar em *overfitting*, ou seja, os modelos PLS são gerados com sobreajuste (Ferreira *et al.*, 1999). Na validação cruzada conhecida como *leave-one-out* (LOO), retira-se uma amostra do conjunto de dados de calibração e constrói-se um modelo de calibração multivariada. Utiliza-se este modelo para prever a concentração dos analitos na amostra previamente excluída do conjunto de calibração. Subsequentemente, a partir do erro desta estimativa de concentração, calcula-se o desvio padrão de previsão da validação cruzada

(*Root Mean Square Error Prediction of Cross Validation* - RMSECV). Assim, com o RMSECV determina-se o número de componentes principais, também chamado de variáveis latentes (do inglês, *Variables Latents* - LV).

O número ótimo de LV de cada modelagem corresponde à melhor previsão de concentrações sem que haja *overfitting*. O cálculo do i -ésimo RMSECV é geralmente realizado a partir da equação 20.

$$RMSECV^i = \sqrt{\frac{1}{nd_{cal.}} \sum_{j=1}^{nd_{cal.}} \left(c_j^{(i)} - \hat{c}_j^{(i)} \right)^2} \quad (20)$$

Em que, $nd_{cal.}$ corresponde ao número de dados do conjunto de calibração, $c_k^{(i)}$ e $\hat{c}_k^{(i)}$ são a referência e o valor predito das concentrações usadas na validação cruzada da amostra j e analito i .

Além de definir o número de LV, o RMSECV fornece a precisão do modelo PLS na previsão da concentração dos dados de calibração. Embora este erro seja essencial à análise dos modelos, a acurácia na quantificação das concentrações ainda deve ser validada a partir da previsão de dados não relacionados ao conjunto de calibração. Assim, para o teste do modelo, considera-se o novo conjunto de dados de validação e calcula-se o desvio padrão de previsão (*Root Mean Square Error of Prediction* - RMSEP). O RMSEP deve ser obtido analogamente ao RMSECV, dado que $\hat{c}_k^{(i)}$, na equação 20, admite o valor predito pelo modelo PLS (Brereton, 2007).

Há ainda outros três parâmetros de performance considerados neste trabalho: o coeficiente de determinação (R^2); a razão entre o desvio padrão da concentração no conjunto de dados de teste e o RMSEP (do inglês, *Residual Predictive Deviation* - RPD); a razão entre a faixa de validade do modelo e o RMSEP (do inglês, *Range Error Ratio* - RER).

O R^2 corresponde ao coeficiente de determinação do ajuste linear entre as concentrações de referência e as previstas pelo modelo PLS.

O parâmetro RPD foi calculado a partir da equação 21.

$$RPD = \frac{DP_{cal.}}{RMSECV} \quad (21)$$

Em que DP_{cal} é o desvio padrão da concentração no conjunto de dados de calibração. Valores de RPD altos equivalem a modelos de boa qualidade, visto que estes possuem o desvio padrão de previsão (RMSECV) inferior ao desvio padrão da média de concentrações na calibração. Ou seja, se $RPD=1$, tem-se que a calibração pela regressão linear PLS não oferece previsões de concentração melhores do que o desvio padrão dos dados e, conseqüentemente, a aplicação do modelo PLS não é justificável. Neste trabalho, somente $RPD \geq 2$ foram considerados satisfatórios na quantificação dos analitos analisados.

O parâmetro de performance RER foi calculado pela equação 22.

$$RER = \frac{range}{RMSECV} \quad (22)$$

No qual, *range* é a faixa de variação na concentração dos dados teste. Quanto maior o valor assumido pelo RER, maior a precisão do modelo. Por exemplo, $RER=10$ significa que o erro padrão da resposta do modelo foi inferior a um décimo do intervalo de calibração. Somente modelos PLS que apresentaram $RER \geq 10$ foram considerados adequados no monitoramento dos analitos estudados.

Salienta-se que modelos de previsão viáveis devem apresentar baixos RMSECV e RMSEP. Adicionalmente, quanto maiores forem os parâmetros RPD e RER, melhor será a eficácia da calibração pela regressão PLS. Embora alguns limiares sejam sugeridos para RPD e RER (Chang *et al.*, 2001), estes não se baseiam em qualquer base estatística (Bellon-Maurel *et al.*, 2010). Em vez disso, a aceitação destes níveis depende essencialmente da área de pesquisa e aplicação dos modelos de monitoramento.

4.5.4. Outliers: Teste Estatístico de Leverage e Resíduo de Student

Outliers são, por definição, amostras que destoam da média obtida no conjunto de amostras totais aferidas (Brereton, 2007). A presença de *outliers* na calibração/ validação pela metodologia da calibração multivariada pode acarretar em diversos problemas, dentre eles a falta de acuracidade da calibração, ou seja, há erros de validação do modelo elevados ou este prevê concentrações sem significado físico (Ferreira *et al.*, 1999).

Há várias abordagens para identificar a influência de uma amostra do conjunto de experimentos no modelo de regressão, como por exemplo, o teste estatístico de *Leverage* (Zewotir e Galpin, 2007).

O *Leverage* pode ser interpretado geometricamente como a distância de uma amostra ao centroide do conjunto de dados. Por definição, este teste foi aplicado a partir do cálculo do *Leverage* das n amostras, que corresponde à diagonal da matriz $\mathbf{H}_{n \times n}$ (h_{ii} , com $h_{ii} > 0$ para qualquer amostra) (Ferreira *et al.*, 1999). A equação 23 apresenta o cálculo realizado para a obtenção da matriz \mathbf{H} .

$$\mathbf{H} = \mathbf{T} \cdot \mathbf{T}^t \quad (23)$$

Onde a matriz \mathbf{T} é a matriz scores (autovetores), cujas dimensões $n \times d$, em que n é o número de amostras e d corresponde aos coeficientes ajustados no modelo de calibração.

A fim de identificar as amostras anômalas, definiu-se o *Leverage* crítico (h_{crit}) do conjunto de experimentos pela equação 24.

$$h_{crit} = 3 \cdot \frac{LV}{nd_{cal}} \quad (24)$$

Na qual, LV é o número de variáveis latentes e nd_{cal} é o número de dados experimentais considerados no modelo de calibração. Então, quando $h_{ii} > h_{crit}$ tem-se que a i -ésima amostra é um *outliers*. Portanto, estas amostras foram removidas do conjunto de calibração. Além disso, se $h_{ii} > 1$, pode-se afirmar que houve *over-fitting* do modelo de calibração (Ferreira *et al.*, 1999).

É possível ainda que seja feita a análise dos resíduos de Student (RS). Este parâmetro é calculado a partir da equação 25.

$$RS = \frac{(c_k^{(i)} - \hat{c}_k^{(i)})}{Lresc_i \sqrt{1 - h_{ii}}} \quad (25)$$

Onde, $Lresc_i = \sqrt{\frac{(c_k^{(i)} - \hat{c}_k^{(i)})^2}{(nd_{cal} - 1)(1 - h_{ii})}}$. O $Lresc_i$ corresponde ao resíduo da amostra i corrigido pela *Leverage*. Nesta análise, obtêm-se a influência de cada amostra em particular

na calibração dos dados a partir de um gráfico de *Resíduos de Student* por amostra (Ferreira *et al.*, 1999).

Ao considerar os resíduos de Student na identificação de *outliers*, analisa-se um gráfico de *RS versus Leverage*, que permite verificar quais amostras destoam do conjunto de dados. Assim, amostras que apresentem resíduos altos e baixos *Leverage* devem ser excluídas do conjunto de dados e o modelo de calibração reconstruído (Ferreira *et al.*, 1999). Neste estudo, esta metodologia foi empregada na identificação de possíveis amostras anômalas no conjunto total de dados.

5. RESULTADOS E DISCUSSÃO

Nas seções subsequentes foi apresentado o resultado da metodologia de monitoramento proposta aplicada em conjuntos espectrais de calibração formados a partir de amostras sintéticas e reais do hidrolisado do pré-tratamento hidrotérmico. O estudo de amostras sintéticas para os modelos de regressão linear multivariada visou verificar como a mistura de alguns analitos pode interferir no desempenho da análise multivariada de dados. Com os resultados desta análise, foram propostos modelos baseados em conjuntos de calibração reais. As regiões com comprimentos de onda relativos ao NIR e ao UV-Vis foram avaliadas nos métodos espectroscópicos. Os seguintes pré-processamentos espectrais foram aplicados em dados modelos (M_i): absorvâncias centradas na média (MC), retirada de regiões espectrais (NL) bem como a primeira (1ª Deriv.) e segunda (2ª Deriv.) derivadas. A técnica quimiométrica de calibração multivariada utilizada em todos os modelos propostos foi a regressão por Mínimos Quadrados Parciais (PLS). A acurácia na previsão de concentrações dos modelos PLS foram avaliadas a partir dos seguintes parâmetros de performance: RMSECV, RMSEP, R^2 , RPD e RER.

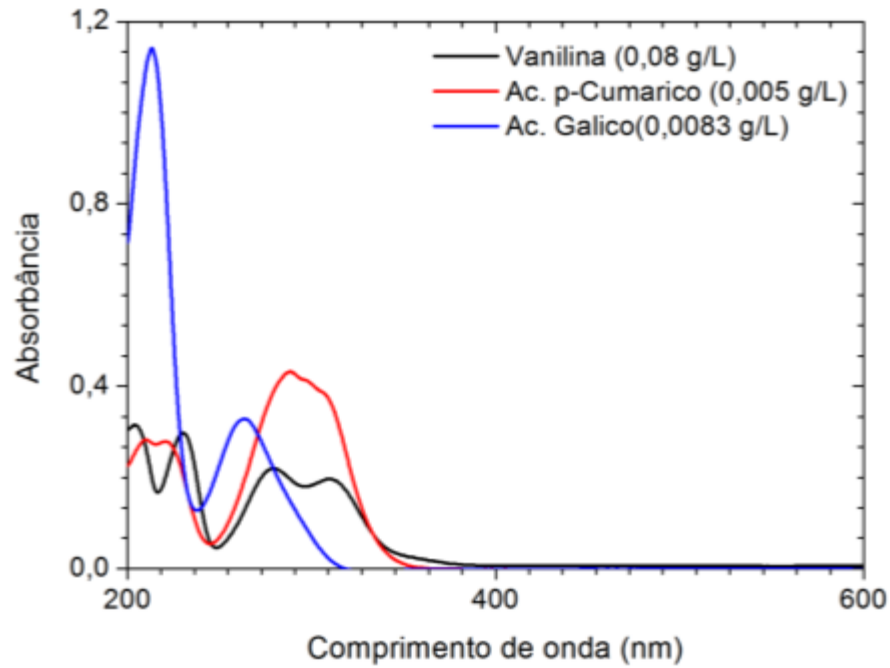
5.1. Conjunto de Dados Sintéticos (S)

Amostras de hidrolisado do pré-tratamento LHW constituem um sistema químico complexo e, conseqüentemente, a acurácia da calibração multivariada pode sofrer interferências. A utilização de conjuntos de calibração formados por amostras sintéticas pode fornecer quais são os possíveis fatores que prejudicam os parâmetros de desempenho PLS. Martelo-Vidal e Vázquez (2014b) reportaram que emprego de misturas sintéticas em modelagens PLS-UV-Vis- NIR mostrou-se satisfatório na quantificação dos principais elementos da fermentação do vinho como, por exemplo, ácidos orgânicos. Além disso, o desempenho dos modelos PLS pode ser aprimorado ao considerar diferentes pré-processamentos espectrais e ao realizar a calibração a partir de bandas de absorção específicas (Martelo-Vidal e Vázquez, 2014a,b; Milligan *et al.*, 2014). Baseado nestas informações, analisou-se o efeito de diferentes pré-processamentos espectrais nos parâmetros de desempenho dos modelos PLS para conjuntos de calibração/ teste sintéticos com o intuito de verificar a melhor quantificação dos principais componentes do hidrolisado do pré-tratamento LHW.

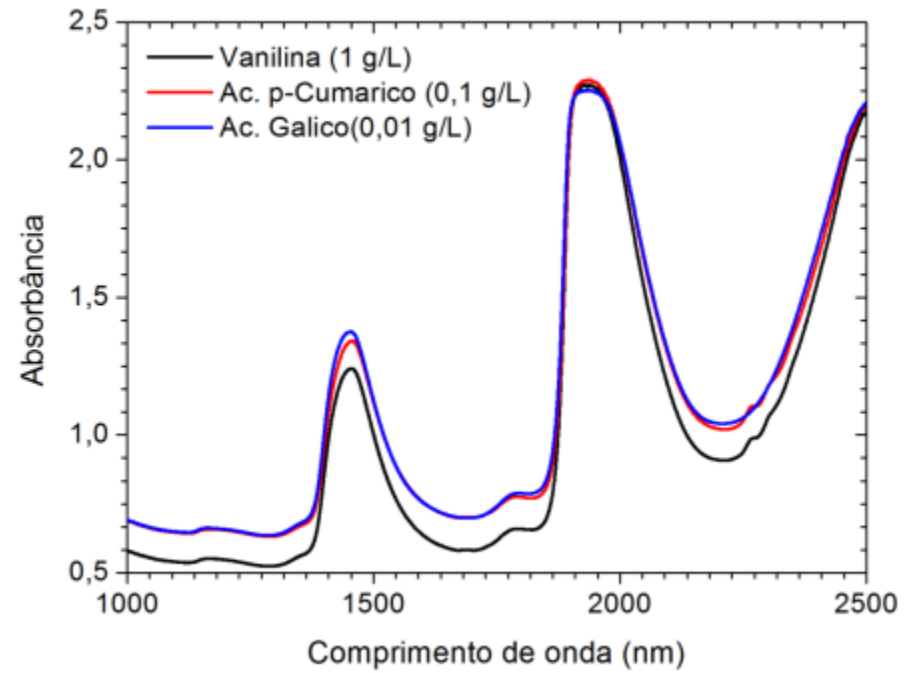
Os inibidores dos bioprocessos da produção de etanol 2G, tais como, o furfural, o hidroximetilfurfural, a vanilina, os ácidos felúrico, p-cumárico e gálico, merecem destaque na inibição/desativação de enzimas celulolíticas/ hemicelulolíticas, bem como na redução da produtividade de etanol durante a fermentação 2G (Delgenes *et al.*, 1996; Larsson *et al.*, 2000; Michelin *et al.*, 2016; Qin *et al.*, 2016; Ximenes *et al.*, 2011). Dentre os componentes do hidrolisado, selecionaram-se três que mais prejudicam tanto a hidrólise quanto a fermentação para compor as soluções puras e misturas do conjunto de dados S. Considerando-se o efeito negativo tanto nos processos de hidrólise quanto de fermentação, foram selecionados a vanilina (VAN), o ácido gálico (GAL) e o ácido p-cumárico (APC) como os principais compostos químicos para compor as amostras sintéticas (vide seções 3.2 - 3.4). Salienta-se que somente os compostos fenólicos foram considerados nesta análise preliminar, visto que estes analitos afetam negativamente todas as etapas subsequentes da produção do etanol 2G. A aquisição de amostras sintéticas e de espectros foi detalhada nas seções 4.1, 4.3 e 4.4. A Figura 7 ilustra os espectros de amostras S puras de VAN, APC e GAL no NIR e no UV-Vis.

Seguida da obtenção dos espectros, o primeiro tratamento de dados realizado foi o teste estatístico de Leverage e resíduos de Student, para os quais não foi considerado nenhum pré-processamento (NP) espectral. A partir destes testes, selecionou-se o melhor conjunto de amostras para calibração e validação dos modelos de previsão PLS-UV-Vis-NIR subsequentes, visto que todas possíveis amostras anômalas foram retiradas dos dados.

Figura 7 – Gráficos dos espectros originais no UV-Vis (a) e NIR (b) das amostras sintéticas puras.



a)



b)

Nas Tabelas 5 e 6, foram apresentados faixas de validade, médias e desvio padrão dos dados de concentração considerados nos modelos PLS-UV-Vis-NIR, bem como o número de amostras (NA) utilizadas na calibração e validação dos modelos S.

Tabela 5 - Faixa de validade, média e desvio padrão dos dados considerados nos modelos PLS-UV-Vis para dado número de amostras (NA) utilizadas na calibração e validação dos modelos sintéticos.

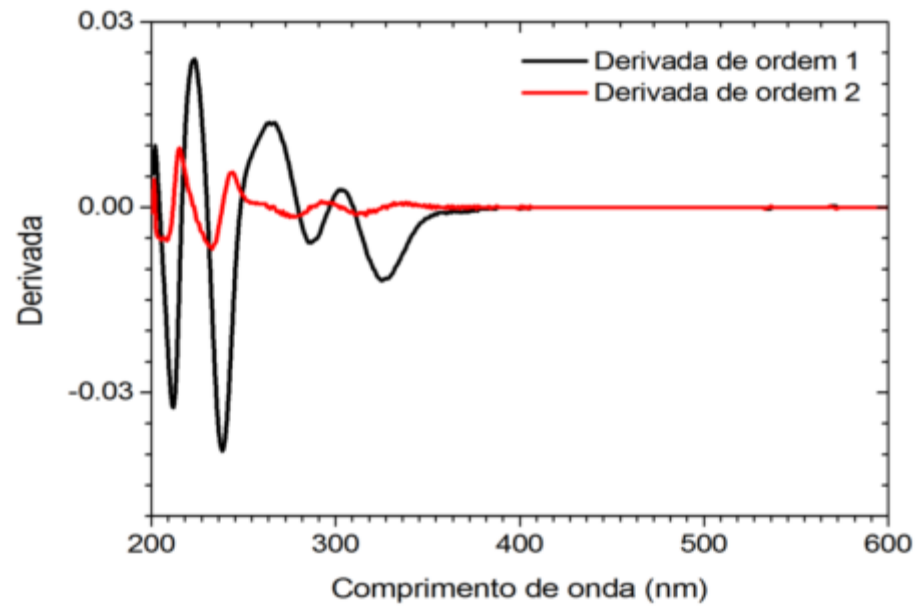
Analitos	Concentrações no conjunto de calibração (g/L) NA= 103				Concentrações no conjunto de teste (g/L) NA=9			
	Min.	Max.	Média	DP	Min.	Max.	Média	DP
VAN	0	0,050	0,0023	0,0084	0	0,012	0,0015	0,0040
GAL	0	0,018	0,0018	0,0042	0	0,010	0,0017	0,0033
APC	0	0,003	0,0002	0,0006	0	0,000	0,0000	0,0000

Tabela 6 - Faixa de validade, média e desvio padrão dos dados considerados nos modelos PLS-NIR para dado número de amostras (NA) utilizadas na calibração e validação dos modelos sintéticos.

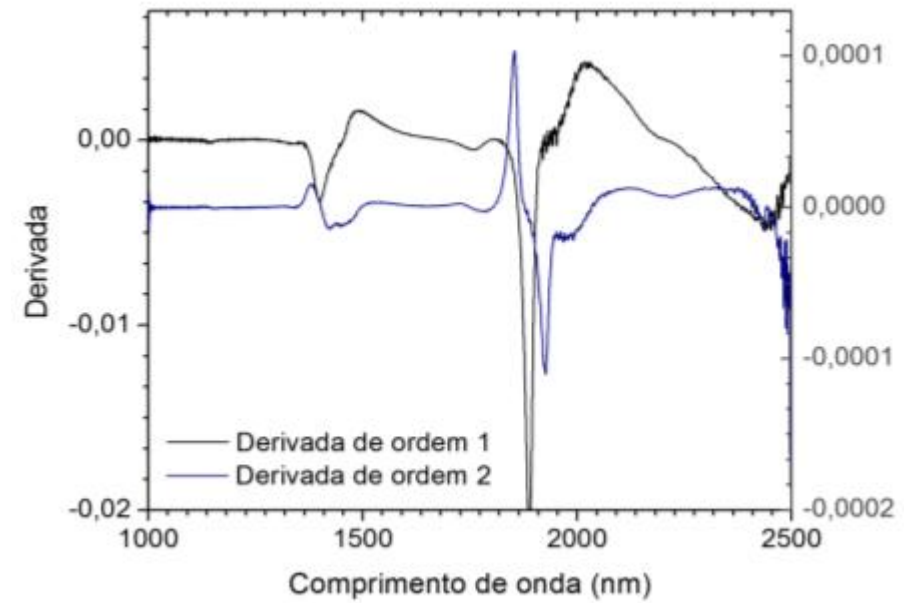
Analitos	Concentrações no conjunto de calibração (g/L) NA= 62				Concentrações no conjunto de teste (g/L) NA=16			
	Min.	Max.	Média	DP	Min.	Max.	Média	DP
VAN	0	2,000	0,118	0,3950	0	1,000	0,072	0,2483
GAL	0	2,000	0,110	0,3768	0	1,400	0,094	0,3489
APC	0	0,190	0,017	0,0417	0	0,180	0,018	0,0464

Salienta-se que, as Tabela 5 e 6 apresentam os conjuntos de dados de calibração e validação após a retirada de *outliers*. No caso do PLS-UV-Vis foram identificadas 54 amostras anômalas e, para o PLS-NIR, 41. Após a retirada de *outliers*, diferentes pré-processamentos foram aplicados nos conjuntos espectrais do UV-Vis e NIR a fim de definir o melhor modelo de previsão característico de cada analito. Em todas as modelagens, consideraram-se combinações de três tipos de pré-processamentos usualmente aplicados em espectros de amostras líquidas: absorbâncias centradas na média (MC) e derivadas (1ª Deriv. e 2ª Deriv., ambas com aproximação por polinômio de 2º grau e janela de 11 pontos). A Figura 8 ilustra as derivadas consideradas nos espectros estudados.

Figura 8 - Gráficos dos espectros das amostras de vanilina no UV-Vis (a) e NIR (b) do conjunto S pré-processados com a primeira e segunda derivadas.



a)



b)

Os parâmetros de desempenho utilizados para avaliar os modelos foram RMSECV, RMSEP, R^2 , RER e RPD. Embora todos estes critérios sejam importantes para definir a previsão PLS-UV-Vis-NIR mais adequada, a escolha dos melhores modelos foi baseada, primordialmente, no menor desvio padrão do erro de previsão da validação cruzada LOO (RMSECV). O RMSECV foi utilizado para definir o modelo mais conveniente, sobretudo devido à clareza de interpretação, já que este foi apresentado em percentual da concentração máxima de cada analito. As Tabelas 7, 8 e 9 apresentam os resultados de todos os modelos baseados no conjunto de dados S.

Tabela 7 - Modelos PLS-UV-Vis⁷ obtidos para o conjunto de amostras sintéticas.

Modelo Sintético	Pré-processamento dos espectros ⁸	Compostos Químicos	LV	RMSE		R ²		RMSECV.C _{MAX} ⁻¹ (%)	RER	RPD
				CV (g/L)	P (g/L)	CV	Ajuste			
PLS-UV-Vis- MS ₀	NP	VAN	3	0,0083	0,0036	0,02	0,05	16,57	6,0	1,0
		GAL	2	0,0018	0,0014	0,83	0,85	10,14	9,9	2,3
		APC	4	0,0004	0,0002	0,53	0,63	11,67	8,6	1,4
PLS-UV-Vis- MS ₁	MC; 1 ^a Deriv.	VAN	5	0,0076	0,0055	0,11	0,16	15,28	6,5	1,1
		GAL	3	0,0015	0,0005	0,87	0,88	8,42	11,9	2,8
		APC	5	0,0004	0,0002	0,54	0,68	11,72	8,5	1,4
PLS-UV-Vis- MS ₂	MC; 2 ^a Deriv.	VAN	4	0,0082	0,0030	0,05	0,15	16,49	6,1	1,0
		GAL	2	0,0016	0,0005	0,86	0,87	8,69	11,5	2,7
		APC	3	0,0004	0,0003	0,50	0,56	11,82	8,5	1,4

⁷ Vide Tabela 5 para conferir a validação dos modelos PLS-UV-Vis sintéticos.

⁸ NP: não pré-processado; MC: centragem na média; 1^a Deriv.: primeira derivada; 2^a Deriv.: segunda derivada.

Tabela 8 - Modelos PLS-UV-Vis⁹ obtidos para o conjunto de amostras sintéticas pré-tratadas por MC e 1ª Derivada (PLS-UV-Vis-MS₁).

Compostos Químicos (validade do modelo)	Faixa do espectro considerada (nm)	LV	RMSE		R ²		RMSECV.C _{MAX} ⁻¹ (%)	RER	RPD
			CV (g/L)	P (g/L)	CV	Ajuste			
VAN (0-50 mg/L)	280-310	1	0,0082	0,0034	0,02	0,05	16,30	6,1	1,0
GAL (0-18 mg/L)	240-250	2	0,0020	0,0008	0,78	0,80	10,91	9,2	2,1
APC (0-3,3 mg/L)	290-300	3	0,0004	0,0003	0,43	0,49	12,62	7,9	1,3

⁹ Vide Tabela 5 para conferir a validação dos modelos PLS-UV-Vis sintéticos

Tabela 9 - Modelos PLS-NIR¹⁰ obtidos para o conjunto de amostras sintéticas.

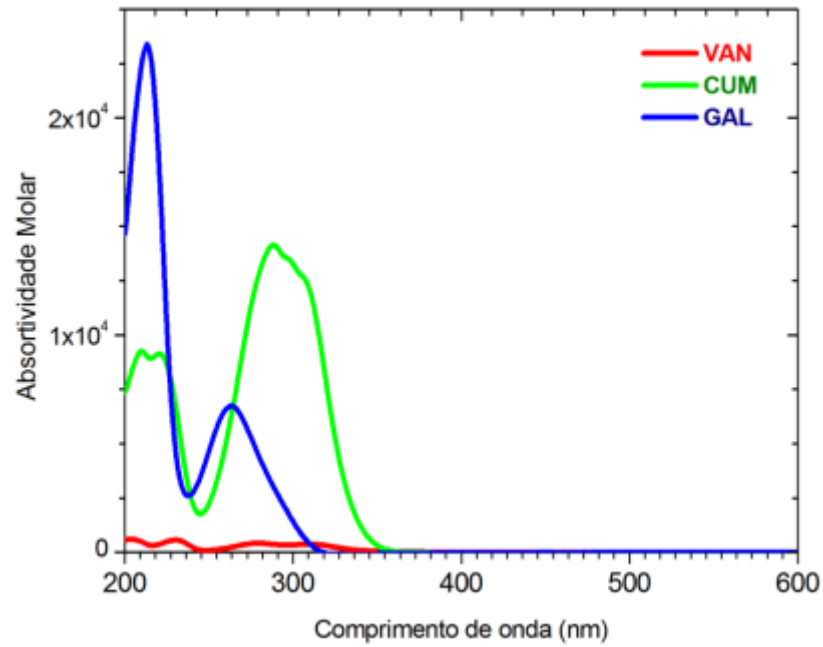
Modelo Sintético	Pré-processamento dos espectros ¹¹	Compostos Químicos	LV	RMSE		R ²		RMSECV.C _{MAX} ⁻¹ (%)	RER	RPD
				CV (g/L)	P (g/L)	CV	Ajuste			
PLS-NIR- MS ₀	NP	VAN	3	0,3918	0,2689	0,03	0,12	19,59	5,1	1,0
		GAL	3	0,3722	0,3233	0,04	0,15	18,61	5,4	1,0
		APC	3	0,0403	0,0379	0,08	0,20	21,19	4,7	1,0
PLS-NIR- MS ₁	NL [1.389 – 1.538] - [1.887 – 2105] nm	VAN	3	0,3917	0,2693	0,03	0,12	19,58	5,1	1,0
		GAL	1	0,3801	0,3386	0,08	0,09	19,01	5,3	1,0
		APC	3	0,0402	0,0379	0,08	0,20	21,17	4,7	1,0
PLS-NIR- MS ₂	MC; 1 ^a Deriv.; NL [1.389 – 1.538] - [1.887 – 2.105] nm	VAN	2	0,4099	0,3166	0,03	0,28	20,50	4,9	1,0
		GAL	5	0,3235	0,2616	0,29	0,87	16,17	56,2	1,2
		APC	2	0,0399	0,0393	0,09	0,19	21,00	4,8	1,0
PLS-NIR- MS ₃	MC; 2 ^a Deriv.; NL [1.389 – 1.538] - [1.887 – 2.105] nm	VAN	1	0,3985	0,2914	0,05	0,43	19,92	5,0	1,0
		GAL	2	0,3487	0,3108	0,17	0,77	17,43	5,6	1,1
		APC	1	0,0408	0,0431	0,09	0,39	21,47	4,7	1,0

¹⁰ Vide Tabela 6 para conferir a validação dos modelos PLS-NIR sintéticos¹¹ NP: não pré-processado; MC: centragem na média; 1^a Deriv.: primeira derivada; 2^a Deriv.: segunda derivada.

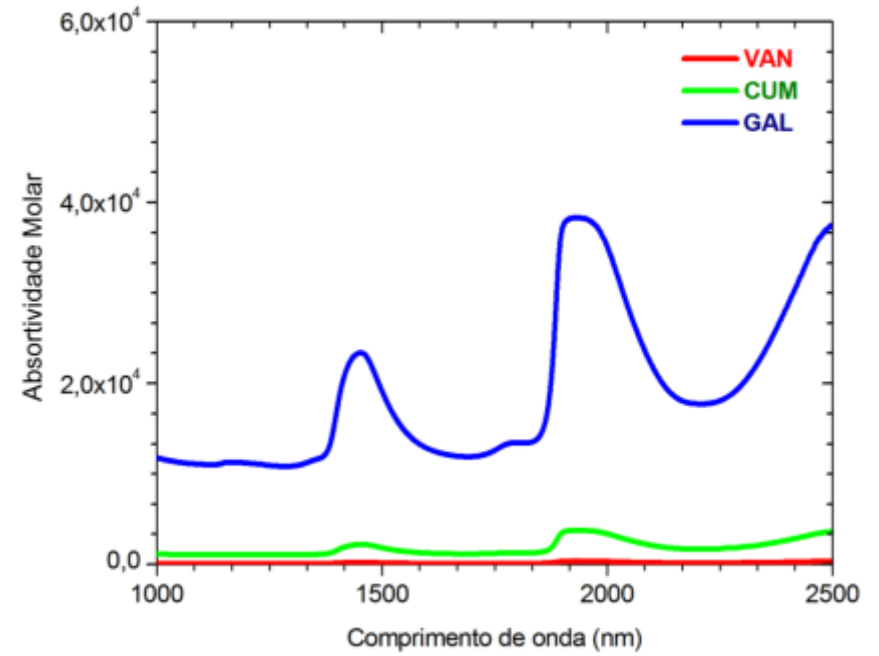
A priori, analisaram-se modelos baseados nos espectros na região do ultravioleta-visível, as características dos conjuntos de calibração e teste (S) foram apresentadas na Tabela 5. Os pré-processamentos, as variáveis latentes e os parâmetros de desempenho de cada modelagem PLS-UV-Vis foram organizados nas Tabelas 7 e 8.

A influência de cada pré-processamento na previsão de concentrações dos analitos foi avaliada, o resultado obtido foi apresentado na Tabela 7. Nestes modelos, observou-se que os desvios padrão de previsão dos analitos VAN e APC foram significativamente superiores aos apresentados pelo GAL. Este resultado pode ser justificado ao realizar a análise da absorvidade molar relativa dos analitos no UV-Vis (Figura 9.a). Na Figura 9.a. observa-se que a VAN apresenta a menor absorvidade no meio sintético em relação ao demais componentes químicos. Este efeito pode ser verificado também na quantificação do APC cuja absorvidade foi inferior à apresentada pelo GAL. Como resultado da baixa absorvidade relativa, os modelos de previsão sofrem alterações significativas em sua acurácia, assim os analitos com maior absorvidade (como por exemplo, o GAL) apresentam menor desvio padrão de previsão (RMSECV).

Figura 9 – Gráficos de absorvidade molar¹² no UV-Vis (a) e NIR (b) dos compostos químicos estudados nas amostras sintéticas.



a)



b)

¹² Somente analitos considerados em amostras sintéticas. Dado que a absorvidade molar é definida por: $\epsilon = \frac{Abs}{C \cdot d} [=] \frac{L}{mol \cdot cm}$. Sendo Abs a absorbância, C a concentração do analito na amostra e d o caminho óptico.

Os resultados da Tabela 7 mostraram ainda que não houve melhora representativa na quantificação da VAN ao aplicar as modelagens PLS-UV-Vis- MS_1 ou PLS-UV-Vis- MS_2 frente à referência (PLS-UV-Vis- MS_0). Observou-se que o menor RMSECV para a VAN ocorreu no M_1 , admitindo valor igual a 15,28%. Ao quantificar o GAL e APC, verificou-se este parâmetro de performance foi reduzido em até 1,72% ao aplicar a derivada primeira com absorbâncias centradas na média (PLS-UV-Vis- MS_1 em relação ao PLS-UV-Vis- MS_0). Desta forma, o RMSECV mínimo observado para estes analitos foi de 8,42% para o GAL e de 11,72% para o APC. Portanto, foi possível definir o PLS-UV-Vis- MS_1 como a modelagem ótima na previsão das concentrações dos três compostos químicos considerados. Este resultado foi baseado apenas na redução do RMSECV. Vale ressaltar que apenas a calibração PLS-UV-Vis para o ácido gálico (GAL) proporcionou um modelo aplicável na previsão da concentração no meio sintético, isto porque, além de RMSECV admitindo baixos valores, o R^2 na CV ($\sim 0,90$), o RPD (~ 3) e o RER (~ 12) foram satisfatórios.

Após definir o pré-processamento espectral adequado na previsão das concentrações dos analitos (PLS-UV-Vis- MS_1), consideraram-se diferentes regiões espectrais para cada analito a fim de aperfeiçoar os modelos PLS-UV-Vis. A análise de gráficos de derivadas primeira e segunda, similares aos ilustrados na Figura 8.a.b (VAN), tornou possível definir regiões de comprimentos de onda com bandas de absorção características a cada composto químico. A Tabela 8 apresenta as regiões características da VAN, GAL e APC, bem como o resultado das modelagens com espectros cujos pré-processamentos foram MC e 1ª Deriv. (PLS-UV-Vis- MS_1). Ao realizar a seleção de comprimentos de onda foi possível observar que os modelos PLS-UV-Vis não foram otimizados, ou seja, não houve melhora nos parâmetros de desempenho para nenhum analito. Observou-se que o RMSECV aumentou e o RER também reduziu.

Na etapa subsequente, analisou-se a viabilidade para quantificação dos analitos a partir de modelos PLS-NIR. Os resultados dos modelos de referência (PLS-NIR- MS_0), bem como os demais modelos PLS-NIR foram dispostos na Tabela 9. Na região espectral do infravermelho próximo, bandas de absorção específicas podem afetar negativamente o processo de calibração. A fim evitar esta interferência, a literatura propõe a utilização da região de 1.613 até 1.754 (referente ao grupo C-H) para a calibração de compostos orgânicos, tais como açúcares. Além do mais, sugere também que a região entre 2.080 e 2.350 favorece a previsão

de concentrações de alcoóis (grupo O-H), como o glicerol, xilitol e o etanol (Morita *et al.*, 2011; Nishii *et al.*, 2012). Morita *et al.* (2011) considerou ainda que as bandas características da água (1.800 até 2.200 nm) podem prejudicar a quantificação de alguns analitos e removeu esta região espectral na calibração dos modelos PLS-NIR. Neste contexto, faz-se necessário enfatizar que, as soluções sintéticas possuem etanol em sua composição, desta forma, mesmo em pequenas concentrações, o solvente etanol-água poderia interferir na calibração dos modelos PLS-NIR. Portanto, a fim de garantir que a calibração multivariada fosse baseada primordialmente nas absorvâncias dos analitos e não na proporção de diluição dos solventes, nos modelos PLS-NIR subsequentes ao MS_0 (MS_1 , M_2 e MS_3), retirou-se simultaneamente as bandas de absorção referentes às bandas da água e do grupo R-OH (1.389–1.538 e 1.887–2.105 nm – vide modelos na Tabela 9).

Nos modelos PLS-NIR se observou que os erros de previsão RMSECV atingiram valores em torno de 16,0 a 22,0%, o que garante que os modelos foram inviáveis frente à previsão de concentrações baseadas em espectros da região do UV-Vis. O gráfico da absorvância molar na região do NIR (Figura 9.b) mostra a influência acentuada de determinadas bandas de absorção na região do NIR, tais como as referentes à água e ao grupo R-OH (1.389–1.538 e 1.887–2.105 nm), que sobressaíram a qualquer outra banda de absorção. A retirada destas regiões favoreceu uma pequena redução do RMSECV para a previsão das concentrações de VAN e APC (vide Tabela 9 - PLS-NIR- MS_0 frente ao PLS-NIR- MS_1). Embora os modelos PLS-NIR tenham apresentado baixa acuracidade na quantificação dos analitos, a modelagem com menor desvio padrão de previsão das concentrações do APC, VAN e GAL foi o PLS-NIR- MS_2 (MC; 1ªDeriv. e NL [1.389–1.538 e 1.887–2.105 nm]).

Em suma, verificou-se que o emprego dos pré-processamentos MC e 1ªDeriv. favoreceu, para determinados analitos, a redução do RMSECV e, conseqüentemente, acarretou na melhora dos parâmetros de desempenho para os modelos PLS-NIR, bem como PLS-UV-Vis. Desta forma, a acurácia dos modelos PLS obtida com o conjunto de amostras sintéticos mostrou que a calibração multivariada de dados pode ser aperfeiçoada desde que pré-processamentos espectrais específicos sejam aplicados ao conjunto de dados. Além disso, o estudo de amostras sintéticas para os modelos de regressão linear multivariada mostrou que a seleção de comprimentos de onda no UV-Vis não foi justificável. As modelagens PLS-NIR foram insatisfatórias na quantificação de todos analitos estudados em meio sintético.

Adicionalmente, tanto os modelos PLS-UV-Vis quanto os modelos PLS-NIR não foram adequados na quantificação de APC e VAN (RMECV superiores a 11% e RPD \sim 1). A quantificação das concentrações que minimiza o RMSECV para a VAN e para o APC corresponde ao PLS-UV-Vis- MS_1 , cujo RMECV foi de, aproximadamente, 15 e 11%, nesta ordem. O valor referente ao RPD destes modelos de previsão foi de 1,0, sugerindo que as previsões não poderiam ser aplicadas no monitoramento das concentrações em soluções sintéticas. Já para a quantificação de GAL, o melhor modelo foi o PLS-UV-Vis- MS_1 cujo RMECV foi, aproximadamente, 8% e RPD igual a \sim 3. Posto que a precisão foi de 0,08 g/L (RER \sim 12), o PLS-UV-Vis- MS_1 propicia a quantificação do GAL em meio sintético. A variação na viabilidade dos modelos PLS sugere que interferentes nas soluções (mistura dos analitos, por exemplo) podem prejudicar a previsão das concentrações de certos compostos químicos, como ocorreu no caso dos modelos obtidos para a VAN e APC.

5.2. Conjunto de Dados Reais do Pré-tratamento Hidrotérmico (LHW)

Embora as amostras sintéticas permitam avaliar a relação entre as absorvâncias e os modelos de previsão PLS, esta análise ainda não é suficiente para quantificar os analitos em amostras reais de pré-tratamento da biomassa lignocelulósica. Isto por que o hidrolisado do bagaço da cana-de-açúcar apresenta uma mistura complexa entre diversos compostos químicos além de apresentar variações no pH de acordo com as condições do processo, implicando alterações significativas na absorvância dos compostos químicos (Khani *et al.*, 2016). Segundo Ewanick *et al.* (2013) e Fragoso *et al.* (2011), estas alterações geradas pela complexidade da solução impõem limitações na quantificação e discriminação de compostos específicos presentes no meio considerado. Assim sendo, a fim de viabilizar a quantificação dos principais inibidores no processo real de pré-tratamento LHW, foram avaliados conjuntos de calibração contendo amostras advindas de hidrolisados variados. Consideraram-se como amostras do processo real diferentes diluições do hidrolisado derivado de três condições de severidade para o pré-tratamento do bagaço de cana-de-açúcar (vide Tabela 4). As concentrações obtidas no HPLC de cada analito no hidrolisado dos pré-tratamentos foram apresentadas na Tabela 10.

Tabela 10 – Concentrações dos analitos no hidrolisado LHW aferidas por HPLC e o desvio padrão de amostragem ($DP^{ESTIMADO}$ - considerando o desvio padrão (DP) médio de 4 amostras distintas em cada hidrolisado) associado à quantificação concomitantemente às diluições das amostras.

Componentes do Hidrolisado	Concentrações e Desvio Padrão de Amostragem (g/L)				
	LHW 01	LHW 01'	LHW 02	LHW 03	$DP^{ESTIMADO}$
VAN	0,033	0,029	0,013	0,213	0,001
FEL	0,420	0,240	0,281	0,236	0,004
APC	0,044	0,022	0,007	0,016	0,001
HMF	0,100	0,100	0,009	1,700	0,003
FUR	2,150	2,600	0,194	7,500	0,051

Estudos realizados por Khani *et al.* (2016) sugerem a modificação de amostras reais diluídas com a adição de pequenos volumes de soluções sintéticas, as quais devem ser adicionadas com o objetivo de descorrelacionar a concentração do conjunto de dados de calibração e teste diluídos utilizados para obter os modelos PLS. Neste contexto, no presente trabalho, foram realizadas adições de amostras sintéticas definidas no laboratório. As soluções padrão foram preparadas de acordo com a metodologia definida na seção 4.1.

Após a aquisição das amostras e varredura dos espectros, realizou-se o primeiro tratamento de dados, o teste estatístico de Leverage em conjunto com a análise resíduos de Student. Os *outliers* foram selecionados a partir dos melhores modelos PLS obtidos sem nenhum processamento espectral (NP). As possíveis amostras anômalas de cada analito foram retiradas, totalizando 11 *outliers* de 160 dados (apenas ~7,0% do total). O conjunto de dados resultante desta análise foi utilizado na obtenção dos modelos PLS subsequentes. Na Tabela 11 foram apresentados a variação nas concentrações mínima e máxima de cada analito considerado, bem como o número de amostras, a média e o desvio padrão dos conjuntos de calibração e teste dos dados reais do pré-tratamento hidrotérmico (LHW).

Tabela 11 - Faixa de validade, média, desvio padrão da média e a quantidade de amostras utilizadas na calibração e validação dos modelos PLS-UV-Vis- NIR com dados LHW.

Compostos Químicos	Concentrações no conjunto de calibração (g/L) (NA= 135)				Concentrações no conjunto de teste (g/L) (NA= 13)			
	Min.	Max.	Média	DP	Min.	Max.	Média	DP
VAN	0,000	0,213	0,026	0,045	0,000	0,165	0,055	0,060
FEL	0,000	0,601	0,101	0,143	0,000	0,519	0,177	0,191
APC	0,000	0,071	0,009	0,014	0,000	0,061	0,014	0,017
HMF	0,000	1,700	0,099	0,252	0,000	1,150	0,174	0,323
FUR	0,000	7,500	0,750	1,274	0,075	5,650	1,542	1,678

Seguida da retirada de *outliers*, diferentes pré-processamentos foram aplicados nos conjuntos espectrais do UV-Vis e NIR a fim de definir o menor erro de previsão (RMSEP) de cada analito. Os parâmetros utilizados para caracterizar e avaliar cada modelo foram apresentados nas Tabela 12 e 13. Assim como para as amostras sintéticas, consideraram-se a seleção de comprimentos de onda e os seguintes pré-processamentos: MC, 1ª Deriv. e 2ª Deriv..

Os primeiros modelos analisados foram baseadas nos espectros na região do NIR (vide Tabela 12). Analogamente ao pré-processamento espectral do conjunto de dados sintéticos, além dos pré-processamentos citados, na região do NIR, foi realizada ainda a seleção de algumas regiões espectrais para as calibrações PLS-NIR correspondentes aos dados LHW (PLS-NIR- M_1 , PLS-NIR- M_2 e PLS-NIR- M_3 apresentados na Tabela 12). Ressalta-se que a seleção de regiões espectrais na região do NIR foi realizada com a finalidade de garantir que não houvesse interferência do solvente etanol-água na calibração multivariada dos modelos PLS-NIR. Portanto, para PLS-NIR- M_1 , M_2 e M_3 dos dados reais, retiraram-se simultaneamente as bandas de absorção referentes às bandas da água e as do grupo R-OH (1.389–1.538 e 1.887–2.105 nm – vide a performance dos modelos na Tabela 12).

Tabela 12 - Modelos PLS-NIR obtidos para o conjunto de amostras LHW.

Modelo LHW	Pré-processamento dos espectros ¹³	Compostos Químicos	LV	RMSE		R ²		RMSECV. C _{MAX} ⁻¹ (%)	RER	RPD
				CV (g/L)	P (g/L)	CV	Ajuste			
PLS-NIR- M ₀	MC	VAN	4	0,033	0,059	0,47	0,53	15,48	6,5	1,4
		FEL	4	0,095	0,127	0,56	0,62	15,81	6,3	1,5
		APC	4	0,011	0,018	0,39	0,45	15,50	6,5	1,3
		HMF	3	0,243	0,329	0,07	0,10	14,27	7,0	1,0
		FUR	4	1,100	1,638	0,27	0,39	14,67	6,8	1,2
PLS-NIR- M ₁	MC; NL [1.389 – 1.538] - [1.887 – 2.105] nm	VAN	3	0,035	0,058	0,42	0,50	16,31	6,1	1,3
		FEL	4	0,094	0,113	0,57	0,63	15,61	6,4	1,5
		APC	4	0,011	0,017	0,34	0,41	16,10	6,2	1,2
		HMF	4	0,252	0,349	0,07	0,18	14,83	6,7	1,0
		FUR	3	1,209	1,072	0,14	0,29	16,12	6,2	1,1
PLS-NIR- M ₂	1 ^a Deriv.; MC; NL [1.389 – 1.538] - [1.887 – 2.105] nm	VAN	4	0,033	0,056	0,46	0,55	15,65	6,4	1,4
		FEL	3	0,092	0,107	0,59	0,63	15,34	6,5	2,6
		APC	7	0,012	0,020	0,39	0,67	16,31	6,1	1,2
		HMF	6	0,170	0,306	0,54	0,77	9,97	10,0	1,5
		FUR	4	1,207	1,352	0,18	0,47	16,09	6,2	1,1
PLS-NIR- M ₃	2 ^a Deriv.; MC; NL [1.389 – 1.538] - [1.887 – 2.105] nm	VAN	5	0,038	0,064	0,36	0,67	17,67	5,7	1,2
		FEL	2	0,096	0,101	0,59	0,55	15,90	6,3	1,5
		APC	2	0,012	0,016	0,34	0,30	16,58	6,0	1,2
		HMF	2	0,251	0,338	0,02	0,06	14,76	6,8	1,0
		FUR	2	1,226	1,686	0,07	0,13	16,35	6,1	1,0

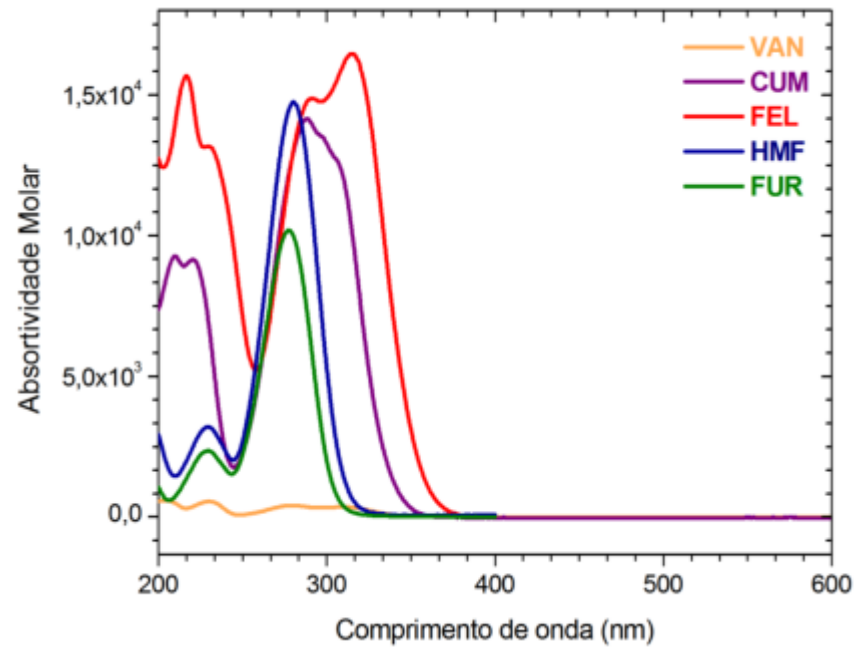
¹³ NP: não pré-processado ; MC: centragem na média; 1^a Deriv.: primeira derivada; 2^a Deriv.: segunda derivada.

A Tabela 12 mostra que os desvios padrão dos modelos PLS-NIR variaram entre 10,0 e 18,0% da concentração máxima de cada analito ($RMSECV \cdot C_{MAX}^{-1}$), o que implica modelos de baixa qualidade na quantificação de todos os analitos, ou seja, baixo RER. Esta tabela de resultados mostrou que os M_i apresentam RMSECV altos, o que acarretou em previsões com desvio padrão de previsão na mesma ordem de grandeza do desvio padrão da média dos dados ($SD_{média}$), ou seja, valores para o parâmetro RPD próximos a 1,0.

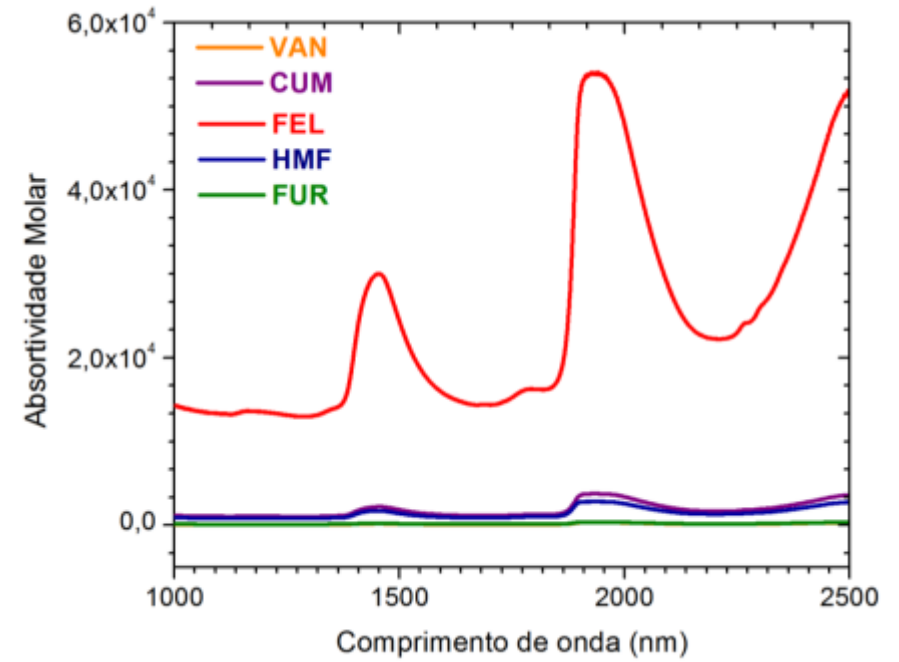
Verificou-se ainda que a retirada das bandas de absorção referentes à água e ao grupo R-OH prejudicou a calibração dos modelos PLS-NIR ao invés de minimizar o desvio padrão de previsão (vide resultados de PLS-NIR- M_0 e PLS-NIR- M_1). Ao observar o desvio padrão de previsão do furfural (composto químico que não possui nenhum grupo R-OH em sua fórmula molecular- vide Figura 6), foi possível notar que houve piora do modelo PLS-NIR- M_2 frente ao PLS-NIR- M_0 . Este resultado indicou que as calibrações estavam diretamente relacionadas com a absorbância do solvente utilizado no preparo das soluções-mãe (água:etanol). Concluiu-se que a variação na concentração dos analitos não foi detectada com a mesma intensidade na absorbância do NIR do que a variação na concentração do etanol nas amostras LHW, o que levou à calibração baseada praticamente na variância apresentada nas bandas de absorção do solvente.

Além disso, a absorptividade molar na região do NIR (vide Figura 10.b) mostra que as bandas de absorção da água, bem como as referentes ao grupo R-OH são extremamente representativas frente às referentes aos *overtones* dos grupos R-CH_i considerados nas calibrações PLS-NIR- M_1 , M_2 e M_3 . Este fato justificou novamente a baixa qualidade dos modelos PLS-NIR para os analitos estudados no conjunto de dados LHW. Desta forma, esta metodologia não foi adequada para a quantificação de analitos em meio real pelo PLS-NIR. Recomenda-se, por exemplo, que em trabalhos futuros outro solvente seja utilizado ou mesmo que a concentração de etanol seja conhecida e acompanhe a proporção de mistura dos solventes.

Figura 10 - Gráficos de absorptividade molar¹³ no UV-Vis (a) e NIR (b) de amostras sintéticas para os compostos analisados nos modelos LHW.



a)



b)

¹³ Somente analitos analisados nas amostras reais de hidrolisado. Dado que a absorptividade molar é definida por: $\epsilon = \frac{Abs}{c \cdot d} [=] \frac{L}{mol \cdot cm}$. Sendo Abs a absorbância, C a concentração do analito na amostra e d o caminho óptico.

No presente estudo, também foi proposto estudar os modelos PLS-UV-Vis obtidos para diferentes pré-tratamentos (MC; 1ª Deriv.; 2ª Deriv.). Como a seleção de regiões espectrais no UV-Vis nas amostras sintéticas não proporcionou melhora significativa nos modelos de previsão, optou-se por utilizar toda a faixa de absorção significativa para os compostos químicos estudados, que compreende a região espectral de 200 - 325 nm (vide Figura 10.a). Salienta-se que, embora esta faixa corresponda à absorção no ultravioleta (vide Figura 1), a nomenclatura UV-Vis foi mantida para todos os modelos.

A Tabela 13 apresenta os modelos PLS-UV-Vis e os respectivos pré-tratamentos espectrais. Além disso, absorvividade molar de cada analito, bem como a absorbância relativa a cada amostra de pré-tratamento hidrotérmico podem ser visualizadas nas Figura 10 e 11, respectivamente.

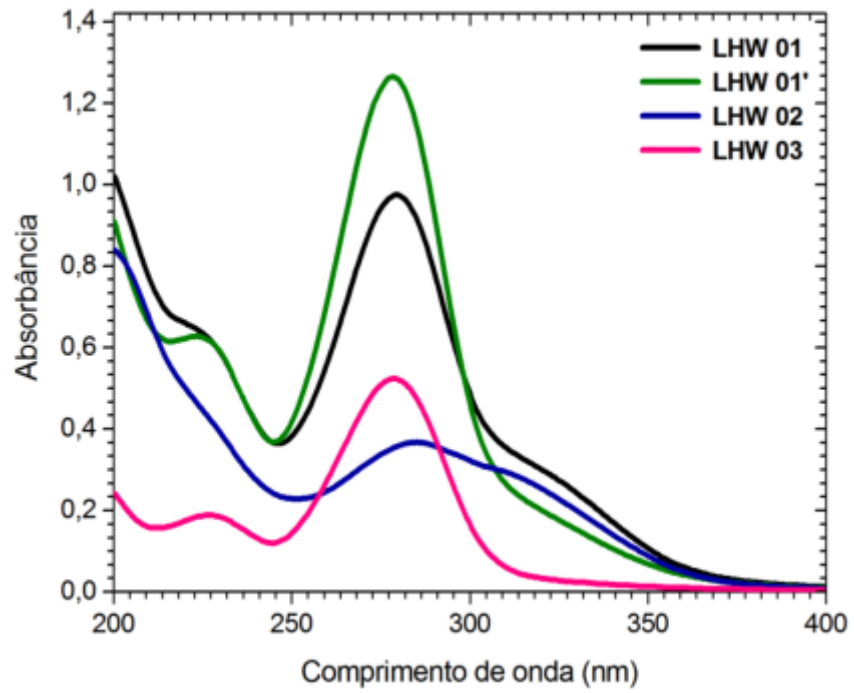
No PLS-UV-Vis observou-se que os modelos PLS-UV-Vis- M_1 e PLS-UV-Vis- M_2 apresentaram melhora significativa na quantificação dos analitos frente ao modelo PLS-NIR de maior precisão, visto que tais modelos exibiram RPD iguais ou superiores a 2,0 além de $11,0 \leq RER \leq 35,0$ para todos os compostos químicos analisados (VAN, FEL, APC, HMF e FUR).

Tabela 13 - Modelos PLS-UV-Vis obtidos para o conjunto de amostras LHW.

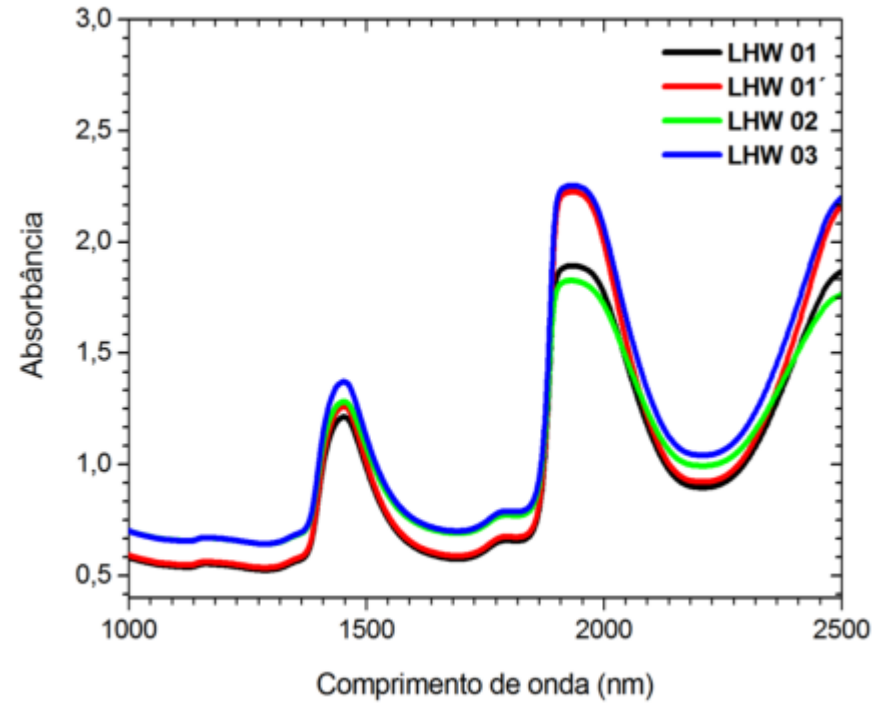
Modelo LHW	Pré-tratamento dos espectros ¹⁴	Compostos Químicos	LV	RMSE		R ²		RMSECV.C _{MAX} ⁻¹ (%)	RER	RPD
				CV (g/L)	P (g/L)	CV	Ajuste			
PLS-UV-Vis- M ₀	[325 – 600] nm	VAN	6	0,017	0,012	0,86	0,93	7,93	12,6	2,7
		FEL	4	0,044	0,053	0,91	0,93	7,26	13,8	3,3
		APC	3	0,010	0,015	0,49	0,57	14,11	7,1	1,4
		HMF	4	0,053	0,049	0,96	0,97	3,14	31,8	4,7
		FUR	2	0,279	0,375	0,95	0,96	3,72	26,9	4,6
PLS-UV-Vis- M ₁	1 ^a Deriv.; MC; NL [325 – 600] nm	VAN	4	0,017	0,013	0,86	0,92	7,87	12,7	2,7
		FEL	4	0,041	0,049	0,92	0,94	6,75	14,8	3,5
		APC	5	0,008	0,013	0,67	0,73	11,44	8,7	1,7
		HMF	4	0,053	0,041	0,96	0,97	3,09	32,4	4,8
		FUR	2	0,282	0,383	0,95	0,96	3,76	26,6	4,5
PLS-UV-Vis- M ₂	2 ^a Deriv.; MC; NL [325 – 600] nm	VAN	4	0,015	0,015	0,89	0,97	7,06	14,2	3,0
		FEL	5	0,045	0,050	0,90	0,95	7,56	13,2	3,2
		APC	10	0,006	0,013	0,80	0,94	8,93	11,2	2,2
		HMF	6	0,048	0,041	0,97	0,99	2,83	35,4	5,2
		FUR	2	0,290	0,383	0,95	0,96	3,86	25,9	4,4

¹⁴ NP: não pré-processado; MC: centragem na média; 1^a Deriv.: primeira derivada; 2^a Deriv.: segunda derivada.

Figura 11 - Gráficos dos espectros pré-processados com AMM no UV-Vis (a) e NIR (b) das amostras LHW.



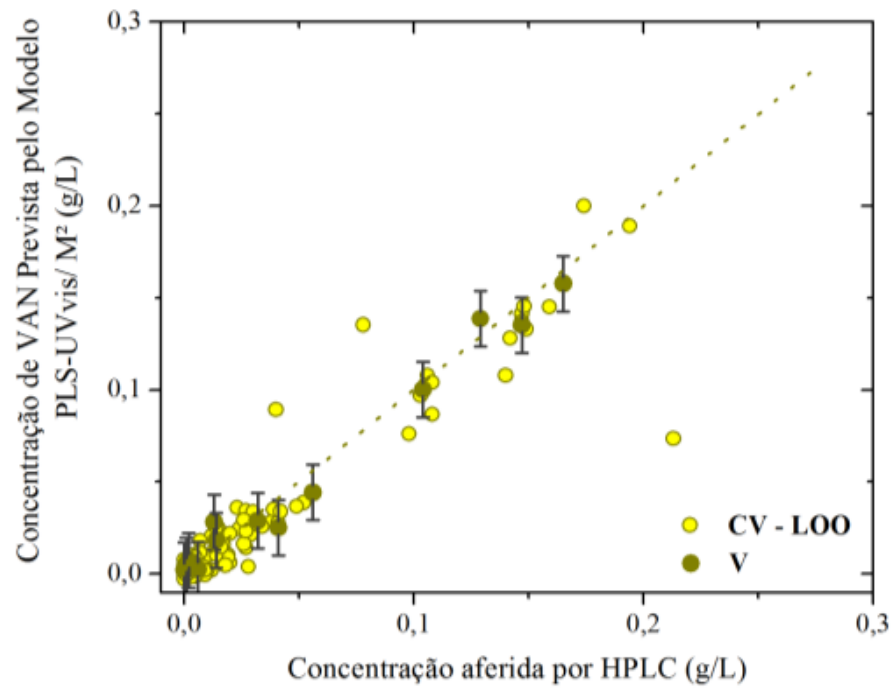
a)



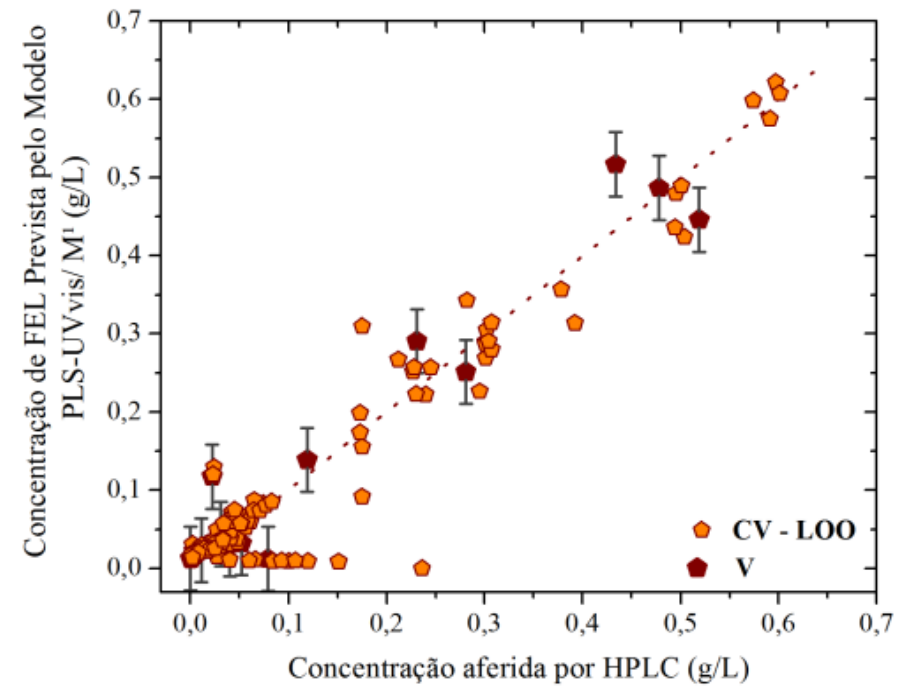
b)

No caso do PLS-UV-Vis, constatou que cada analito especifica o pré-tratamento dos espectros. Desta forma, o modelo de quantificação adequado da VAN, APC e HMF corresponde ao PLS-UV-Vis- M_2 , que foi definido pela maior precisão (RER ~ 14 para a VAN, RER ~ 11 para APC e RER ~ 35 para o HMF) e RMSECV mínimo, correspondendo, respectivamente, a 0,015 (7,06%), 0,006 (8,93%) e 0,048 g/L (2,83%). Os parâmetros de performance RPD e R^2 confirmaram a viabilidade do PLS-UV-Vis- M_2 . A quantificação da vanilina apresentou RPD ~ 3 , mostrando que o desvio padrão da previsão (RMSECV) foi 3 vezes menor que o desvio das concentração ($DP_{\text{média}}$). Enquanto que o R^2 na CV foi de 0,89 (VAN), indicando que o modelo define 89% dos dados previstos a partir da referência (concentrações do HPLC). Na previsão de concentrações do HMF obtiveram-se valores ainda mais promissores sendo que o RPD foi de 5 e o R^2 na CV igual a 0,97. Em contrapartida, os parâmetros de performance do ácido APC não foram tão promissores quanto os dos modelos de quantificação para a VAN e HMF, mas ainda assim foram satisfatórios. O PLS-UV-Vis- M_2 do APC apresentou desvio padrão de previsão na validação cruzada de, aproximadamente, 9,0%, RPD de 2 e o R^2 na CV igual a 0,80. Portanto, os parâmetros de performance referentes à previsão da VAN, APC e HMF no modelo PLS-UV-Vis- M_2 garantem que estes compostos químicos podem ser quantificados e monitorados com acurácia durante os processos de pré-processamento hidrotérmico da biomassa da cana-de-açúcar. O modelo PLS-UV-Vis- M_0 destacou-se por prever com acuracidade a concentração do FUR, cujo RMSEP foi de 0,375 g/L (3,72%). Nesta quantificação obtiveram-se RPD de 5, RER de 27 e R^2 igual a 0,95 na CV dos dados. A previsão das concentrações de FEL nas amostras LHW destacou-se no PLS-UV-Vis- M_1 , este modelo apresentou desvio padrão de previsão na validação cruzada de 6,75%, RER de 15, RPD de 4 e o R^2 na CV igual a 0,92. O gráfico correspondente à absortividade molar no UV-Vis dos analitos (vide Figura 10.a) não forneceu informações preliminares sobre os modelos de previsão PLS-UV-Vis, visto que todos os compostos químicos possuem absorbâncias significativas nas regiões estudadas. A Figura 12 apresenta os gráficos da validação cruzada LOO (vide Figura 12.a.b.c.d.e) e teste dos modelos para os modelos de previsão da VAN (PLS-UV-Vis- M_2), do FEL (PLS-UV-Vis- M_1), do APC (PLS-UV-Vis- M_2), do HMF (PLS-UV-Vis- M_2), bem como do FUR (PLS-UV-Vis- M_0).

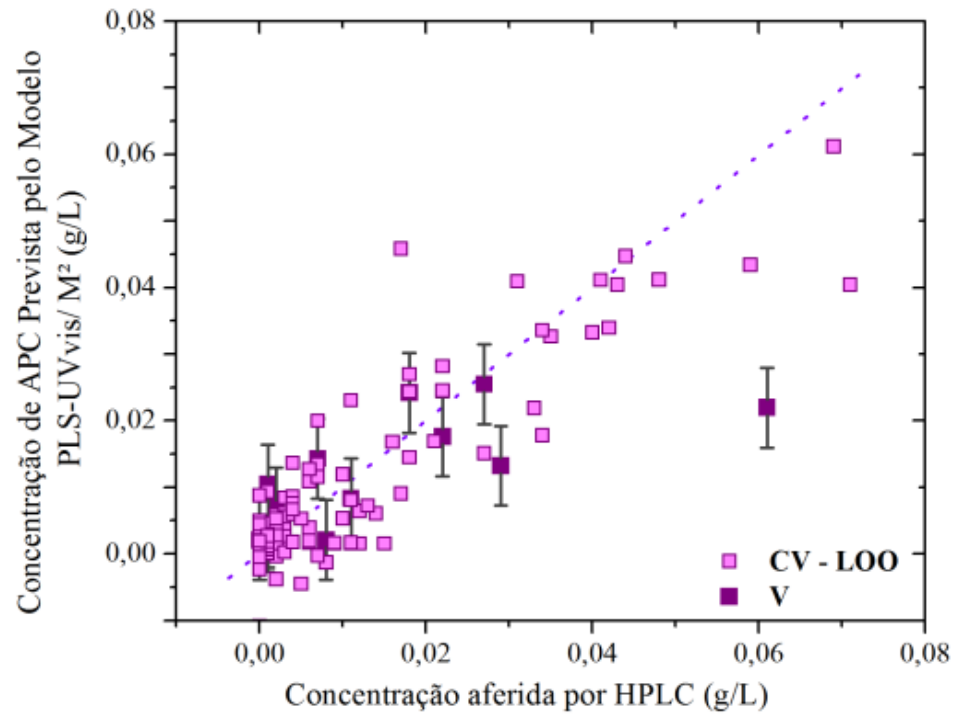
Figura 12 – Gráficos de validação cruzada (CV-LOO) dos melhores modelos PLS, dado a bisettriz de referência. Validação (V) dos modelos com a barra do desvio padrão de previsão na CV (RMSECV). Posto que, para cada analito, tem-se: VAN, PLS-UV-Vis- M_2 (a); FEL, PLS-UV-Vis- M_1 (b); APC, PLS-UV-Vis- M_2 (c); HMF, PLS-UV-Vis- M_2 (d); FUR, PLS-UV-Vis- M_0 (e).



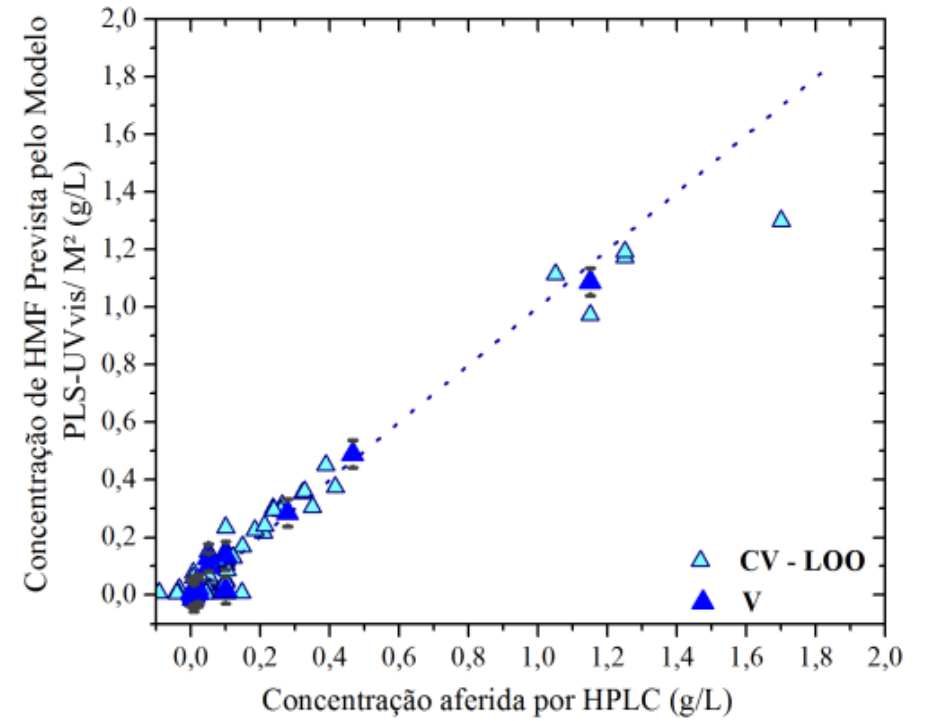
a)



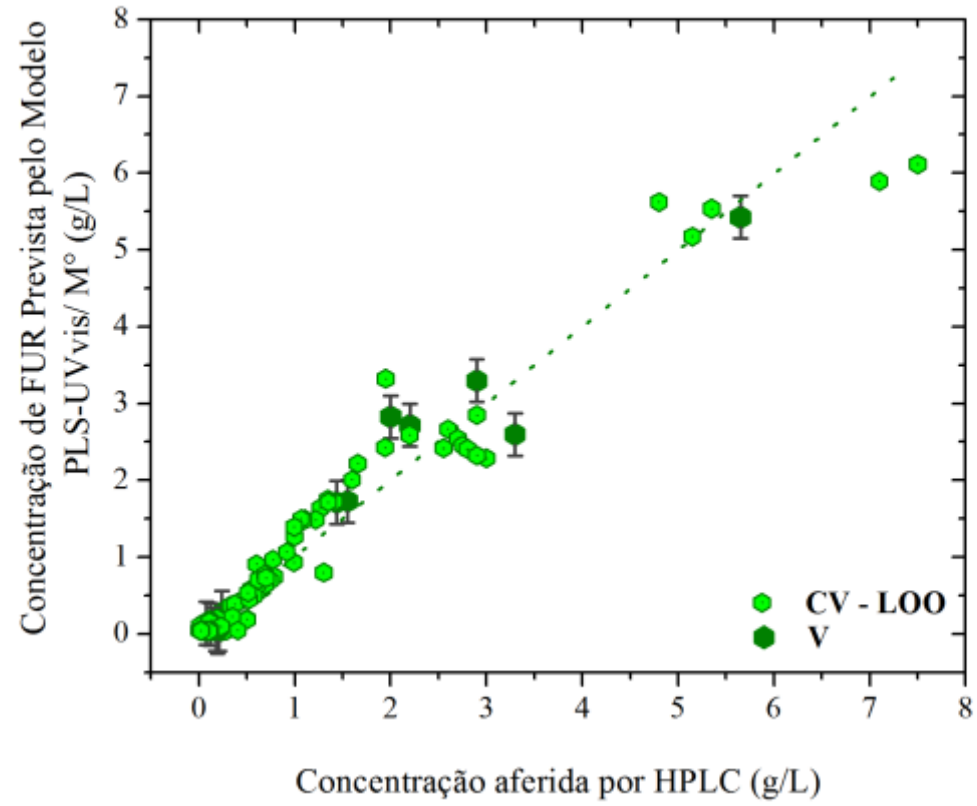
b)



c)



d)



e)

Resumindo os resultados, as análises realizadas definiram que a metodologia definida para os modelos PLS-UV-Vis são adequados para quantificar os analitos característicos do hidrolisado LHW. Destaca-se que os valores de RPD obtidos para a VAN, FEL, HMF e FUR foram superiores a 2,0, caracterizando modelos que podem ser utilizados na quantificação destes analitos em meio real (Martelo-Vidal e Vázquez, 2014). Ressalta-se também que a menor precisão das previsões pelo PLS-UV-Vis foi de, aproximadamente, 9,0% da concentração máxima, enquanto que a maior foi de 2,83%. A Tabela 14 mostra o resumo dos resultados apresentados na Figura 13.

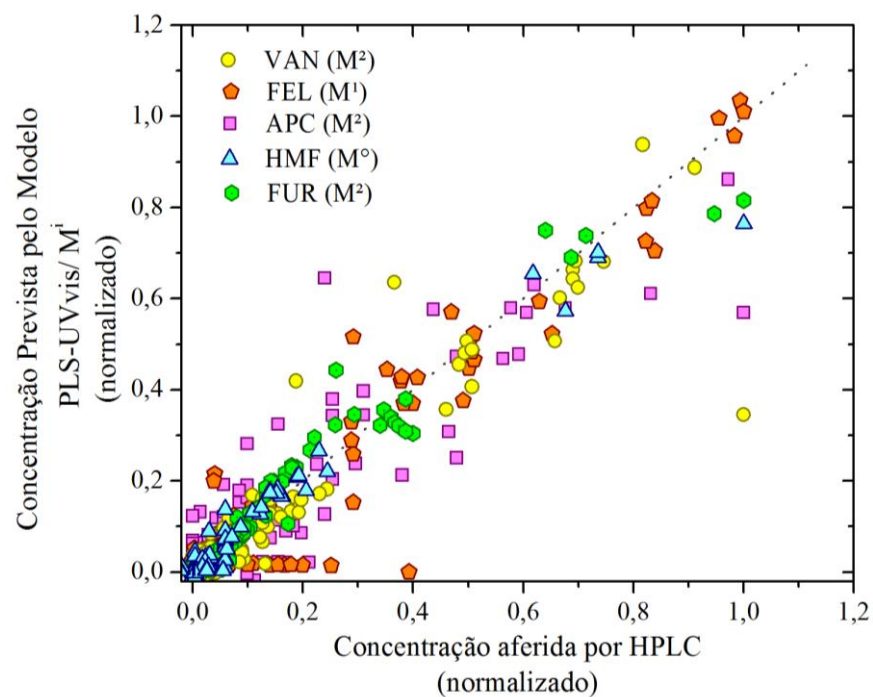
Tabela 14- Resumo dos modelos PLS-UV-Vis¹⁵ aplicados no monitoramento *off-line* dos analitos presentes no hidrolisado advindo do tratamento hidrotérmico da biomassa de cana-de-açúcar.

Modelo LHW	Pré-processamento dos espectros¹⁶	Compostos Químicos	LV	RMSECV (g/L)	RMSEP (g/L)	RMSECV. C_{MAX}⁻¹ (%)
PLS-UV-Vis- M ₀	MC; NL [325 – 600] nm	FUR	2	0,279	0,375	3,72
PLS-UV-Vis-M ₁	1 ^a Deriv.; MC; NL [325 – 600] nm	FEL	4	0,041	0,049	6,75
PLS-UV-Vis- M ₂	2 ^a Deriv.; MC; NL [325 – 600] nm	VAN APC HMF	4 10 6	0,015 0,006 0,048	0,015 0,013 0,041	7,06 8,93 2,83

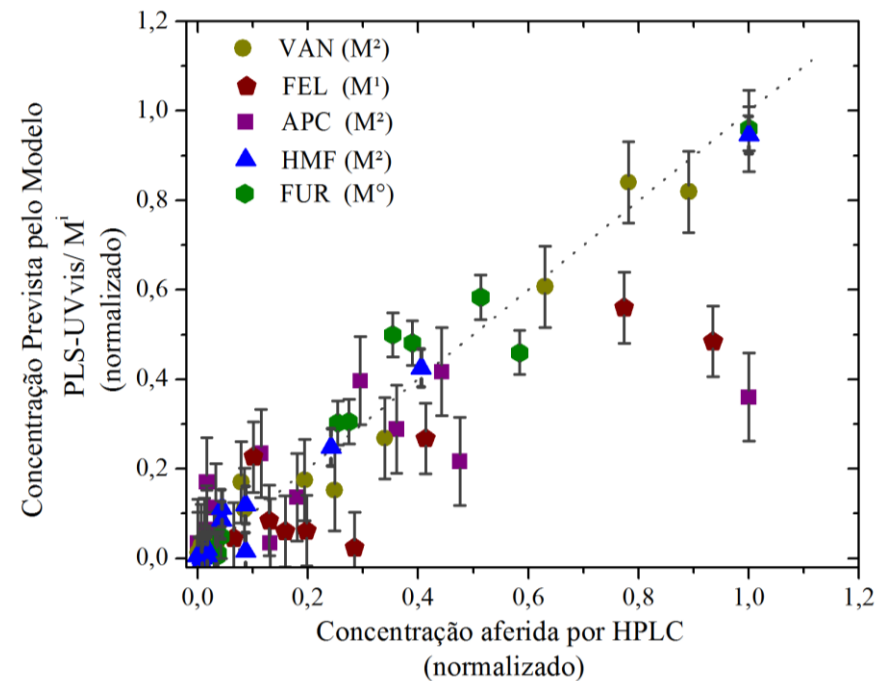
¹⁵ Vide Tabela 11 para verificar a faixa de validade dos modelos PLS-UV-Vis baseado em amostras reais do hidrolisado.

¹⁶ **NP**: não pré-processado; **MC**: centragem na média; **1^a Deriv.**: primeira derivada; **2^a Deriv.**: segunda derivada.

Figura 13- Resumo dos melhores modelos PLS-UV-Vis de cada analito¹⁷ para a validação cruzada (LOO) (a) e para a validação externa de cada modelo, apresentando o desvio padrão de previsão (b) (vide Tabela 14).



a)



b)

¹⁷ Concentrações normalizadas com a concentração máxima de cada composto químico considerado nos modelos PLS-UV-Vis (vide Tabela 11)

6. CONSIDERAÇÕES FINAIS

A principal contribuição deste trabalho foi mostrar a viabilidade do monitoramento dos compostos fenólicos e dos furaldeídos a partir da regressão PLS associada a duas técnicas de espectroscopia, NIR e UV-Vis. O estudo de amostras sintéticas para os modelos de regressão linear multivariada mostrou que aqueles baseados em dados sintéticos foram adequados na previsão da concentração do GAL, apresentando RMECV de, aproximadamente, 8,4% e RPD igual a 3,0 (PLS-UV-Vis- M_1). Já a quantificação de APC e VAN por modelos PLS-UV-Vis/NIR não foram satisfatórios. Ao considerar o conjunto de dados reais, os resultados garantiram ainda que os modelos PLS-UV-Vis de maior acurácia podem ser utilizados no monitoramento *off-line* de compostos fenólicos e furaldeídos durante o pré-tratamento LHW do bagaço de cana-de-açúcar em três condições de operação diferentes. O modelo para os analitos VAN e FEL apresentou acurácia satisfatória na previsão de concentrações (RPD ~3,0 e RPD ~4,0, respectivamente), bem como RMSECV mínimo, correspondendo a 7,1% para a VAN e 6,8% para o FEL. A quantificação do HMF em amostras reais foi definida pelo modelo cujo RMSEP foi de 2,8% e RPD de 5,0. Adicionalmente, a previsão de concentrações ótima do APC forneceu valor de 8,9% para o RMSECV, com RPD de 2,0. Enquanto que no modelo de previsão do FUR, observaram-se RMSECV de 3,7% e RPD igual a 5,0. Sumariamente, a quantificação de inibidores do hidrolisado LHW por PLS-UV-Vis aplicada na produção de etanol 2G correlaciona diferentes condições de operação do pré-tratamento com a eficiência da hidrólise enzimática e da fermentação alcoólica. Salienta-se que o resultado deste trabalho contribui na realização da interface entre a quimiometria aplicada e a sua implementação em processos de larga escala envolvidos na produção do etanol 2G. Desta forma, propõe-se que, em trabalhos futuros, seja desenvolvido o protocolo de controle da etapa de pré-tratamento com base no modelo PLS-UV-Vis desenvolvido neste trabalho. Finalmente, a partir de protocolos de controle/otimização de processos baseados na quantificação PLS-UV-Vis é possível minimizar o impacto negativo dos compostos fenólicos e furaldeídos durante a obtenção do etanol 2G, pois viabiliza-se a implementação de estratégias de processo baseadas na concentração destes inibidores.

7. REFERÊNCIAS

ABDI, H. Partial least squares regression and projection on latent structure regression (PLS Regression). **WIREs Comp Stat**, v. 2, n. 1, p. 97-106, 2010. Disponível em: <<http://dx.doi.org/10.1002/wics.51>>.

ACOSTA-ESTRADA, B. A.; GUTIÉRREZ-URIBE, J. A.; SERNA-SALDÍVAR, S. O. Bound phenolics in foods, a review. **Food Chemistry**, v. 152, p. 46-55, 6/1/ 2014. ISSN 0308-8146. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0308814613017305>>.

BANERJEE, N.; BHATNAGAR, R.; VISWANATHAN, L. Inhibition of glycolysis by furfural in *Saccharomyces cerevisiae*. **European journal of applied microbiology and biotechnology**, v. 11, n. 4, p. 226-228, 1981/12/01 1981. ISSN 0171-1741. Disponível em: <<http://dx.doi.org/10.1007/BF00505872>>.

BAO, Y. et al. Measurement of Soluble Solid Contents and pH of White Vinegars Using VIS/NIR Spectroscopy and Least Squares Support Vector Machine. **Food and Bioprocess Technology**, v. 7, n. 1, p. 54-61, 2014// 2014. ISSN 1935-5149. Disponível em: <<http://dx.doi.org/10.1007/s11947-013-1065-0>>.

BELLON-MAUREL, V., FERNANDEZ-AHUMADA, E., PALAGOS, B., ROGER, J.M., MCBRATNEY, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. **Trac-Trend. Anal. Chem.** v. 29, p. 1073-1081, 2010.

BELGUIDOUM, K. et al. HPLC coupled to UV-vis detection for quantitative determination of phenolic compounds and caffeine in different brands of coffee in the Algerian market. **Journal of the Taiwan Institute of Chemical Engineers**, v. 45, n. 4, p. 1314-1320, 7//2014. ISSN 1876-1070. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1876107014000972>>.

BLANCO, M.; PEINADO, A. C.; MAS, J. Analytical monitoring of alcoholic fermentation using NIR spectroscopy. **Biotechnology and Bioengineering**, v. 88, n. 4, p. 536-542, 2004. ISSN 1097-0290. Disponível em: <<http://dx.doi.org/10.1002/bit.20214>>.

BLANCO, M.; VILLARROYA, I. NIR spectroscopy: a rapid-response analytical tool. **TrAC Trends in Analytical Chemistry**, v. 21, n. 4, p. 240-250, 4// 2002. ISSN 0165-9936. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0165993602004041>>.

BRERETON, R. G. **Applied Chemometrics for Scientists**. . England: John Wiley e Sons Ltd, 2007.

CARRASCO, C. et al. SO₂ catalyzed steam pretreatment and fermentation of enzymatically hydrolyzed sugarcane bagasse. **Enzyme Microb Technol**, v. 46, 2010. Disponível em: <<http://dx.doi.org/10.1016/j.enzmictec.2009.10.016>>.

CHANG, C.W., LAIRD, D.A., MAUSBACH, M.J., HURBURGH, C.R.. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. **Soil Sci. Soc. Am. J.** v. 65, p. 480-490, 2001.

CHEN, S.-F. et al. Compositional Analysis of Water-Soluble Materials in Corn Stover. **Journal of Agricultural and Food Chemistry**, v. 55, n. 15, p. 5912-5918, 2007/07/01 2007. ISSN 0021-8561. Disponível em: <<http://dx.doi.org/10.1021/jf0700327>>.

COZZOLINO, D. The Role of Visible and Infrared Spectroscopy Combined with Chemometrics to Measure Phenolic Compounds in Grape and Wine Samples. **Molecules**, v. 20, n. 1, p. 726, 2015. ISSN 1420-3049. Disponível em: <<http://www.mdpi.com/1420-3049/20/1/726>>.

COZZOLINO, D. et al. Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy. **Analytica Chimica Acta**, v. 513, n. 1, p. 73-80, 6/18/2004. ISSN 0003-2670. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0003267003011541>>.

DELGENES, J. P.; MOLETTA, R.; NAVARRO, J. M. Effects of lignocellulose degradation products on ethanol fermentations of glucose and xylose by *Saccharomyces cerevisiae*, *Zymomonas mobilis*, *Pichia stipitis*, and *Candida shehatae*. **Enzyme and Microbial Technology**, v. 19, n. 3, p. 220-225, 1996/08/15 1996. ISSN 0141-0229. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0141022995002375>>.

DRENNEN JK, K. E., LODDER RA. **Advances and perspectives in near-infrared spectroscopy**. . p.22: 443–475. 1991

EWANICK, S. M. et al. Real-time understanding of lignocellulosic bioethanol fermentation by Raman spectroscopy. **Biotechnology for Biofuels**, v. 6, n. 1, p. 1-8, 2013. ISSN 1754-6834. Disponível em: <<http://dx.doi.org/10.1186/1754-6834-6-28>>.

FERREIRA, M. M. C. et al. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, v. 22, p. 724-731, 1999. ISSN 0100-4042. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40421999000500016&enrm=iso>.

FLORENCIO, C.; BADINO, A. C.; FARINAS, C. S. Soybean protein as a cost-effective lignin-blocking additive for the saccharification of sugarcane bagasse. **Bioresource Technology**, v. 221, p. 172-180, 12// 2016. ISSN 0960-8524. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0960852416312962>>.

FRAGOSO, S. et al. Application of FT-MIR Spectroscopy for Fast Control of Red Grape Phenolic Ripening. **Journal of Agricultural and Food Chemistry**, v. 59, n. 6, p. 2175-2183, 2011/03/23 2011. ISSN 0021-8561. Disponível em: <<http://dx.doi.org/10.1021/jf104039g>>.

FRIZON, C. N. T. et al. Determination of total phenolic compounds in yerba mate (*Ilex paraguariensis*) combining near infrared spectroscopy (NIR) and multivariate analysis. **LWT -**

Food Science and Technology, v. 60, n. 2, Part 1, p. 795-801, 3// 2015. ISSN 0023-6438. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0023643814006653>>.

GUO, F. et al. Differences in the adsorption of enzymes onto lignins from diverse types of lignocellulosic biomass and the underlying mechanism. **Biotechnol Biofuels**, v. 7, 2014. Disponível em: <<http://dx.doi.org/10.1186/1754-6834-7-38>>.

HAALAND, D. M.; THOMAS, E. V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. **Analytical Chemistry**, v. 60, n. 11, p. 1193-1202, 1988/06/01 1988. ISSN 0003-2700. Disponível em: <<http://dx.doi.org/10.1021/ac00162a020>>.

HIMMEL, M. E.; RUTH, M. F.; WYMAN, C. E. Cellulase for commodity products from cellulosic biomass. **Current Opinion in Biotechnology**, v. 10, n. 4, p. 358-364, 8// 1999. ISSN 0958-1669. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0958166999800652>>.

HONGQIANG, L.; HONGZHANG, C. Near-infrared spectroscopy with a fiber-optic probe for state variables determination in solid-state fermentation. **Process Biochemistry**, v. 43, n. 5, p. 511-516, 5// 2008. ISSN 1359-5113. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1359511308000305>>.

JOLLIFFE, I. T. **Principal Components in Regression Analysis**. Verlag New York: Springer, 1986. 150-405.

JÖNSSON, L.; ALRIKSSON, B.; NILVEBRANT, N. Bioconversion of lignocellulose: inhibitors and detoxification. **Biotechnol Biofuels**, v. 6, 2013// 2013. Disponível em: <<http://dx.doi.org/10.1186/1754-6834-6-16>>.

JÖNSSON, L. J.; MARTÍN, C. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. **Bioresource Technology**, v. 199, p. 103-112, 1// 2016. ISSN 0960-8524. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0960852415014042>>.

KABEL, M. A. et al. Effect of pretreatment severity on xylan solubility and enzymatic breakdown of the remaining cellulose from wheat straw. **Bioresource Technology**, v. 98, n. 10, p. 2034-2042, 7// 2007. ISSN 0960-8524. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0960852406004032>>.

KADLEC, P.; GABRYS, B.; STRANDT, S. Data-driven Soft Sensors in the process industry. **Computers e Chemical Engineering**, v. 33, n. 4, p. 795-814, 4/21/ 2009. ISSN 0098-1354. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0098135409000076>>.

KNAUF, M., MONIRUZZAMAN, M. Lignocellulosic biomass processing: A perspective. **International Sugar Journal**, v. 106 (1263), p. 147-150, 2004.

KHANI, R.; RAHMANIAN, R.; MOTLAGH, N. V. UV–Visible Spectrometry and Multivariate Calibration as a Rapid and Reliable Tool for Simultaneous Quantification of Ternary Mixture of Phenolic Acids in Fruit Juice Samples. **Food Analytical Methods**, v. 9, n. 5, p. 1112-1119, 2016. ISSN 1936-976X. Disponível em: <<http://dx.doi.org/10.1007/s12161-015-0287-3>>.

KHODDAMI, A.; WILKES, M.; ROBERTS, T. Techniques for Analysis of Plant Phenolic Compounds. **Molecules**, v. 18, n. 2, p. 2328, 2013. ISSN 1420-3049. Disponível em: <<http://www.mdpi.com/1420-3049/18/2/2328>>.

KIM, Y. et al. Hydrolysis-determining substrate characteristics in liquid hot water pretreated hardwood. **Biotechnol Bioeng**, v. 112, 2015// 2015. Disponível em: <<http://dx.doi.org/10.1002/bit.25465>>.

KIM, Y., XIMENES, E., MOSIER, N.S., LADISCH, M.R. 2011a. Soluble inhibitors/deactivators of cellulase enzymes from lignocellulosic biomass. **Enzyme Microb Technol**, v. 48, 2011// 2011. Disponível em: <<http://dx.doi.org/10.1016/j.enzmictec.2011.01.007>>.

KO, J. et al. Compounds inhibiting the bioconversion of hydrothermally pretreated lignocellulose. **Applied Microbiology and Biotechnology**, v. 99, n. 10, p. 4201-4212, 2015/05/01 2015. ISSN 0175-7598. Disponível em: <<http://dx.doi.org/10.1007/s00253-015-6595-0>>.

KO, J. K. et al. Effect of liquid hot water pretreatment severity on properties of hardwood lignin and enzymatic hydrolysis of cellulose. **Biotechnology and Bioengineering**, v. 112, n. 2, p. 252-262, 2015. ISSN 1097-0290. Disponível em: <<http://dx.doi.org/10.1002/bit.25349>>.

LARSSON, S. et al. Influence of lignocellulose-derived aromatic compounds on oxygen-limited growth and ethanolic fermentation by *Saccharomyces cerevisiae*. **Appl Biochem Biotech**, v. 84–86, 2000// 2000. Disponível em: <<http://dx.doi.org/10.1385/ABAB:84-86:1-9:617>>.

LEHNINGER, A.L. **Princípios de bioquímica**. 2 ed., São Paulo, Brasil, Sarvier, 1986.

LI, X. et al. Biomass gasification in a circulating Fluidized bed. **Biomass and Bioenergy**, v. 26, 2004. Disponível em: <[http://dx.doi.org/10.1016/S0961-9534\(03\)00084-9](http://dx.doi.org/10.1016/S0961-9534(03)00084-9)>.

LIEBMANN, B.; FRIEDL, A.; VARMUZA, K. Determination of glucose and ethanol in bioethanol production by near infrared spectroscopy and chemometrics. **Analytica Chimica Acta**, v. 642, n. 1–2, p. 171-178, 5/29/ 2009. ISSN 0003-2670. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0003267008018308>>.

LOURENÇO, N. D. et al. Bioreactor monitoring with spectroscopy and chemometrics: a review. **Analytical and Bioanalytical Chemistry**, v. 404, n. 4, p. 1211-1237, 2012/09/01 2012. ISSN 1618-2642. Disponível em: <<http://dx.doi.org/10.1007/s00216-012-6073-9>>.

MARTELO-VIDAL, M. J.; VÁZQUEZ, M. Determination of polyphenolic compounds of red wines by UV–VIS–NIR spectroscopy and chemometrics tools. **Food Chemistry**, v. 158, p. 28-34, 9/1/ 2014a. ISSN 0308-8146. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0308814614002659>>.

MARTELO-VIDAL, M. J.; M. VÁZQUEZ, 2014b. **Evaluation of ultraviolet, visible, and near infrared spectroscopy for the analysis of wine compounds.**: Czech Journal of Food Sciences. 32: 37-47 p. 2014b.

MARTIN, C. et al. Comparison of the fermentability of enzymatic hydrolyzates of sugarcane bagasse pretreated by steam explosion. **Appl Environ Microbiol**, v. 98–100, 2002.

MARTINS, L. H. D. S.; RABELO, S. C.; COSTA, A. C. D. Effects of the pretreatment method on high solids enzymatic hydrolysis and ethanol fermentation of the cellulosic fraction of sugarcane bagasse. **Bioresource Technology**, v. 191, p. 312-321, 9// 2015. ISSN 0960-8524. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0960852415006884>>.

MICHELIN, M. et al. Effect of phenolic compounds from pretreated sugarcane bagasse on cellulolytic and hemicellulolytic activities. **Bioresource Technology**, v. 199, p. 275-278, 1// 2016. ISSN 0960-8524. Disponível em: <www.sciencedirect.com/science/article/pii/S0960852415012262>.

MILLIGAN, M. et al. Semisynthetic model calibration for monitoring glucose in mammalian cell culture with in situ near infrared spectroscopy. **Biotechnology and Bioengineering**, v. 111, n. 5, p. 896-903, 2014. ISSN 1097-0290. Disponível em: <<http://dx.doi.org/10.1002/bit.25161>>.

MONRROY, M. et al. Fourier transformed near infrared (FT-NIR) spectroscopy for the estimation of parameters in pretreated lignocellulosic materials for bioethanol production. **Journal of Chemical Technology e Biotechnology**, v. 90, n. 7, p. 1281-1289, 2015. ISSN 1097-4660. Disponível em: <<http://dx.doi.org/10.1002/jctb.4427>>.

MORITA, H. et al. A new screening method for recombinant *Saccharomyces cerevisiae* strains based on their xylose fermentation ability measured by near infrared spectroscopy. **Analytical Methods**, v. 6, n. 17, p. 6628-6634, 2014. ISSN 1759-9660. Disponível em: <<http://dx.doi.org/10.1039/C4AY00785A>>.

MOSIER, N. et al. Features of promising technologies for pretreatment of lignocellulosic biomass. **Bioresource Technology**, v. 96, n. 6, p. 673-686, 4// 2005. ISSN 0960-8524. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0960852404002536>>.

NELLES, O. **Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models** 1 Ed. Germany: Springer, 2001.

NISHII, T. et al. Selection of the NIR Region for a Regression Model of the Ethanol Concentration in Fermentation Process by an Online NIR and Mid-IR Dual-Region Spectrometer and 2D Heterospectral Correlation Spectroscopy. **Analytical Sciences**, v. 28, n. 12, p. 1165-1170, Dec 2012. ISSN 0910-6340. Disponível em: < <Go to ISI>://WOS:000312977200008 >.

OWEN, T. **Fundamentals of modern UV-visible spectroscopy**. Germany: Agilent Technologies, 2000. 136.

OVEREND, R. P.; CHORNET, E.; GASCOIGNE, J. A. Fractionation of lignocellulosics by steam-aqueous pretreatments. **Philosophical Transaction of the Royal Society A**, v.321(1561):523-536,1987.

PALMQVIST, E.; HAHN-HÄGERDAL, B. Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition. **Bioresource Technol**, v. 74, 2000// 2000. Disponível em: < [http://dx.doi.org/10.1016/S0960-8524\(99\)00161-3](http://dx.doi.org/10.1016/S0960-8524(99)00161-3) >.

PINTO, A. S. S. et al. Monitoring of the cellulosic ethanol fermentation process by near-infrared spectroscopy. **Bioresource Technology**, v. 203, p. 334-340, 3// 2016. ISSN 0960-8524. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0960852415016995>>.

PINTO, A. S. S. et al.. Monitoramento da Produção De Etanol Celulósico por Espectroscopia no Infravermelho Próximo. In: XX CONGRESSO BRASILEIRO DE ENGENHARIA QUÍMICA, 2014a, Florianópolis/SC.

PINTO, A. S. S. et al.. Quantificação de Glicose e Etanol na Fermentação Alcoólica por Espectroscopia no Infravermelho Próximo (NIR). In: SIMPÓSIO NACIONAL DE INSTRUMENTAÇÃO AGROPECUÁRIA 2014b, São Carlos/SP.

PIPPO, W. A. Energy Recovery from Sugarcane-Trash in the Light of 2nd Generation Biofuels. Part 1: Current Situation and Environmental Aspects. **Waste Biomass Valor**, v. 2, p. 1-16, 2011.

QIN, L. et al. Inhibition of lignin-derived phenolic compounds to cellulase. **Biotechnology for Biofuels**, v. 9, n. 1, p. 70, 2016// 2016. ISSN 1754-6834. Disponível em: < <http://dx.doi.org/10.1186/s13068-016-0485-2> >.

RIBEIRO, M. P. A. et al. Multivariate calibration methods applied to the monitoring of the enzymatic synthesis of ampicilin. **Chemometrics and Intelligent Laboratory Systems**, v. 90, n. 2, p. 169-177, 2/15/ 2008. ISSN 0169-7439. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0169743907001955> >.

RINNAN, Å.; BERG, F. V. D.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC Trends in Analytical Chemistry**, v. 28, n. 10, p. 1201-1222, 11// 2009. ISSN 0165-9936. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0165993609001629> >.

RODRÍGUEZ-ZÚÑIGA, U. et al. Fast Determination of the Composition of Pretreated Sugarcane Bagasse Using Near-Infrared Spectroscopy. **BioEnergy Research**, v. 7, n. 4, p. 1441-1453, 2014/12/01 2014. ISSN 1939-1234. Disponível em: <<http://dx.doi.org/10.1007/s12155-014-9488-7>>.

SAMPAIO, P. N. et al. In situ near infrared spectroscopy monitoring of cyprosin production by recombinant *Saccharomyces cerevisiae* strains. **Journal of Biotechnology**, v. 188, p. 148-157, Oct 20 2014. ISSN 0168-1656. Disponível em: <WOS:000344183700022>.

SAINI, J. K.; PATEL, A. K.; ADSUL, M.; SINGHANIA, R. R. Cellulase adsorption on lignin: A roadblock for economic hydrolysis of biomass. **Renewable Energy**, v. 98 p. 29-42, 2016.

SARRAGUÇA, M. et al. Quantitative monitoring of an activated sludge reactor using on-line UV-visible and near-infrared spectroscopy. **Analytical and Bioanalytical Chemistry**, v. 395, n. 4, p. 1159-1166, 2009/10/01 2009. ISSN 1618-2642. Disponível em: <<http://dx.doi.org/10.1007/s00216-009-3042-z>>.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627-1639, 1964/07/01 1964. ISSN 0003-2700. Disponível em: <<http://dx.doi.org/10.1021/ac60214a047>>.

SCHNEIDERMAN, S. J. et al. Quantifying second generation ethanol inhibition: Design of Experiments approach and kinetic model development. **Bioresource Technology**, v. 179, p. 219-226, 3// 2015. ISSN 0960-8524. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0960852414017015>>.

SIESLER, H. W. et al. **Near-Infrared Spectroscopy: Principles, Instruments, Applications**. 1st. German: John Wiley e Sons, 2002.

SILLS, D. L.; GOSSETT, J. M. Using FTIR to predict saccharification from enzymatic hydrolysis of alkali-pretreated biomasses. **Biotechnology and Bioengineering**, v. 109, n. 2, p. 353-362, 2012. ISSN 1097-0290. Disponível em: <<http://dx.doi.org/10.1002/bit.23314>>.

SUN, S. et al. The strong association of condensed phenolic moieties in isolated lignins with their inhibition of enzymatic hydrolysis. **Green Chemistry**, v. 18, n. 15, p. 4276-4286, 2016. ISSN 1463-9262. Disponível em: <<http://dx.doi.org/10.1039/C6GC00685J>>.

SWIERENGA, H. et al. Development of robust calibration models in near infra-red spectrometric applications. **Analytica Chimica Acta**, v. 411, n. 1-2, p. 121-135, 5/1/ 2000. ISSN 0003-2670. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0003267000007182>>.

TAHERZADEH, M. J. et al. Physiological effects of 5-hydroxymethylfurfural on *Saccharomyces cerevisiae*. **Applied Microbiology and Biotechnology**, v. 53, n. 6, p. 701-708, 2000. ISSN 1432-0614. Disponível em: <<http://dx.doi.org/10.1007/s002530000328>>.

TEJIRIAN, A.; XU, F. Inhibition of enzymatic cellulolysis by phenolic compounds. **Enzyme Microb Technol**, v. 48, 2011// 2011. Disponível em: <<http://dx.doi.org/10.1016/j.enzmictec.2010.11.004>>.

TURKIA, H. et al. Capillary electrophoresis for the monitoring of phenolic compounds in bioprocesses. **Journal of Chromatography A**, v. 1278, p. 175-180, 2/22/ 2013. ISSN 0021-9673. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0021967313000423>>.

VAN DER POL, E. C. et al. By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. n. 1432-0614 (Electronic).

VASCONCELLOS, V. M. et al. Production efficiency versus thermostability of (hemi)cellulolytic enzymatic cocktails from different cultivation systems. **Process Biochemistry**, ISSN 1359-5113, 2015. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1359511315300453>>.

WONG, D. S. Feruloyl esterase. **Applied Biochemistry and Biotechnology**, v. 133, n. 2, p. 87-112, 2006/05/01 2006. ISSN 0273-2289. Disponível em: <<http://dx.doi.org/10.1385/ABAB%3A133%3A2%3A87>>.

WYMAN, C.E. Biomass ethanol: Technical progress, opportunities, e commercial challenges. **Annual Review of Energy e the Environment**, v. 24, p. 189-226, 1999.

XIMENES, E. et al. Inhibition of cellulases by phenols. **Enzyme Microb Technol**, v. 46, 2010// 2010. Disponível em: <<http://dx.doi.org/10.1016/j.enzmictec.2009.11.001>>.

XIMENES, E. et al. Deactivation of cellulases by phenols. **Enzyme Microb Technol**, v. 48, 2011// 2011. Disponível em: <<http://dx.doi.org/10.1016/j.enzmictec.2010.09.006>>.

ZENG, Y. et al. Lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels. **Curr Opin Biotechnol**, v. 27, 2014// 2014. Disponível em: <<http://dx.doi.org/10.1016/j.copbio.2013.09.008>>.

ZEWOTIR, T.; GALPIN, J. A unified approach on residuals, leverages and outliers in the linear mixed model. **TEST**, v. 16, n. 1, p. 58-75, 2007/05/01 2007. ISSN 1133-0686. Disponível em: <<http://dx.doi.org/10.1007/s11749-006-0001-2>>.

ZHA, Y. et al. Identifying inhibitory compounds in lignocellulosic biomass hydrolysates using an exometabolomics approach. **BMC Biotechnology**, v. 14, n. 1, p. 22, 2014// 2014. ISSN 1472-6750. Disponível em: <<http://dx.doi.org/10.1186/1472-6750-14-22>>.

ZHANG, Y.-H. P.; LYND, L. R. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. **Biotechnology and Bioengineering**, v. 88, n. 7, p. 797-824, 2004. ISSN 1097-0290. Disponível em: <<http://dx.doi.org/10.1002/bit.20282>>.

ZHANG, Y. H. P. What is vital (and not vital) to advance economically-competitive biofuels production. **Process Biochem**, v. 46, 2011// 2011. Disponível em: <<http://dx.doi.org/10.1016/j.procbio.2011.08.005>>.

ZHENG, H.-Z.; HWANG, I.-W.; CHUNG, S.-K. Enhancing polyphenol extraction from unripe apples by carbohydrate-hydrolyzing enzymes. **Journal of Zhejiang University. Science. B**, Hangzhou, v. 10, n. 12, p. 912-919, 07/02/received 10/19/accepted 2009. ISSN 1673-1581/1862-1783. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2789526/>>.