

UNIVERSIDADE FEDERAL DE SÃO CARLOS– UFSCAR  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA– CCET  
DEPARTAMENTO DE ENGENHARIA QUÍMICA– DEQ  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA– PPGEQ

HOMERO AIDAR NETO

**USO DE TÉCNICA DE ESPECTROSCOPIA  
NIR NO DESENVOLVIMENTO DE  
SOFTSENSOR PARA MONITORAMENTO  
*ON-LINE* DA FERMENTAÇÃO ALCOÓLICA  
INDUSTRIAL**

São Carlos  
2024



HOMERO AIDAR NETO

**USO DE TÉCNICA DE ESPECTROSCOPIA  
NIR NO DESENVOLVIMENTO DE  
SOFTSENSOR PARA MONITORAMENTO  
*ON-LINE* DA FERMENTAÇÃO ALCOÓLICA  
INDUSTRIAL**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Química do Centro de Ciências Exatas e de Tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Engenharia Química.

Área de concentração: Pesquisa e Desenvolvimento de Processos Químicos

Orientador: Prof. Dr. Marcelo Perencin de Arruda Ribeiro

São Carlos

2024





**UNIVERSIDADE FEDERAL DE SÃO CARLOS**

Centro de Ciências Exatas e de Tecnologia  
Programa de Pós-Graduação em Engenharia Química

---

**Folha de Aprovação**

---

Defesa de Dissertação de Mestrado do candidato Homero Aidar Neto, realizada em 22/02/2024.

**Comissão Julgadora:**

Prof. Dr. Marcelo Perencin de Arruda Ribeiro (UFSCar)

Profa. Dra. Arianne Silveira Sbrice Pinto (DU)

Prof. Dr. Charles Dayan Farias de Jesus (UFSCar)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Engenharia Química.

*Este trabalho é dedicado às crianças adultas que,  
quando pequenas, sonharam em se tornar cientistas.*

---

# Agradecimentos

---

Dedico este trabalho primeiramente à minha família e namorada que sempre me apoiaram em minhas decisões.

Agradeço ao meu orientador Dr. Marcelo Perencin de Arruda Ribeiro pelo imenso suporte e paciência nesses dois anos.

A todos os meus amigos de São Carlos e de fora que me deram suporte presencial e virtual durante toda essa jornada.

Aos professores e membros do grupo da Fermentec que me estimularam a navegar nesse grande tema de pesquisa: fermentação alcoólica.





---

# Apoio Financeiro

---

Especiais agradecimentos à Fermentec Tecnologias em Açúcar e Álcool LTDA (Piracicaba), à FINEP (Financiadora de Estudos e Projetos) e ao MCTIC (Ministério da Ciência, Tecnologia, Inovações e Comunicações) com recursos do FNDCT (Fundo Nacional de Desenvolvimento Científico e Tecnológico) pelo apoio financeiro (UNIFEI PROEX nº 23088.017409/2022-61 e UFSCar 14.708 – ProEx nº 012827/2022-18). Este estudo foi financiado também pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - código 001 e 88887.701760/2022-00 - e pela FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) - Processo 2016/10.636-8.



*“Somos essencialmente profissionais do sentido. Educamos, quando ensinamos com sentido. Educar é impregnar de sentido a vida. A profissão docente está centrada na vida, no bem querer.”*  
*(Prof. Gilberto Teixeira)*



---

# Resumo

---

O controle e o monitoramento automatizado de bioprocessos são ferramentas fundamentais para a indústria biotecnológica moderna. A fermentação alcoólica, por exemplo, é um processo biotecnológico já bem consolidado na indústria sucroalcooleira, tendo alta produtividade e microrganismo robusto. No entanto, um fator ainda pode ser melhorado: o monitoramento do bioprocessos. O monitoramento da fermentação alcoólica tipicamente é realizado com técnicas *at-line* e *off-line*, como Cromatografia Líquida de Alta Eficiência (HPLC), o que implica em atraso para verificar o fim da fermentação e/ou algum distúrbio. A espectroscopia NIR pode ser utilizada para monitorar bioprocessos, inferindo em tempo real variáveis chave na fermentação, como a concentração residual de açúcar. A grande dificuldade no uso dessa técnica é a presença de ruído intenso nas medições, necessitando de pré-tratamentos para seu uso em um softsensor. O uso de um modelo fenomenológico de fermentação pode reduzir o ruído associado dos dados experimentais, melhorando a estimativa das variáveis da fermentação pela técnica quimiométrica. Diante destes aspectos, foi desenvolvida uma metodologia para a inferência de concentração celular, de concentração de substrato e de concentração de produto em uma fermentação alcoólica. A metodologia se baseia na suavização dos dados experimentais das amostras por modelo cinético fermentativos. Tais dados foram correlacionados com espectros NIR para ajuste do modelo quimiométrico PLSR (*Partial Least Squares Regression*) e construção do sensor virtual. Três fermentações bateladas alimentadas foram utilizadas para construção do conjunto de calibração e três bateladas foram utilizadas para validação do sensor.

**Palavras-chave:** fermentação alcoólica, monitoramento, espectroscopia NIR, softsensor.



---

# Abstract

---

Automated control and monitoring of bioprocesses are fundamental tools for the modern biotechnology industry. Chemical introduction, for example, is a biotechnological process already well established in the sugar and alcohol industry, with high productivity and robust microorganisms. However, one factor can still be improved: bioprocess monitoring. Detailed fermentation monitoring is carried out with *at-line* and *off-line* techniques, such as High Performance Liquid Chromatography (HPLC), which implies a delay to verify the end of fermentation and/or some disturb. NIR spectroscopy can be used to monitor bioprocesses, inferring key variables in fermentation in real time, such as residual sugar concentration. A major difficulty in using this technique is the presence of intense noise at frequencies, requiring pre-treatment for its use in a soft sensor. The use of a phenomenological interference model can reduce the noise associated with experimental data, improving the estimation of interference variations using the chemometric technique. Given these aspects, a methodology was developed for inferring cell concentration, substrate concentration and product concentration in a prepared fermentation. The methodology is based on smoothing experimental data from samples using a fermentative kinetic model. Such data were correlated with NIR spectra to adjust the chemometric model PLSR (*Partial Least Squares Regression*) and construct the virtual sensor. Three fed-batch fermentations were used to construct the detection set and three batches were used to validate the sensor.

**Keywords:** alcoholic fermentation, monitoring, NIR spectroscopy, softsensor.





---

# Lista de ilustrações

---

Figura 1 – Produção mundial de etanol no período de 2007 a 2021. . . . .	36
Figura 2 – Fluxograma simplificado de produção de etanol 1G e 2G. . . . .	38
Figura 3 – Via metabólica simplificada da fermentação alcoólica para a <i>Saccharomyces cerevisiae</i> . . . . .	39
Figura 4 – Fluxograma do processo Melle-Boinot operando em fermentação contínua. . . . .	41
Figura 5 – Ilustração esquemática do processo Melle-Boinot simplificado aplicado para uma dorna de fermentação operada de forma semicontínua. . . . .	42
Figura 6 – Diagrama esquemático de um sensor virtual. . . . .	47
Figura 7 – Exemplo de um gráfico de resíduos de Student <i>versus Leverage</i> : linhas pretas pontilhadas (limites para o Resíduo de Student), linha vermelha pontilhada (valor de <i>Leverage</i> crítico do conjunto de calibração ( $h_{crit}$ )), círculos azuis (amostras de calibração), cruces vermelhas (amostras de teste). . . . .	55
Figura 8 – Princípio da validação cruzada <i>k-fold</i> com 5 grupos. . . . .	57
Figura 9 – Otimização do número de variáveis latentes pela minimização do <i>Root Mean Square Error of Cross Validation (RMSECV)</i> . . . . .	58
Figura 10 – Otimização do número de variáveis latentes pelo Teste F de Osten: linha azul com círculo azul (valores de $F_{(k)}^{estat}$ ), linha vermelha tracejada (valor de $F_{5\%}$ ). . . . .	60
Figura 11 – Regiões de espectroscopia e comprimentos de onda característicos. . . . .	61
Figura 12 – Medição real de absorção. . . . .	63
Figura 13 – Exemplos de espectro na região de comprimento de onda na faixa de 2245 - 2368 nm. <b>(a)</b> : Espectros brutos, <b>(b)</b> : espectros suavizados por uma média de janela móvel com largura 5, <b>(c)</b> : espectros suavizados por uma média de janela móvel com largura 11. . . . .	65

Figura 14 – Estimativa da primeira derivada por Savitzky-Golay. Uma janela de 7 pontos e um polinômio de segunda ordem são usados para suavização. . . . .	66
Figura 15 – O espectro da amostra (pontos azuis) plotado em relação a um espectro de referência selecionado. Os termos de correção escalar são encontrados como a interceptação e a inclinação da linha preta, que é encontrada a partir do ajuste da regressão de mínimos quadrados em todos os pontos. . . . .	68
Figura 16 – Aparato experimental completo instalado para conduzir as fermentações sob monitoramento dos Sistemas NIR-Online X-One (3, 12 e 14), incluindo ainda a estrutura de suporte (8), biorreator e camisa (9), sistema de recirculação de meio reacional (7 e 13), de fornecimento de meio de alimentação (4 e 6), de controle de temperatura (15) e computadores para monitoramento do processo (1, 3 e 5). . . . .	71
Figura 17 – Diagrama esquemático do sistema experimental utilizado para realização dos ensaios fermentativos. . . . .	72
Figura 18 – Espectrofotômetro NIR-Online X-One. . . . .	76
Figura 19 – Espectrofotômetro NIR-Online X-One montado no sistema experimental. . . . .	77
Figura 20 – Suavização dos dados experimentais das <b>Ferm 1</b> utilizando o modelo cinético fermentativo. . . . .	81
Figura 21 – Suavização dos dados experimentais das <b>Ferm 2</b> utilizando o modelo cinético fermentativo. . . . .	82
Figura 22 – Suavização dos dados experimentais das <b>Ferm 3</b> utilizando o modelo cinético fermentativo. . . . .	83
Figura 23 – Suavização dos dados experimentais das <b>Ferm 4</b> utilizando o modelo cinético fermentativo. . . . .	84
Figura 24 – Suavização dos dados experimentais das <b>Ferm 5</b> utilizando o modelo cinético fermentativo. . . . .	85
Figura 25 – Suavização dos dados experimentais das <b>Ferm 6</b> utilizando o modelo cinético fermentativo. . . . .	86
Figura 26 – Espectros na região do <i>Near-infrared</i> (NIR) obtidos pelo Sistema NIR-Online X-One durante as Fermentações de 1 a 6. . . . .	87
Figura 27 – Regiões de concentração de etanol ( $C_p$ ), açúcares redutores totais ( $C_s$ ) e células ( $C_x$ ) para as diferentes fermentações: Ferm 1 – vermelho; Ferm 2 – azul; Ferm 3 – verde; Ferm 4 – preto; Ferm 5 – amarelo; Ferm 6 – ciano. . . . .	88
Figura 28 – Efeito do pré-tratamento <i>Extended Multiplicative Scatter Correction</i> (EMSC) nos espectros do conjunto de calibração. . . . .	89
Figura 29 – Efeito do pré-tratamento EMSC nos espectros do conjunto de validação externa. . . . .	89

Figura 30 – Otimização da complexidade do modelo com dados experimentais para a variável $C_x$ . . . . .	91
Figura 31 – Otimização da complexidade do modelo com dados experimentais para a variável $C_s$ . . . . .	91
Figura 32 – Otimização da complexidade do modelo com dados experimentais para a variável $C_p$ . . . . .	92
Figura 33 – Otimização da complexidade do modelo com dados experimentais suavizados para a variável $C_x$ . . . . .	92
Figura 34 – Otimização da complexidade do modelo com dados experimentais suavizados para a variável $C_s$ . . . . .	93
Figura 35 – Otimização da complexidade do modelo com dados experimentais suavizados para a variável $C_p$ . . . . .	93
Figura 36 – Gráfico de validação cruzada para o componente $C_x$ do modelo com dados experimentais otimizado e inferência do grupo de validação externa (da esquerda para direita, Ferm 4 e Ferm 5, e abaixo, Ferm 6) . . . . .	94
Figura 37 – Gráfico de validação cruzada para o componente $C_s$ do modelo com dados experimentais otimizado e inferência do grupo de validação externa (da esquerda para direita, Ferm 4 e Ferm 5, e abaixo, Ferm 6) . . . . .	95
Figura 38 – Gráfico de validação cruzada para o componente $C_p$ do modelo com dados experimentais otimizado e inferência do grupo de validação externa (da esquerda para direita, Ferm 4 e Ferm 5, e abaixo, Ferm 6) . . . . .	96
Figura 39 – Gráfico de validação cruzada para o componente $C_x$ do modelo com dados experimentais suavizados otimizado e inferência do grupo de validação externa (da esquerda para direita, Ferm 4 e Ferm 5, e abaixo, Ferm 6). . . . .	97
Figura 40 – Gráfico de validação cruzada para o componente $C_s$ do modelo com dados experimentais suavizados otimizado e inferência do grupo de validação externa (da esquerda para direita, Ferm 4 e Ferm 5, e abaixo, Ferm 6). . . . .	98
Figura 41 – Gráfico de validação cruzada para o componente $C_x$ do modelo com dados experimentais suavizados otimizado e inferência do grupo de validação externa (da esquerda para direita, Ferm 4 e Ferm 5, e abaixo, Ferm 6). . . . .	99



---

## Lista de tabelas

---

Tabela 1 – Modelos cinéticos com e sem inibição. . . . .	46
Tabela 2 – Diferentes cenários presentes nos ensaios fermentativos realizados. . . .	74
Tabela 3 – Condições e parâmetros dos ensaios fermentativos realizados . . . . .	80
Tabela 4 – Modelos PLS obtidos com dados experimentais (da esquerda para direita em relação ao RMSEP: Ferm 4, Ferm 5 e Ferm 6 ). . . . .	100
Tabela 5 – Modelos PLS obtidos com dados experimentais suavizados (da esquerda para direita em relação ao RMSEP: Ferm 4, Ferm 5 e Ferm 6 ). . . . .	100



---

# Lista de siglas

---

$Abs$  Absorbância

$A$  matriz da resposta do instrumento (espectros de calibração)

$A_{test}$  matriz da resposta do instrumento (espectros de teste)

**COP-21 da Convenção do Clima** 21<sup>a</sup> Conferência das Partes da Convenção-Quadro das Nações Unidas sobre a Mudança do Clima

$c$  concentração do analito

**CLS** *Classical Least Squares*

$C$  matriz de concentração de componentes

$c_{test,i}$  concentração de referência na predição da amostra  $i$  do conjunto de teste

$c_{test/pred,i}$  concentração predita da amostra  $i$  do conjunto de teste

$c_{cab,i}$  concentração de referência da amostra  $i$  do conjunto de calibração

$c_{cab/pred,i}$  concentração predita por validação cruzada da amostra  $i$  do conjunto de calibração

$C_x$  concentração de células em g/L

$C_s$  concentração de substrato, em g/L

$C_p$  concentração de etanol em g/L

$d$  caminho óptico

**DO** Densidade ótica

**EMSC** *Extended Multiplicative Scatter Correction*

$e$  Absortividade molar

$E_A$  matriz dos erros espectrais não ajustados pelo modelo

$E_C$  matriz dos erros de concentração não ajustados pelo modelo

$E$  erro residual nos valores não ajustados pelo modelo *Principal Component Analysis* (PCA)

$E_j^2$  erros quadráticos de estimativa entre as concentrações de referência do grupo de validação  $j$  e as concentrações preditas pelo modelo ajustado com o grupo de treinamento

$E_{trein,j}$  erros quadráticos de estimativa entre as concentrações de referência do grupo de treinamento  $j$  e as concentrações preditas pelo modelo ajustado com o grupo de treinamento



$F_{(k)}^{estat}$  estatística de teste calculada a partir de um teste estatístico da distribuição F para  $k$  variáveis latentes

**GEE** Gases de Efeito Estufa

$h$  número de variáveis latentes retidas no modelo PLS

**HPLC** *High Performance Liquid Chromatography*

$H_{cab}$  vetor que contém os valores de *Leverage* para as amostras de calibração

$H_{test}$  vetor que contém os valores de *Leverage* para as amostras de teste

$h_{crit}$  *Leverage* crítico do conjunto de calibração

$H_{cab}(i)$  valor de *Leverage* da amostra  $i$  do conjunto de calibração

$H_0$  hipótese nula

$H_1$  hipótese alternativa

$I$  Intensidade da radiação emitida

$I_0$  Intensidade da radiação incidente

**ILS** *Inverse Least Squares*

$k$  constante de proporcionalidade

$K$  matriz de parâmetros que representam o produto entre a absortividade específica de vezes o comprimento do caminho óptico de dado componente puro

$K_{opt}$  matriz de parâmetros otimizados para a regressão *Classical Least Squares* (CLS)

$k$  número de grupos na validação cruzada *k-fold*

**LD** linearmente dependentes

**LV** *Latent Variables*

$L_A$  matriz de *loadings* (autovetores) de matriz da resposta do instrumento (espectros de calibração) ( $A$ )

$L_C$  matriz de *loadings* (autovetores) de matriz de concentração de componentes ( $C$ )

**MSPC** *Multivariate Statistical Process Control*

**MSC** *Multiplicative Scatter Correction*

**MLR** *Multiple Linear Regression*

**NIR** *Near-infrared*

$n_{cal}$  número de amostras do conjunto de calibração

$n_c$  número de componentes químicos

$n_{ond}$  número de comprimentos de onda discretos

$n_{test}$  número de amostras do conjunto de teste

$n_{cal/trei}$  número de amostras do conjunto de calibração utilizadas alternadamente nos grupos de treinamento para validação cruzada

$n_{simples}$  número de parâmetros do modelo simples

$n_{complexo}$  número de parâmetros do modelo complexo

$p$  número de componentes principais retidos no modelo PCA

**PCA** *Principal Component Analysis*

**PLSR** *Partial Least Squares Regression*

**Proálcool** Programa Nacional do Álcool

**PC** *Principal Components*

**PCR** *Principal Component Regression*

*PLS1* modelos PLS univariado

*PLS2* modelos PLS multivariado

$P$  matriz de coeficientes a ser ajustada pelo modelo inverso

$P_{opt}$  matriz de parâmetros otimizados para o modelo inverso

$P_{adic}$  vetor dos parâmetros adicionais no modelo complexo comparado com o modelo simples

$Q$  matriz de autovalores (*scores*)

**RMSECV** *Root Mean Square Error of Cross Validation*

*RMSEP* Root Mean Square Error of Prediction

*RMSEcal* Root Mean Square Error of calibration

$R_A$  erro residual em  $A$  nos valores não ajustados pelo modelo PLS

$R_C$  erro residual em  $C$  nos valores não ajustados pelo modelo PLS

$RS_{cab,i}$  Resíduo de Student da amostra  $i$  do conjunto de calibração

$RS_{test,i}$  Resíduo de Student da amostra  $i$  do conjunto de teste

$RMSECV_k^2$  Root Mean Square Error of Cross Validation for  $k$  latent variables

$RMSECV_{k+1}^2$  Root Mean Square Error of Cross Validation for  $k + 1$  latent variables

**SUPERSYS-HCDC** *A supervision and control tool based on artificial intelligence for high cell density cultivations*

**SNV** *Standard Normal Variate*

**SG** Savitzky-Golay

**SPA** *Successive Projections Algorithm*

$Tr$  Transmissão

$T$  matriz de *scores* (autovalores) de  $A$

$T_{test}$  matriz de *scores* (autovalores) de  $A_{test}$

$U$  matriz de *scores* (autovalores) de  $C$

$W$  matriz de pesos de  $A$

$var_{cab}$  estimativa da variância dos resíduos para o conjunto de calibração

$Y$  matriz de autovetores (*loadings*)



---

# Sumário

---

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>31</b>
1.1	Objetivos . . . . .	33
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA . . . . .</b>	<b>35</b>
<b>2.1</b>	<b>Fermentação Alcoólica . . . . .</b>	<b>35</b>
2.1.1	Produção mundial de Etanol . . . . .	35
2.1.2	Etanol no Brasil . . . . .	36
2.1.3	Processo Industrial . . . . .	37
2.1.4	Bioquímica da Fermentação Alcoólica . . . . .	39
2.1.5	Modos de operação dos processos fermentativos . . . . .	40
2.1.5.1	Processos contínuos . . . . .	40
2.1.5.2	Processos descontínuos . . . . .	41
2.1.5.3	Processos semicontínuos . . . . .	42
2.1.6	Influência da Temperatura na Fermentação Alcoólica . . . . .	43
2.1.7	Modelagem da Fermentação Alcoólica . . . . .	44
<b>2.2</b>	<b>Sensor Virtual (<i>softsensor</i>) para fermentação . . . . .</b>	<b>47</b>
<b>2.3</b>	<b>Calibração Multivariada Clássica em Quimiometria . . . . .</b>	<b>48</b>
2.3.1	Fundamento teórico e técnicas . . . . .	48
2.3.2	Deteção de <i>Outliers</i> (pontos anômalos) . . . . .	53
2.3.2.1	Resíduos de Student e <i>Leverage</i> . . . . .	53
2.3.3	Otimização da complexidade do modelo quimiométrico . . . . .	55
2.3.3.1	RMSECV e RMSEcal . . . . .	58
2.3.3.2	Teste F de Osten . . . . .	59
<b>2.4</b>	<b>Espectroscopia de Infravermelho Próximo (NIR) . . . . .</b>	<b>61</b>
2.4.1	Fundamentos . . . . .	61
2.4.2	Sensores Virtuais baseados em dados . . . . .	62
2.4.2.1	Modelos estatísticos multivariados . . . . .	62

2.4.3	Pré-tratamento dos dados espectrométricos . . . . .	63
2.4.3.1	Média Móvel . . . . .	64
2.4.3.2	Derivadas . . . . .	64
2.4.3.3	MSC . . . . .	66
2.4.3.4	SNV . . . . .	67
<b>3</b>	<b>MATERIAIS E MÉTODOS . . . . .</b>	<b>69</b>
<b>3.1</b>	<b>Fermentações de <i>S. cerevisiae</i> em Biorreator . . . . .</b>	<b>69</b>
3.1.1	Aparato Experimental e Testes Analíticos . . . . .	69
3.1.2	Ensaio Fermentativo . . . . .	73
3.1.3	Uso do Modelo Matemático Fermentativo para suavização dos dados experimentais . . . . .	75
<b>3.2</b>	<b>Monitoramento do processo com NIR . . . . .</b>	<b>76</b>
<b>3.3</b>	<b>Tratamento Matemático dos dados . . . . .</b>	<b>77</b>
3.3.1	Pré-processamento dos Espectros . . . . .	77
3.3.2	Calibração Multivariada e Otimização da Complexidade do modelo . . . .	78
<b>4</b>	<b>RESULTADOS E DISCUSSÃO . . . . .</b>	<b>79</b>
<b>5</b>	<b>CONCLUSÃO . . . . .</b>	<b>101</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>103</b>



---

# Capítulo 1

## Introdução

---

A crescente demanda energética e a preocupação com as emissões de Gases de Efeito Estufa (GEE) impulsionam a busca por fontes alternativas de energia renováveis, sustentáveis e economicamente viáveis (JOSHI et al., 2017). O etanol é um biocombustível com grande potencial para substituir os combustíveis fósseis e reduzir de forma significativa a emissão de GEE (KANG et al., 2015).

O Brasil é o segundo maior produtor mundial de etanol (27% da produção mundial), mantendo-se atrás apenas dos Estados Unidos (RFA, 2022). Em 2022, foram produzidos 30,75 bilhões de litros de etanol no Brasil, sendo que o etanol combustível correspondeu a 27,53 bilhões de litros dessa produção (89,5%) (USDABR, 2024). Em comparação com 2022, estima-se que em 2023 foram produzidos 32,95 bilhões de litros de etanol no Brasil, um aumento de 7%. A expectativa de aumento está relacionada com o aumento da produção de etanol de milho, preço mais elevado do petróleo e o incentivo do programa RenovaBio, possibilitando que a produção de etanol dobre até 2030 (STAFF, 2018).

No Brasil, a principal matéria-prima empregada para produzir etanol é a cana-de-açúcar, sendo que o etanol obtido a partir da cana reduz em cerca de 89% a emissão de GEE ( $\text{CO}_2$ ,  $\text{CH}_4$  e  $\text{NO}_2$ ) em comparação à gasolina (CANA, 2018). Nas indústrias sucroalcooleiras, a produção de etanol pode ser dividida nas seguintes etapas: extração, tratamento, fermentação e cogeração de vapor e energia (PEREIRA; BADINO; CRUZ, 2018; BAEYENS et al., 2015).

A maior parte das destilarias emprega o processo batelada alimentada para a fermentação com reciclo de leveduras, processo conhecido como Melle-Boinot (BASSO; BASSO, 2018). A grande vantagem desse processo é que ele permite operar a fermentação com altas concentrações de células de levedura, evitando o consumo de açúcares para crescimento e brotação de leveduras no início de cada ciclo, o que reduz o tempo e os custos associados

com preparação do inóculo e aumenta a eficiência geral do processo fermentativo (NEITZEL et al., 2020). Iniciando com uma alta concentração de células de levedura, diminui-se o tempo de fermentação e aumenta-se a produtividade (BASSO; BASSO, 2018).

A fermentação alcoólica apresenta uma grande limitação ao longo do processo: variáveis como concentração de células, concentração de substrato e concentração de etanol não costumam ser medidas *on-line* por causa da dificuldade, alto custo e/ou não existência dos sensores associados, necessitando de técnicas *at-line* e/ou *offline* para sua determinação, como Cromatografia Líquida de Alta Eficiência, *High Performance Liquid Chromatography* (HPLC).

No entanto, o uso de HPLC inviabiliza o monitoramento em tempo real, impedindo que se otimize e controle o bioprocessamento industrialmente devido ao tempo despendido na análise das amostras. Para a fermentação alcoólica, busca-se direcionar o processo para que ocorra aumento do teor alcoólico do vinho produzido e diminuição do volume de vinhaça. O conhecimento em tempo real do perfil de açúcares e de etanol dentro da dorna fermentativa permite que se otimize as variáveis manipuladas de processo, como a vazão de alimentação de mosto e a vazão de água de refrigeração, deslocando o sistema de situações desfavoráveis, a exemplo de altas temperaturas, inibição do metabolismo celular pelo etanol (SONEGO, 2016; SANTOS; ALUÍZIO; CALDAS, 2015; RIVERA et al., 2006).

Considerando estes aspectos, a aplicação de técnicas que permitam o monitoramento do processo a partir de informações diretas e adquiridas de maneira rápida, precisa e reprodutível torna-se crítica (BILSKEY et al., 2020; PINTO, 2017). Uma forma de estimar, em tempo real, seria por meio de *softsensors* (sensores virtuais). Sensores virtuais são modelos matemáticos que estimam, em tempo real, a variável desejada (variável predita) a partir de dados medidos do sistema (variáveis preditoras) (LOTUFO; GARCIA, 2008; FORTUNA et al., 2006). Para o ajuste do modelo matemático do sensor virtual, deve-se determinar o conjunto de dados históricos da variável predita por meio de alguma técnica analítica, como, por exemplo, o HPLC, e correlacionar com o conjunto de dados históricos das variáveis preditoras selecionadas.

A técnica de espectroscopia NIR, espectroscopia do Infravermelho próximo, é uma tecnologia madura que tem sido amplamente utilizada para monitorar vários processos de fermentação em escalas de laboratório e piloto (GARGALO et al., 2022). Ela é baseada na propriedade de absorção de energia de uma molécula em uma determinada região do espectro eletromagnético, notadamente em comprimentos de onda de 780 a 2500 nm (PASQUINI, 2003). Sua principal vantagem em relação às análises químicas convencionais é a capacidade de quantificar multiconstituintes com o mínimo, ou nenhuma, preparação de amostra, de forma rápida e não destrutiva. Para seu uso, no entanto, necessita-se realizar a calibração multivariada que engloba a definição do modelo e ajuste de seus parâmetros.

A Regressão por Mínimos Quadrados Parciais, *Partial Least Squares Regression* (PLSR), é umas das principais técnicas de análise multivariada encontradas na literatura e vem se mostrando eficaz no monitoramento de diversos componentes em meios reacionais complexos (PAULA, 2020; PINTO, 2017; SVENDSEN, 2016; RIBEIRO, 2007). O uso de espectros NIR associados à PLSR já vem sendo aplicado na quantificação de compostos presentes no meio de fermentação alcoólica (FULGÊNCIO et al., 2023; NASCIMENTO et al., 2017; BLANCO; PEINADO; MAS, 2004).

Apesar da vasta aplicação de técnicas de espectroscopia NIR associadas à análise multivariada na quantificação de analitos distintos na fermentação alcoólica com meio definido, não foram encontrados estudos na literatura que abordassem o uso de espectros NIR associados à calibração multivariada para monitoramento de uma fermentação alcoólica com meio complexo. O meio complexo, diferentemente do meio definido que, no geral, possui composição química definida e controlada, apresenta corpos estranhos ao meio (fibras, gomas, leveduras selvagens, contaminação) (PACHECO, 2010). No caso do Brasil, onde 70% das unidades de processamento de cana-de-açúcar são usinas com destilarias anexas (produzem tanto açúcar quanto etanol) (CAVALETT et al., 2012), o meio complexo é resultado do uso de melão e/ou de caldo de cana como mosto de fermentação.

Além disso, a falta de reprodutibilidade e o grande erro experimental inerentes aos processos biológicos, (LOPES; GABRIEL; BORGES, 2011), podem dificultar a criação de um sensor virtual com boa capacidade preditiva. As técnicas padrões para acompanhamento da fermentação alcoólica, como HPLC e leitura de densidade ótica (DO), podem apresentar ruído em suas medições, tornando os dados provenientes dessas técnicas inviáveis para construção de um modelo quimiométrico baseado em espectroscopia eficiente.

Uma possível solução seria o uso de um modelo fenomenológico de fermentação ajustado com os dados experimentais dos ensaios. Isso poderia reduzir tais efeitos, além de gerar mais dados para a calibração do sensor. No entanto, erros de modelagem podem acarretar em erro sistemático, prejudicando seu desempenho na estimação das variáveis preditas. Estudos avaliando a suavização dos dados experimentais por modelos fermentativos e seu impacto em modelos de calibração multivariada baseados em espectroscopia NIR parecem ainda não terem sido abordados na literatura.

Portanto, neste trabalho, foi proposto desenvolver um *softsensor* baseado na técnica de espectroscopia NIR para monitoramento *on-line* de uma fermentação alcoólica industrial.

## 1.1 Objetivos

O presente estudo teve como objetivo geral desenvolver um *softsensor* baseado na técnica de espectroscopia NIR para monitoramento em tempo real de uma fermentação alcoólica industrial.

Os objetivos específicos do trabalho são:

- a) Realização de ensaios de fermentação alcoólica com meio industrial para coleta de dados experimentais e de espectros;
- b) Uso de um modelo fenomenológico de fermentação para suavizar dados experimentais dos ensaios fermentativos;
- c) Definição de uma metodologia para tratamento dos dados provenientes dos espectros;
- d) Avaliação/validação do desempenho do sensor virtual com metodologia clássica (dados experimentais) e com metodologia proposta (uso do modelo fenomenológico para alisar dados experimentais).

---

## Capítulo 2

# Revisão Bibliográfica

---

O intuito deste trabalho foi desenvolver um *softsensor* baseado na técnica de **espectroscopia NIR** para acompanhamento *on-line* de uma **fermentação alcoólica industrial**. Assim, para realização desse objetivo, nas seções subsequentes, foram detalhadas as áreas de conhecimento necessárias para seu êxito. Primeiramente, apresenta-se uma revisão sobre a fermentação alcoólica, destacando, principalmente, o processo industrial e a modelagem matemática.

Em seguida, foram descritos os principais embasamentos das técnicas quimiométricas clássicas utilizadas para construção dos modelos de calibração multivariada baseados em espectroscopia NIR.

Por último, foi discutida a Espectroscopia de Infravermelho Próximo (NIR), dando ênfase em trabalhos da literatura que construíram sensores virtuais baseados nessa técnica para fermentação. Ademais, a importância do pré-tratamento dos dados espectrométricos foi revisada, destacando-se as principais técnicas utilizadas na literatura e o efeito delas para diminuição de erros aleatórios nos espectros.

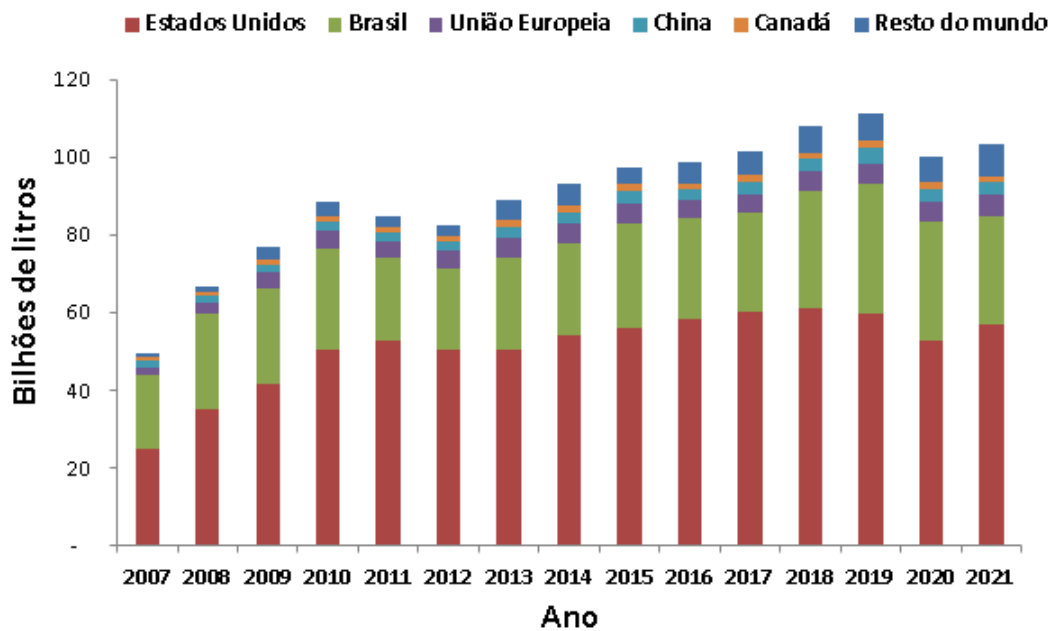
## 2.1 Fermentação Alcoólica

### 2.1.1 Produção mundial de Etanol

O Brasil é o segundo maior produtor mundial de bioetanol (RFA, 2022). Cerca de 82% de todo etanol produzido no mundo é oriundo do Brasil e dos Estados Unidos, que utilizam como principais matérias-primas a cana-de-açúcar e o milho, respectivamente (AFDC, 2021). De maneira geral, a produção global de etanol continua crescendo, mas houve uma redução de 10% em 2020 devido a pandemia do COVID-19, conforme apresentado

na Figura 1.

Figura 1 – Produção mundial de etanol no período de 2007 a 2021.



Adaptado de AFDC (2021)

A tendência de crescimento na produção global de etanol é respaldada por fatores como políticas de incentivo ao aumento da porcentagem em volume de etanol anidro na gasolina do Brasil (18 a 27%) e de outros países, diminuição da emissão de gás carbônico assinada por diversas nações participantes da 21<sup>a</sup> Conferência das Partes da Convenção-Quadro das Nações Unidas sobre a Mudança do Clima (COP-21 da Convenção do Clima), alto custo da eletricidade e de investimentos para implantação de sistemas *plug-in* de carros híbridos (VANDENBERGHE et al., 2022).

### 2.1.2 Etanol no Brasil

A indústria sucroalcooleira nacional passou por diversas inovações tecnológicas que proporcionaram a ampliação da sua capacidade produtiva. Em 20 de Fevereiro de 1931, o Decreto nº 19.717 estabeleceu a aquisição obrigatória de etanol anidro, na proporção de 5% da gasolina importada. Em 1933, foi criado o Instituto do Açúcar e do Alcool (IAA), órgão governamental responsável pelo controle da produção de açúcar e etanol, fixando os preços e os regimes de produção, além do monopólio na comercialização externa do açúcar (LOPES; GABRIEL; BORGES, 2011).

Com a Crise do Petróleo na década de 1970, em que os preços dos barris atingiram valores bastante elevados, o Brasil importava cerca de 80% de suas necessidades. As soluções propostas foram o aumento da produção interna de petróleo e concomitantemente a utilização do etanol como combustível automotivo. Dessa forma, o governo instituiu

o Programa Nacional do Álcool (Proálcool) que fornecia incentivos e subsídios para a expansão de destilarias, o desenvolvimento de veículos movidos a etanol, a mistura de etanol anidro com gasolina (E20) e a pesquisa e desenvolvimento de produção de etanol (GRASSI; PEREIRA, 2019; LOPES; GABRIEL; BORGES, 2011).

Como consequência, houve um significativo desenvolvimento dos setores sucroalcooleiro brasileiro, e em 20 anos, a produção de etanol aumentou mais de 28 vezes, a frota de carros a etanol chegou a 4,5 milhões e, por último, 10% da energia total da matriz energética brasileira teve origem dos produtos da cana-de-açúcar (OLIVEIRA, 2002).

Desde 2003, com a introdução dos veículos *flex-fuel* (carro biocombustível), que pode ser abastecido com etanol hidratado ou gasolina em qualquer proporção, a demanda por etanol cresceu ainda mais. O recrudescimento do preço do barril de petróleo que saltou de US\$ 30 em 2003 e alcançou mais de US\$ 130 no auge da crise de 2008 contribuiu ainda mais para a consolidação da tecnologia *flex-fuel* (GOMES, 2020).

Ademais, a preocupação com as mudanças climáticas e os efeitos no meio ambiente estimularam autoridades brasileiras a tomar medidas concretas para reduzir emissões de CO<sub>2</sub>. Nesse sentido, o Governo Federal criou o programa RenovaBio em dezembro de 2016 com o objetivo de estimular a produção de biocombustíveis no Brasil (Lei nº 13.576/2017). Pretende-se aumentar a eficiência energética dos processos e reduzir as emissões de gases causadores do efeito estufa em 37% até 2025 e para 43% até 2030 (considerando os níveis de emissões de 2005 como linha de base), conforme firmado pelo Brasil na COP-21 (GRASSI; PEREIRA, 2019).

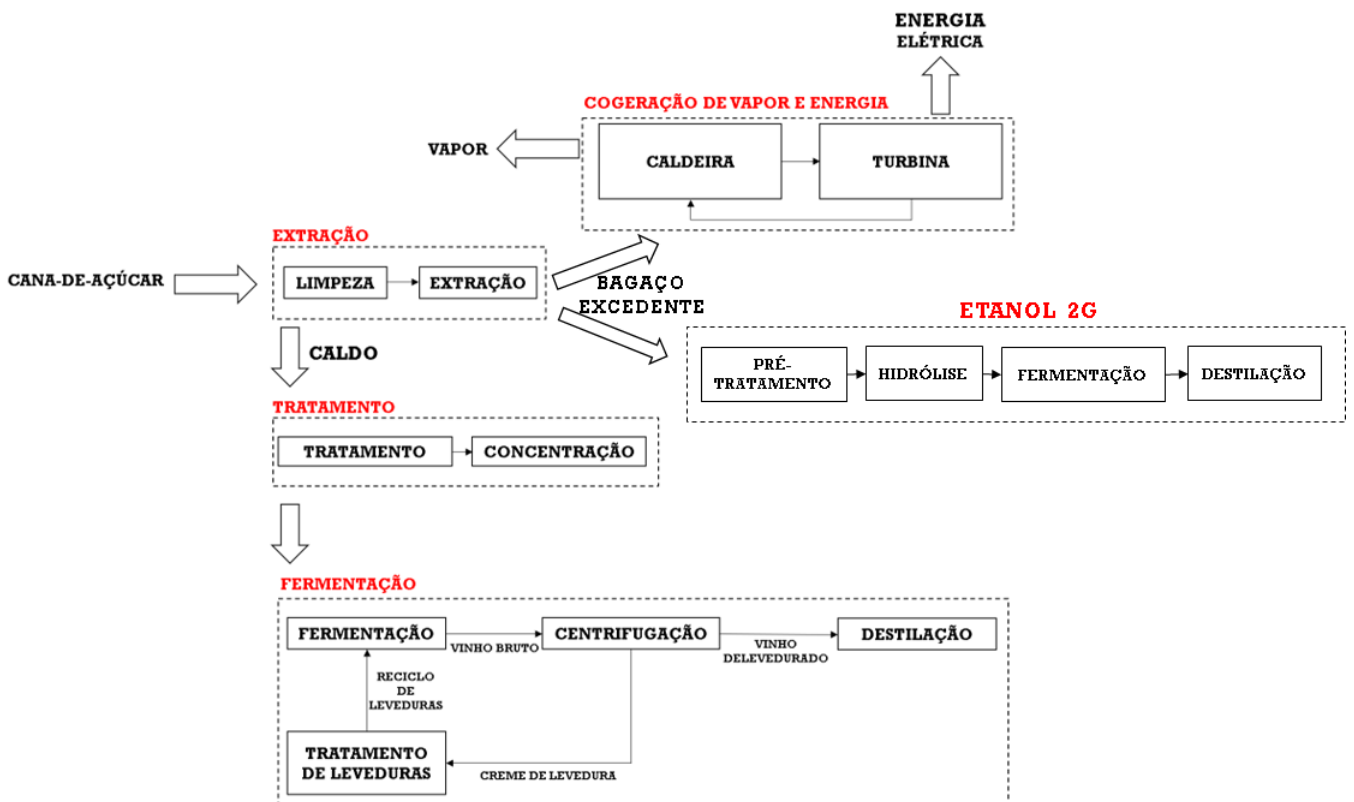
### 2.1.3 Processo Industrial

Atualmente, o etanol pode ser obtido em unidades industriais de primeira (1G - etanol, milho, etc.) ou segunda geração (2G - materiais lignocelulósicos). É possível integrar a produção de etanol de primeira e segunda geração (MARTÍNEZ-JIMENEZ et al., 2022) a partir da cana-de-açúcar. Na produção do etanol 2G, ocorre o aumento das despesas relativas ao custo de capital (adição de etapas ao processo) e ao custo de operação (uso de enzimas hidrolíticas na etapa de hidrólise da biomassa) (LONGATI et al., 2018). Vários são os desafios a serem superados, considerando as duas principais etapas adicionais do processo de produção de etanol 2G, pré-tratamento, hidrólise.

O pré-tratamento do bagaço, necessário para uma ação mais efetiva das enzimas hidrolíticas, aumenta a presença de compostos tóxicos, como aldeídos, ácidos orgânicos e fenóis. Devido à presença desses inibidores, um efeito estressante é imposto à levedura, influenciando na produção de etanol (BASSO; BASSO, 2018; HAHN-HÄGERDAL et al., 2006). Há ainda a concorrência entre o etanol de segunda geração e a produção de eletricidade pelo uso do bagaço (MALUF; GURGEL, 2016). Por isso, na maioria das usinas brasileiras, somente se produz o etanol de primeira geração (BASSO; BASSO, 2018; LONGATI et al., 2018).

Aproximadamente, 70% das unidades de processamento de cana-de-açúcar do Brasil são usinas com destilarias anexas (produzem tanto açúcar quanto etanol); as demais são usinas (produzem apenas açúcar) e destilarias autônomas (produzem apenas etanol) (CAVALETT et al., 2012). A produção de etanol 1G em uma destilaria autônoma brasileira pode ser dividida em quatro principais etapas: extração, tratamento, fermentação e cogeração de vapor e energia (PEREIRA; BADINO; CRUZ, 2018). A Figura 2 apresenta um fluxograma simplificado de produção de etanol 1G e 2G.

Figura 2 – Fluxograma simplificado de produção de etanol 1G e 2G.



Adaptado de Pereira, Badino e Cruz (2018)

A sinergia entre a produção de açúcar e bioetanol foi responsável pela consolidação do setor sucroalcooleiro. A cogeração de bioeletricidade e calor, que são obtidos da queima de bagaço da cana-de-açúcar e/ou queima da palha, permite a autossuficiência energética e subsistência das usinas de cana-de-açúcar 1G. Desde 1999, usinas de cana-de-açúcar podem comercializar sua energia excedente, o que torna o processo ainda mais sustentável e lucrativo (GRASSI; PEREIRA, 2019).

Embora a produção de etanol de primeira geração já seja uma tecnologia bem consolidada, é possível implementar melhorias para elevar o rendimento e reduzir os custos como: aplicação de melhores práticas agrônômicas (cultivares mais produtivas e resistentes), seleção de linhagens e melhoria das características de microrganismos que garantam alta

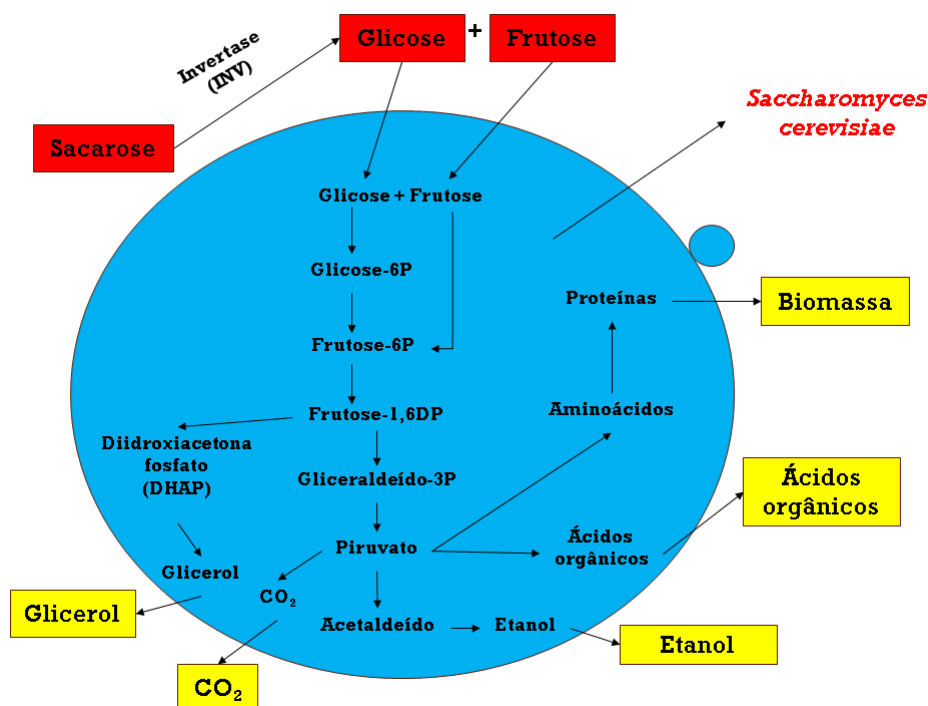


eficiência fermentativa, controle inteligente da temperatura durante a fermentação (diminuindo o consumo de água e a toxicidade do etanol), diminuição da contaminação bacteriana, desenvolvimento de novos processos fermentativos empregando matérias-primas alternativas, etc (LOPES; GABRIEL; BORGES, 2011; PACHECO, 2011).

### 2.1.4 Bioquímica da Fermentação Alcoólica

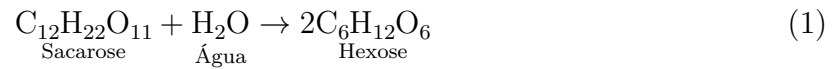
A geração de energia sem a cadeia de transporte de elétrons é chamada de fermentação (SHULER; KARGI, 2006). Existem três tipos de fermentação: a alcoólica, a lática e a acética. A fermentação alcoólica é o processo de transformação de açúcares fermentescíveis (glicose, frutose, etc.) presentes em uma solução em etanol e dióxido de carbono ( $\text{CO}_2$ ) por meio de leveduras em condição anaeróbia. A levedura mais utilizada nas usinas brasileiras é a *Saccharomyces cerevisiae*, sendo ela um microrganismo anaeróbio facultativo, ou seja, tem a habilidade de se ajustar metabolicamente, tanto em condições de aerobiose como de anaerobiose. A transformação do açúcar em etanol e  $\text{CO}_2$  envolve 12 reações em sequência ordenada, cada qual catalisada por uma enzima específica. O objetivo principal da levedura ao metabolizar o açúcar pela via anaeróbia é gerar energia (ATP) para a manutenção das atividades vitais e do crescimento microbiano (LIMA et al., 2002). A Figura 3 apresenta um esquema simplificado da via metabólica da fermentação alcoólica.

Figura 3 – Via metabólica simplificada da fermentação alcoólica para a *Saccharomyces cerevisiae*.

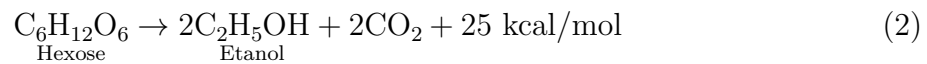


Adaptado de Almeida (2019)

Na fermentação do caldo de cana em que a sacarose é o açúcar predominante, existe uma primeira fase que consiste na inversão da sacarose, ou seja, o seu desdobramento nas hexoses glicose e frutose pela ação de uma enzima, a invertase, produzida pela levedura, conforme apresentado na Equação 1 (LOPES; GABRIEL; BORGES, 2011).



Na segunda fase, ocorre a conversão das hexoses em etanol, gás carbônico e liberação de energia, como mostrado na equação simplificada de Gay-Lussac, Equação 2, o que evidencia o caráter exotérmico da fermentação alcoólica (LOPES; GABRIEL; BORGES, 2011).



A fermentação alcoólica, portanto, origina diversos produtos e subprodutos que podem ter grande impacto na análise espectroscópica.

### 2.1.5 Modos de operação dos processos fermentativos

Existem diferentes modos de operação para a condução da fermentação. Tanto produtoras de aguardente com unidades industriais de maior porte quanto pequenas unidades produtoras de álcool utilizam o processo fermentativo de cortes em que o fermento é multiplicado em uma dorna. Esse método consiste em repartir o conteúdo da dorna quando aproximadamente metade do açúcar presente se transforma em álcool e ir completando seu volume com mosto (LOPES; GABRIEL; BORGES, 2011).

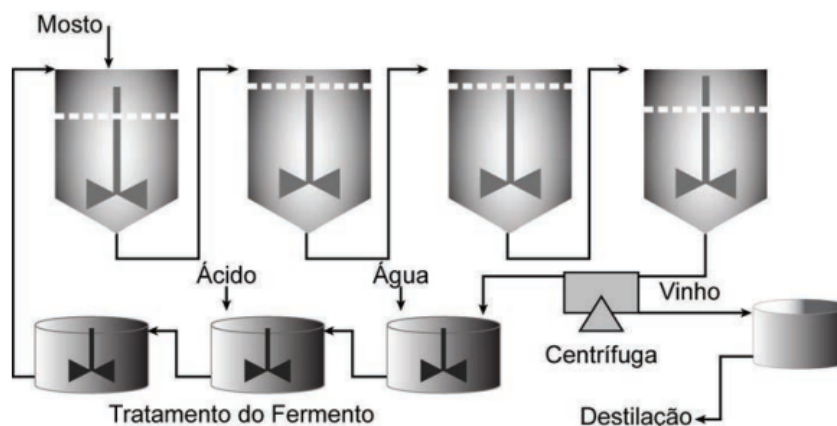
No caso das destilarias de álcool de porte comercial, o processo fermentativo utilizado é o de separação e reutilização do fermento, denominado processo Melle-Boinot. Tal processo pode ser conduzido de forma contínua, descontínua ou semicontínua. No Brasil, cerca de 80% das destilarias utilizam o processo semicontínuo, também denominado de batelada alimentada (BASSO; BASSO, 2018).

#### 2.1.5.1 Processos contínuos

O processo de fermentação contínua caracteriza-se por possuir uma alimentação contínua de meio de cultura a uma determinada vazão constante, sendo o volume de reação mantido constante através da retirada contínua de caldo fermentado. Isso faz com que o sistema opere em regime permanente, condição na qual as variáveis de estado (concentração de células, de substrato limitante e de produto) permanecem constantes ao longo do tempo de operação do sistema (SCHMIDELL et al., 2001). Os fermentadores podem ser operados de forma individual ou acoplados em série, podendo ter ou não a recirculação de

células de levedura. A fermentação Melle-Boinot pode ser realizada de forma contínua, em que o mosto é misturado ao fermento na primeira dorna e será reaproveitado para as demais dornas num processo contínuo até chegar à última dorna (SONEGO, 2016; LOPES; GABRIEL; BORGES, 2011). A Figura 4 apresenta um fluxograma do processo de fermentação contínua.

Figura 4 – Fluxograma do processo Melle-Boinot operando em fermentação contínua.



Adaptado de Lopes, Gabriel e Borges (2011)

Entre as grandes vantagens do modo contínuo de operação estão: redução de tempos mortos, levando a um aumento na produtividade; obtenção de um meio fermentado mais uniforme, o que facilita o processo de recuperação do produto; maior facilidade de utilização de controladores; redução de mão-de-obra empregada no processo. Já como desvantagens pode-se citar: maior possibilidade de ocorrência de contaminações, por se tratar de um sistema essencialmente aberto; dificuldades de operação em estado estacionário em determinadas situações (formação de espuma, crescimento do microrganismo nas paredes do reator, etc.) (SCHMIDELL et al., 2001).

### 2.1.5.2 Processos descontínuos

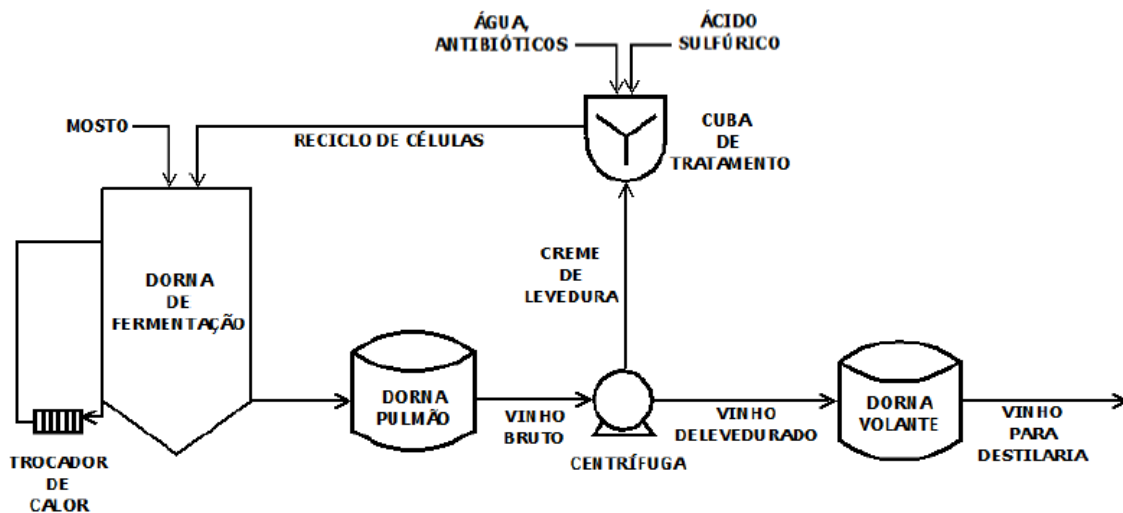
Em um processo descontínuo, também conhecido como em batelada, as etapas de enchimento do tanque de fermentação, limpeza, inoculação e operações de descarga para o meio fermentado são realizados de forma intermitente. Dessa forma, o inóculo (levedura) é colocado na cuba de fermentação, adiciona-se o mosto e, ao final da fermentação, o mosto fermentado (vinho) segue para a seção de separação (SANTOS; ALUÍZIO; CALDAS, 2015). O risco de contaminação nesse processo é menor devido à esterilização da dorna e à adição de novo inóculo a cada ciclo de fermentação. Entretanto, a elevada concentração de substrato na alimentação pode acarretar efeitos inibitórios ou formação de subprodutos indesejados, diminuindo assim, os índices de produtividade e rendimento da fermentação.

Ademais, para cada batelada, o reator tem que ser limpo, o mosto e o inóculo devem ser carregados ao sistema para dar início ao processo fermentativo, demandando tempo de operação (BORGES, 2008; MAIORELLA; WILKE; BLANCH, 1981).

### 2.1.5.3 Processos semicontínuos

No processo semicontínuo, também conhecido como em batelada alimentada, consiste na adição de um ou mais nutrientes no fermentador durante o cultivo e os produtos formados permanecem no fermentador até o final do processo. A fermentação Melle-Boinot semicontínua, caracterizada pelo reciclo de células, é a mais utilizada nas destilarias de álcool de porte comercial. Nesse processo, a alimentação dura cerca de 5h e ao término de cada fermentação etanólica (9 a 12h) o meio fermentado é separado por centrifugação em duas partes: vinho delevedurado e creme de levedura. O vinho delevedurado é enviado para destilaria e o creme de leveduras (suspensão concentrada em leveduras) é encaminhado para os pré-fermentadores para diluição e tratamento ácido. Por fim, o creme de levedura tratado é reciclado à dorna, a qual é alimentada por uma corrente de mosto iniciando um novo ciclo de fermentação (GOMES, 2020; SONEGO, 2016; LOPES; GABRIEL; BORGES, 2011). A etapa de fermentação etanólica para uma dorna operada em modo semicontínuo é ilustrada na Figura 5.

Figura 5 – Ilustração esquemática do processo Melle-Boinot simplificado aplicado para uma dorna de fermentação operada de forma semicontínua.



Adaptado de Gomes (2020)

Na batelada alimentada, pode-se realizar o controle da vazão de alimentação do substrato com a finalidade de controlar sua concentração no biorreator e deslocar o metabolismo microbiano para uma determinada via de interesse, na qual ocorre maior formação do produto de interesse. A vazão de alimentação pode ser constante ou variar com o

tempo e a adição de substrato pode ser de forma contínua ou intermitente (SONEGO, 2016; SCHMIDELL et al., 2001).

As principais vantagens associadas a esse processo são: altas eficiências de fermentação e produtividade, menor risco de inibição por substrato e produtos, menor tempo de fermentação, vinho com maior teor alcoólico, economia de açúcar, devido a menor reprodução celular, menor consumo de ácido sulfúrico e de antibióticos, etc (BASSO; BASSO, 2018; LOPES et al., 2016; SONEGO, 2016; LALUCE et al., 2016).

### 2.1.6 Influência da Temperatura na Fermentação Alcoólica

Como visto na Equação 2, o calor liberado na conversão da glicose ( $C_6H_{12}O_6$ ) em etanol ( $C_2H_5OH$ ) e dióxido de carbono ( $CO_2$ ) é de 25 kcal/mol de glicose consumida. Ademais, leveduras são microrganismos mesófilos, ou seja, desenvolvem-se melhor em ambientes com temperatura moderada. Logo, devido a essa característica exotérmica, a temperatura da fermentação deve ser controlada para garantir a produtividade esperada e evitar a proliferação de bactérias contaminantes, sendo que estas são favorecidas por temperaturas elevadas (SONEGO, 2016; SANTOS; ALUÍZIO; CALDAS, 2015).

A temperatura ótima para a fermentação com leveduras varia de 26 a 35°C. No entanto, a temperatura de fermentação em unidades industriais produtoras de etanol pode ultrapassar 40°C em casos de controle ineficiente da temperatura (SANTOS; ALUÍZIO; CALDAS, 2015). Em temperaturas elevadas, a levedura torna-se mais vulnerável à toxicidade exercida pelo etanol produzido, ocasionando queda na viabilidade celular (RIVERA et al., 2006).

O tempo de fermentação é dependente da temperatura de operação do sistema. Fermentações realizadas em temperaturas inferiores às utilizadas em processos industriais (abaixo de 30°C) aumentam o teor alcoólico do vinho produzido, além de diminuir o volume de vinhaça e o consumo de vapor na destilação (VELOSO et al., 2019). Esse aumento é decorrente da ampliação da resistência do microrganismo ao etanol, embora também retarde a cinética da reação, aumentando o tempo de fermentação, devido a alterações da célula que afetam a membrana celular e atividades metabólicas enzimáticas envolvendo reações bioquímicas (DEED; DEED; GARDNER, 2015). Já fermentações realizadas em temperaturas mais próximas de processos industriais bem controlados (30 a 35°C) ocorre diminuição do teor alcoólico e maior produção de vinhaça, mas menor tempo de fermentação (DIAS et al., 2012).

Na maioria das destilarias brasileiras, o resfriamento das dornas nas usinas é realizado utilizando-se trocadores de calor de placas externos às dornas, onde o vinho em fermentação é bombeado ao trocador, retornando em circuito fechado (PINTO, 2014). O sistema de resfriamento é, em geral, composto por torres de resfriamento que suprem utilidade fria a uma temperatura dependente da temperatura de bulbo úmido das condições climáticas locais (SONEGO, 2016; MAGAZONI, 2011).

### 2.1.7 Modelagem da Fermentação Alcoólica

O uso da modelagem matemática pode diminuir os custos do desenvolvimento do processo de fermentação alcoólica, pois reduz o trabalho experimental desnecessário ao permitir o estudo dos vários parâmetros do processo e as interações destes por meio de simulação. Além disso, fornece uma melhor compreensão do processo, o que é útil para uma etapa posterior de otimização e controle do processo fermentativo (SONEGO, 2016; RIVERA et al., 2006).

Segundo Schmidell et al. (2001), os modelos cinéticos utilizados na fermentação podem ser classificados quanto ao número de componentes usados na representação celular, estruturados e não estruturados, e quanto à heterogeneidade da população dos microrganismos, segregados e não segregados.

- **Estruturados:** as células são descritas em maiores detalhes, considerando, por exemplo, componentes intracelulares, permitindo descrever o estado das células, sua adaptação às mudanças do meio ambiente e as principais reações bioquímicas envolvidas. Requerem conhecimento sobre cinética de reação, termodinâmica, transporte e propriedades físicas;
- **Não estruturados:** o material celular é representado por uma única variável, por exemplo, a massa celular ou o número de células, sem considerar variações de componentes intracelulares, ou usar tais variações na previsão do comportamento cinético do processo. Baseam-se em balanços de massa e energia combinados com expressões para cinética;
- **Segregados:** as células são consideradas discretas, como indivíduos de uma população heterogênea, com distribuição de idade, de tamanho e de propriedades celulares;
- **Não segregados:** a população celular é considerada homogênea, isto é, todas as células apresentam o mesmo comportamento.

No geral, os modelos segregados e estruturados oferecem uma descrição mais detalhada do comportamento cinético do processo fermentativo que os não segregados e os não estruturados, mas à custa de maior complexidade e maior esforço computacional requerido. No caso de modelos segregados, características específicas do ciclo celular de um microrganismo podem ser incorporadas ao modelo cinético para validar e ampliar sua aplicação (SCHMIDELL et al., 2001; BAILEY; OLLIS, 1986).

Os modelos cinéticos podem ainda ser classificados em dois grupos: modelos sem inibição e modelos que consideram inibição por substrato, células e/ou produto. Alguns desses modelos, que também são não segregados e não estruturados, são apresentados na Tabela 1 (BADINO; CRUZ, 2012), em que  $\mu$  é a velocidade específica de crescimento celular ( $h^{-1}$ );  $\mu_{max}$  é a velocidade máxima específica de crescimento celular ( $h^{-1}$ );  $C_S$  é a

concentração de substrato na solução ( $g.L^{-1}$ );  $K_S$  é a constante de saturação ( $g.L^{-1}$ );  $u$  é uma constante cinética do modelo de MOSER (1958);  $C_X$  é a concentração de células na solução ( $g.L^{-1}$ );  $K_{SX}$  é uma constante cinética do modelo de CONTOIS (1959);  $K_{IS}$  é a constante de inibição pelo substrato ( $g.L^{-1}$ );  $v$  é uma constante do modelo de WU et al. (1988);  $K_{IP}$  é a constante de inibição pelo produto ( $g.L^{-1}$ );  $C_P$  é a concentração de produto na solução ( $g.L^{-1}$ );  $C_P^*$  é o valor de concentração de produto em que o crescimento cessa ( $g.L^{-1}$ );  $n$  é a constante do modelo de LEVENSPIEL (1980);  $C_X^*$  é o valor de concentração de células em que o crescimento cessa ( $g.L^{-1}$ );  $S_m$  é a constante de inibição do substrato ( $g.L^{-1}$ ).

Tabela 1 – Modelos cinéticos com e sem inibição.

Modelo	Inibição	Proposto por
$\mu = \mu_{max} \left( \frac{C_S}{K_S + C_S} \right)$	Não	MONOD (1942)
$\mu = \mu_{max} \left( \frac{C_S^u}{K_S + C_S^u} \right)$	Não	MOSER (1958)
$\mu = \mu_{max} \left( \frac{C_S}{K_{SX} \cdot C_X + C_S} \right)$	Não	CONTOIS (1959)
$\mu = \mu_{max} \left( \frac{C_S}{K_S + C_S + \frac{C_S^2}{K_{IS}}} \right)$	Substrato	ANDREWS (1968)
$\mu = \mu_{max} \left( \frac{C_S}{K_S + C_S + C_S \cdot \left( \frac{C_S}{K_{IS}} \right)^v} \right)$	Substrato	WU et al. (1988)
$\mu = \mu_{max} \left( \frac{C_S}{K_S + C_S} \right) \cdot e^{-(K_{IP} \cdot C_P)}$	Produto	AIBA et al. (1968)
$\mu = \mu_{max} \left( \frac{C_S}{K_S + C_S} \right) \cdot \left( 1 - \frac{C_P}{C_P^*} \right)^n$	Produto	LEVENSPIEL et al. (1980)
$\mu = \mu_{max} \left( \frac{C_S}{K_S + C_S} \right) \cdot \left( \frac{K_{IP}}{K_{IP} + C_P} \right)$	Produto	HOPPE & HANSFORD (1982)
$\mu = \mu_{max} \left( \frac{C_S}{K_S + C_S} \right) \cdot \left( 1 - \frac{C_X}{C_X^*} \right)^m$	Célula	LEE et al. (1983)
$\mu = \mu_{max} \left( \frac{C_S}{K_S + C_S + \frac{C_S^2}{K_{IS}}} \right) \cdot \left( 1 - \frac{C_P}{C_P^*} \right)^n$	Substrato Produto	ANDREWS (1968)/ LEVENSPIEL (1980)
$\mu = \mu_{max} \left( \frac{C_S}{K_S + C_S + \frac{C_S^2}{K_{IS}}} \right) \cdot \left( \frac{K_{IP}}{K_{IP} + C_P} \right)$	Substrato Produto	ANDREWS (1968)/ HOPPE & HANSFORD (1982)
$\mu = \mu_{max} \left( \frac{C_S}{K_S + C_S} \right) \cdot \exp \left( \frac{-C_S}{S_m} \right) \cdot \left( 1 - \frac{C_P}{C_P^*} \right)$	Substrato Produto	AIBA et al. (1968)/ TYAGI & GHOSE (1982)

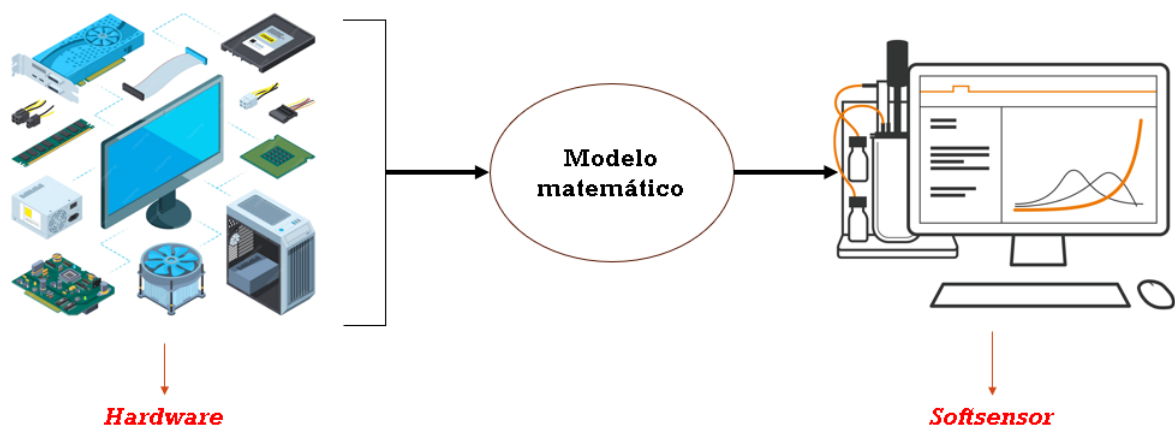


## 2.2 Sensor Virtual (*softsensor*) para fermentação

Os processos de fermentação têm ganhado cada vez mais sensores para monitoramento. O monitoramento tem sido feito com medidas *on-line* de variáveis físicas (pressão, velocidade do impedor, massa) e algumas variáveis físico-químicas (pH, potencial de redução), onde variáveis mais específicas como concentração de substrato e produto são dificilmente monitoradas. Um desafio com o monitoramento e controle da fermentação é que, para muitas variáveis importantes do processo, há poucos ou nenhum sensor adequado para medições *online* (Eliasson Lantz et al., 2010).

Dessa forma, sensores virtuais podem ser uma boa alternativa para melhorar o controle e monitoramento da fermentação. Sensores virtuais são modelos matemáticos que estimam, em tempo real, a variável desejada a partir de dados medidos da planta (LO-TUFO; GARCIA, 2008; FORTUNA et al., 2006). Eles se baseiam em usar informações *on-line* para estimar variáveis de um cultivo que seriam difíceis de medir diretamente. Apresentam dois componentes principais (Figura 6). O primeiro componente consiste nos sensores de *hardware* que irão fornecer informações das variáveis indiretas que apresentam relação com a variável de interesse. O segundo componente será um modelo ou uma correlação matemática que descreve a interdependência entre os dados dos sensores *in situ* com a variável-alvo (MESQUITA, 2017; ZHANG, 2009).

Figura 6 – Diagrama esquemático de um sensor virtual.



Muitos dos sensores por softwares propostos podem ser agrupados em três classes: (I) sensores baseados em modelos (precisam de um conhecimento fenomenológico prévio do processo); (II) sensores baseados em dados (usam informações de sistemas similares para inferir valores); (III) sensores híbridos (englobam recursos das classes I e II).

Os sensores baseados em dados (II) ou modelo empíricos utilizam um grande banco de dados de processo, não necessitando de um detalhamento extenso do sistema. Ademais, embora vários modelos mecanísticos tenham sido desenvolvidos para bioprocessos ao longo

dos anos, quando aplicados a sistemas altamente não lineares e multidimensionais, eles são frequentemente imprecisos por causa da rigidez das equações diferenciais. Dessa forma, modelos empíricos são uma boa alternativa para desenvolvimento de sensores para fermentação (ZHANG, 2009). Tais sensores podem ser divididos em (ZHANG, 2009):

- **Modelo com redes neurais artificiais:** redes neurais artificiais têm a capacidade de representar sistemas multivariáveis, reconhecer padrões e modelar estruturas, sendo muito utilizadas para modelar bioprocessos. A desvantagem das redes é o seu aspecto de "caixa preta", ou seja, é difícil interpretar e analisar o modelo criado;
- **Modelo com lógica *fuzzy* (nebulosa):** a lógica nebulosa é capaz de lidar com sistemas que não possuem ou requerem informações exatas e quantitativas, aproximando-se da linguagem humana. Assim, é adequada para bioprocessos, onde os dados reais são imprecisos e muitas vezes corrompidos por ruído, contaminantes;
- **Modelos estatísticos multivariados:** quando há compreensão limitada do processo, mas uma grande quantidade de dados de processo disponíveis de uma variedade de condições de operação, é possível aplicar técnicas de Controle Estatístico de Processo Multivariado, *Multivariate Statistical Process Control* (MSPC), a exemplo de PCA, *Partial Least Squares Regression* (PLSR), *Principal Component Regression* (PCR), e *Successive Projections Algorithm* (SPA).

Por último, existem os sensores baseados em modelos híbridos. Uma vez que tanto os modelos mecânicos quanto os empíricos têm certas desvantagens, modelos híbridos combinam as propriedades e características destes dois na fermentação. Os modelos híbridos combinam os benefícios da modelagem mecânica e empírica para produzir sistemas que, frequentemente, são mais precisos e extrapolam melhor as condições de processo (Eliasson Lantz et al., 2010; ZHANG, 2009).

## 2.3 Calibração Multivariada Clássica em Quimiometria

### 2.3.1 Fundamento teórico e técnicas

Grande parte dos métodos de calibração clássicos aplicados à análise espectrofotométrica tem como base a Lei de Lambert-Beer, que diz que a quantidade de luz absorvida é proporcional ao número de moléculas que a absorvem. Com base nisso, define-se a transmitância ( $Tr$ ) pela Equação 3:

$$Tr = \frac{I}{I_0} = e^{(-k \cdot d \cdot c)} \quad (3)$$

onde Transmitância ( $Tr$ ), Intensidade da radiação incidente ( $I_0$ ), Intensidade da radiação emitida ( $I$ ), constante de proporcionalidade ( $k$ ), caminho óptico ( $d$ ), concentração do analito ( $c$ ) são as variáveis envolvidas.

A versão linearizada da Equação 3 encontra-se abaixo:

$$Abs = -\log_{10}T = \log_{10} \left( \frac{I_0}{I} \right) = e \cdot d \cdot c \quad (4)$$

em que Absorbância ( $Abs$ ) e Absortividade molar ( $e$ ) são definidas.

A Equação 4 é conhecida por Lei de Lambert-Beer. A dependência linear entre concentração e absorbância de um analito para um determinado comprimento de onda é a base para métodos quantitativos.

O modelo de regressão por Mínimos Quadrados Clássico, CLS, pode ser utilizado para ajustar a Equação 4. Considerando o número de amostras do conjunto de calibração ( $n_{cal}$ ), número de comprimentos de onda discretos ( $n_{ond}$ ) e número de componentes químicos ( $n_c$ ), a representação matricial desta regressão pode ser escrita pela Equação 5 (HAALAND; THOMAS, 1988):

$$A = CK + E_A \quad (5)$$

onde a  $A$ ,  $n_{cal} \times n_{ond}$ , a  $C$ ,  $n_{cal} \times n_c$ , a matriz de parâmetros que representam o produto entre a absortividade específica de vezes o comprimento do caminho óptico de dado componente puro ( $K$ ),  $n_c \times n_{ond}$ , e a matriz dos erros espectrais não ajustados pelo modelo ( $E_A$ ),  $n_{cal} \times n_{ond}$ , são as variáveis envolvidas. Então,  $K$  é uma matriz de parâmetros ajustáveis, que pode ser calculada utilizando um método de mínimos quadrados clássico para minimizar  $\|E_A^T \times E_A\|$  (NELLES, 2020):

$$K_{opt} = (C^T C)^{-1} C^T A \quad (6)$$

A matriz de parâmetros otimizados para a regressão CLS ( $K_{opt}$ ), Equação 6, minimiza a variância do erro na absorbância,  $E_A$ . No entanto, sua maior desvantagem na predição de concentrações está na necessidade do conhecimento de todos os componentes químicos presentes (incluindo componentes interferentes) e a calibração destes na região espectral de interesse (RIBEIRO, 2007).

A previsão da concentração pode ser melhorada utilizando-se a Lei de Lambert-Beer Inversa, na qual, tem-se:

$$C = AP + E_C \quad (7)$$

em que a matriz dos erros de concentração não ajustados pelo modelo ( $E_C$ ),  $n_{cal} \times n_c$ , a matriz de coeficientes a ser ajustada pelo modelo inverso ( $P$ ),  $n_{ond} \times n_c$ , são as variáveis envolvidas. A matriz de parâmetros otimizados para o modelo inverso ( $P_{opt}$ ) minimiza o erro  $\|E_C^T \times E_C\|$  na previsão de  $C$ , assim:

$$P_{opt} = (A^T A)^{-1} A^T C \quad (8)$$

Este método é denominado Mínimos Quadrados Inverso, *Inverse Least Squares* (ILS), também conhecido como Regressão Linear Múltipla, *Multiple Linear Regression* (MLR). A maior vantagem desse método é que, se as colunas de  $E_C$  forem independentes, a análise de cada analito pode ser feita separadamente:

$$c = Ap + e_c \quad (9)$$

onde  $c$  é o vetor  $n_{cal} \times 1$  com concentrações de apenas um componente e  $e_c$ ,  $n_{ond} \times 1$ , o vetor de erro residual nos valores de concentração não ajustados pelo modelo. O vetor  $p$ ,  $n_{ond} \times 1$ , pode ser otimizado utilizando mínimos quadrados durante a calibração (Equação 10) e a concentração do componente de interesse será dada por (Equação 11):

$$p_{opt} = (A^T A)^{-1} A^T c \quad (10)$$

$$c_{pred} = ap_{opt} \quad (11)$$

onde  $c_{pred}$  é a concentração predita de um analito em uma amostra ( $1 \times 1$ ) e  $a$  é a o vetor de absorbâncias da amostra medido pelo instrumento ( $1 \times n_{ond}$ ).

Portanto, utilizando ILS, uma análise espectral qualitativa pode ser realizada, mesmo se a concentração de apenas um componente for conhecida na mistura de calibração. Entretanto, os componentes não incluídos na análise devem estar presentes e devem ser implicitamente modelados durante a calibração (HAALAND; THOMAS, 1988).

A matriz resultante  $A^T A$  tem dimensão  $n_{ond} \times n_{ond}$ . Se  $A$  tem posto matricial, *rank*, total  $n_{ond}$ , o que requer que  $n_{cal} \geq n_{ond}$ , é possível inverter a matriz  $A^T A$  para que  $P_{opt}$  seja ajustável. Entretanto, mesmo que  $n_{ond}$  de  $A$  não exceda  $n_{cal}$ , podem surgir problemas de multicolinearidade quando  $n_{ond}$  aumenta. A multicolinearidade faz com que  $A^T A$  seja singular ( $rank(A) \leq n_{cal}, n_{ond}$ ), ou seja, os componentes da matriz resultante podem formar um conjunto de dados linearmente dependentes (LD). Consequentemente,  $A^T A$  não admite inversa e, portanto, o parâmetro  $P_{opt}$  não pode ser ajustado pela regressão ILS (NELLES, 2020).

Diferentes metodologias de regressão multivariada foram propostas para tratar dos problemas de singularidade da matriz  $A^T A$ . A maior parte delas divide-se em duas grandes categorias (RIBEIRO, 2007; ARAÚJO et al., 2001). O primeiro grupo inclui os métodos que reduzem a singularidade da  $A^T A$ , escolhendo  $n_{ond}'$  comprimentos de onda específicos (colunas de  $A$ ) em vez de usar o espectro completo. A vantagem desta abordagem é que a matriz resultante  $n_{cal} \times n_{ond}'$  é também uma matriz de absorbâncias, não ocorrendo modificação do significado físico da Equação 7. 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. A abordagem deste grupo tem a vantagem de utilizar apenas parte da informação completa, ou seja, considera-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  $A$  (BRERETON, 2007).

O Algoritmo de Projeções Sucessivas, SPA, é um exemplo de método pertencente ao primeiro grupo. Este método se baseia na seleção progressiva, que consiste em recolher  $n$  variáveis minimamente redundantes da matriz  $A$ . A partir de uma coluna inicial ( $a_0$ ) de  $A$ , o algoritmo determina qual coluna,  $a_1$ , tem a maior projeção no subespaço,  $S_0$ , ortogonal à  $a_0$ . Na próxima iteração, SPA é restrito ao subespaço  $S_0$  (de dimensão  $n_{cal} - 1$ ). A projeção do  $a_1$  sobre  $S_0$  é a nova coluna inicial ( $a_0'$ ). As iterações são realizadas até  $n$  variáveis de  $A$  ( $a_{ns}$ ) sejam escolhidas. Então, ILS é aplicado usando uma nova matriz  $A'$ , construída com as  $n$  colunas escolhidas de  $A$ .

Há ainda o método de Regressão por Componentes Principais, PCR, que pertence ao segundo grupo. Nesta metodologia, aplica-se a Análise de Componentes Principais, PCA, na matriz  $A$ , em seguida utilizam-se as componentes de  $A$  como variáveis independentes no modelo ILS para a previsão da matriz  $C$ .

No PCA, a matriz  $A$  é projetada para um novo conjunto de eixos de coordenadas ortonormal ( $\alpha$ , autovetores de  $A^T A$ ), que são calculados a fim de maximizar a variância dos dados ( $A^T A$ ) ao longo dos eixos (ABDI, 2010). Então, encontra-se uma função linear que maximize a variância em  $A$ , expressada pela combinação linear de cada comprimento de onda ( $n_{ond}$ ). A Equação 12 expressa matematicamente a combinação linear entre variáveis independentes:

$$A = QY^T + E \quad (12)$$

onde a matriz de autovalores (*scores*) ( $Q$ ),  $n_{cal} \times p$ , a matriz de autovetores (*loadings*) ( $Y$ ),  $n_{ond} \times p$ , e o erro residual nos valores não ajustados pelo modelo PCA ( $E$ ),  $n_{cal} \times n_{ond}$ , são as variáveis envolvidas. As colunas da matriz  $Q$  também são chamadas de Componentes Principais, *Principal Components* (PC). Cada coluna de  $Q$  maximiza a variância de  $A$ , sendo que cada uma é linearmente independente e ortonormal à outra. É possível

determinar  $n_{ond}$  componentes principais, mas é desejável que haja a redução das variáveis independentes. Assim, a maior variância de  $A$  deve ser representada pelo número de componentes principais retidos no modelo PCA ( $p$ ), onde  $p \leq n_{ond}$  (JOLLIFFE, 2010).

O PCR proporciona modelos com previsões razoáveis de concentração, 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  $A$ , não considerando a estimativa ótima da matriz  $C$ . Ou seja, não há garantia de que as PC escolhidas para representar a absorvância ( $A$ ) serão significativos para a previsão da concentração ( $C$ ) (ABDI, 2010). O SPA também apresenta este inconveniente, no qual o critério relevante de cada entrada baseia-se apenas nos dados de entrada ( $A$ ). 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 (PINTO, 2017).

A solução deste problema está na utilização da Regressão por Mínimos Quadrados Parciais, PLSR. No PLSR, determina-se o conjunto de *scores* (conhecidas como Variáveis Latentes, *Latent Variables* (LV)) que permite a decomposição de  $A$  e  $C$  simultaneamente (Equação 13). Desta forma, ambas as matrizes  $A$  e  $C$  são projetadas em um novo conjunto comum às duas, a fim de maximizar a covariância  $A^T C$  (HAALAND; 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.

$$\begin{aligned} A &= TL_A^T + R_A \\ C &= UL_C^T + R_C \end{aligned} \quad (13)$$

As variáveis envolvidas são: número de variáveis latentes retidas no modelo PLS ( $h$ ); a matriz de *scores* (autovalores) de  $A$  ( $T$ ),  $n_{cal} \times h$ ; a matriz de *loadings* (autovetores) de  $A$  ( $L_A$ ),  $n_{ond} \times h$ ; o erro residual em  $A$  nos valores não ajustados pelo modelo PLS ( $R_A$ ),  $n_{cal} \times n_{ond}$ ; a matriz de *scores* (autovalores) de  $C$  ( $U$ ),  $n_{cal} \times h$ ; a matriz de *loadings* (autovetores) de  $C$  ( $L_C$ ),  $n_c \times h$ ; o erro residual em  $C$  nos valores não ajustados pelo modelo PLS ( $R_C$ ),  $n_{cal} \times n_c$ .

Deve-se determinar o número correto de LV ( $h \leq n_{ond}$ ) para que haja a redução do número de variáveis independentes e construção de um modelo de calibração robusto e eficaz. É possível criar modelos com mais de uma variável dependente ( $n_c > 1$ ), sendo que existe uma distinção em termos de algoritmo entre modelos PLS multivariado (*PLS2*) e modelos PLS univariado (*PLS1*). O algoritmo NIPALS-PLS é considerado o algoritmo mais robusto para calibrar modelos PLS em quimiometria, lidando tanto com casos univariados quanto com casos multivariados (ABDI, 2010).

Em decorrência destas peculiaridades, a técnica PLSR frente às demais técnicas é vantajosa, sendo o método mais reportado na literatura para a calibração multivariada

espectrométrica (JIANG et al., 2018; KHANI; RAHMANIAN; MOTLAGH, 2016; TAYLOR et al., 2009; WOLD; SJÖSTRÖM; ERIKSSON, 2001; HAALAND; THOMAS, 1988).

### 2.3.2 Detecção de *Outliers* (pontos anômalos)

*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 em modelos multivariados pode acarretar em diversos problemas, dentre eles a falta de acuracidade da calibração, ou seja, há grandes erros de validação do modelo ou este prevê concentrações sem significado físico (FERREIRA et al., 1999). Esses valores anômalos podem ser causados por problemas no instrumento de processo e em sua operação, e na preparação de amostras (LIU; CAI; SHAO, 2008). Dessa forma, a detecção de *outliers* tem sido um tema bastante explorado em estudos de calibração multivariada.

Um grande número de métodos tem sido proposto para detectar *outliers* em diferentes tipos de conjuntos de dados. Esses métodos são classificados em dois grupos: diagnóstico padrão e regressão robusta (LIANG; KVALHEIM, 1996). Para os métodos de diagnóstico padrão, são identificados e retirados os pontos discrepantes do conjunto de calibração e depois ajusta-se o restante dos dados com o modelo multivariado escolhido. Já métodos de regressão robusta são construídos de modo que forneçam resultados confiáveis mesmo com *outliers* presentes nos dados (WALCZAK; MASSART, 1998).

*Leverage*, *Studentized Residuals* (Resíduos de Student) são alguns exemplos de métodos de diagnóstico padrão para detecção de *outliers*.

#### 2.3.2.1 Resíduos de Student e *Leverage*

O *Leverage* pode ser interpretado geometricamente como a distância de uma amostra ao centroide do conjunto de dados, denotando sua influência no modelo de regressão (FERREIRA et al., 1999). Pode ser calculado por meio da Equação 14 para cada amostra do conjunto de calibração:

$$H_{cab} = \text{diag} \left( T \left( T^T T \right)^{-1} T^T \right) \quad (14)$$

onde o vetor que contém os valores de *Leverage* para as amostras de calibração ( $H_{cab}$ ),  $n_{cal} \times 1$ , está presente. É possível também determinar o valor de *Leverage* para amostras de teste, ou seja, amostras que não foram utilizadas no ajuste do modelo, como mostrado na Equação 15.

$$H_{test} = \text{diag} \left( T_{test} \left( T_{test}^T T_{test} \right)^{-1} T_{test}^T \right) \quad (15)$$

As variáveis envolvidas são: o número de amostras do conjunto de teste ( $n_{test}$ ); a matriz da resposta do instrumento (espectros de teste) ( $A_{test}$ ),  $n_{test} \times n_{ond}$ ; a matriz de *scores* (autovalores) de  $A_{test}$  ( $T_{test}$ ),  $n_{test} \times h$ ; a matriz de pesos de  $A$  ( $W$ ),  $n_{ond} \times h$ ; o vetor que contém os valores de *Leverage* para as amostras de teste ( $H_{test}$ ),  $n_{test} \times 1$ .

Um valor de *Leverage* pequeno indica que a amostra em questão influencia pouco na construção do modelo de calibração. Por outro lado, se as medidas experimentais de uma amostra são diferentes das outras do conjunto de calibração, ela provavelmente terá uma alta influência no modelo (FERREIRA et al., 1999). Para identificar as amostras anômalas, define-se o  $h_{crit}$ , como mostrado na Equação 16:

$$h_{crit} = 3 \cdot \frac{h}{n_{cal}} \quad (16)$$

onde é estabelecido que se alguma amostra de  $H_{cab}$  ou  $H_{test} > h_{crit}$ , ela é considerada suspeita e deve ser analisada.

Outra técnica interessante para identificação de possíveis *outliers* no conjunto de dados é por meio dos resíduos das concentrações, calculados por validação cruzada (tema tratado na seção "**Otimização da complexidade do modelo quimiométrico**") para o conjunto de calibração e pelo erro de predição para o conjunto de teste. Para obter a influência de cada amostra do conjunto de calibração, utiliza-se o Resíduo de Student da amostra  $i$  do conjunto de calibração ( $RS_{cab,i}$ ), como mostrado na Equação 17:

$$\begin{aligned} var_{cab} &= \sum_{i=1}^{n_{cal}} \frac{c_{cab,i} - c_{cab/pred,i}}{(n_{cal} - h)} \\ RS_{cab,i} &= \frac{c_{cab,i} - c_{cab/pred,i}}{\sqrt{var_{cab} \cdot (1 - H_{cab}(i))}} \end{aligned} \quad (17)$$

onde, as variáveis envolvidas são: a estimativa da variância dos resíduos para o conjunto de calibração ( $var_{cab}$ ); a concentração de referência da amostra  $i$  do conjunto de calibração ( $c_{cab,i}$ ); a concentração predita por validação cruzada da amostra  $i$  do conjunto de calibração ( $c_{cab/pred,i}$ ); o valor de *Leverage* da amostra  $i$  do conjunto de calibração ( $H_{cab}(i)$ ). Pode-se determinar também o Resíduo de Student da amostra  $i$  do conjunto de teste ( $RS_{test,i}$ ), como mostrado na Equação 18:

$$RS_{test,i} = \frac{c_{test,i} - c_{test/pred,i}}{\sqrt{var_{cab}}} \quad (18)$$

onde, as variáveis envolvidas são: a concentração de referência na predição da amostra  $i$  do conjunto de teste ( $c_{test,i}$ ); a concentração predita da amostra  $i$  do conjunto de teste ( $c_{test/pred,i}$ ).

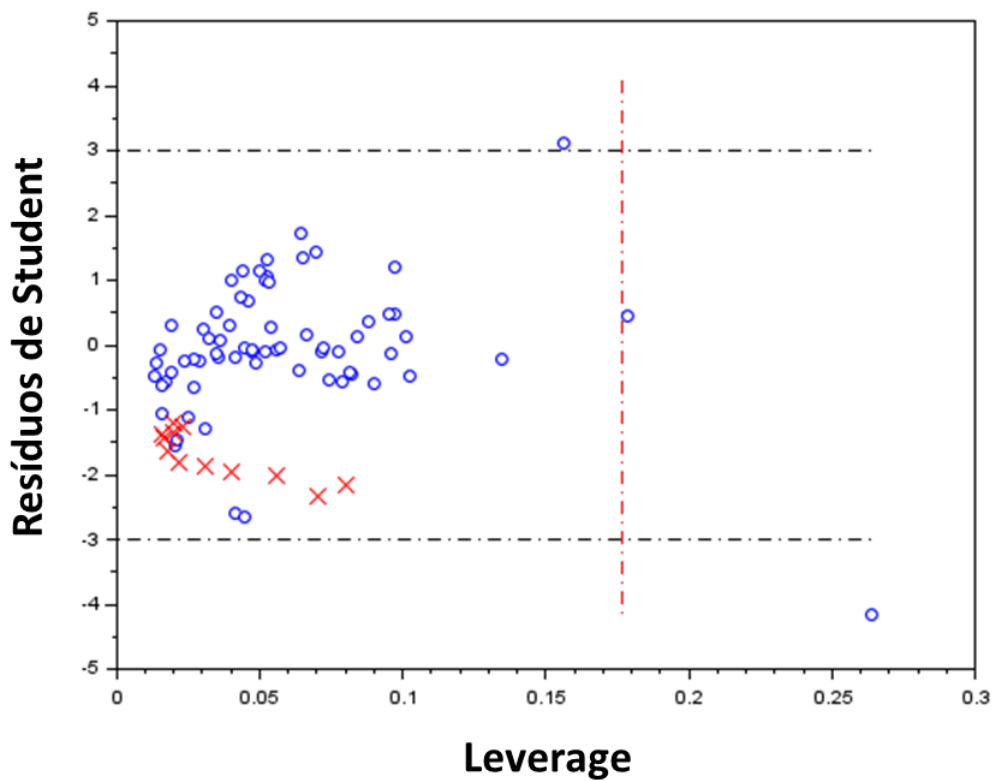
Supondo-se que os Resíduos de Student são normalmente distribuído, pode-se aplicar um teste  $t$  como indicativo, para determinar o intervalo de confiança a 95%. Como os



Resíduos de Student são definidos em unidades de desvio padrão do valor médio, os valores de  $RS_{test,i}$  ou  $RS_{cab,i} \pm 3$  são considerados suspeitos e devem ser analisados.

Dessa forma, analisa-se o gráfico de  $RS_{cab,i}$  e  $RS_{test,i}$  versus *Leverage* para verificar as amostras que destoam do conjunto de dados. Assim, amostras com altos resíduos e altos valores de *Leverage* devem ser excluídas do conjunto de dados e o modelo de calibração reconstruído (PINTO, 2017). A Figura 7 apresenta um exemplo de um gráfico de resíduos de Student versus *Leverage*.

Figura 7 – Exemplo de um gráfico de resíduos de Student versus *Leverage*: linhas pretas pontilhadas (limites para o Resíduo de Student), linha vermelha pontilhada (valor de  $h_{crit}$ ), círculos azuis (amostras de calibração), cruces vermelhas (amostras de teste).



### 2.3.3 Otimização da complexidade do modelo quimiométrico

A otimização da complexidade do modelo quimiométrico é um aspecto crucial no desenvolvimento de modelos robustos e eficazes para análise de dados químicos. Em quimiometria, a complexidade do modelo refere-se à quantidade de informação contida no modelo em relação ao número de variáveis independentes e ao número de parâmetros

do modelo. Uma complexidade excessiva pode levar a modelos superajustados, que não conseguem generalizar bem para novos dados, enquanto uma complexidade muito baixa pode resultar em modelos simplificados demais, incapazes de capturar a complexidade dos dados subjacentes.

Dessa forma, um aspecto complicado no uso de métodos de análise multivariada é a escolha do número de Variáveis Latentes (*Latent Variables*, LV). Para métodos exploratórios, como PCA, deseja-se determinar quantas LV's devem ser selecionadas para compor o modelo e quais contêm apenas ruído. Para métodos preditivos, como PLSR, deve-se garantir que toda a variabilidade de interesse é incluída para a previsão, sem introduzir variabilidade que levaria a uma redução na qualidade das previsões para amostras diferentes daquelas usadas na criação do modelo multivariado (RUTLEDGE; ROGER; LESNOFF, 2021).

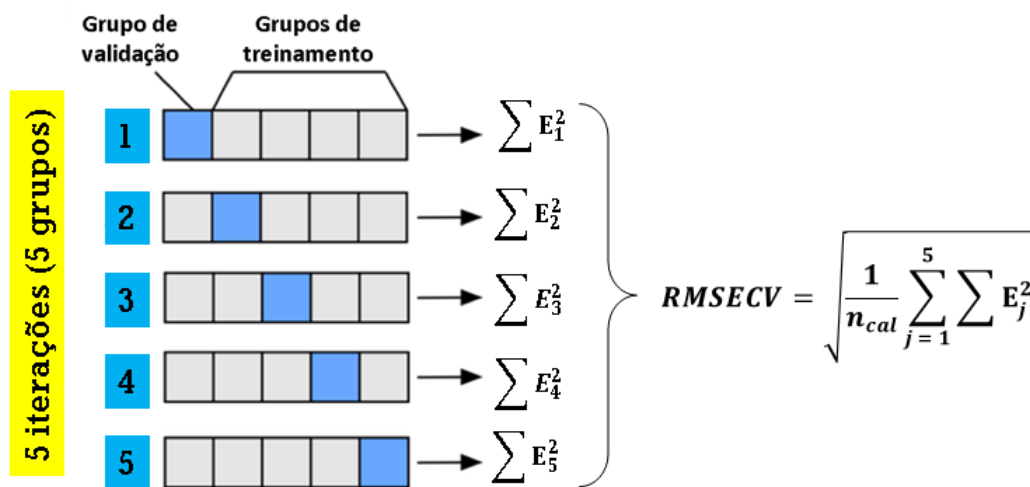
Diferentes métodos têm sido propostos para determinação do número de LV de um modelo multivariado. Esses métodos são classificados em duas categorias: análise do erro de predição e caracterização do modelo (RUTLEDGE; ROGER; LESNOFF, 2021). Para os métodos de análise do erro de predição, são utilizadas as predições de amostras que não foram incluídas para construção do modelo multivariado, como, por exemplo, um conjunto de amostras teste ou amostras removidas sequencialmente do conjunto de calibração para realização da **validação cruzada**. Antes de dar prosseguimento à descrição de alguns métodos de análise do erro de predição, será descrito o procedimento de validação cruzada.

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, 2020). Esta escolha é baseada na avaliação cuidadosa da magnitude dos erros de previsão de um dado modelo de calibração, visto que definir um número de regressores inadequado pode acarretar em *overfitting*, ou seja, os modelos são gerados com sobreajuste (FERREIRA et al., 1999). O princípio da validação cruzada *k-fold* (*k*-grupos) é apresentado na Figura 8. Primeiramente, o conjunto de calibração é dividido em *k* grupos (no caso da Figura 8, 5 partes). A cada iteração, retira-se um bloco de amostras (grupo de validação) do conjunto de calibração e o utiliza como conjunto de teste, enquanto os blocos restantes (grupo de treinamento) são usados para criação do modelo que é empregado para prever as concentrações do conjunto de teste. Os erros quadráticos de estimativa entre as concentrações de referência do grupo de validação e as concentrações preditas pelo modelo ajustado com o grupo de treinamento ( $E_1^2$ ) são calculados e somados. Esse procedimento é repetido *k* vezes, alternando o grupo de validação e determina-se então o valor da Raiz do Erro Quadrático Médio da Validação Cruzada (*RMSECV*), dado pela Equação 19:

$$RMSECV = \sqrt{\frac{1}{n_{cal}} \sum_{j=1}^k \sum E_j^2} \quad (19)$$

onde, as variáveis envolvidas são: o número de grupos na validação cruzada  $k$ -fold ( $k$ ), os erros quadráticos de estimativa entre as concentrações de referência do grupo de validação  $j$  e as concentrações previstas pelo modelo ajustado com o grupo de treinamento ( $E_j^2$ ).

Figura 8 – Princípio da validação cruzada  $k$ -fold com 5 grupos.



O  $RMSECV$  é uma aproximação do desvio padrão do erro de validação, podendo ser utilizado como uma aproximação do desvio padrão amostral (RIBEIRO, 2007). Embora este erro seja essencial à análise dos modelos, outras métricas podem ainda ser utilizadas, como a Raiz do Erro Quadrático Médio de Calibração, *Root Mean Square Error of calibration* ( $RMSEcal$ ), e a Raiz do Erro Quadrático Médio da Previsão, *Root Mean Square Error of Prediction* ( $RMSEP$ ), representadas nas Equações 20 e 21, respectivamente:

$$RMSEcal = \sqrt{\frac{1}{n_{cal}/trei} \sum_{j=1}^k \sum E_{trein,j}^2} \quad (20)$$

$$RMSEP = \sqrt{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (c_{test,i} - c_{test/pred,i})^2} \quad (21)$$

onde, as variáveis envolvidas são: o número de amostras do conjunto de calibração utilizadas alternadamente nos grupos de treinamento para validação cruzada ( $n_{cal}/trei$ ); os erros quadráticos de estimativa entre as concentrações de referência do grupo de treinamento  $j$

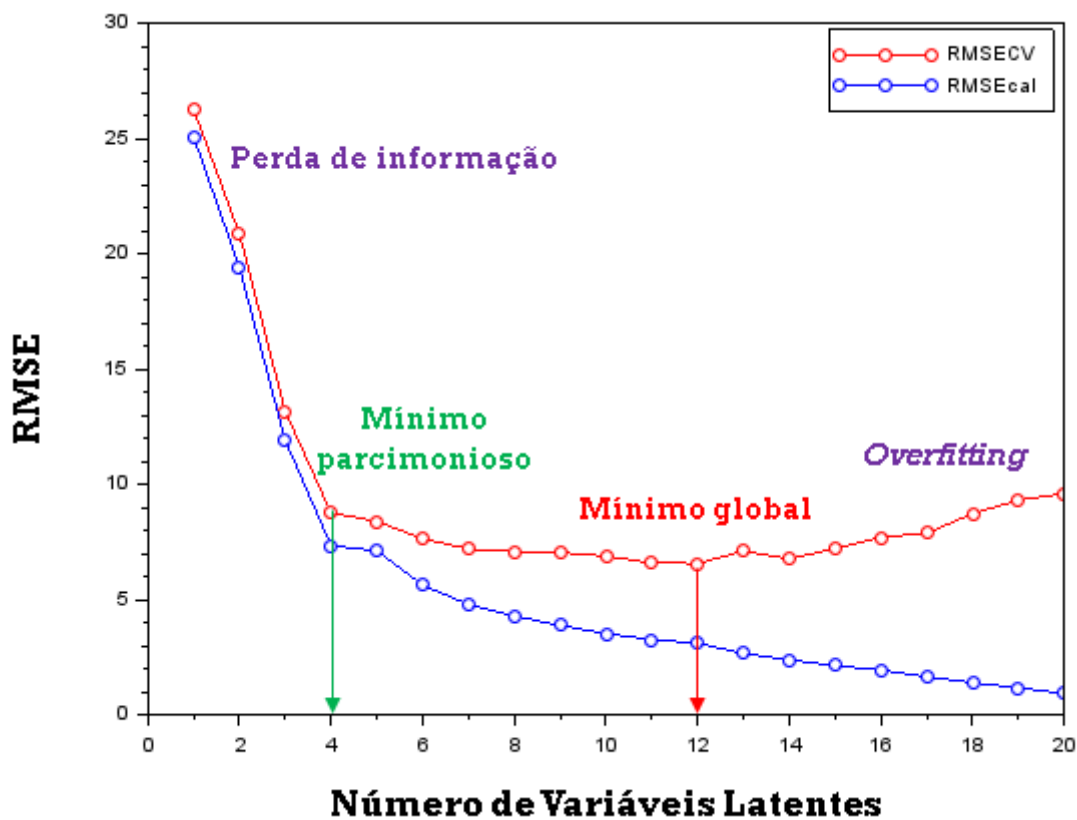
e as concentrações preditas pelo modelo ajustado com o grupo de treinamento ( $E_{trein,j}$ ). O  $RMSE_{cal}$  tem como objetivo estimar o desvio padrão do erro de ajuste,  $\alpha$ , quantificando os erros residuais para as amostras usadas na criação dos modelos. Já o  $RMSEP$  tem a função de realmente validar o modelo construído, analisando a acurácia na previsão das concentrações de amostras não relacionadas ao conjunto de calibração.

Tendo todos esses conceitos definidos, serão expostos alguns métodos de análise do erro de predição para otimização da complexidade, sendo eles:  $RMSECV$  e  $RMSE_{cal}$ , **Teste F de Osten**.

### 2.3.3.1 $RMSECV$ e $RMSE_{cal}$

A métrica  $RMSECV$  é uma aproximação do desvio padrão de validação. Dessa forma, plotando-se a evolução dos valores de  $RMSECV$  versus o número de Variáveis Latentes, é possível ver que geralmente há um valor de  $RMSECV$  mínimo. Esse valor corresponde a um equilíbrio entre informação e ruído (BISSETT, 2015), indicando o número otimizado de variáveis latentes a ser utilizado no modelo, como mostrado na Figura 9.

Figura 9 – Otimização do número de variáveis latentes pela minimização do  $RMSECV$ .



Ao observar a Figura 9, nota-se que a partir de 4 variáveis latentes, o valor de  $RMSECV$  não diminui mais tão intensamente (aproximadamente 78% do valor do  $RMSECV$  mínimo global). No entanto, o valor de  $RMSE_{cal}$  continua com uma tendência de queda,

aproximando-se de zero e distanciando-se do  $RMSECV$ . Isso é um indicativo de que o aumento do número de variáveis latentes no modelo não melhorará sua capacidade preditiva (NELLES, 2020).

Dessa forma, ao utilizar a métrica  $RMSECV$  para otimização da complexidade de um modelo quimiométrico, deve-se ter parcimônia na escolha do número de regressores. Embora o uso do  $RMSECV$  mínimo global seja recomendado, é importante avaliar a variação do  $RMSECV$  à medida que se aumenta o número de variáveis latentes, observando se há a melhora real da capacidade preditiva do modelo.

### 2.3.3.2 Teste F de Osten

A ideia principal para a aplicação de testes estatísticos para modelar a complexidade de modelos é assumir dois modelos com complexidade diferente e decidir se o modelo mais complexo faz uso significativo de seus parâmetros adicionais. O modelo simples tem  $n_{simples}$  parâmetros, e o segundo modelo, mais complexo possui  $n_{complexo} > n_{simples}$  parâmetros. Assume-se que o modelo complexo é idêntico ao modelo simples se os  $n_{complexo} - n_{simples}$  parâmetros adicionais são igualados a zero (NELLES, 2020). Assim, as seguintes hipóteses são formuladas, como mostrado na Equação 22:

$$\begin{aligned} H_0 : P_{adic} &= 0 \text{ (Modelo simples melhor)} \\ H_1 : P_{adic} &\neq 0 \text{ (Modelo complexo melhor)} \end{aligned} \quad (22)$$

onde as variáveis envolvidas são: o número de parâmetros do modelo simples ( $n_{simples}$ ); o número de parâmetros do modelo complexo ( $n_{complexo}$ ); a hipótese nula ( $H_0$ ); a hipótese alternativa ( $H_1$ ); vetor dos parâmetros adicionais no modelo complexo comparado com o modelo simples ( $P_{adic}$ ). A chamada hipótese nula  $H_0$  expressa o fato que os parâmetros adicionais são zero, ou seja, não possuem diferença estatisticamente significativa de zero. Se isso estiver correto, o modelo complexo não faz uso de seus parâmetros adicionais e, portanto, o modelo simples deve ser selecionado. Um teste estatístico decidirá se a hipótese nula deve ser aceita ou rejeitada.

O Teste F de Osten pode ser utilizado para comparar modelos quimiométricos PLS com diferentes números de LV (OSTEN, 1988). Utilizando a métrica  $RMSECV^2$ , que é a variância de previsão da validação cruzada, é possível comparar modelos. Tendo os valores de  $RMSECV$  calculados para diferentes números de LV, calcula-se o chamado  $F_{(k)}^{estat}$ , como mostrado na Equação 23:

$$F_{(k)}^{estat} = \frac{\left( \frac{RMSECV_k^2 - RMSECV_{k+1}^2}{n_{ond}} \right)}{\left( \frac{RMSECV_{k+1}^2}{n_{cal} \cdot n_{ond} - (k+1) \cdot n_{ond}} \right)} \quad (23)$$

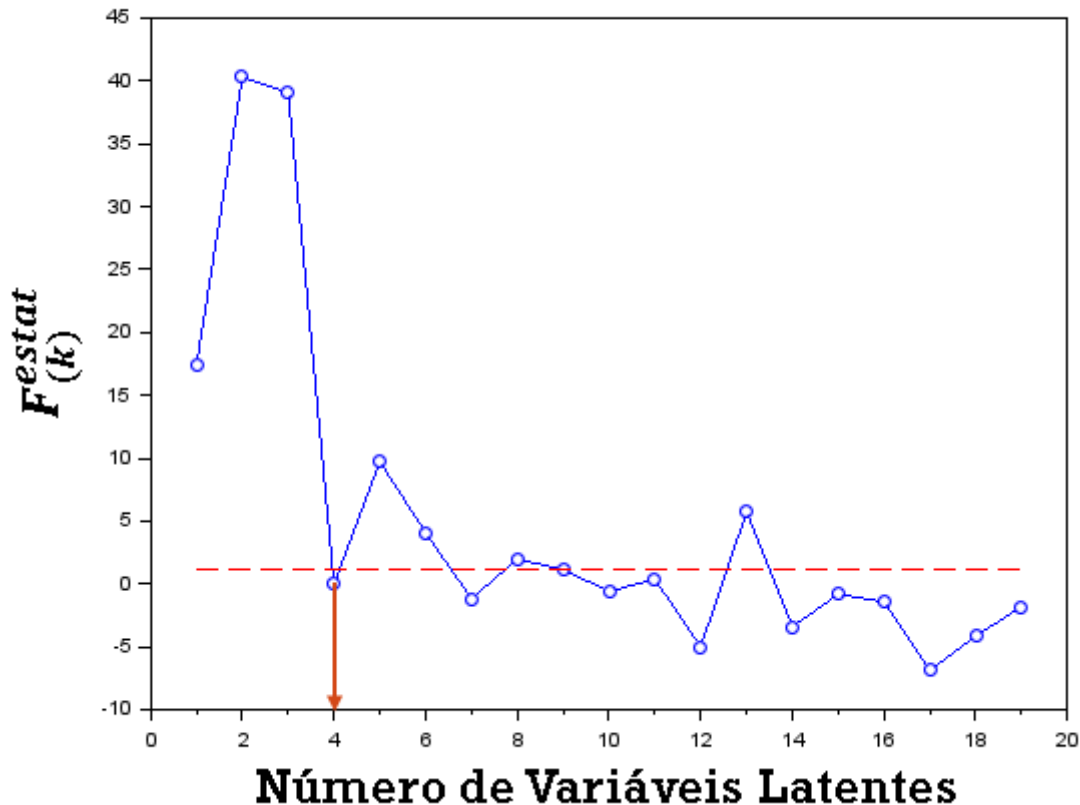
onde, as variáveis envolvidas são: a estatística de teste calculada a partir de um teste estatístico da distribuição F para  $k$  variáveis latentes (estatística de teste calculada a partir de um teste estatístico da distribuição F para  $k$  variáveis latentes ( $F_{(k)}^{estat}$ )); a Raiz do Erro Quadrático Médio da Validação Cruzada para  $k$  variáveis latentes, Root Mean Square Error of Cross Validation for  $k$  latent variables ( $RMSECV_k^2$ ); a Raiz do Erro Quadrático Médio da Validação Cruzada para  $k + 1$  variáveis latentes, Root Mean Square Error of Cross Validation for  $k + 1$  latent variables ( $RMSECV_{k+1}^2$ ).

A hipótese nula  $H_0$  é aceita com significância  $\alpha$  se (Equação 24):

$$F_{(k)}^{estat} \leq F_{\alpha}(n_{ond}, n_{cal} \cdot n_{ond} - (k + 1) \cdot n_{ond}) \quad (24)$$

onde  $F_{\alpha}(n_{ond}, n_{cal} \cdot n_{ond} - (k + 1) \cdot n_{ond})$  é a estatística de teste tabelada para uma distribuição F com significância  $\alpha$  e graus de liberdade  $n_{ond}$  e  $n_{cal} \cdot n_{ond} - (k + 1) \cdot n_{ond}$ . A Figura 10 apresenta o perfil de  $F_{(k)}^{estat}$  versus o número de Variáveis Latentes e o valor de  $F_{5\%}(n_{ond}, n_{cal} \cdot n_{ond} - (k + 1) \cdot n_{ond})$ .

Figura 10 – Otimização do número de variáveis latentes pelo Teste F de Osten: linha azul com círculo azul (valores de  $F_{(k)}^{estat}$ ), linha vermelha tracejada (valor de  $F_{5\%}$ ).



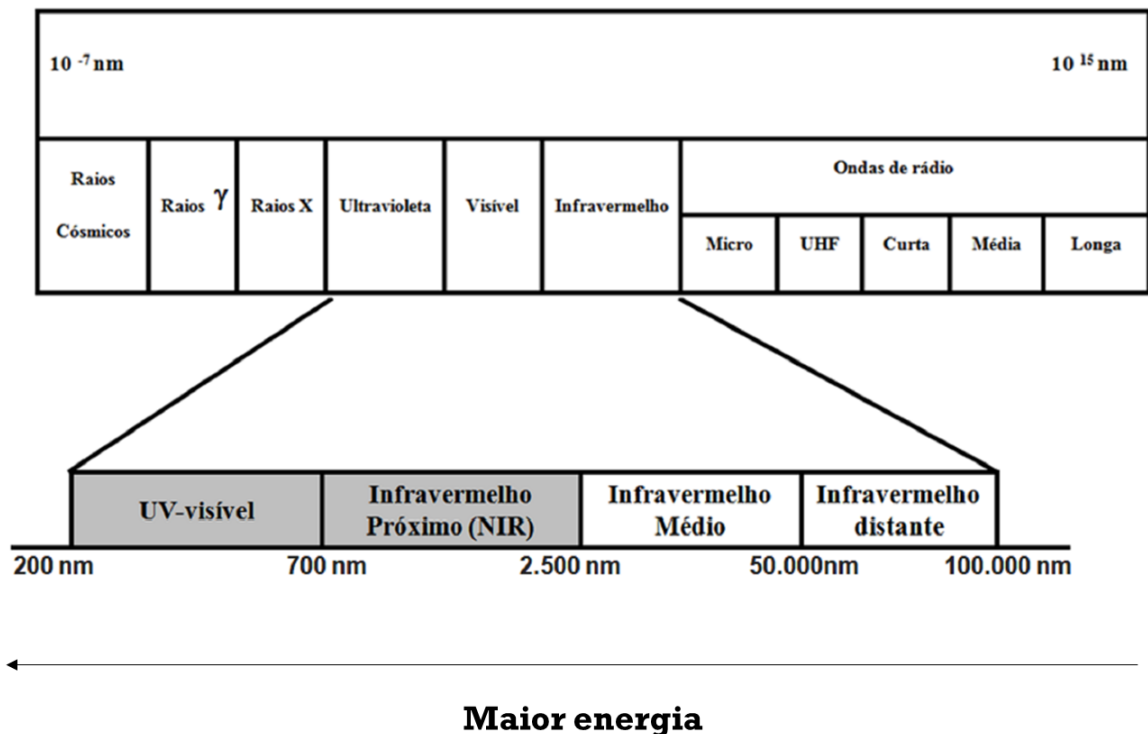
Ao observar a Figura 10, nota-se que a partir de 4 variáveis latentes,  $F^{estat} < F_{5\%}(n_{ond}, n_{cal} \cdot n_{ond} - (k + 1) \cdot n_{ond})$ , o que é indicativo de que não é estatisticamente significativo utilizar um modelo com 5 variáveis latentes em comparação com um modelo de 4 variáveis latentes, não rejeitando a hipótese nula.

## 2.4 Espectroscopia de Infravermelho Próximo (NIR)

### 2.4.1 Fundamentos

A região do infravermelho é dividida em três grupos principais, próximo (NIR), médio e distante, onde cada região é definida por um intervalo de comprimentos de ondas e possui características próprias. A região do infravermelho próximo corresponde ao intervalo de 13000 a 4000  $\text{cm}^{-1}$  (700 – 2500 nm). Associada a técnicas de quimiometria, como a calibração multivariada, a espectroscopia além de ser rápida e não destrutiva, possibilita determinar diversos componentes em meio reacional (PASQUINI, 2003). A Figura 11 apresenta as regiões de espectroscopia e comprimentos de onda característicos.

Figura 11 – Regiões de espectroscopia e comprimentos de onda característicos.



Adaptado de Pinto (2017)

O principal fundamento da espectroscopia no Infravermelho Próximo (NIR) está relacionado às ligações químicas e seus níveis de vibração, sendo assim uma espectroscopia

vibracional. Basicamente, ela atua na vibração molecular quantificando a razão entre a radiação incidente e a radiação absorvida em uma amostra, valores dependentes da estrutura eletrônica do material, ou seja, sua composição (PAULA, 2020; PINTO, 2017). A energia da radiação eletromagnética, transmitida pelo equipamento, é relacionada com a frequência de vibração dos átomos da amostra, que pode ser calculada pela equação de Bohr (Equação 25):

$$E = h \cdot \nu \quad (25)$$

Onde  $h$  é a constante de Planck ( $h = 6,626 \times 10^{-34}$  J.s) e  $\nu$  é a frequência de vibração.

Durante a vibração dos átomos, pode ocorrer a transição de elétrons entre os níveis de energia, que podem ser interpretados por modelos de oscilação harmônica (mudança entre níveis consecutivos) ou anarmônica (mudança entre níveis não adjacentes). Assim, surgem os sobretons, que são as bandas de absorção originadas nessa transição eletrônica (PAULA, 2020; PASQUINI, 2003). Também podem surgir as bandas de combinação que são originadas em moléculas que possuem dois ou mais tipos de vibração, conhecidas como poliatômicas (BLANCO; VILLARROYA, 2002).

Por conseguinte, os sobretons e as bandas de combinação, contidos no NIR, podem ser usados para quantificar a concentração de compostos orgânicos, pois é possível identificar ligações como O-H, C-H, C-O e N-H (ROGER; MALLET; MARINI, 2022; CEN; HE, 2007). A espectroscopia NIR tem o potencial e capacidade para ser utilizada na caracterização e no monitoramento da fermentação alcoólica.

## 2.4.2 Sensores Virtuais baseados em dados

Como já discutido, sensores virtuais são compostos por técnicas computacionais que proporcionam estimativas em tempo real para variáveis de processo que não podem ser medidas de maneira direta nestas condições por razões tecnológicas ou econômicas (KADLEC; GABRYS; STRANDT, 2009). Na literatura, há diversos exemplos de monitoramento de bioprocessos utilizando a espectroscopia NIR associada a modelos estatísticos multivariados.

### 2.4.2.1 Modelos estatísticos multivariados

Exemplos de aplicação de modelos estatísticos multivariados em sistemas fermentativos podem ser encontrados com frequência na literatura. Rodrigues et al. (2018) utilizaram técnica de espectroscopia no Infravermelho médio com transformada de Fourier (Fourier transform mid-infrared (FT-MIR)) associada com regressão de mínimos quadrados parciais para monitorar a fermentação de bioetanol, identificando os seguintes componentes: sacarose, glicose, frutose, etanol, glicerol, células. As dissertações de Pinto



(2017) e Paula (2020) abordaram o monitoramento de compostos presentes na produção de etanol de 2ª geração por meio de métodos espectrométricos, obtendo boa relação modelo - sistema. Svendsen (2016) em sua tese de doutorado estudou o uso de técnicas de espectroscopia do Infravermelho próximo e fluorescência para monitoramento da fermentação láctica, ressaltando sua importância para melhor entendimento do processo.

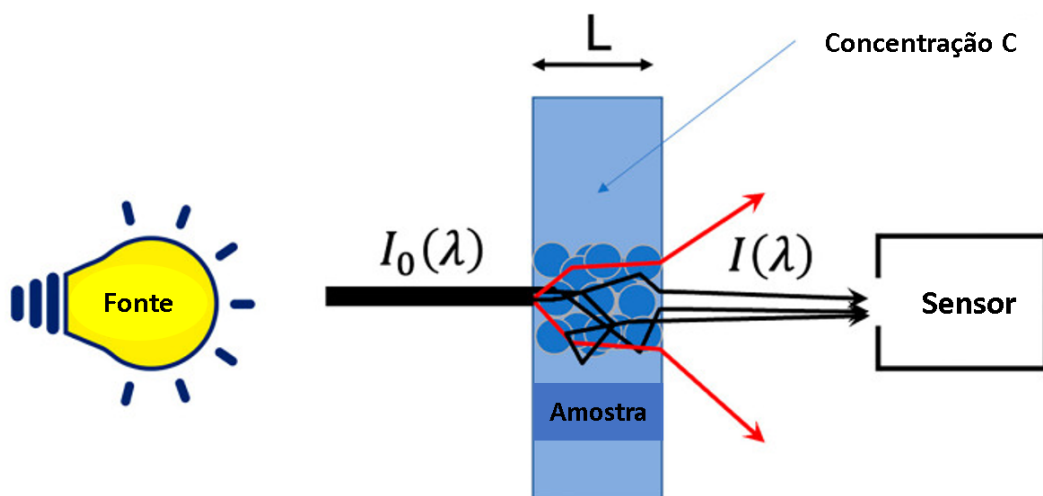
### 2.4.3 Pré-tratamento dos dados espectrométricos

A Lei de Lambert-Beer sugere uma relação linear entre a absorbância do espectro e a concentração dos constituintes do meio analisado. A validade dessa Lei pode ser questionada em sistemas reais de medição, a exemplo de monitoramento em tempo real, onde pode ocorrer dispersão de luz.

Esta dispersão tem duas consequências, como mostrado na Figura 12. A primeira é um alongamento do caminho óptico, que introduz um termo multiplicativo. A segunda é a perda de fótons, que serão contados erroneamente como uma absorção e, portanto, introduz um termo aditivo. A dispersão introduz ainda ruído estocástico na medição da absorbância, adicionando comportamento aleatório ao processo (ROGER; MALLET; MARINI, 2022).

Logo, os métodos de pré-tratamento são usados para remover erros aleatórios ou sistemáticos dos dados obtidos nas análises quimiométricas, possibilitando melhorar a inferência de modelos ajustados a partir desses dados. Eles podem ser divididos em duas categorias: métodos de correção de dispersão e derivadas espectrais (RINNAN; BERG; ENGELSEN, 2009).

Figura 12 – Medição real de absorção.



Adaptado de Roger, Mallet e Marini (2022)

Nos métodos de correção de dispersão, dois pré-tratamentos são mais conhecidos e utilizados: Correção de Dispersão Multiplicativa, *Multiplicative Scatter Correction* (MSC), e Variação Normal Padrão, *Standard Normal Variate* (SNV). Tais técnicas são projetadas para reduzir a variabilidade entre amostras devido à dispersão, além de ajustar mudanças de linha de base.

As derivadas espectrais têm a capacidade de remover tanto efeitos aditivos quanto efeitos multiplicativos nos espectros e têm sido usadas em espectroscopia analítica por décadas (RINNAN; BERG; ENGELSEN, 2009). É comum que se utilize suavização para não reduzir a relação sinal-ruído nos espectros derivados (RINNAN; BERG; ENGELSEN, 2009).

Dentre os métodos existentes, destaca-se que neste trabalho foram testados os métodos de média móvel, SNV, MSC e derivada de Savitzky-Golay (SG).

### 2.4.3.1 Média Móvel

A média móvel, designada na Equação 26 como  $\bar{a}$ , é utilizada para o alisamento dos dados espectrais. Isso faz com que as absorbâncias em um determinado intervalo sejam centralizadas, sendo conservada a variação dessas bandas. Esse método introduz um único parâmetro,  $m$ , que consiste no tamanho da janela móvel, definida como um ponto central e  $k$  pontos anteriores e posteriores, conforme a Equação 27 abaixo (ZAKAMULIN, 2017). A Figura 13 apresenta o efeito da média móvel com diferentes tamanhos de janela móvel em um espectro.

$$\bar{a} = \frac{x_{n-k} + \dots + x_n + \dots + x_{n+k}}{m} = \frac{1}{m} \sum_{i=-k}^k x_{n+i} \quad (26)$$

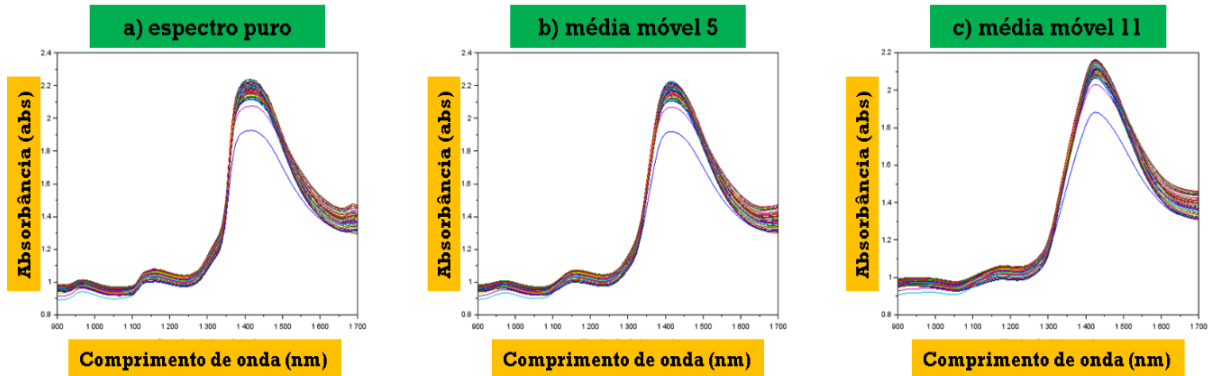
$$m = 2k + 1 \quad (27)$$

### 2.4.3.2 Derivadas

As derivadas são usadas para corrigir a linha de base de dados espectrométricos, removendo qualquer inclinação ou deslocamento. O efeito *offset* é um deslocamento da linha de base de uma quantidade constante de todo espectro. O princípio por trás do método é que os pontos de inflexão em picos próximos tornam-se pontos nas derivadas. Isso significa que os picos que se sobrepõem em um espectro normal têm maior probabilidade de serem resolvidos no espectro derivado correspondente. A segunda derivada pode ser utilizada para remover efeito de inclinação de picos sobrepostos (PAULA, 2020; SVENDSEN, 2016; BRERETON, 2007).

O método mais básico para derivação são as diferenças finitas: a primeira derivada é estimada como a diferença entre dois pontos de medição espectral subsequentes; a derivada

Figura 13 – Exemplos de espectro na região de comprimento de onda na faixa de 2245 - 2368 nm. (a): Espectros brutos, (b): espectros suavizados por uma média de janela móvel com largura 5, (c): espectros suavizados por uma média de janela móvel com largura 11.



de segunda ordem é então estimada calculando a diferença entre dois pontos sucessivos do espectros de derivada de primeira ordem:

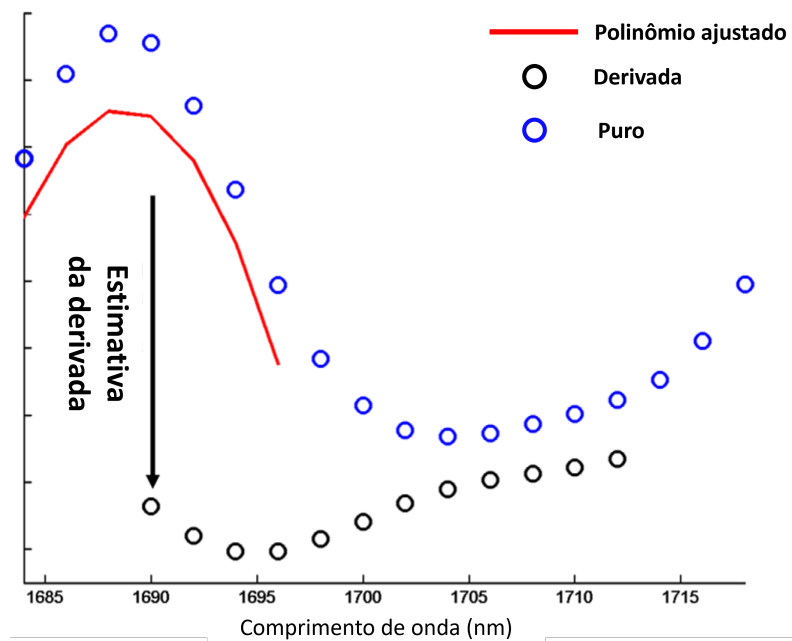
$$x'_i = x_i - x_{i-1} \quad (28)$$

$$x''_i = x'_i - x'_{i-1} = x'_{i-1} - 2 \cdot x_i + x_{i+1} \quad (29)$$

onde  $x'_i$  denota a primeira derivada e  $x''_i$  denota a segunda derivada no ponto  $i$  (comprimento de onda).

Este método é simples, mas, infelizmente, não é viável para medições mais reais devido à inflação de ruído, devendo ser evitado na prática (RINNAN; BERG; ENGELSEN, 2009). Em razão disso, para um bom uso de técnicas de derivação, é necessário utilizar suavização para não reduzir em excesso a relação sinal-ruído nos espectros corrigidos. Savitzky e Golay (1964) popularizaram um método para derivação numérica de um vetor que inclui uma etapa de suavização. Para encontrar a derivada no ponto central  $i$ , um polinômio é ajustado em uma janela simétrica com dados brutos (Figura 14). Quando os parâmetros para este polinômios são calculados, a derivada de qualquer ordem de esta função pode ser facilmente encontrada analiticamente e esse valor é subsequentemente usado como a estimativa da derivada para este ponto central. Esta operação é aplicada a todos os pontos do espectro sequencialmente. O número de pontos usados para calcular o polinômio (tamanho da janela) e o grau do polinômio ajustado são parâmetros de sintonia (RINNAN; BERG; ENGELSEN, 2009).

Figura 14 – Estimativa da primeira derivada por Savitzky-Golay. Uma janela de 7 pontos e um polinômio de segunda ordem são usados para suavização.



Adaptado de Rinnan, Berg e Engelsen (2009)

### 2.4.3.3 MSC

MSC é uma das técnicas mais utilizadas para espectroscopia NIR (RINNAN; BERG; ENGELSEN, 2009). Este pré-tratamento é capaz de minimizar deslocamentos da linha de base e do efeito multiplicativo. O conceito por trás do MSC é que imperfeições (por exemplo, efeito de dispersão indesejável) serão removidos da matriz de dados antes da modelagem dos dados. O MSC compreende duas etapas:

1. Estimativa dos coeficientes de correção (contribuições aditivas e multiplicativas)

$$x_{org} = b_0 + b_{ref,1} \cdot x_{ref} + e \quad (30)$$

2. Corrigindo o espectro

$$x_{corr} = \frac{x_{org} - b_0}{b_{ref,1}} = x_{ref} + \frac{e}{b_{ref,1}} \quad (31)$$

onde  $x_{org}$  é o espectro de amostra original medido pelo espectrômetro NIR,  $x_{ref}$  é o espectro de referência usado para pré-processamento de todo o conjunto de dados,  $e$  é a parte

não modelada do  $x_{org}$ ,  $x_{corr}$  são os espectros corrigidos e  $b_0$  e  $b_{ref,1}$  são parâmetros escalares, que diferem para cada amostra. A Figura 15 ilustra a interpretação dos parâmetros escalares.

Na maioria das aplicações, o espectro médio do conjunto de calibração é usado como espectro de referência, considerando que há uma dificuldade para determinação de um verdadeiro espectro de referência em medições NIR. Isso ocorre em razão dos sinais de diferentes componentes químicos serem fortemente sobrepostos e correlacionados, impedindo a a quantificação da linha de base (AFSETH; KOHLER, 2012).

Em muitos casos, a aproximação de cada espectro pelo seu espectro médio e uma linha de base constante não é suficiente. Logo, a Equação 30 deve ser expandida com mais termos, tornando-se a técnica conhecida como Correção de Dispersão Multiplicativa Estendida, EMSC. Esta expansão pode incluir ajuste polinomial de segunda ou maior ordem ao espectro de referência, ajuste de uma linha de base no eixo do comprimento de onda e uso de conhecimentos *a priori* a partir dos espectros de interesse ou espectrais interferentes. A Equação 32 apresenta a formulação desse pré-tratamento com a expansão:

$$x_{org} = \begin{bmatrix} 1 & x_{ref} & x_{ref}^2 & \lambda & \lambda^2 & x_{int,1} & x_{int,2} & \dots \end{bmatrix} \cdot b + e \quad (32)$$

onde  $\lambda$  é o vetor de correção para o eixo do comprimento de onda,  $x_{int,i}$  é a inclusão de conhecimento *a priori* de informações espectrais desejadas/indesejadas (por exemplo, o espectro de uma espécie interferente conhecida). A Equação 32 pode ser ainda expandida para incluir quaisquer outras correções apropriadas.  $b$  é um conjunto de escalares (coeficientes de correção) dados pela Equação 33:

$$b = [b_0 \quad b_{ref,1} \quad b_{ref,2} \quad b_{\lambda,1} \quad b_{\lambda,2} \quad b_{int,1} \quad b_{int,2} \quad \dots] \quad (33)$$

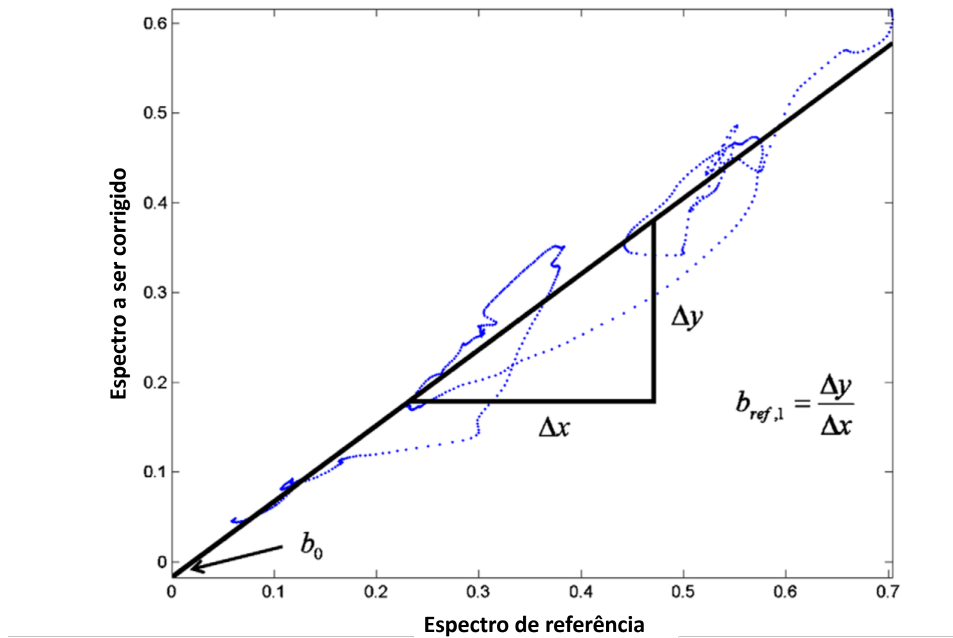
onde  $b_0$  é a correção de interceptação,  $b_{ref,i}$  é a correção de acordo com a  $i$ -ésima ordem da referência,  $b_{\lambda,i}$  é a correção do eixo do comprimento de onda da  $i$ -ésima ordem, e  $b_{int,i}$  é a correção da  $i$ -ésima informação conhecida.

#### 2.4.3.4 SNV

O pré-processamento SNV é provavelmente o segundo método mais aplicado para correção de dispersão de dados NIR (RINNAN; BERG; ENGELSEN, 2009; BARNES; DHANOA; LISTER, 1989). O formato básico para correção de SNV é o mesmo do MSC tradicional:

$$x_{corr} = \frac{x_{org} - a_0}{a_1} \quad (34)$$

Figura 15 – O espectro da amostra (pontos azuis) plotado em relação a um espectro de referência selecionado. Os termos de correção escalar são encontrados como a interceptação e a inclinação da linha preta, que é encontrada a partir do ajuste da regressão de mínimos quadrados em todos os pontos.



Adaptado de Rinnan, Berg e Engelsen (2009)

onde  $a_0$  é o valor médio do espectro amostral a ser corrigido e  $a_1$  é o desvio padrão do espectro amostral.

O fundamento de correção de sinal relacionado com SNV é o mesmo do MSC, exceto que não é necessário um sinal de referência. Em vez disso, cada observação é processada por si só, isolada do restante do conjunto. A falta de necessidade de uma referência comum pode ser uma vantagem prática (RINNAN; BERG; ENGELSEN, 2009).

---

## Capítulo 3

# Materiais e Métodos

---

### 3.1 Fermentações de *S. cerevisiae* em Biorreator

Nesta etapa, foram obtidos dados experimentais para mapear a cinética dos diferentes ensaios fermentativos realizados. As variáveis determinadas por amostragem durante os cultivos foram: concentração de células em g/L ( $C_x$ ), concentração de substrato, em g/L ( $C_s$ ), concentração de etanol em g/L ( $C_p$ ). *A supervision and control tool based on artificial intelligence for high cell density cultivations* (SUPERSYS-HCDC) (HORTA et al., 2014), uma ferramenta de supervisão e controle baseada em inteligência artificial para cultivo de alta densidade celular, desenvolvido *in-house* utilizando o *Labview*®, foi utilizado para monitoramento e controle durante as fermentações.

#### 3.1.1 Aparato Experimental e Testes Analíticos

O aparato experimental utilizado neste trabalho está disposto na Figura 16 e o seu diagrama esquemático do sistema experimental está contido na Figura 17. Todas as fermentações foram realizadas com controle de temperatura mantida a 32°C, por meio de manipulação da temperatura do banho termostaticado (equipamento 15, Figura 16; Figura 17) que circula pela camisa do biorreator (equipamento 9, Figura 16; Figura 17). A temperatura e o pH foram monitorados em tempo real pelo software SUPERSYS-HCDC, mas o controle do pH não foi ativado, para reproduzir as condições de fermentação empregadas nas usinas. O fornecimento do meio de alimentação para ensaios de batelada alimentada (equipamentos 4 e 6, Figura 16; Figura 17) foi realizado por meio de outra bomba peristáltica controlada pelo software SUPERSYS-HCDC com vazão variando entre 0,15 L/h e 0,5 L/h, de acordo com a estratégia de cultivo adotada.

Para promover a homogeneização da suspensão e reproduzir o sistema de recirculação do mosto empregado industrialmente, a mistura do meio reacional durante as fermentações foi realizada por meio do sistema de recirculação (itens 7 e 13 da Figura 16; Figura 17), acoplado ao biorreator. A vazão de recirculação foi mantida em 0,28 L/min para minimizar o tempo de residência do mosto no sistema de recirculação e garantir que o fluido passando pela janela de leitura do acessório NIR-Online X-Cell DN50 (equipamento 8, Figura 16; Figura 17) apresente a mesma composição do meio reacional em um dado instante. Além disso, o sistema de agitação por impelidores disponível no biorreator (Figura 17) foi utilizado em algumas fermentações para avaliar sua influência na mistura do sistema.

Utilizou-se um biorreator de 5 L (item 9 da Figura 16) do tipo tanque agitado para realização das fermentações (equipamento 9, Figura 16). Os métodos analíticos empregados para determinação das variáveis de interesse estão descritos abaixo.

- Concentração celular: foi determinada por leitura de Densidade ótica (DO) através de um espectrofotômetro no comprimento de onda de 600 nm;
- Concentração de substrato e produto: foram obtidas por HPLC utilizando coluna Sugar – Pak<sup>TM</sup> (300 x 6,5 mm) a 80 °C com água ultra pura como fase móvel (0,6 mL/min).

Durante as fermentações, amostras do caldo de cultivo foram retiradas para aferir as variáveis previstas a cada 30 minutos ou 1 hora. As leveduras, os procedimentos para preparo do pé-de-cuba e da suspensão foram gentilmente cedidos pela FERMENTEC LTDA (Piracicaba, Brasil).



Figura 16 – Aparato experimental completo instalado para conduzir as fermentações sob monitoramento dos Sistemas NIR-Online X-One (3, 12 e 14), incluindo ainda a estrutura de suporte (8), biorreator e camisa (9), sistema de recirculação de meio reacional (7 e 13), de fornecimento de meio de alimentação (4 e 6), de controle de temperatura (15) e computadores para monitoramento do processo (1, 3 e 5).

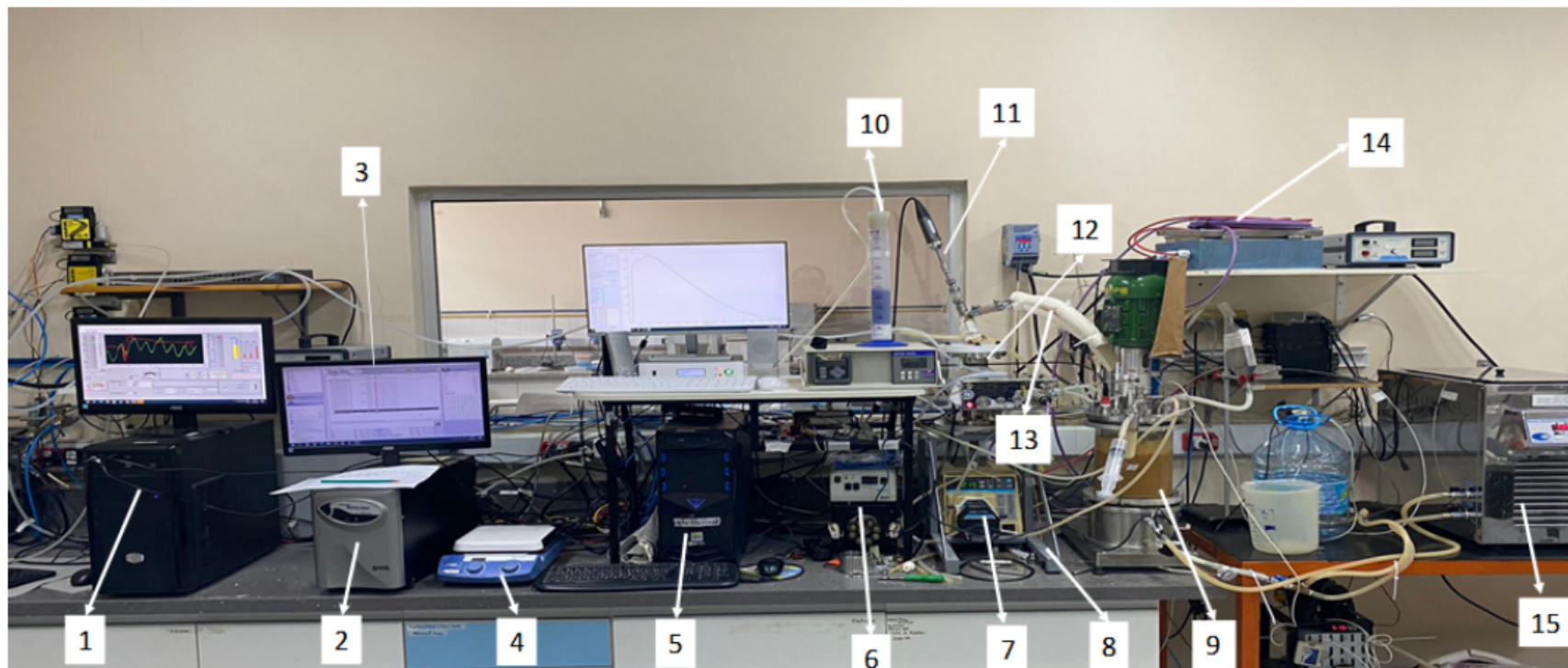
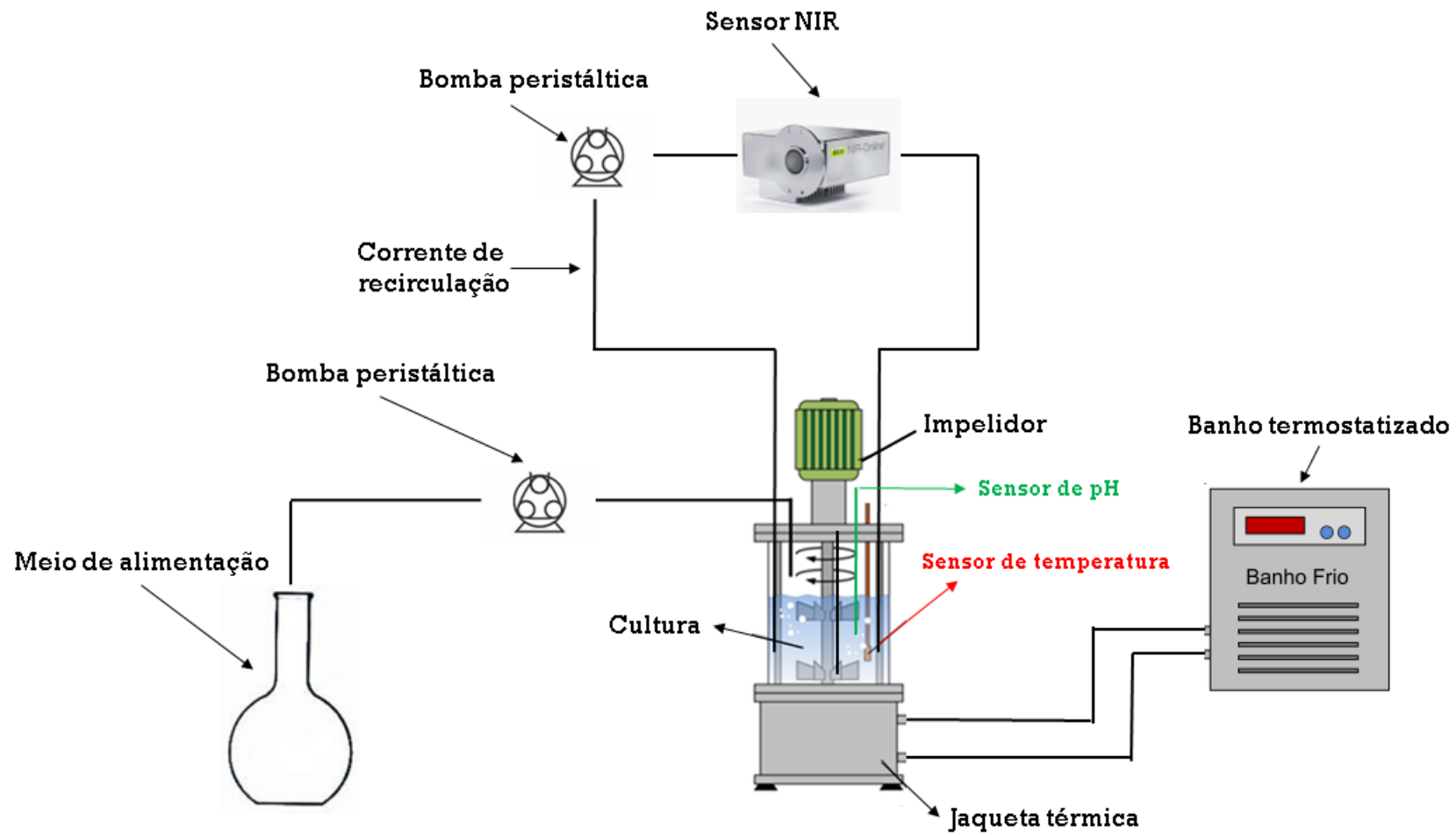


Figura 17 – Diagrama esquemático do sistema experimental utilizado para realização dos ensaios fermentativos.



### 3.1.2 Ensaios Fermentativos

Diversos ensaios fermentativos foram realizados com o objetivo de testar o desempenho do sensor virtual baseado em espectroscopia NIR em diferentes cenários presentes em uma fermentação industrial (Tabela 2). Cinco parâmetros foram testados: modo de operação, presença de agitação, levedura, nível de contaminação e meio de cultivo.

Dois modos de operação da fermentação foram testados: batelada (B) e batelada alimentada (BA). Como a batelada alimentada é o modo mais utilizado para produção industrial de etanol, as três primeiras fermentações BA (Ferm 1, Ferm 2 e Ferm 3) foram idealizadas para ampliar a faixa de valores assumidos pelas variáveis preditas. Dessa forma, a Ferm 1 operou com baixa concentração celular ( $< 5$  g/L), permitindo que ocorresse acúmulo de substrato. Já a Ferm 2 operou com uma concentração celular média ( $> 20$  g/L) e uma alimentação menos concentrada, diminuindo a concentração de substrato. A Ferm 3 compreendeu alta concentração celular ( $> 50$  g/L) e uma alimentação mais concentrada.

Após a realização das três primeiras fermentações, observou-se uma dificuldade na homogeneização do reator com apenas o sistema de recirculação. Logo, nos experimentos seguintes, foi ativado o sistema de agitação por impelidores disponível no biorreator.

Três leveduras *S. cerevisiae* foram utilizadas nas fermentações: Fermel® liofilizada (FL), FT2330L®, FT2858 L®. Estas últimas duas são leveduras floculantes presentes em meio industrial, responsáveis por problemas na centrifugação. O nível de contaminação foi menor nas fermentações em batelada alimentada e maior nas bateladas, como forma de testar a eficiência da calibração nesse meio.

O meio de cultivo formulado a partir de melaço (concentração de ART  $\approx 870$  g/L), cedido por uma usina parceira da FERMENTEC LTDA (Piracicaba, Brasil), foi diluído em água para correção da concentração de substrato e foi utilizado em todas as fermentações, exceto a Ferm 7. Neste ensaio, como forma de testar a robustez da inferência do sensor virtual em um meio diferente, utilizou-se meio formulado a partir de caldo de cana, adquirido de um vendedor local de São Carlos.

Tabela 2 – Diferentes cenários presentes nos ensaios fermentativos realizados.

ID	Modo de operação	Agitação	Levedura	Contaminação (bastonetes/mL)	Meio de cultivo
Ferm 1	BA	Não	FL	$\approx 10^7$	Melaço diluído
Ferm 2	BA	Não	FL	$\approx 10^7$	Melaço diluído
Ferm 3	BA	Não	FL	$\approx 10^7$	Melaço diluído
Ferm 4	B	Sim	FL	$\approx 10^9$	Melaço diluído
Ferm 5	B	Sim	FL/FT2330L/FT2858L	$\approx 10^{10}$	Melaço diluído
Ferm 6	B	Sim	FL	$\approx 10^{10}$	Caldo de cana

### 3.1.3 Uso do Modelo Matemático Fermentativo para suavização dos dados experimentais

O modelo fenomenológico utilizado para suavização dos dados experimentais é baseado no trabalho de Mesquita (2021) e está apresentado abaixo (Equações 35, 36, 37, 38, 39):

$$\frac{dV}{dt} = F \quad (35)$$

$$\frac{dC_x}{dt} = \left( \mu - \frac{F}{V} \right) \cdot C_x \quad (36)$$

$$\frac{dC_s}{dt} = (C_{sm} - C_s) \cdot \frac{F}{V} - \frac{1}{Y_{x/s}} \cdot \mu \cdot C_x \quad (37)$$

$$\frac{dC_p}{dt} = \frac{Y_{P/S}}{Y_{X/S}} \cdot \mu \cdot C_x - \frac{F}{V} \cdot C_p \quad (38)$$

$$\mu = \frac{\mu_{max} \cdot C_s}{k_s + C_s} \cdot \underbrace{\exp\left(-\frac{C_s}{S_m}\right)}_{\text{termo I}} \underbrace{\left(1 - \frac{C_p}{P_m}\right)}_{\text{termo II}} \quad (39)$$

onde  $V$  é o volume total (L),  $F$  é a vazão de alimentação (L/h),  $C_x$  é a concentração de células (g/L),  $\mu$  é a velocidade específica de crescimento ( $\text{h}^{-1}$ ),  $C_s$  é a concentração de substrato (g/L),  $C_{sm}$  é concentração de substrato no mosto (g/L),  $Y_{x/s}$  é o rendimento de células em relação ao substrato,  $C_p$  é a concentração de produto (g/L),  $Y_{p/s}$  é o rendimento do produto em relação ao substrato,  $\mu_{max}$  é constante de crescimento máxima ( $\text{h}^{-1}$ ),  $k_s$  é a constante de saturação (g/L),  $S_m$  é a constante de inibição do substrato (g/L) e  $P_m$  é a constante de inibição do produto (g/L).

Esse modelo cinético assume como hipóteses: não segregação, (população celular homogênea), não estrutura (sem variações de componentes intracelulares), não há morte celular, não há desvio de substrato para manutenção celular, a produção de etanol está associada ao crescimento, há inibição por substrato (termo I) e há inibição por etanol (termo II).

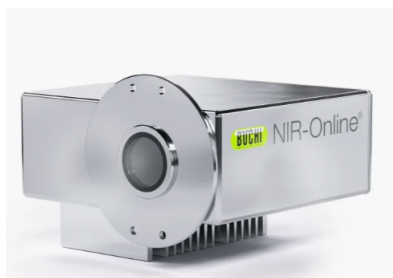
Os dados experimentais foram suavizados a partir de um ajuste matemático ao modelo descrito acima realizado separadamente para cada componente em cada fermentação, o que consiste em minimizar a função objetivo apenas com os dados de um dos componentes. Estes novos dados suavizados foram utilizados como referência das variáveis preditas na calibração do sensor virtual baseado em espectroscopia NIR. O *software* utilizado para ajuste cinético foi o MATLAB<sup>®</sup> (2022a), utilizando o método Levenberg-Marquardt.

## 3.2 Monitoramento do processo com NIR

Medições de espectro foram feitas com o equipamento NIR-Online X-One (Buchi Brasil LTDA) (Figura 18)<sup>1</sup>. O espectrofotômetro NIR foi acoplado ao sistema de recirculação implementado no biorreator (Figura 19), permitindo que se coletassem espectros em tempo real durante o decorrer das fermentações. O sensor utiliza uma lâmpada para produzir radiação infravermelha próxima que interage com as moléculas da amostra. A luz refletida da amostra é coletada através de um conjunto de fibras ópticas que direcionam a luz para o espectrofotômetro NIR. O espectrofotômetro NIR consiste em uma rede de difração para dispersar espacialmente a luz de acordo com o comprimento de onda e uma matriz de fotodiodos de índio-gálio-arsenieto contendo múltiplos elementos.

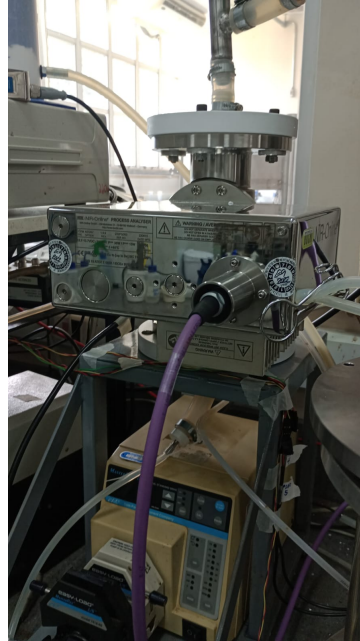
O intervalo de comprimento de onda utilizado foi de 900 a 1700 nm com resolução de 5 nm. Realizaram-se varreduras a cada 10 segundos e espectros de amostra a cada 10 minutos. Espectros de amostra consistem em calcular o espectro médio a partir de diversas varreduras em um determinado intervalo de tempo. Por exemplo, durante a coleta de amostra de caldo de cultivo, selecionava-se a opção de coleta de espectro de amostra do software do sensor NIR, determinando-se um espectro médio associado para esta amostra. Esta opção foi utilizada também para determinar espectros de amostra mesmo sem a retirada de amostra, como forma de expandir o número de espectros para análises futuras.

Figura 18 – Espectrofotômetro NIR-Online X-One.



<sup>1</sup> <https://www.buchi.com/pt/produtos/instrumentos/nir-online-process-analyzer>

Figura 19 – Espectrofotômetro NIR-Online X-One montado no sistema experimental.



### 3.3 Tratamento Matemático dos dados

Os procedimentos de refinamento dos dados, as análises estatísticas e o ajuste do modelo necessários neste desenvolvimento foram implementados no software Scilab® (Scilab 6.1.1).

#### 3.3.1 Pré-processamento dos Espectros

Os espectros obtidos a partir dos ensaios fermentativos foram pré-processados sendo que, para este estudo, foi considerado a técnica EMSC. Aproximou-se cada espectro pelo seu espectro médio, uma linha de base e termos linear e quadrático no eixo do comprimento de onda, como mostrado nas Equações 40 e 41:

$$x_{org} = \begin{bmatrix} 1 & x_{cab,med} & \lambda & \lambda^2 \end{bmatrix} \cdot b + e \quad (40)$$

$$b = [b_0 \quad b_{cab,med} \quad b_{\lambda,1} \quad b_{\lambda,2}] \quad (41)$$

onde  $x_{cab,med}$  é o espectro médio do conjunto de calibração. Para cada espectro, realizou-se o ajuste da equação 40 e corrigiu-se o espectro por meio da Equação 42:

$$x_{corr} = \frac{x_{org} - b_0 - b_{\lambda,1} \cdot \lambda - b_{\lambda,2} \cdot \lambda^2}{b_{ref,1}} \quad (42)$$

Este pré-tratamento é extremamente utilizado em espectroscopia NIR por causa da sua capacidade de minimizar deslocamentos da linha de bases, efeito multiplicativo, variações aleatórias, ressaltando as informações relevantes do espectro.

### 3.3.2 Calibração Multivariada e Otimização da Complexidade do modelo

Neste trabalho, a regressão PLS foi utilizada como modelo para o sensor virtual. Os dados espectroscópicos foram correlacionados com os dados de referência das variáveis preditas. Dois tipos de dados de referência foram utilizados para construção de modelos: dados experimentais e dados experimentais suavizados por modelo cinético fermentativo.

Para otimização da complexidade de modelo, utilizou-se o método de validação cruzada *k-fold* com 5 grupos e calculou-se o desvio padrão do erro de validação cruzada para diferentes números de LV, ou seja, o  $RMSECV_k^2$ . Tendo estes valores calculados, determinou-se o melhor número de variáveis latentes pela comparação das técnicas de otimização  $RMSECV$  e  $RMSEcal$  e Teste F de Osten.



---

## Capítulo 4

# Resultados e Discussão

---

A Tabela 3 apresenta as condições empregadas nas fermentações realizadas. Destaca-se a heterogeneidade de faixas de concentrações de células e de ART do meio de alimentação para as Ferms 1, 2 e 3, o que foi feito propositalmente para ampliar a faixa de valores das variáveis preditas.

As Figuras 20 a 25 apresentam o resultado da suavização dos dados experimentais para cada ensaio fermentativo e cada variável predita ( $C_x$ ,  $C_p$  e  $C_s$ ). Para realizar as suavizações, foi utilizado o modelo fenomenológico descrito anteriormente, no qual se aplicou mínimos quadrados ponderados separadamente para cada componente em cada fermentação para o ajuste do modelo aos dados experimentais. Observa-se que para as Ferms 1 a 3, Figuras 20 a 22, há um desvio grande dos dados experimentais em relação aos dados suavizados. Isso foi atribuído à uma possível falta de homogeneização do reator durante os ensaios, o que foi corrigido nos experimentos seguintes com a ativação do sistema de agitação por impelidor do biorreator.

Os espectros obtidos na região do NIR para as fermentações encontram-se na Figura 26. É possível notar que todos os espectros possuem uma forma similar, com um grande pico ao redor de 1400 nm correspondente ao primeiro sobretom vibracional da água. Este comportamento era esperado e se deve à alta concentração de água no caldo de fermentação. Outra característica importante nos espectros obtidos durante as fermentações está na similaridade dos espectros em uma mesma fermentação. Embora haja variações no espectro dentro de uma mesma fermentação, essas variações, em algumas regiões do infravermelho, são menores que a diferença nos espectros de fermentações diferentes. Isso é resultado das grandes diferenças entre os ensaios.

Tabela 3 – Condições e parâmetros dos ensaios fermentativos realizados

ID	Volume (pé de cuba - L)	Volume final (L)	Conc. celular (pé de cuba - g/L)	Conc. de ART no meio de alimentação (g/L)
Ferm 1	1,5	4,5	4,2	392,8
Ferm 2	1,8	5,6	23,9	273,6
Ferm 3	1,67	5,0	57,1	386,2
Ferm 4	5,0	5,0	13,4	-
Ferm 5	5,0	5,0	12,6	-
Ferm 6	5,0	5,0	18,5	-

Figura 20 – Suavização dos dados experimentais das **Ferm 1** utilizando o modelo cinético fermentativo.

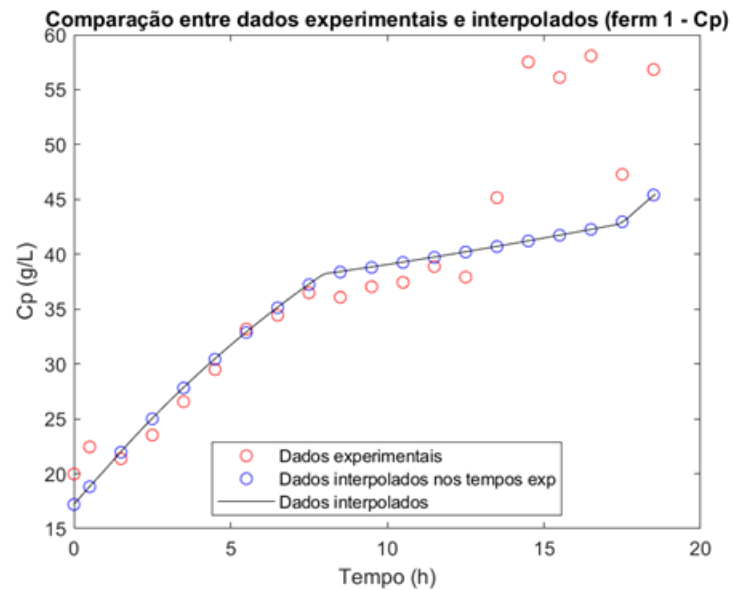
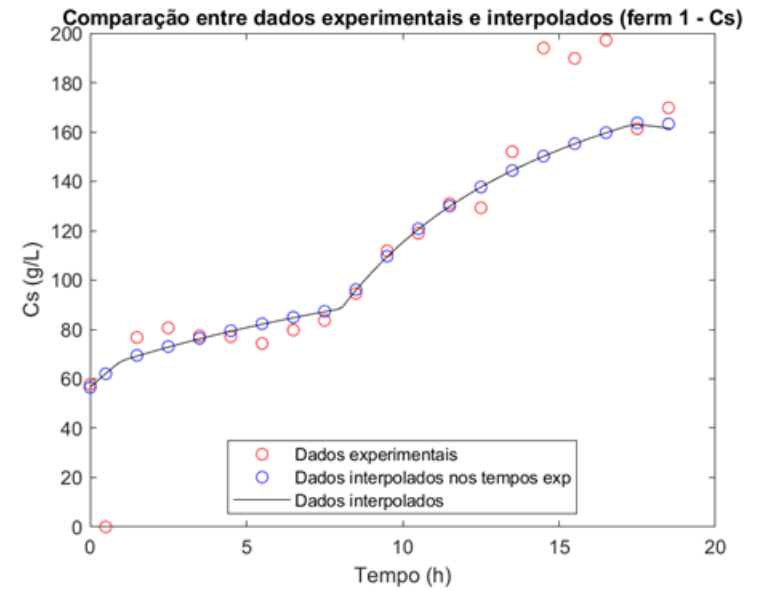
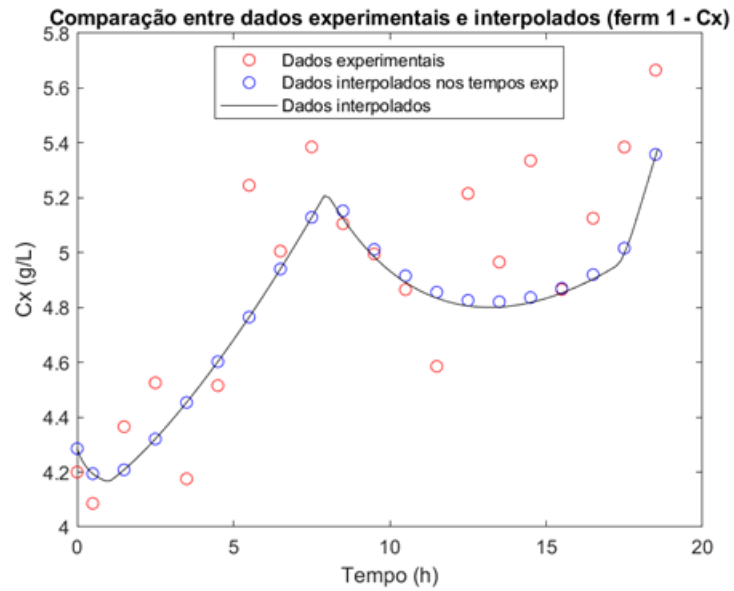


Figura 21 – Suavização dos dados experimentais das **Ferm 2** utilizando o modelo cinético fermentativo.

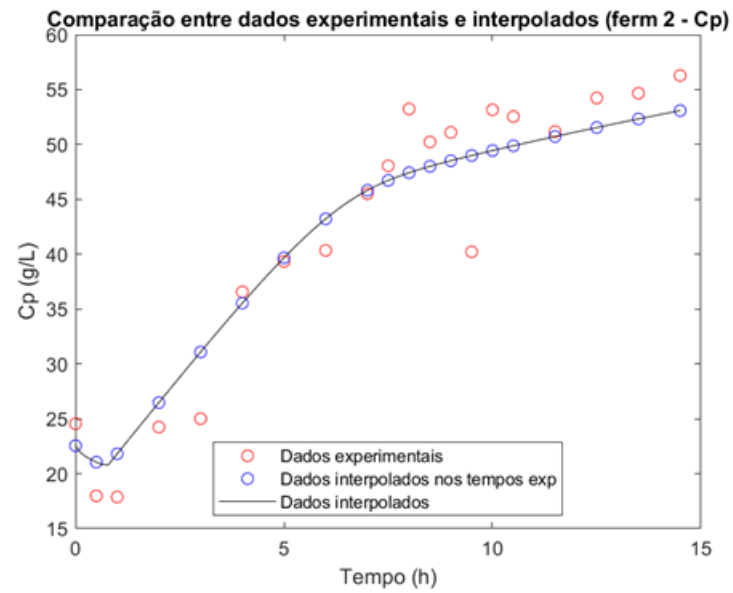
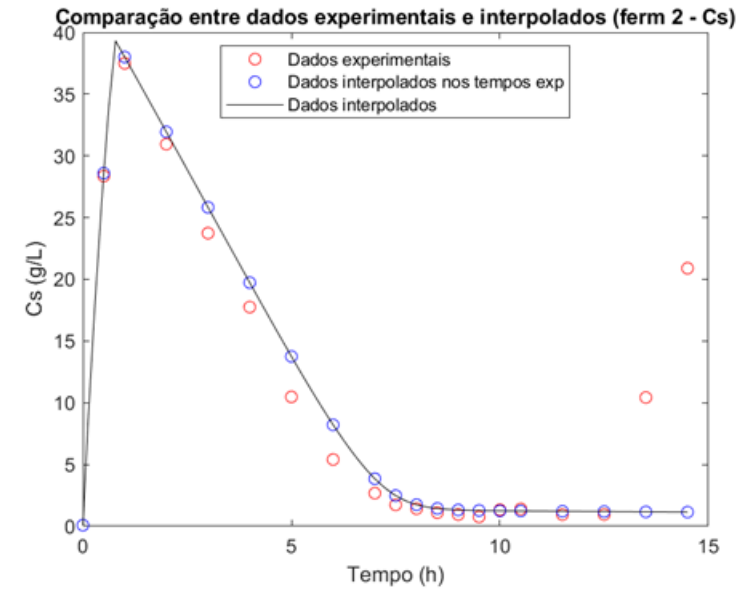
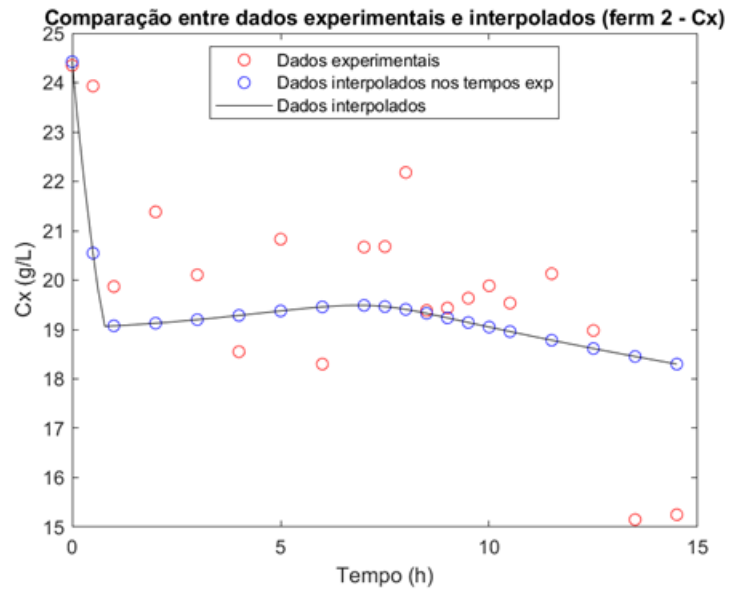


Figura 22 – Suavização dos dados experimentais das **Ferm 3** utilizando o modelo cinético fermentativo.

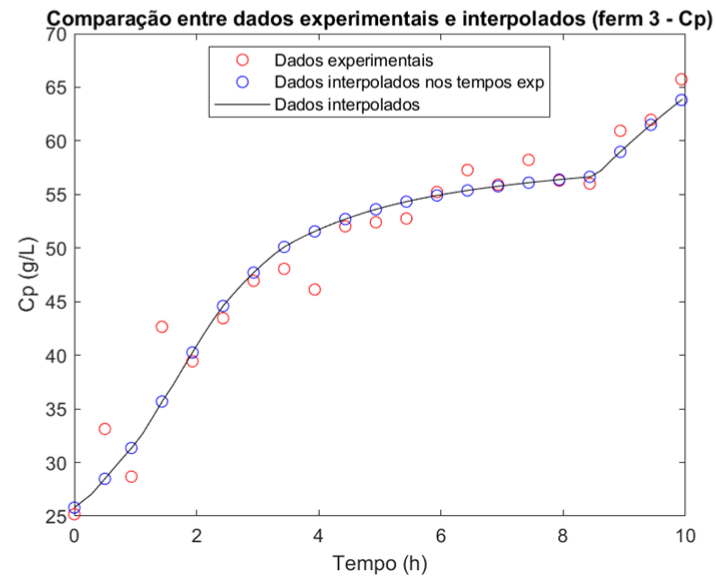
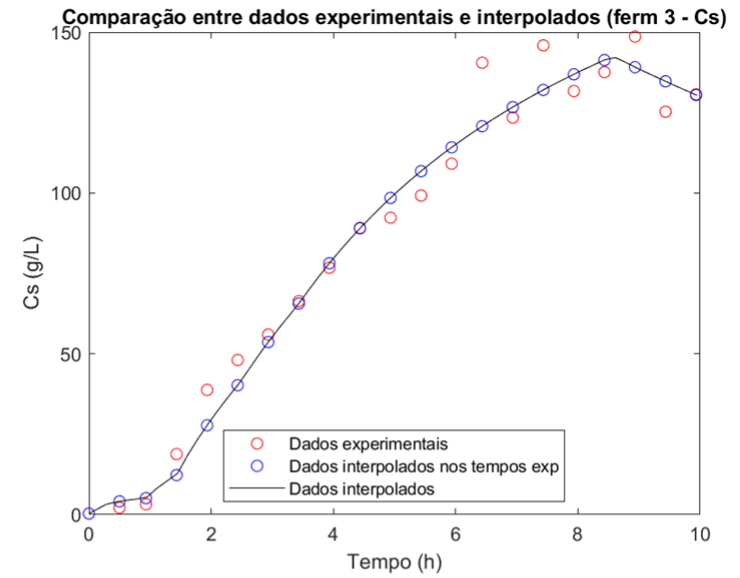
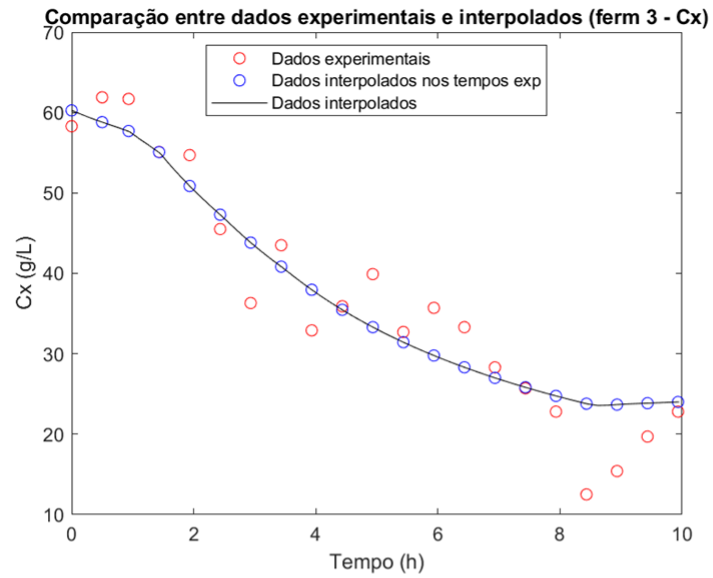


Figura 23 – Suavização dos dados experimentais das **Ferm 4** utilizando o modelo cinético fermentativo.

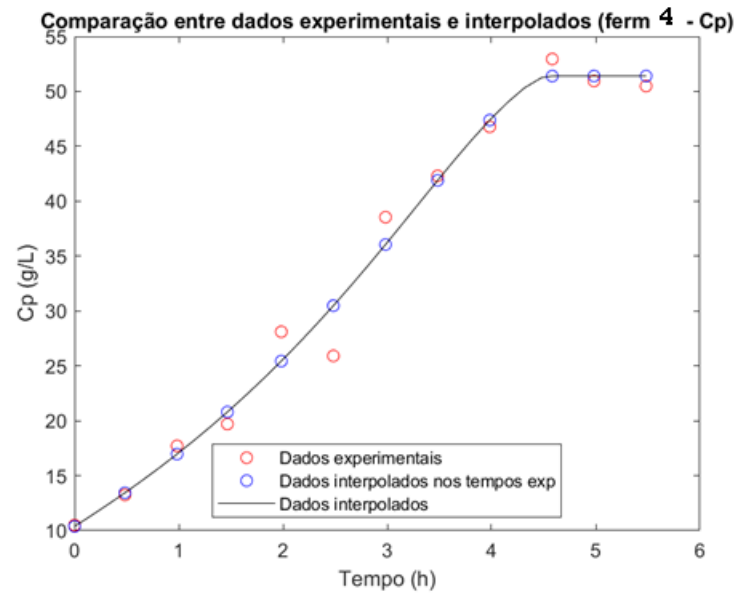
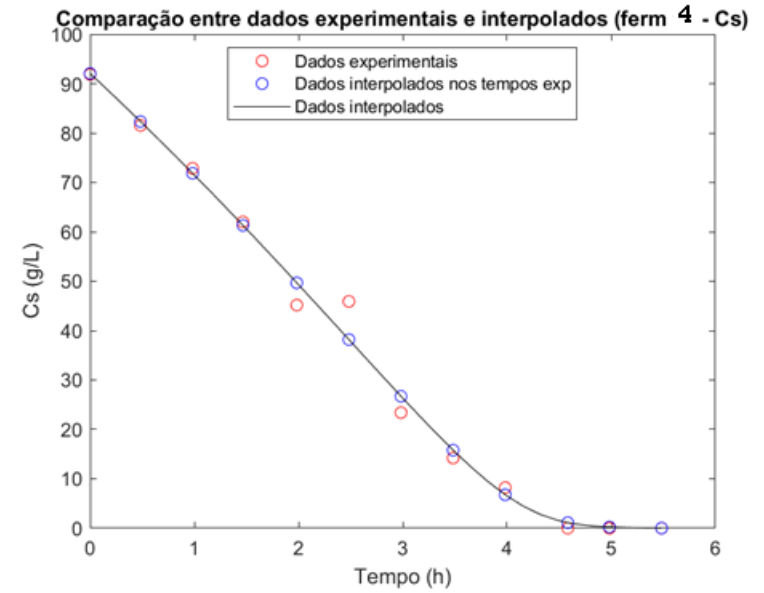
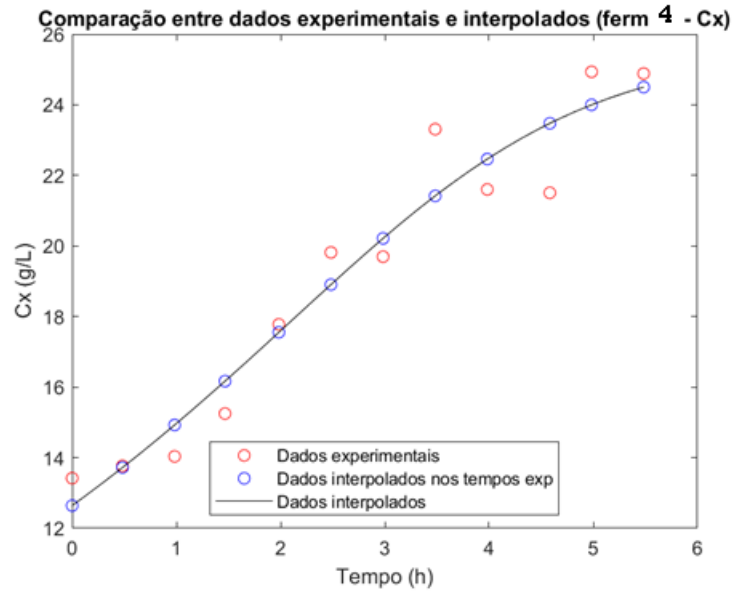


Figura 24 – Suavização dos dados experimentais das **Ferm 5** utilizando o modelo cinético fermentativo.

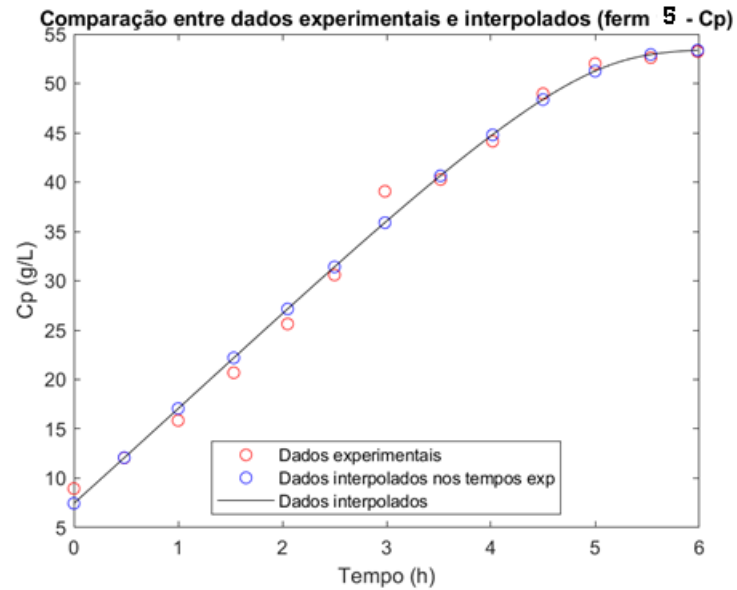
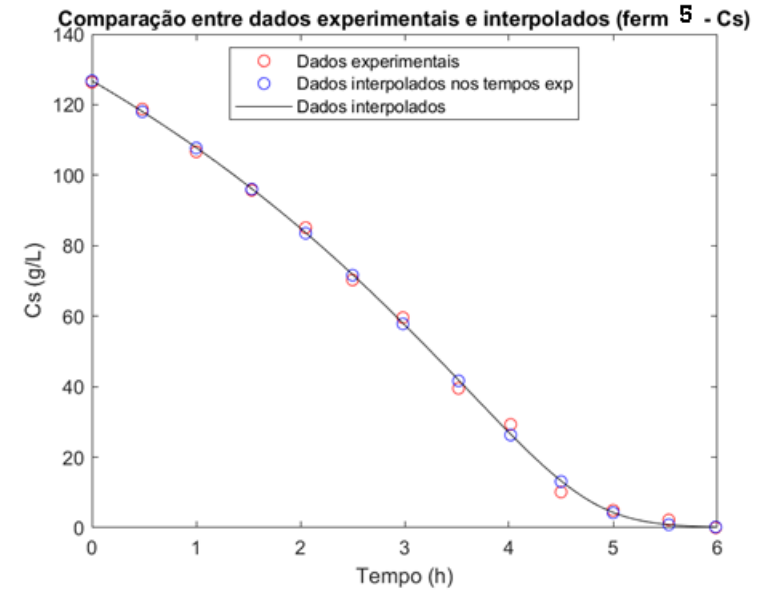
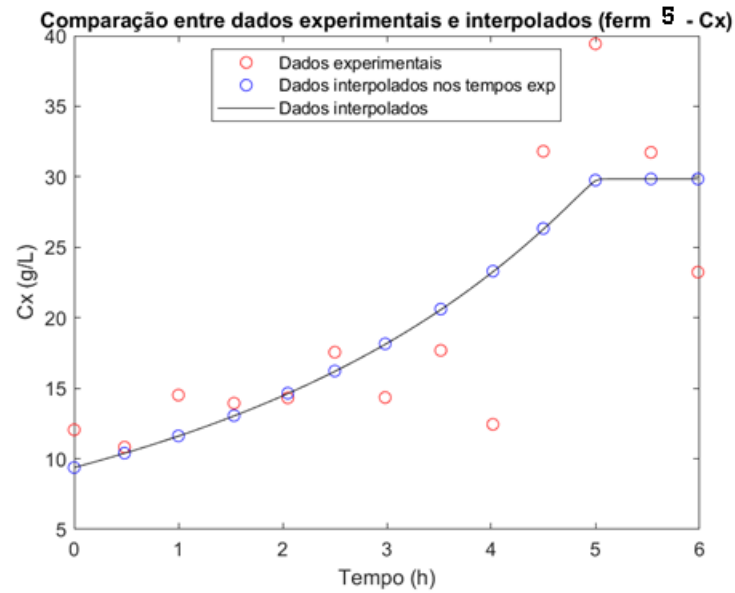


Figura 25 – Suavização dos dados experimentais das **Ferm 6** utilizando o modelo cinético fermentativo.

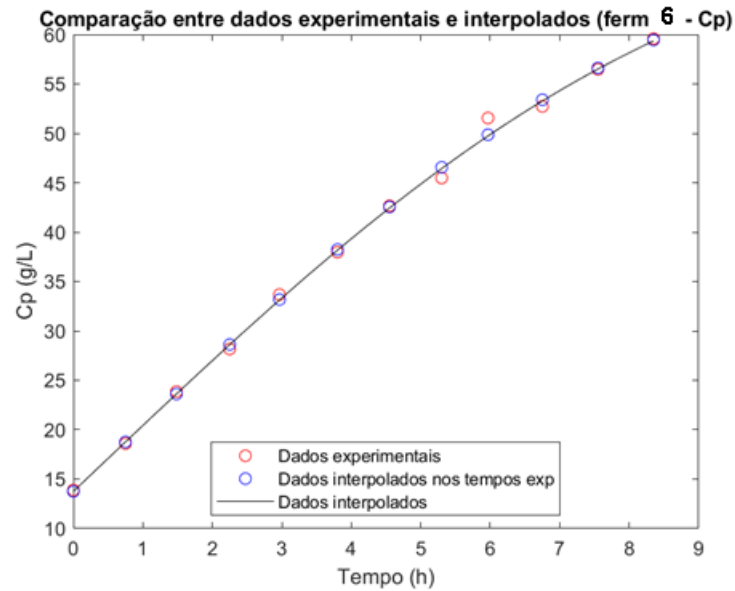
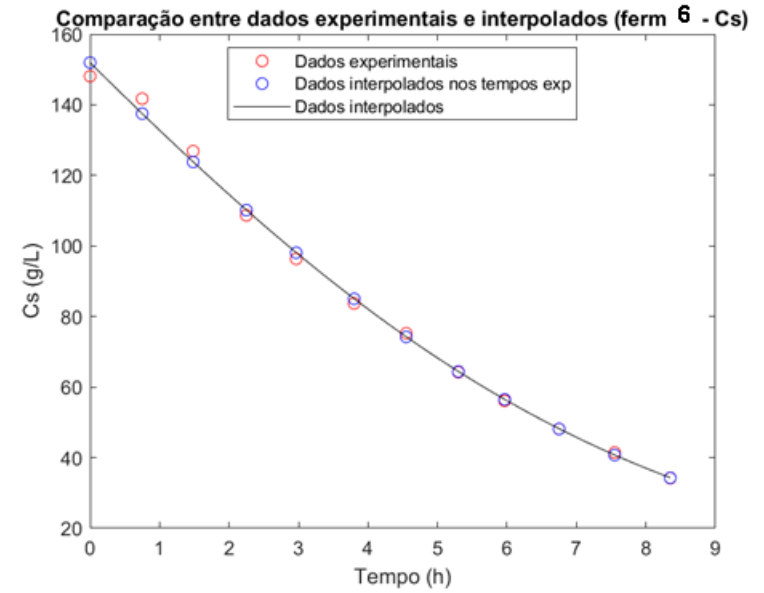
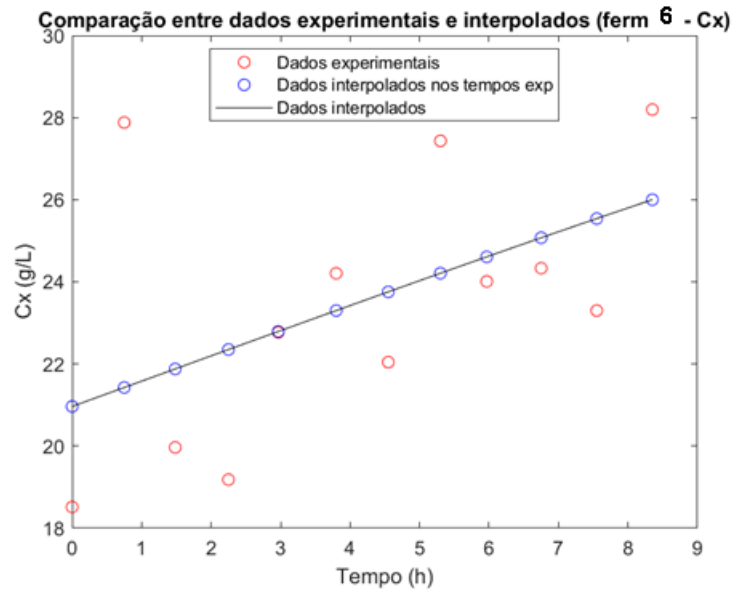
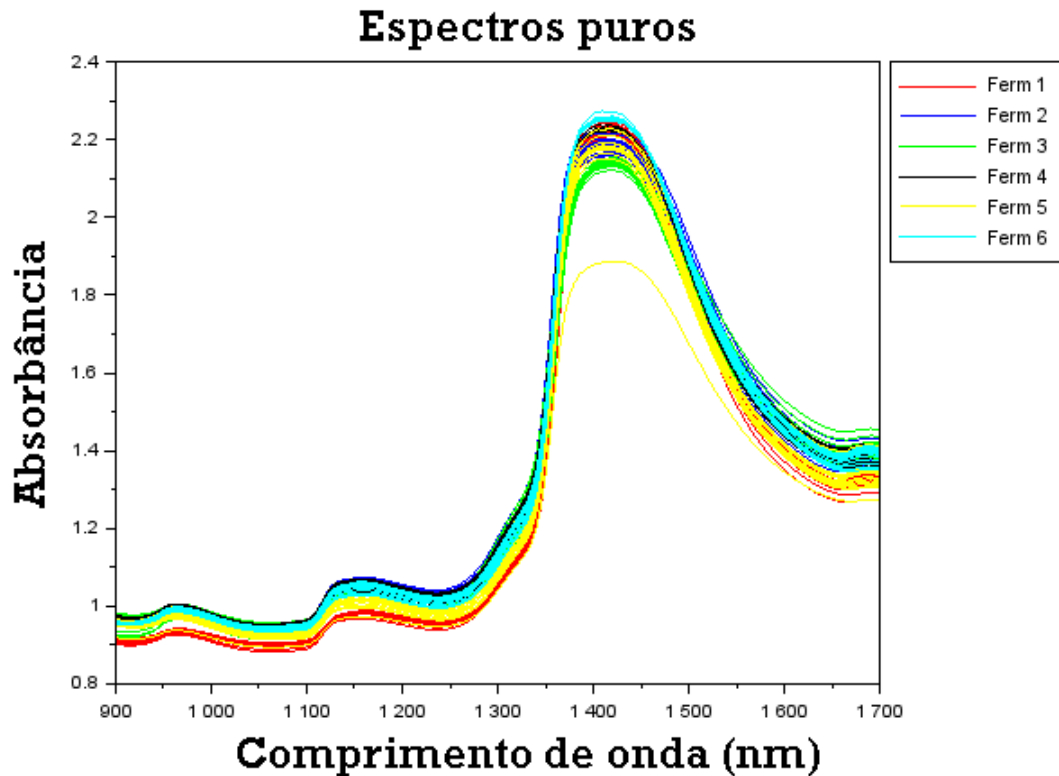




Figura 26 – Espectros na região do NIR obtidos pelo Sistema NIR-Online X-One durante as Fermentações de 1 a 6.



Os ensaios fermentativos foram desenvolvidos para representar diferentes situações que ocorrem em processos industriais. Dessa forma, ao analisar Figura 27, que apresenta as regiões de concentrações de etanol, açúcares redutores totais e células para as fermentações, é possível observar que as fermentações 1 a 3 abrangeram regiões bem distintas do diagrama de fases. Isto é um resultado esperado, considerando que estas fermentações foram idealizadas para serem utilizadas como conjunto de calibração.

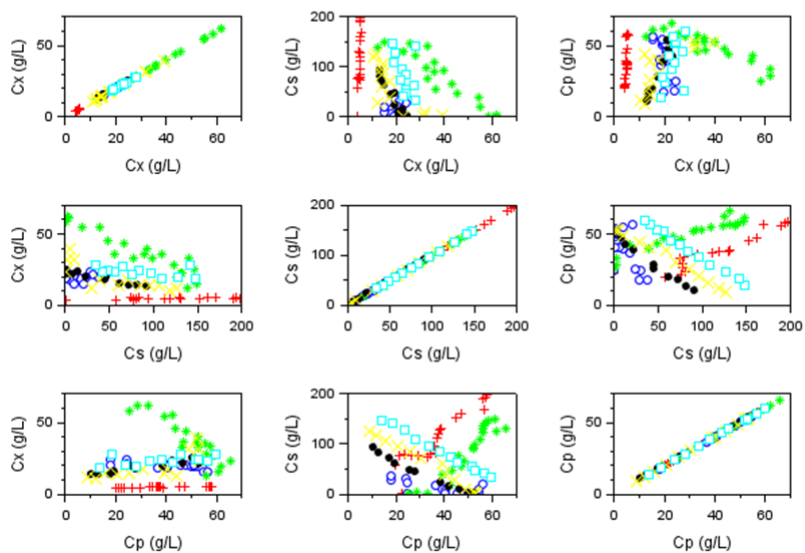
Já as fermentações batelada 4 a 6 tiveram comportamento similar entre si. As diferenças entre elas estão relacionadas com a faixa de concentrações, sendo que a Ferm 4 deteve menores valores de  $C_s$  e a Ferm 6 deteve maiores valores de  $C_s$ .

A suavização dos dados experimentais ficou aparente na Figura 27. Os dados experimentais suavizados reduziram o ruído intrínseco das técnicas analíticas, notadamente nas variáveis  $C_s$  e  $C_p$ .

Tendo esses resultados para essas fermentações, decidiu-se utilizar as fermentações 1 a 3 para construção do conjunto de calibração do modelo quimiométrico PLS e as fermentações 4 a 6 como conjunto de validação externa. Primeiramente, os espectros do conjunto de calibração foram pré-tratados com EMSC, como mostrado na Figura 28. Em seguida, utilizou-se o espectro médio do conjunto de calibração para pré-tratar com EMSC os espectros do conjunto de validação, como mostrado na Figura 29.

Figura 27 – Regiões de concentração de etanol ( $C_p$ ), açúcares redutores totais ( $C_s$ ) e células ( $C_x$ ) para as diferentes fermentações: Ferm 1 – vermelho; Ferm 2 – azul; Ferm 3 – verde; Ferm 4 – preto; Ferm 5 – amarelo; Ferm 6 – ciano.

### Dados experimentais



### Dados experimentais suavizados

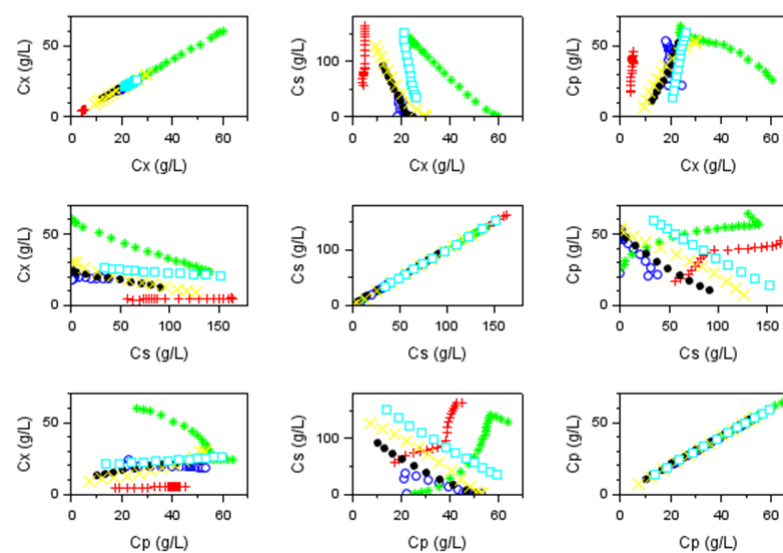


Figura 28 – Efeito do pré-tratamento EMSC nos espectros do conjunto de calibração.

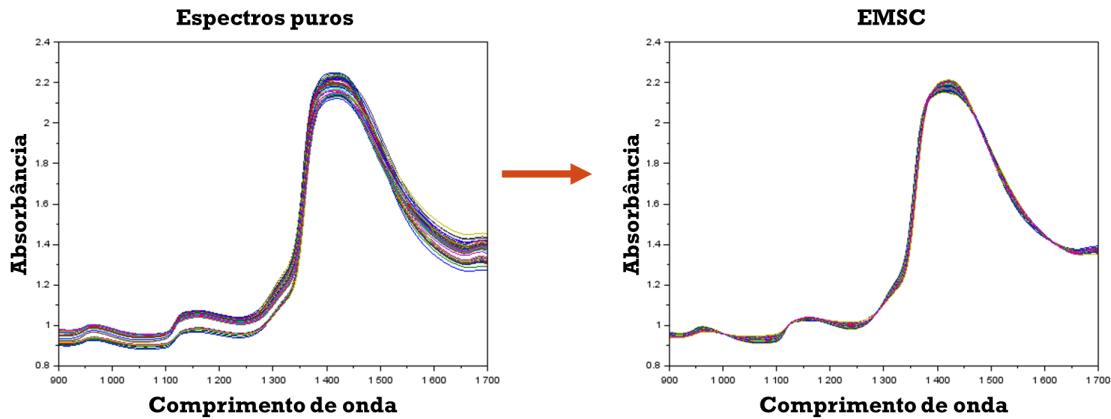
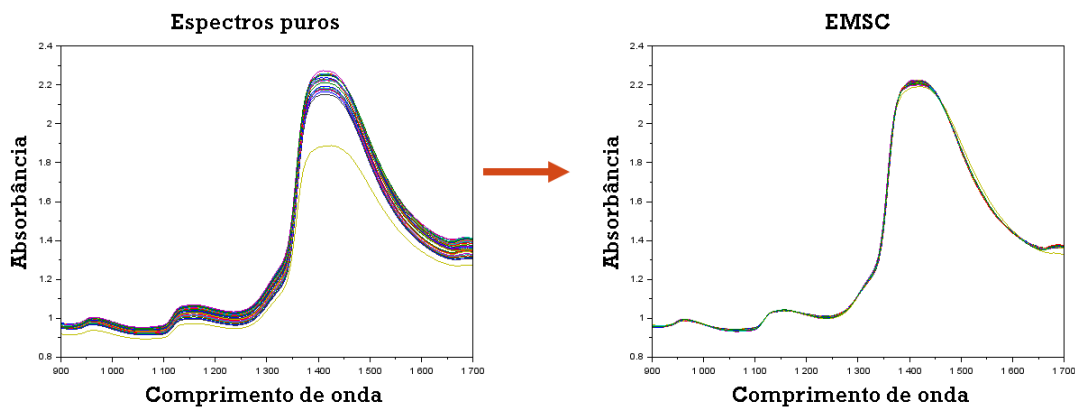


Figura 29 – Efeito do pré-tratamento EMSC nos espectros do conjunto de validação externa.



Criaram-se modelos quimiométricos PLS com dados experimentais e com dados experimentais suavizados. Para determinação da complexidade do modelo, foi utilizada validação cruzada *k-fold* com 10 grupos e análise das relações *RMSECV* e *RMSEcal* e **Teste F de Osten**. As Figuras 30 a 32 apresentam o perfil de *RMSECV* e *RMSEcal* e **Teste F de Osten** versus o número de variáveis latentes para as variáveis  $C_x$ ,  $C_s$  e  $C_p$  dos modelos com dados experimentais, respectivamente. Já as Figuras 33 a 35 apresentam o perfil de *RMSECV* e *RMSEcal* e **Teste F de Osten** versus o número de variáveis latentes para as variáveis  $C_x$ ,  $C_s$  e  $C_p$  dos modelos com dados experimentais suavizados.

Observa-se que, de maneira geral, as duas técnicas para otimização da complexidade do modelo determinaram o mesmo número de variáveis latentes para cada componente. Optou-se por não utilizar o mínimo global do *RMSECV* por uma questão de parcimônia na escolha do número de variáveis latentes e pelo motivo do erro de validação cruzada não reduzir substancialmente com o aumento do número de regressores.

Os modelos otimizados foram utilizados para inferir as concentrações dos grupos de validação externa (Ferm 4, Ferm 5 e Ferm6). As Figuras 36 a 38 apresentam os resultados de validação cruzada para os componente  $C_x$ ,  $C_s$  e  $C_p$ , respectivamente, do modelo com dados experimentais otimizado e inferência do grupo de validação externa. As Figuras 39 a 41 apresentam os resultados de validação cruzada para os componente  $C_x$ ,  $C_s$  e  $C_p$ , respectivamente, do modelo com dados experimentais suavizados otimizado e inferência do grupo de validação externa.

As Tabelas 4 e 5 apresentam os resultados estatísticos para os modelos PLS obtidos com dados experimentais e com dados experimentais suavizados, respectivamente. É possível observar que, de maneira geral, houve melhora na estimativa das concentrações do grupo de validação externa quando se utilizaram dados experimentais suavizados (diminuição do *RMSEP*). Isso pode ser consequência da redução do ruído intrínseco dos dados experimentais do grupo de calibração, notadamente na concentração de substrato. Como mostrado nas Figuras 20 a 22, há um grande número de amostras com valores ruidosos para a variável  $C_s$ , o que pode ter sido amenizado com a suavização pelo modelo cinético fermentativo.

As variáveis  $C_x$  e  $C_p$  detiveram valores de *RMSEP* muito superiores aos de *RMSECV* para as fermentações 5 e 6. No caso da Ferm 5, o uso de leveduras floculantes pode ter impactado na caracterização da concentração de células pela sua capacidade de formar floculos no sistema. Já na Ferm 6, o uso de um meio baseado em caldo de cana, diferente do que foi utilizado para treinamento do sensor virtual, e o número pequeno de amostras de calibração na sua faixa de concentrações (vide Figura 27) podem ter impactado na performance da inferência no grupo de validação externa.

Logo, o desenvolvimento de uma abordagem de suavização de dados de referência a partir de modelo cinético fermentativo para inferência quimiométrica mostrou-se interessante como metodologia para estimação e acompanhamento da fermentação alcoólica. Embora os resultados estatísticos não tenham sido todos promissores para todas as variáveis preditas, estes podem ser justificados pela variabilidade e complexidade da fermentação alcoólica: presença de contaminação, tipo de levedura, modo de condução da fermentação. Algumas melhorias podem impactar no melhor desenvolvimento do sensor: maior conjunto de dados de calibração, expandindo a faixa de valores, e buscando reter o máximo da variabilidade do sistema possível.

Figura 30 – Otimização da complexidade do modelo com dados experimentais para a variável  $C_x$ .

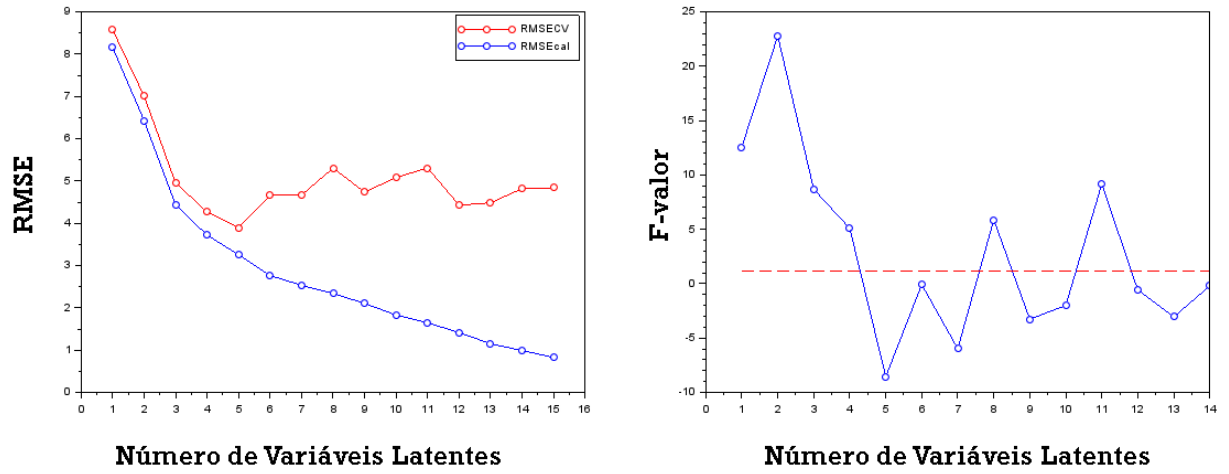


Figura 31 – Otimização da complexidade do modelo com dados experimentais para a variável  $C_s$ .

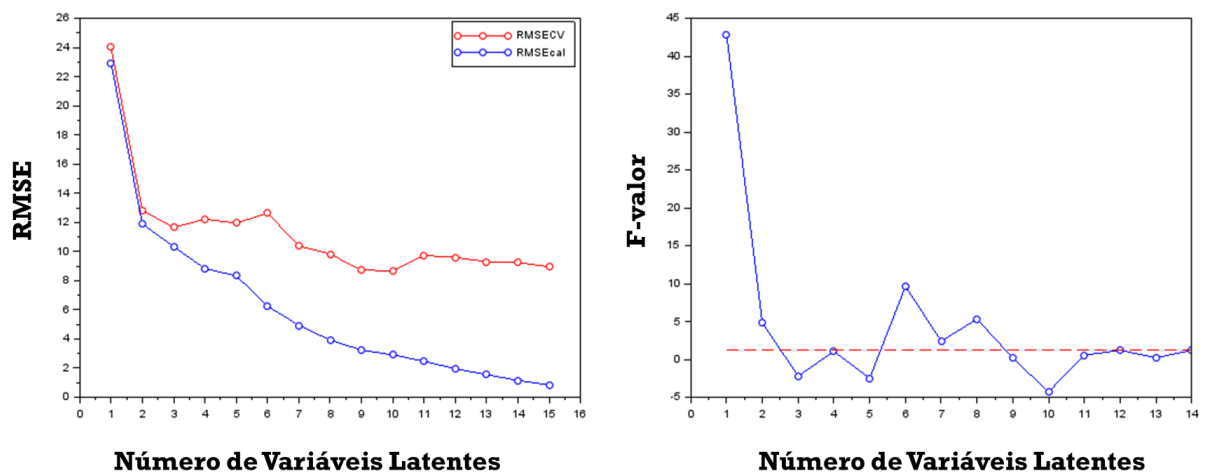


Figura 32 – Otimização da complexidade do modelo com dados experimentais para a variável  $C_p$ .

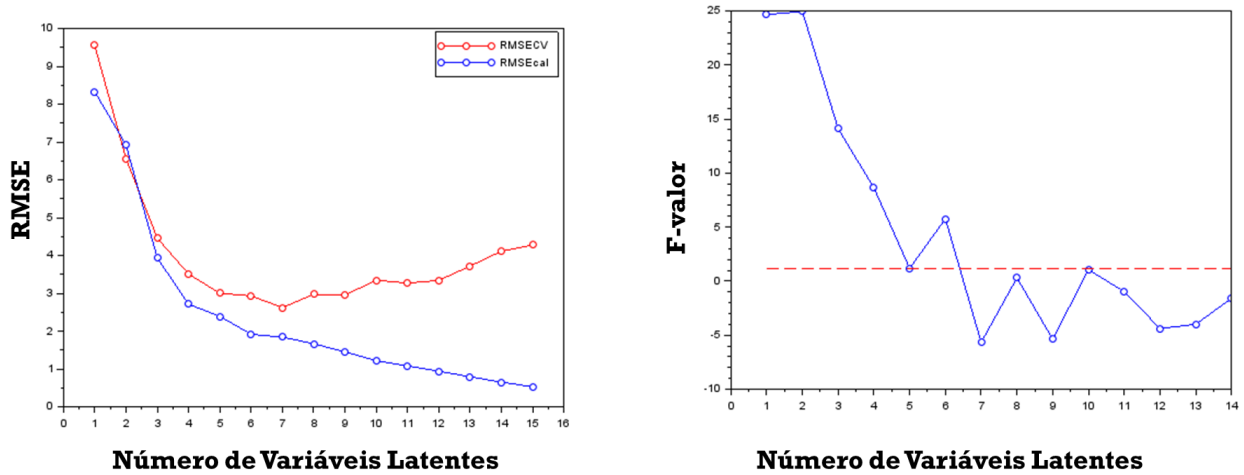


Figura 33 – Otimização da complexidade do modelo com dados experimentais suavizados para a variável  $C_x$ .

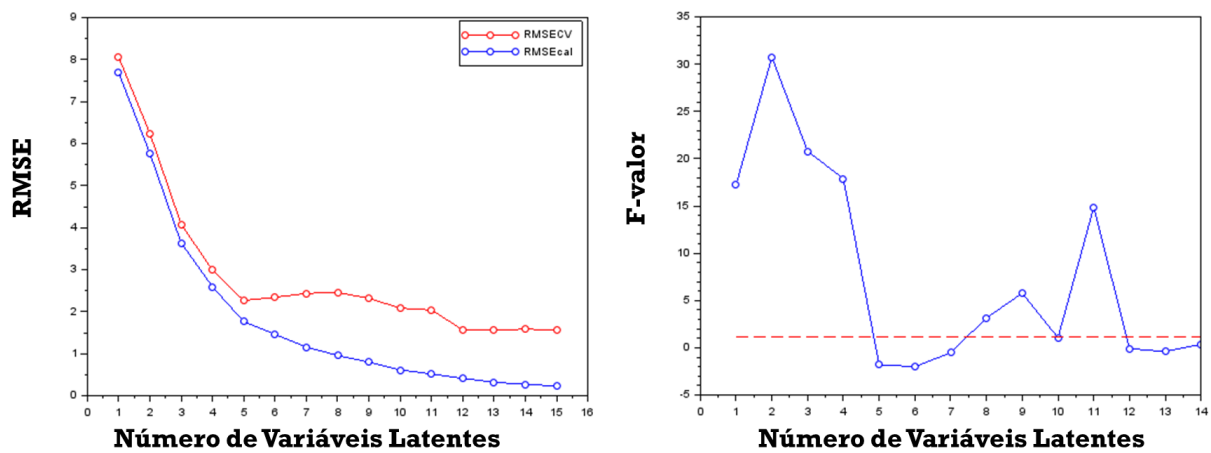


Figura 34 – Otimização da complexidade do modelo com dados experimentais suavizados para a variável  $C_s$ .

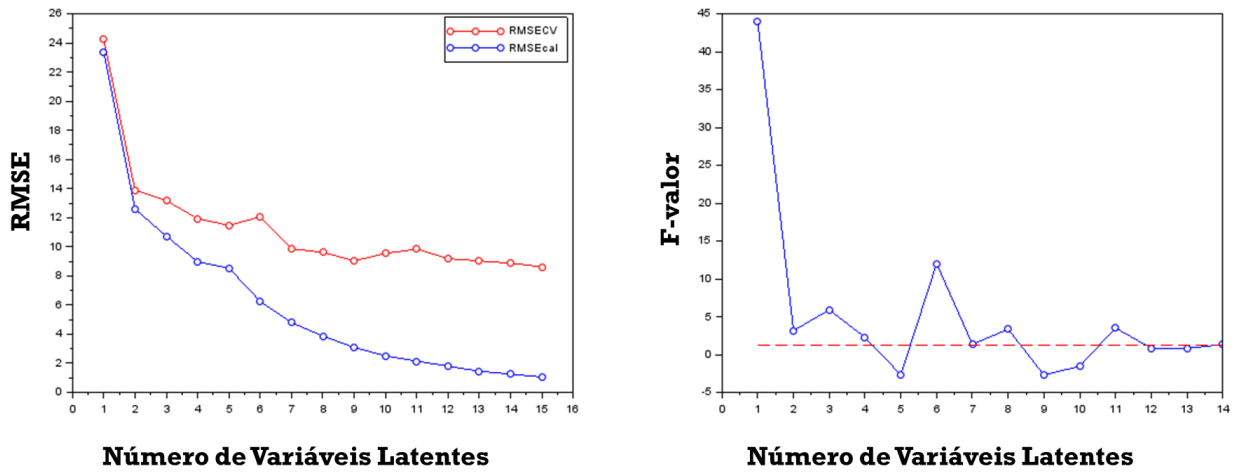


Figura 35 – Otimização da complexidade do modelo com dados experimentais suavizados para a variável  $C_p$ .

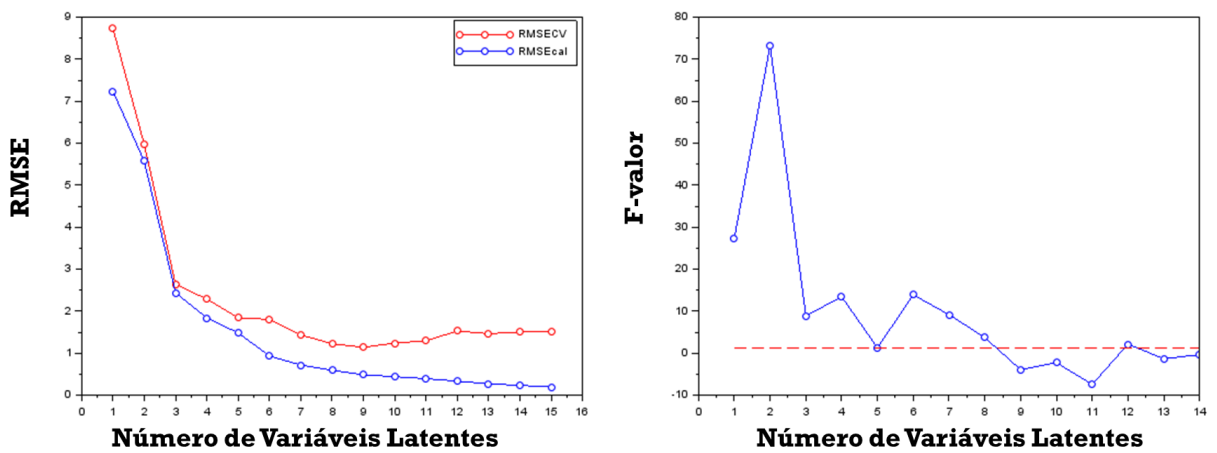


Figura 36 – Gráfico de validação cruzada para o componente  $C_x$  do modelo com dados experimentais otimizado e inferência do grupo de validação externa (da esquerda para direita, Ferm 4 e Ferm 5, e abaixo, Ferm 6)

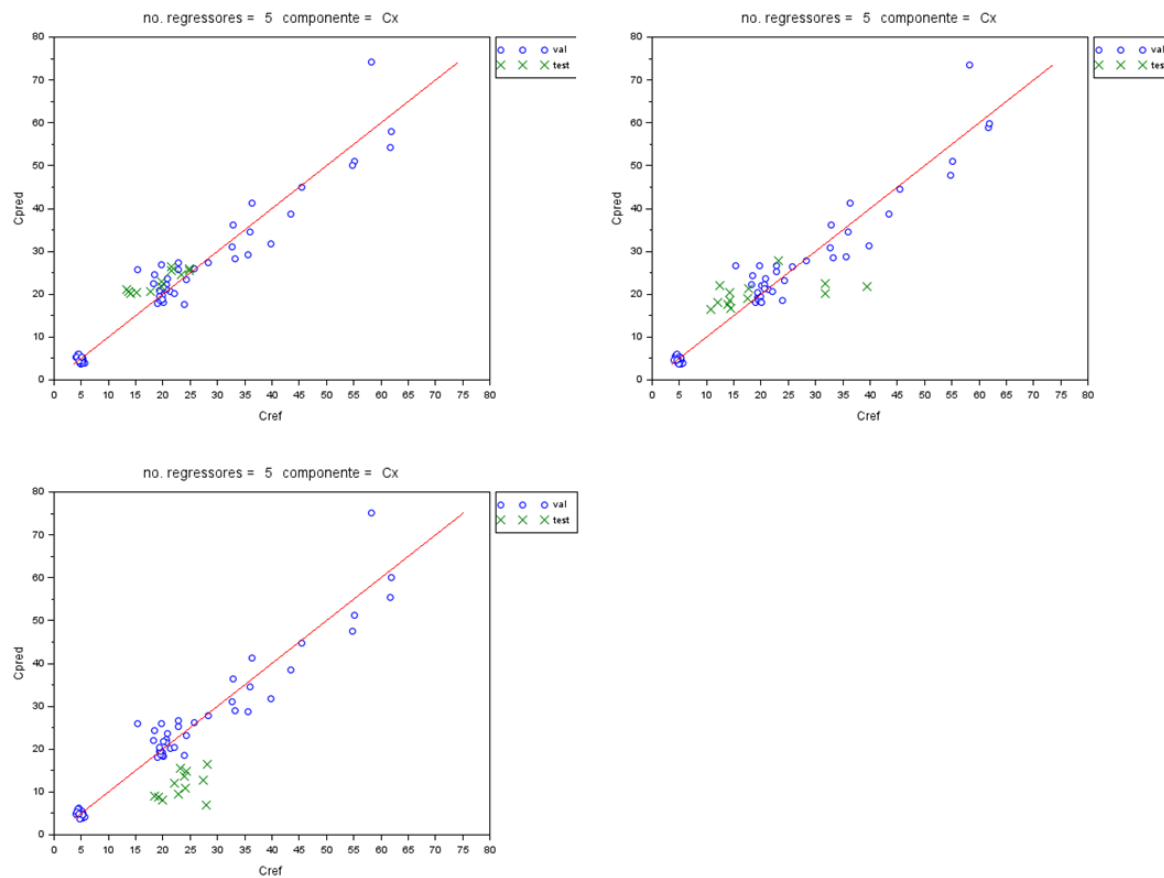




Figura 37 – Gráfico de validação cruzada para o componente  $C_s$  do modelo com dados experimentais otimizado e inferência do grupo de validação externa (da esquerda para direita, Ferm 4 e Ferm 5, e abaixo, Ferm 6)

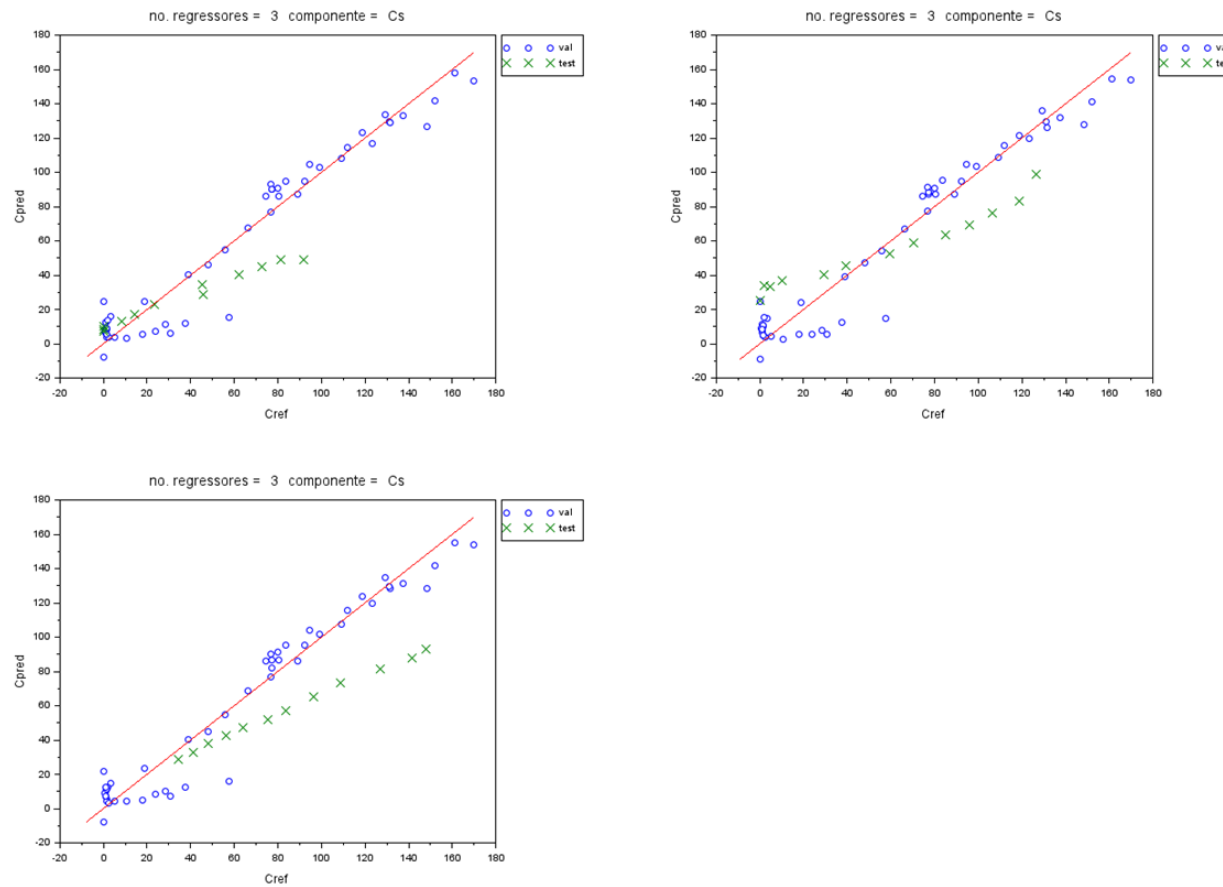


Figura 38 – Gráfico de validação cruzada para o componente  $C_p$  do modelo com dados experimentais otimizado e inferência do grupo de validação externa (da esquerda para direita, Ferm 4 e Ferm 5, e abaixo, Ferm 6)

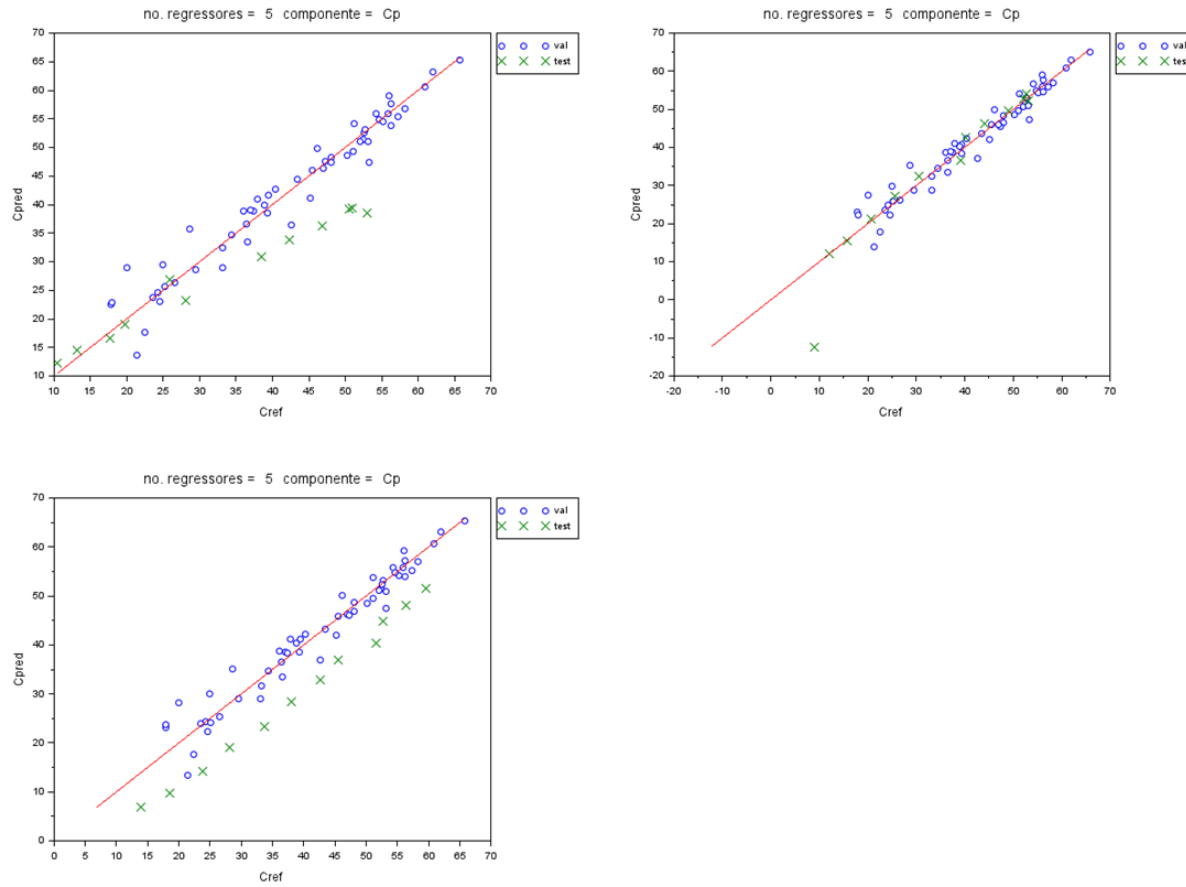


Figura 39 – Gráfico de validação cruzada para o componente  $C_x$  do modelo com dados experimentais suavizados otimizado e inferência do grupo de validação externa (da esquerda para direita, Ferm 4 e Ferm 5, e abaixo, Ferm 6).

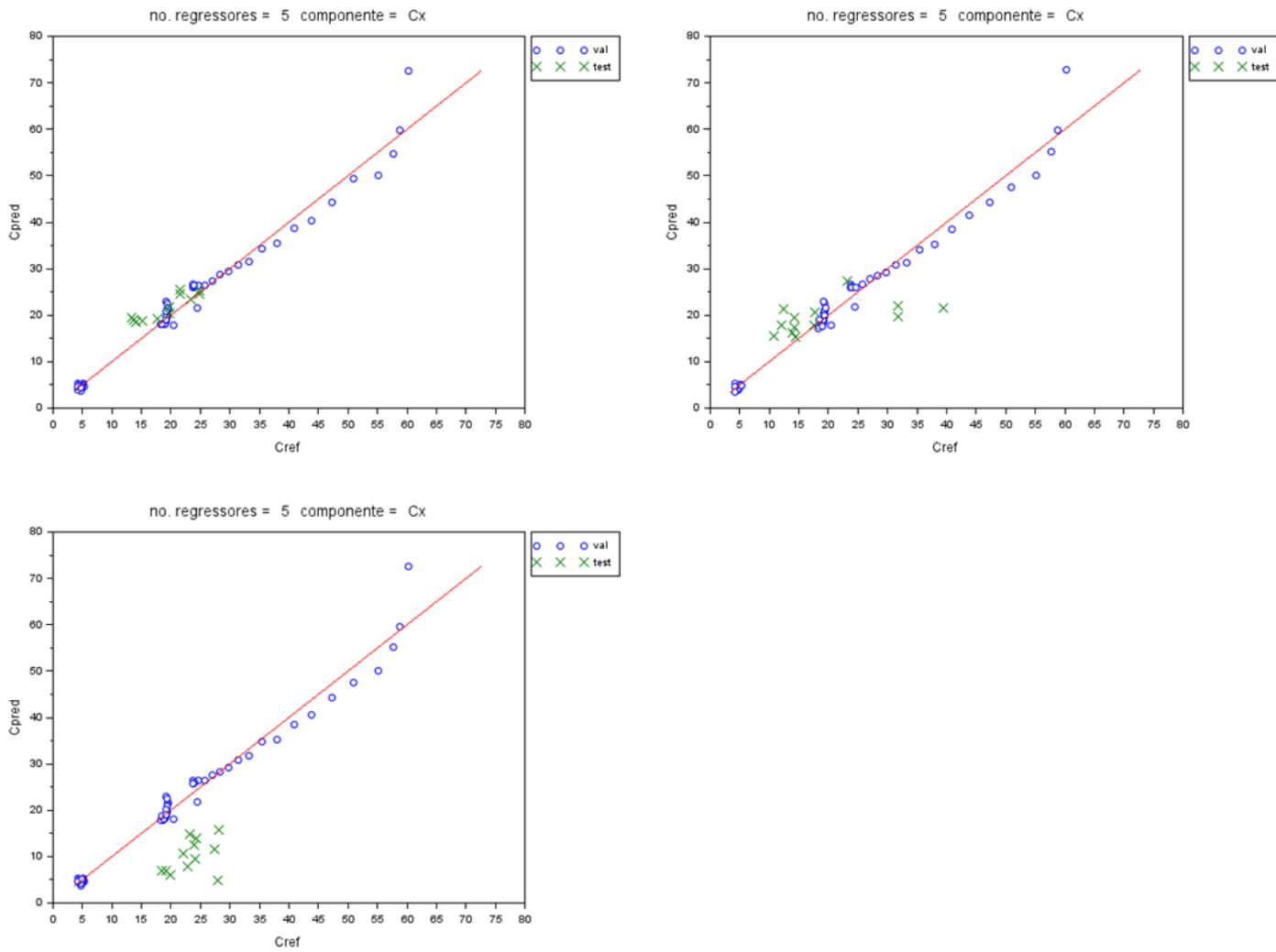


Figura 40 – Gráfico de validação cruzada para o componente  $C_s$  do modelo com dados experimentais suavizados otimizado e inferência do grupo de validação externa (da esquerda para direita, Ferm 4 e Ferm 5, e abaixo, Ferm 6).

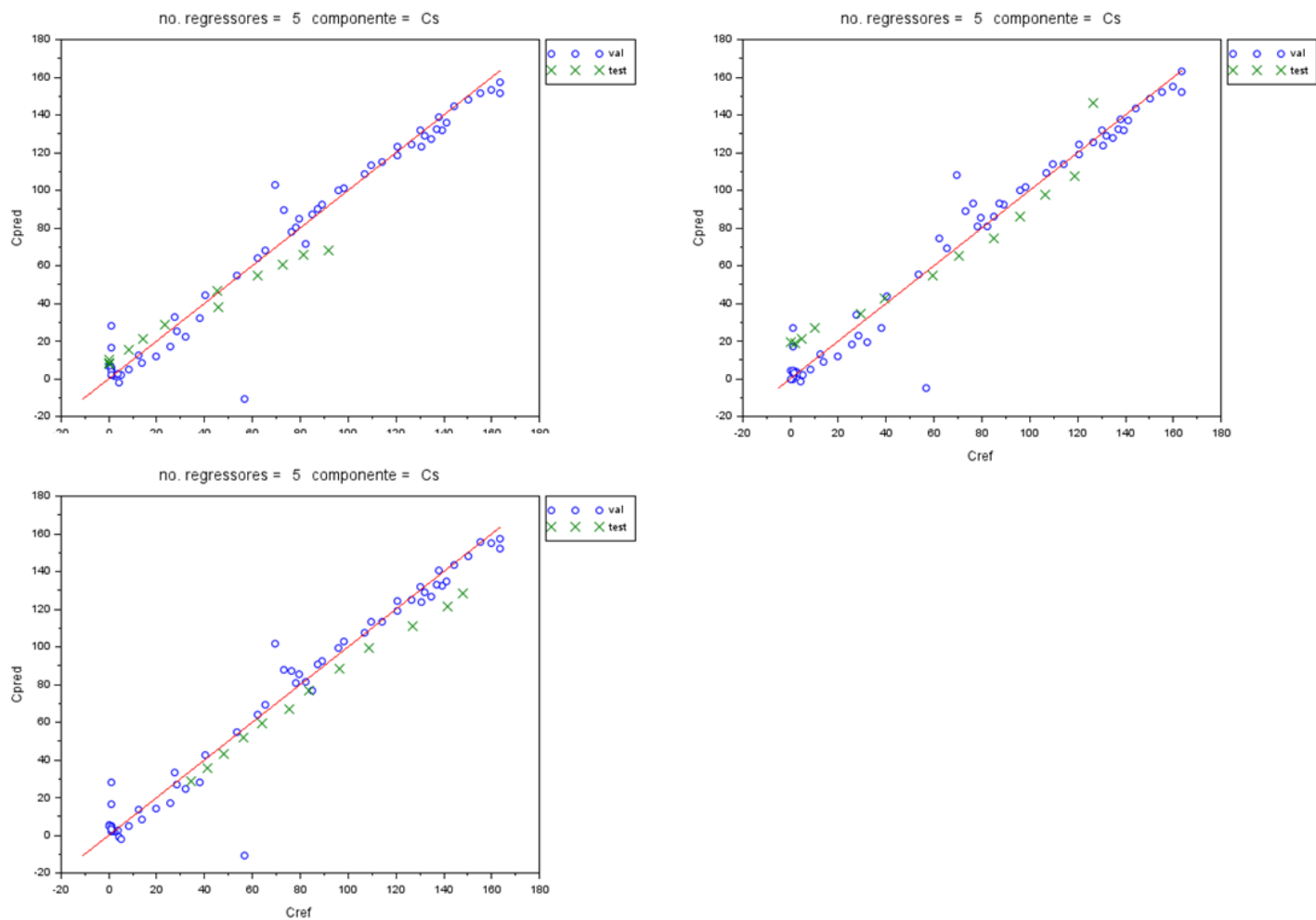


Figura 41 – Gráfico de validação cruzada para o componente  $C_x$  do modelo com dados experimentais suavizados otimizado e inferência do grupo de validação externa (da esquerda para direita, Ferm 4 e Ferm 5, e abaixo, Ferm 6).

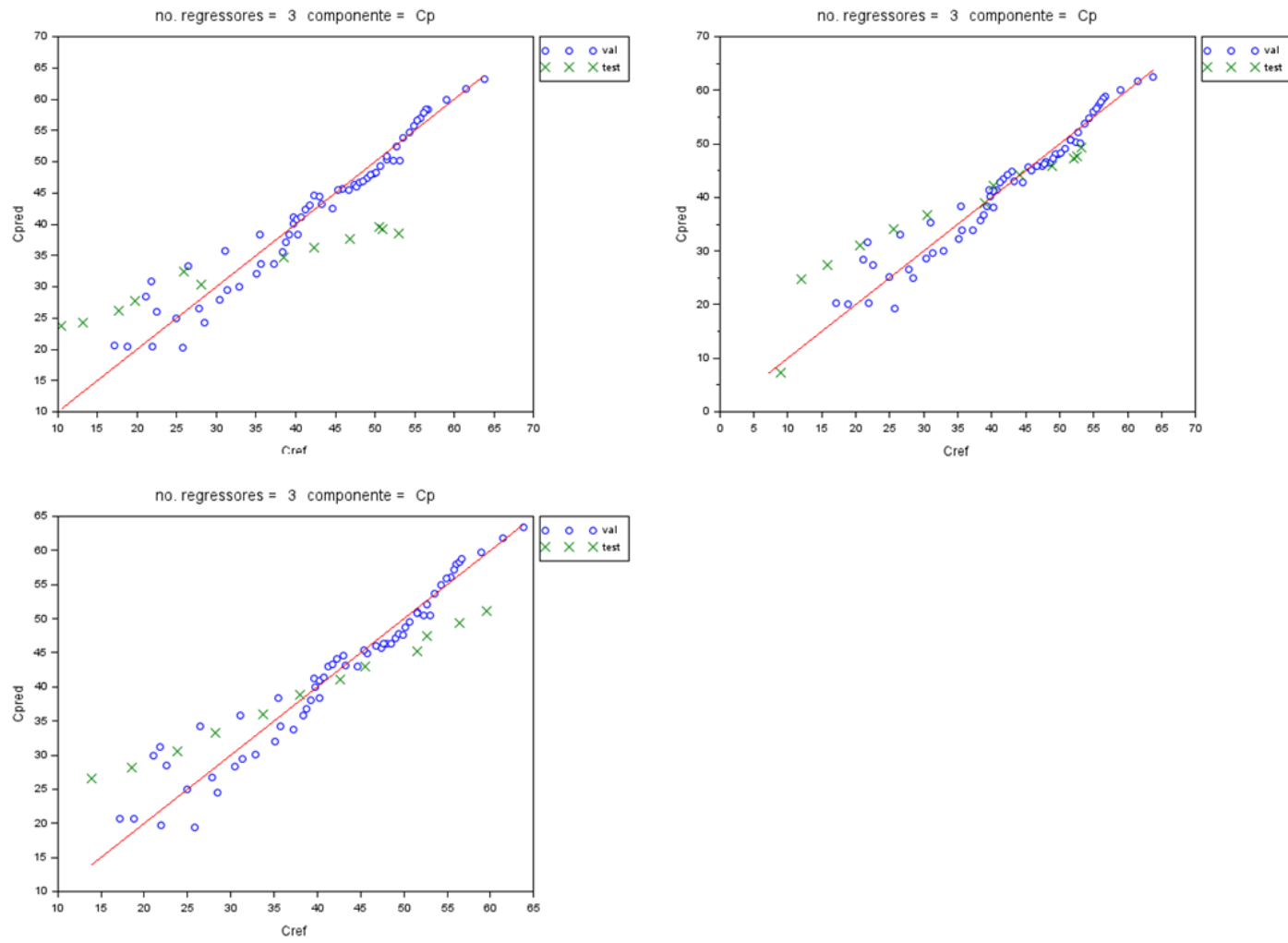


Tabela 4 – Modelos PLS obtidos com dados experimentais (da esquerda para direita em relação ao RMSEP: Ferm 4, Ferm 5 e Ferm 6 ).

Componente	LV	RMSECV (g/L)	RMSEP (g/L)
$C_x$	5	3,95	4,41 - 7,84 - 12,37
$C_s$	3	11,86	11,33 - 14,13 - 13,67
$C_p$	5	3,18	7,88 - 6,07 - 9,10

Tabela 5 – Modelos PLS obtidos com dados experimentais suavizados (da esquerda para direita em relação ao RMSEP: Ferm 4, Ferm 5 e Ferm 6 ).

Componente	LV	RMSECV (g/L)	RMSEP (g/L)
$C_x$	5	2,25	3,28 - 7,67 - 13,79
$C_s$	5	11,75	10,92 - 12,68 - 11,02
$C_p$	3	2,6	9,70 - 6,74 - 6,64

---

## Capítulo 5

### Conclusão

---

A principal contribuição deste trabalho foi mostrar a viabilidade de abordagem de suavização de dados de referência a partir de modelo cinético fermentativo para construção de sensor virtual baseado em espectroscopia NIR. O uso de um modelo cinético fermentativo para suavização de dados experimentais conseguiu reduzir o erro de predição da variável  $C_s$  em todas as fermentações do grupo de validação externa. As variáveis  $C_x$  e  $C_p$  não tiveram redução significativa do erro de predição com o modelo baseado nessa metodologia.

Como trabalho futuro, recomenda-se utilizar um maior conjunto de dados de calibração, expandindo a faixa de valores, e buscando reter o máximo da variabilidade do sistema possível (contaminação, leveduras, meio de cultivo). Outras metodologias incluem: seleção de comprimentos de onda, métodos de detecção de *outliers*, uso de redes neurais.





---

## Referências

---

ABDI, H. Partial least squares regression and projection on latent structure regression (pls regression). **Wiley interdisciplinary reviews: computational statistics**, Wiley Online Library, v. 2, n. 1, p. 97–106, 2010.

AFDC. **ALTERNATIVE FUELS DATA CENTER: Global Ethanol Production**. São Paulo, Brasil, 2021. Disponível em: <<https://afdc.energy.gov/data/10331>>. Acesso em: 22.11.2022.

AFSETH, N. K.; KOHLER, A. Extended multiplicative signal correction in vibrational spectroscopy, a tutorial. **Chemometrics and Intelligent Laboratory Systems**, Elsevier, v. 117, p. 92–99, 2012.

ALMEIDA, L. P. **Avaliação da transferência de calor em fermentações alcoólicas convencional e extrativa**. Dissertação (Dissertação de Mestrado) — Universidade Federal de São Carlos, 2019.

ARAÚJO, M. C. U. et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. **Chemometrics and intelligent laboratory systems**, Elsevier, v. 57, n. 2, p. 65–73, 2001.

BADINO, A. C.; CRUZ, A. J. G. da. Reatores químicos e biológicos. **Coleção UAB-UFSCar, São Carlos**, 2012.

BAEYENS, J. et al. Challenges and opportunities in improving the production of bio-ethanol. **Progress in Energy and Combustion Science**, Elsevier, v. 47, p. 60–88, 2015.

BAILEY, J. E.; OLLIS, D. F. **Biochemical Engineering Fundamentals**. 2<sup>a</sup>. ed. McGraw-Hill College, 1986. 928 p. ISBN 978-0070032125. Disponível em: <<https://lead.to/amazon/com/?op=bt&la=en&cu=usd&key=0070032122>>.

BARNES, R.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. **Applied spectroscopy**, SAGE Publications Sage UK: London, England, v. 43, n. 5, p. 772–777, 1989.

BASSO, T. O.; BASSO, L. C. Peculiaridades e tendências da fermentação alcoólica no brasil. **Revista Brasileira de Engenharia Química**, Associação Brasileira de Engenharia Química, v. 34, p. 5–11, 2018.

- BILSKEY, S. R. et al. A comparative analysis of methods for quantitation of sugars during the corn-to-ethanol fermentation process. **SLAS TECHNOLOGY: Translating Life Sciences Innovation**, SAGE Publications Sage CA: Los Angeles, CA, v. 25, n. 5, p. 494–504, 2020.
- BISSETT, A. C. **Improvements to PLS Methodology**. Tese (Tese de Doutorado) — The University of Manchester (United Kingdom), 2015.
- BLANCO, M.; PEINADO, A. C.; MAS, J. Analytical monitoring of alcoholic fermentation using nir spectroscopy. **Biotechnology and bioengineering**, Wiley Online Library, v. 88, n. 4, p. 536–542, 2004.
- BLANCO, M.; VILLARROYA, I. Nir spectroscopy: a rapid-response analytical tool. **TrAC Trends in Analytical Chemistry**, v. 21, n. 4, p. 240–250, 2002. ISSN 0165-9936. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0165993602004041>>.
- BORGES, P. **Otimização dinâmica da fermentação alcoólica no processo em batelada alimentada**. 2008. Dissertação (Mestrado) — Universidade Federal de Uberlândia, 2008.
- BRERETON, R. G. Sequential methods. In: \_\_\_\_\_. **Applied Chemometrics for Scientists**. John Wiley & Sons, Ltd, 2007. cap. 4, p. 111–144. ISBN 9780470057780. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470057780.ch4>>.
- CANA, N. **NOVA CANA: Possíveis melhorias nas usinas de produção de etanol**. Paraná, Brasil, 2018. Disponível em: <<https://www.novacana.com/etanol/melhorias-usinas-producao-etanol>>. Acesso em: 19.11.2022.
- CAVALETT, O. et al. Environmental and economic assessment of sugarcane first generation biorefineries in brazil. **Clean Technologies and Environmental Policy**, Springer, v. 14, n. 3, p. 399–410, 2012.
- CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. **Trends in Food Science & Technology**, v. 18, n. 2, p. 72–83, 2007. ISSN 0924-2244. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0924224406002524>>.
- DEED, R. C.; DEED, N. K.; GARDNER, R. C. Transcriptional response of *saccharomyces cerevisiae* to low temperature during wine fermentation. **Antonie Van Leeuwenhoek**, Springer, v. 107, n. 4, p. 1029–1048, 2015.
- DIAS, M. et al. Improving bioethanol production – comparison between extractive and low temperature fermentation. **Applied Energy**, v. 98, p. 548–555, 2012. ISSN 0306-2619. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0306261912003169>>.
- Eliasson Lantz, A. et al. 12 - online monitoring of fermentation processes in lignocelluloses-to-bioalcohol production. In: WALDRON, K. (Ed.). **Bioalcohol Production**. Woodhead Publishing, 2010, (Woodhead Publishing Series in Energy). p. 315–339. ISBN 978-1-84569-510-1. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9781845695101500128>>.

- FERREIRA, M. et al. Quimiometria i: calibração multivariada, um tutorial. **Química Nova**, SciELO Brasil, v. 22, p. 724–731, 1999.
- FORTUNA, L. et al. **Soft Sensors for Monitoring and Control of Industrial Processes (Advances in Industrial Control)**. Hardcover. Springer, 2006. 289 p. ISBN 978-1846284793. Disponível em: <<https://lead.to/amazon/com/?op=bt&la=en&cu=usd&key=1846284791>>.
- FULGÊNCIO, A. C. da C. et al. Combining portable nir spectroscopy and multivariate calibration for the determination of ethanol in fermented alcoholic beverages by a multi-product model. **Talanta Open**, Elsevier, v. 7, p. 100180, 2023.
- GARGALO, C. L. et al. On-line monitoring of process parameters during fermentation. **Current Developments in Biotechnology and Bioengineering**, Elsevier, p. 117–164, 1 2022.
- GOMES, B. C. **Análise térmica da fermentação alcoólica convencional e extrativa com arraste por dióxido de carbono**. Dissertação (Dissertação de Mestrado) — Universidade Federal de São Carlos, 2020.
- GRASSI, M. C.; PEREIRA, G. A. Energy-cane and renovabio: Brazilian vectors to boost the development of biofuels. **Industrial Crops and Products**, Elsevier B.V., v. 129, p. 201–205, 3 2019. ISSN 09266690.
- 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**, ACS Publications, v. 60, n. 11, p. 1193–1202, 1988.
- HAHN-HÄGERDAL, B. et al. Bio-ethanol—the fuel of tomorrow from the residues of today. **Trends in biotechnology**, Elsevier, v. 24, n. 12, p. 549–556, 2006.
- HORTA, A. et al. A supervision and control tool based on artificial intelligence for high cell density cultivations. **Brazilian Journal of Chemical Engineering**, SciELO Brasil, v. 31, p. 457–468, 2014.
- JIANG, H. et al. Monitoring alcohol concentration and residual glucose in solid state fermentation of ethanol using ft-nir spectroscopy and 11-pls regression. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, Elsevier, v. 204, p. 73–80, 2018.
- JOLLIFFE, I. Principal components in regression analysis. **Principal Component Analysis; Jolliffe, IT, Ed.; Springer Series in Statistics**, p. 129–155, 2010.
- JOSHI, G. et al. Challenges and opportunities for the application of biofuel. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 79, p. 850–866, 2017.
- KADLEC, P.; GABRYS, B.; STRANDT, S. Data-driven soft sensors in the process industry. **Computers & Chemical Engineering**, v. 33, n. 4, p. 795–814, 2009. ISSN 0098-1354. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0098135409000076>>.
- KANG, Q. et al. Hybrid operation of the bio-ethanol fermentation. **Separation and Purification Technology**, Elsevier, v. 149, p. 322–330, 2015.

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**, Springer, v. 9, p. 1112–1119, 2016.

LALUCE, C. et al. Fermentation of sugarcane juice and molasses for ethanol production. **Sugarcane-based biofuels and bioproducts**, John Willey & Sons Hoboken, New York, p. 53–86, 2016.

LIANG, Y.-Z.; KVALHEIM, O. M. Robust methods for multivariate analysis—a tutorial review. **Chemometrics and Intelligent Laboratory Systems**, Elsevier, v. 32, n. 1, p. 1–10, 1996.

LIMA, U. de A. et al. **Biotecnologia Industrial - Volume 3 (PROCESSOS FERMENTATIVOS E ENZIMÁTICOS)**. 1ª edição. ed. EDGARD BLUCHER, 2002. ISBN 978-8521202806. Disponível em: <<https://lead.to/amazon/com/?op=bt&la=en&cu=usd&key=8521202806>>.

LIU, Z.; CAI, W.; SHAO, X. Outlier detection in near-infrared spectroscopic analysis by using monte carlo cross-validation. **Science in China Series B: Chemistry**, Springer, v. 51, n. 8, p. 751–759, 2008.

LONGATI, A. A. et al. Defining research and development process targets through retro-techno-economic analysis: The sugarcane biorefinery case. **Bioresource Technology**, Elsevier Ltd, v. 263, p. 1–9, 9 2018. ISSN 18732976.

LOPES, C. H.; GABRIEL, A.; BORGES, M. Produção de etanol a partir da cana de açúcar. **Tecnologia de produção de Etanol. São Carlos: Universidade Federal de São Carlos**, 2011.

LOPES, M. L. et al. Ethanol production in brazil: a bridge between science and industry. **brazilian journal of microbiology**, SciELO Brasil, v. 47, p. 64–76, 2016.

LOTUFO, F. A.; GARCIA, C. Sensores virtuais ou soft sensors: Uma introdução. In: **Proc. of the 7th Brazilian Conference on Dynamics, Control and Applications, Presidente Prudente**. São Paulo, Brasil: [s.n.], 2008. p. 1–9. Disponível em: <[https://www.researchgate.net/publication/228413086\\_Sensores\\_Virtuais\\_ou\\_Soft\\_Sensors\\_Uma\\_introducao](https://www.researchgate.net/publication/228413086_Sensores_Virtuais_ou_Soft_Sensors_Uma_introducao)>.

MAGAZONI, F. C. **Análise Dinâmica de um chiller de absorção de brometo de lítioágua em processo de resfriamento de dorna de fermentação alcoólica**. Tese (Tese de Doutorado) — Universidade Federal de Santa Catarina, 2011.

MAIORELLA, B.; WILKE, C. R.; BLANCH, H. Alcohol production and recovery. In: SPRINGER. **Bioenergy**. University of California Berkeley, California 94720, 1981. p. 43–92.

MALUF, G.; GURGEL, A. C. Competição entre o etanol de segunda geração e a bioeletricidade pelo uso do bagaço de cana de açúcar. **ANPEC - Associação Nacional dos Centros de Pós-Graduação em Economia**, 2016.

MARTÍNEZ-JIMENEZ, F. D. et al. Integration of first- and second-generation ethanol production: Evaluation of a mathematical model to describe sucrose and xylose co-fermentation by recombinant *Saccharomyces cerevisiae*. **Renewable Energy**, Elsevier Ltd, v. 192, p. 326–339, 6 2022. ISSN 18790682.

MESQUITA, T. J. B. **Controle de biorreator com base em modelo de fluxos metabólicos: sintonia das condições de microaeração para maximização da produção de etanol por *S. cerevisiae***. Dissertação (Dissertação de Mestrado) — Universidade Federal de São Carlos, 2017.

MESQUITA, T. J. B. **Estruturas de controle de biorreator baseadas em fluxos metabólicos para fermentações micro-aeradas**. Tese (Doutorado) — Universidade Federal de São Carlos, 2021.

NASCIMENTO, R. J. A. d. et al. Real time and in situ near-infrared spectroscopy (nirs) for quantitative monitoring of biomass, glucose, ethanol and glycerine concentrations in an alcoholic fermentation. **Brazilian Journal of Chemical Engineering**, SciELO Brasil, v. 34, p. 459–468, 2017.

NEITZEL, T. et al. Impact of the melle-boinot process on the enhancement of second-generation ethanol production by *Spathaspora passalidarum*. **Renewable Energy**, Elsevier, v. 160, p. 1206–1216, 2020.

NELLES, O. **Nonlinear System Identification: From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes**. Hardcover. Springer, 2020. 1253 p. ISBN 978-3030474386. Disponível em: <<https://lead.to/amazon/com/?op=bt&la=en&cu=usd&key=3030474380>>.

OLIVEIRA, J. A. P. de. The policymaking process for creating competitive assets for the use of biomass energy: the brazilian alcohol programme. **Renewable and Sustainable Energy Reviews**, Pergamon, v. 6, p. 129–140, 1 2002. ISSN 1364-0321.

OSTEN, D. W. Selection of optimal regression models via cross-validation. **Journal of Chemometrics**, Wiley Online Library, v. 2, n. 1, p. 39–48, 1988.

PACHECO, T. F. **Fermentação alcoólica com leveduras de características floculantes em reator tipo torre com escoamento ascendente**. Dissertação (Dissertação de Mestrado) — Universidade Federal de Uberlândia, 2010.

PACHECO, T. F. **Produção de etanol: primeira ou segunda geração?** Embrapa Agroenergia (Brasília, Brasil), 2011. Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/32985/1/CITE-04.pdf>>. Acesso em: 28.11.2022.

PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. **Journal of the Brazilian chemical society**, SciELO Brasil, v. 14, p. 198–219, 2003.

PAULA, A. R. d. **Monitoramento da hidrólise enzimática de bagaço de cana-de-açúcar por métodos espectrofotométricos**. Dissertação (Dissertação de Mestrado) — Universidade Federal de São Carlos, 2020.

PEREIRA, R. D.; BADINO, A. C.; CRUZ, A. J. Assessing the performance of industrial ethanol fermentation unit using neural networks. **Computer Aided Chemical Engineering**, Elsevier, v. 43, p. 175–180, 1 2018. ISSN 1570-7946.

- PINTO, A. S. S. **Monitoramento por espectroscopia dos compostos fenólicos e furaldeídos gerados no processamento de biomassa lignocelulósica**. Dissertação (Dissertação de Mestrado) — Universidade Federal de São Carlos, 2017.
- PINTO, C. R. d. S. **Remoção de etanol por stripping empregando dióxido de carbono**. Tese (Tese de Doutorado) — Universidade Federal de São Carlos, 2014.
- RFA. **Renewable Fuels Association: World Fuel Ethanol Production**. Washington D.C., Estados Unidos, 2022. Disponível em: <<https://ethanolrfa.org/markets-and-statistics/annual-ethanol-production>>. Acesso em: 18.11.2022.
- RIBEIRO, M. P. d. A. **Operação ótima de reator para síntese enzimática de ampicilina com cristalização simultânea dos produtos**. Tese (Tese de Doutorado) — Universidade Federal de São Carlos, 2007.
- 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**, Elsevier, v. 28, n. 10, p. 1201–1222, 2009.
- RIVERA, E. C. et al. Evaluation of optimization techniques for parameter estimation: Application to ethanol fermentation considering the effect of temperature. **Process Biochemistry**, v. 41, n. 7, p. 1682–1687, 2006. ISSN 1359-5113. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1359511306000699>>.
- RODRIGUES, K. C. et al. Real-time monitoring of bioethanol fermentation with industrial musts using mid-infrared spectroscopy. **Industrial & Engineering Chemistry Research**, ACS Publications, v. 57, n. 32, p. 10823–10831, 2018.
- ROGER, J.-M.; MALLET, A.; MARINI, F. Preprocessing nir spectra for aquaphotomics. **Molecules**, MDPI, v. 27, n. 20, p. 6795, 2022.
- RUTLEDGE, D. N.; ROGER, J.-M.; LESNOFF, M. Different methods for determining the dimensionality of multivariate models. **Frontiers in Analytical Science**, Frontiers, v. 1, p. 754447, 2021.
- SANTOS, F.; ALUÍZIO, B.; CALDAS, C. (Ed.). **Sugarcane: Agricultural Production, Bioenergy and Ethanol**. Kindle edition. Academic Press, 2015. 801 p. Disponível em: <<https://lead.to/amazon/com/?op=bt&la=en&cu=usd&key=B00Y2H1IFS>>.
- SAVITZKY, A.; GOLAY, M. J. Smoothing and differentiation of data by simplified least squares procedures. **Analytical chemistry**, ACS Publications, v. 36, n. 8, p. 1627–1639, 1964.
- SCHMIDELL, W. et al. **Biotecnologia Industrial - Volume 2 (ENGENHARIA BIOQUÍMICA)**. 1ª edição. ed. EDGARD BLUCHER, 2001. ISBN 978-8521202790. Disponível em: <<https://lead.to/amazon/com/?op=bt&la=en&cu=usd&key=8521202792>>.
- SHULER, M. L.; KARGI, F. **Bioprocess Engineering: Basic Concepts**. 1nd. ed. Prentice Hall, 2006. Paperback. ISBN 9788177587401. Disponível em: <<http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0134782151>>.

SONEGO, J. L. S. **Estudo da produção de etanol de sacarose por fermentação extrativa utilizando arraste com dióxido de carbono**. Tese (Tese de Doutorado) — Universidade Federal de São Carlos, 2016.

STAFF, R. **Brazil ethanol output could double by 2030 - gov't study**. Rio de Janeiro, Brasil, 2018. Disponível em: <<https://www.reuters.com/article/brazil-ethanol/brazil-ethanol-output-could-double-by-2030-govt-study-idUSL1N1U70U1>>. Acesso em: 22.01.2024.

SVENDSEN, C. **On-Line Monitoring of Fermentation Processes by Near Infrared and Fluorescence Spectroscopy: Elucidating and Exploring Process Dynamics**. Tese (Doutorado), 2016.

TAYLOR, M. et al. Partial least squares regression as a powerful tool for investigating large combinatorial polymer libraries. **Surface and Interface Analysis: An International Journal devoted to the development and application of techniques for the analysis of surfaces, interfaces and thin films**, Wiley Online Library, v. 41, n. 2, p. 127–135, 2009.

USDABR. **United States Department of Agriculture Brazil**. São Paulo, Brasil, 2024. Disponível em: <<https://usdabrazil.org.br/>>. Acesso em: 22.01.2024.

VANDENBERGHE, L. P. et al. Beyond sugar and ethanol: The future of sugarcane biorefineries in brazil. **Renewable and Sustainable Energy Reviews**, Elsevier Ltd, v. 167, 10 2022. ISSN 18790690.

VELOSO, I. I. et al. Fed-batch ethanol fermentation at low temperature as a way to obtain highly concentrated alcoholic wines: Modeling and optimization. **Biochemical Engineering Journal**, v. 141, p. 60–70, 2019. ISSN 1369-703X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1369703X18303620>>.

WALCZAK, B.; MASSART, D. Multiple outlier detection revisited. **Chemometrics and intelligent laboratory systems**, Elsevier, v. 41, n. 1, p. 1–15, 1998.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109–130, 2001. ISSN 0169-7439. PLS Methods. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169743901001551>>.

ZAKAMULIN, V. **Market Timing with Moving Averages: The Anatomy and Performance of Trading Rules (New Developments in Quantitative Trading and Investment)**. Kindle edition. Palgrave Macmillan, 2017. 461 p. Disponível em: <<https://lead.to/amazon/com/?op=bt&la=en&cu=usd&key=B077GB3T34>>.

ZHANG, H. Software sensors and their applications in bioprocess. In: **Computational intelligence techniques for bioprocess modelling, supervision and control**. [S.l.]: Springer, 2009. p. 25–56.