



UNIVERSIDADE FEDERAL DE SÃO CARLOS
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA
QUÍMICA

Daniel de Castro Souza

**Modelagem *Fuzzy*-Cinética e Monitoramento
Espectrofotométrico da Hidrólise Enzimática do
Bagaço de Cana-de-Açúcar em Sistema de
Reatores com Separação de Fases**

São Carlos - SP

2024

Daniel de Castro Souza

**Modelagem *Fuzzy*-Cinética e Monitoramento
Espectrofotométrico da Hidrólise Enzimática do Bagaço
de Cana-de-Açúcar em Sistema de Reatores com
Separação de Fases**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Química para obtenção do título de Mestre em Engenharia Química.

Orientação Prof. Dr. Marcelo Perencin de Arruda Ribeiro

São Carlos - SP

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 Daniel de Castro Souza, realizada em 27/02/2024.

Comissão Julgadora:

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

Prof. Dr. Vitor Badiale Furlong (AU)

Prof. Dr. Ruy de Sousa Junior (UFSCar)

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

Dedico este trabalho aos meus pais, em retribuição a todo o suporte dado durante minha formação. Cujo apoio me permitiu explorar um caminho de significado e realização inquantificáveis.

Agradecimentos

Ao Marcelo e ao Vitor, meus orientadores, pela enorme paciência e suporte nessa jornada.
A Bruna, pelo companheirismo, amor e suporte.
A Cíntia, por me guiar por caminhos que não conhecia e por me ajudar a sair do chão.
A todos os meus amigos, cuja companhia foi crucial para a manutenção da minha sanidade.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES) - Código de financiamento 001.

O presente trabalho foi realizado com o apoio do Conselho Nacional de Desenvolvimento Tecnológico (CNPq) - Bolsa # 130467/2022-7.

O presente trabalho foi realizado com o apoio da Fundação de Apoio à Pesquisa de São Paulo (FAPESP) - projeto # 2016/10636-8.

*“Não quero olhar para trás um dia... e desejar ter feito diferente.
Aqui e agora, preciso saber que fiz tudo o que pude.
(Aerith Gainsborough)”*

Resumo

A diminuição das reservas de combustíveis fósseis aliada à crescente demanda energética mundial fomenta a busca por alternativas. Neste escopo, uma possível solução pode surgir a partir do uso de biocombustíveis, como o etanol, produzido com base em fontes renováveis. No Brasil, em alinhamento com tais esforços, este álcool já é produzido em larga escala a partir da fermentação do caldo-de-cana. Este processo pode ter seu rendimento incrementado através da utilização do próprio bagaço restante da cana-de-açúcar como substrato em fermentação alcoólica. O etanol produzido dessa forma é denominado etanol de segunda geração (2G). Para possibilitar tal processo, o bagaço deve ser pré-tratado, expondo suas fibras e permitindo sua hidrólise. A hidrólise é responsável por quebrar as moléculas de celulose em monossacarídeos que possam ser consumidos pelas leveduras durante a fermentação. A hidrólise enzimática ocorre em condições amenas e libera menos compostos que inibem a fermentação, o que, apesar do alto custo da enzima, a torna mais interessante em relação à hidrólise química. Com base neste contexto, o presente projeto visou estudar um sistema reacional não convencional construído no grupo de pesquisa, utilizando-se de dados experimentais de cromatografia líquida e de espectroscopia de absorção do infravermelho obtidos anteriormente. Devido à complexidade do sistema reacional, a aplicação de modelos fenomenológicos é impraticável, e abordagens semi-mecanísticas apresentam-se como alternativa. Portanto, um modelo fuzzy-cinético baseado na cinética de Michaelis-Menten foi desenvolvido visando obter um sistema equacional capaz de descrever o processo para diferentes condições de alimentação. Este estudo apresenta resultados significativos, destacando a criação de um modelo cinético inovador para a hidrólise enzimática do bagaço de cana-de-açúcar em reatores com separação de fases. A análise de identificabilidade dos parâmetros do modelo revelou sua robustez na descrição da cinética da reação. Prevendo possíveis desvios entre o modelo e a reação real, optou-se por desenvolver um sistema de monitoramento utilizando espectroscopia na região do infravermelho próximo. Neste trabalho, um modelo preditivo de concentrações com base nos espectros foi desenvolvido através de uma combinação de seleção de variáveis por algoritmo genético e da regressão por mínimos quadrados parciais, obtendo-se um modelo com erro de validação cruzada de $\pm 0,012$ g/L para a celobiose, $\pm 0,650$ g/L para a glicose e $\pm 0,072$ g/L para a xilose.

Palavras-chave: Hidrólise enzimática, Modelagem de processos, Etanol 2G, Modelagem *fuzzy*, Monitoramento espectroscópico.

Abstract

The decrease in fossil fuel reserves combined with the growing global energy demand fosters the search for alternatives. Within this scope, a possible solution may arise from the use of biofuels, such as ethanol, produced based on renewable sources. In Brazil, in alignment with such efforts, this alcohol is already produced on a large scale from the fermentation of sugarcane juice. This process can have its yield increased by using the remaining sugarcane bagasse itself as a substrate in alcoholic fermentation. The ethanol produced in this way is called second-generation ethanol (2G). To enable this process, the bagasse must be pretreated, exposing its fibers and allowing its hydrolysis. Hydrolysis is responsible for breaking the cellulose molecules into monosaccharides that can be consumed by yeast during fermentation. Enzymatic hydrolysis occurs under mild conditions and releases fewer compounds that inhibit fermentation, which, despite the high cost of the enzyme, makes it more attractive compared to chemical hydrolysis. Based on this context, the present project aimed to study a non-conventional reaction system built in the research group, using experimental data from liquid chromatography and infrared absorption spectroscopy obtained previously. Due to the complexity of the reaction system, the application of phenomenological models is impractical, and semi-mechanistic approaches are presented as an alternative. Therefore, a fuzzy-kinetic model based on Michaelis-Menten kinetics was developed aiming to obtain an equation system capable of describing the process for different feed conditions. This study presents significant results, highlighting the creation of an innovative kinetic model for the enzymatic hydrolysis of sugarcane bagasse in reactors with phase separation. The identifiability analysis of the model parameters revealed its robustness in describing the reaction kinetics. Anticipating possible deviations between the model and the real reaction, it was chosen to develop a monitoring system using near-infrared spectroscopy. In this work, a predictive concentration model based on spectra was developed through a combination of variable selection by genetic algorithm and partial least squares regression, resulting in a model with cross-validation errors of $\pm 0,012$ g/L for cellobiose, $\pm 0,650$ g/L for glucose, and $\pm 0,072$ g/L for xylose.

Keywords: Enzymatic hydrolysis, Process modeling, 2G ethanol, Fuzzy modeling, Spectroscopic monitoring.

Lista de ilustrações

Figura 1 – Estrutura da biomassa lignocelulósica (RYTIOJA et al., 2014).	23
Figura 2 – Representação esquemática do sistema de reatores utilizado (FURLONG, 2019).	28
Figura 3 – Projeto do reator de sólidos (esquerda) e protótipo do sistema utilizado (direita) (FURLONG, 2019).	29
Figura 4 – Esquematização das reações de hidrólise consideradas.	35
Figura 5 – Exemplificação do comportamento entre erro de validação/ajuste (Adaptado de (NELLES, 2001)).	41
Figura 6 – Comparação entre o modelo MM-IP e dados experimentais no ensaio de batelada.	45
Figura 7 – Comparação entre o modelo MM-IP e dados experimentais no ensaio de batelada alimentada.	46
Figura 8 – Comparação entre o modelo MM-IP <i>fuzzy</i> e dados experimentais no ensaio de batelada.	46
Figura 9 – Comparação entre o modelo MM-IP <i>fuzzy</i> dados experimentais no ensaio de batelada alimentada.	47
Figura 10 – Matriz de correlação para o modelo MM-IP <i>fuzzy</i>	48
Figura 11 – Valores de sensibilidade média para os conjunto de parâmetros ACS (esquerda) e BCS (direita) no modelo <i>fuzzy</i> linear.	48
Figura 12 – Comparação entre o modelo MM <i>fuzzy</i> sigmoide e dados experimentais no ensaio de batelada.	50
Figura 13 – Comparação entre o modelo MM <i>fuzzy</i> sigmoide e dados experimentais no ensaio de batelada alimentada.	51
Figura 14 – Matriz de correlação para o modelo MM <i>fuzzy</i> sigmoide	51
Figura 15 – Valores de sensibilidade média para os conjunto de parâmetros ACS (esquerda) e BCS (direita) no modelo <i>fuzzy</i> sigmoide.	52
Figura 16 – Comparação entre o modelo MM <i>fuzzy</i> sigmoide e dados experimentais de validação.	53
Figura 17 – Concentração calculada de sólidos reativos ao longo dos experimentos de validação.	54
Figura 18 – Espectros NIR brutos (esquerda) e amostras selecionadas (direita).	55
Figura 19 – Espectros após o corte (esquerda) e após os filtros (direita).	55
Figura 20 – Variáveis selecionadas para os modelos 5 e 6.	56
Figura 21 – Gráficos das concentrações de referência (HPLC) x concentrações preditas para a validação cruzada do modelo 6.	57

Figura 22 – Predição das concentrações de celobiose, glicose e xilose pelo modelo 6 para o ensaio 2 (Batelada).	58
Figura 23 – Predição das concentrações de celobiose, glicose e xilose pelo modelo 6 para o ensaio 3 (Batelada alimentada de ezima).	58
Figura 24 – Predição das concentrações de celobiose, glicose e xilose pelo modelo 6 para o ensaio 4 (Batelada alimentada de enzima).	59
Figura 25 – Predição das concentrações de celobiose, glicose e xilose pelo modelo 6 para o ensaio 5 (Batelada alimentada de bagaço).	59

Lista de tabelas

Tabela 1 – Métodos de pré-tratamento (adaptado de (ZHAO et al., 2022)).	26
Tabela 2 – Condições iniciais estimadas dos ensaios.	34
Tabela 3 – Número total e combinações identificáveis de acordo com o número de parâmetros considerados no modelo MM-IP <i>fuzzy</i>	49
Tabela 4 – Número total e combinações identificáveis de acordo com o número de parâmetros considerados no modelo MM com <i>fuzzy</i> sigmoide.	52
Tabela 5 – Modelos PLS ajustados	56
Tabela 6 – Modelos PLS ajustados com seleção de variáveis	56
Tabela 7 – Parâmetros ajustados para o modelo MM-IP <i>fuzzy</i> linear	69
Tabela 8 – Parâmetros ajustados para o modelo MM <i>fuzzy</i> sigmoide	70

Sumário

1	INTRODUÇÃO	19
2	OBJETIVOS	21
3	REVISÃO BIBLIOGRÁFICA	23
3.1	Bagaço de cana de açúcar como matéria prima para etanol 2G	23
3.1.1	Celulose	24
3.1.2	Hemicelulose	24
3.1.3	Lignina	24
3.2	Pré-tratamento	24
3.3	Hidrólise de biomassa	25
3.3.1	Hidrólise enzimática	27
3.3.2	Estado-da-arte e desafios da hidrólise enzimática	27
3.4	Sistema de reatores	27
3.5	Modelagem matemática	28
3.5.1	Modelo cinético semi-mecanístico	28
3.5.2	Lógica <i>fuzzy</i>	30
3.6	Monitoramento espectroscópico e calibração multivariada	31
3.6.1	Lei de Lambert-Beer	31
3.6.2	Espectroscopia do infravermelho próximo	32
4	MATERIAIS E MÉTODOS	33
4.1	Coleta de dados dos ensaios	33
4.1.1	Condições experimentais	33
4.1.2	Aparato analítico	34
4.2	Modelo Cinético	34
4.2.1	Ajuste dos parâmetros cinéticos	34
4.2.2	Desenvolvimento do modelo cinético	35
4.2.3	Abordagem <i>fuzzy</i> cinética	36
4.2.4	Identificabilidade dos parâmetros	37
4.3	Modelagem espectroscópica	39
4.3.1	Calibração multivariada	39
4.3.1.1	Validação cruzada	40
4.3.2	Pré-tratamento dos dados espectrométricos	41
4.3.2.1	Alisamento por média móvel	42
4.3.2.2	Filtro de Savitzky–Golay com primeira derivada	42

4.3.2.3	Seleção de variáveis por algoritmo genético	42
5	RESULTADOS E DISCUSSÃO	45
5.1	Modelo cinético	45
5.1.1	Ajuste inicial	45
5.1.2	Análise de identificabilidade	48
5.1.3	Reidentificação paramétrica	49
5.1.4	Validação do modelo	53
5.2	Monitoramento espectroscópico	54
5.2.1	Curva de predição de concentrações do modelo PLS	57
6	CONCLUSÕES E PRÓXIMOS PASSOS	61
	REFERÊNCIAS	63
	APÊNDICE A – PARÂMETROS DOS MODELOS <i>FUZZY</i>-CINÉTICOS AJUSTADOS.	69

1 Introdução

A extensiva extração e utilização de combustíveis fósseis configura-se como um dos grandes catalisadores do desenvolvimento e prosperidade da humanidade nos últimos séculos. A natureza finita de tais recursos, aliados ao impacto ambiental associado ao uso destes, traz para a comunidade internacional um debate sobre a sua disponibilidade (CHENEY; HAWKES, 2007). Nesse sentido, é consensual entre a maioria dos cientistas que a extração de petróleo já atingiu seu pico, ou o atingirá em um futuro próximo, diminuindo gradativamente ao longo do século XXI (ABAS; KALAIR; KHAN, 2015). Em meio a esta problemática, o crescimento populacional e a crescente demanda por energia, levam à busca por fontes energéticas renováveis. Uma possível solução para substituir parte dos combustíveis fósseis encontra-se no uso do bioetanol (PATTANAİK et al., 2019).

Tendo em vista essa alternativa, no Brasil já é desenvolvida e aplicada em larga escala a produção de etanol a partir da cana-de-açúcar, devido a programas de incentivo governamentais (BERTRAND; DUSSAP, 2022). Como subproduto deste processo há o bagaço de cana, que consiste em uma biomassa de composição lignocelulósica formada após a moagem (CARVALHO et al., 2013). O bagaço é composto principalmente pelos polímeros lignina, celulose e hemicelulose, que por sua vez são compostos principalmente por carboidratos. Esses compostos, portanto, poderiam ser utilizados para produzir mais bioetanol. Tais substâncias, no entanto, não são fermentadas a partir dos processos de primeira geração, sendo necessário adequar o bagaço para que este possa ser fermentado (SOUZA et al., 2012).

Um dos principais processos para a fermentabilidade é a hidrólise destes polímeros, decompondo-os em monômeros fermentáveis. Para tal, há duas vias: a hidrólise ácida e a enzimática. Destas, a segunda é dita mais vantajosa, por minimizar o impacto ambiental de resíduos, apresentar maior rendimento além de operar sob condições ambientes (LIMAYEM; RICKE, 2012). A via enzimática, no entanto, é descrita como um dos estágios mais críticos para a competitividade do etanol lignocelulósico por apresentar desafios ainda não contornados, dentre eles estão a baixa concentração de glicose no material hidrolisado, alto custo, elevados gasto energético e tempo reacional (CHANDEL et al., 2018; SUAREZ et al., 2014).

Sendo assim, para o estabelecimento da biorefinaria de segunda geração é necessária a otimização deste processo, tanto na operação quanto na arquitetura do reator. Para atacar esta problemática foi desenvolvido por Furlong um sistema de reatores para a hidrólise enzimática, com os objetivos de operação e alimentação automáticas, além de permitir escalabilidade em volume e minimizar o gasto energético através da separação de

fases sólido-líquido (FURLONG, 2019).

A complexidade do sistema e do processo, no entanto dificultam o monitoramento direto dos produtos de interesse, sendo necessárias retiradas de amostras para caracterização, inviabilizando o controle por instrumentação direta. Para contornar tal problema, pode-se aplicar um *soft-sensor*. Este, por sua vez, é composto por um software capaz de prever variáveis de processo através de dados obtíveis de maneira indireta, como espectros de infravermelho e ultravioleta (SAGMEISTER et al., 2013).

Neste cenário, o presente trabalho visa aprimorar o monitoramento dos produtos da hidrólise enzimática de bagaço pré-tratado no sistema de reatores desenvolvido por Furlong, sendo estes: glicose, celobiose e xilose. Para tal, foi investigada a modelagem cinética e a utilização de métodos de regressão a fim de permitir o desenvolvimento de um *softsensor* para futuro uso em estimadores de estado a fim de permitir um melhor controle sobre o processo.

2 Objetivos

Como delineado no capítulo anterior, o objetivo deste projeto é aprimorar a descrição e o monitoramento da hidrólise enzimática ocorrendo sob separação de fases no reator desenvolvido pelo grupo de pesquisa. Para alcançar esse propósito, dados provenientes de ensaios realizados anteriormente no grupo de pesquisa, com diferentes configurações de alimentação, obtidos por meio de cromatografia líquida de alta eficiência (*High Performance Liquid Chromatography*, HPLC) e espectroscopia no infravermelho próximo (*Near infrared*, NIR) foram organizados, processados e analisados para a construção de um sensor virtual para monitorar a concentração dos principais produtos da hidrólise (glicose, celobiose e xilose).

Como objetivos específicos, destacam-se:

- Desenvolvimento de um modelo cinético para descrição da hidrólise no reator sob diferentes cargas de sólidos.
- Análise da identificabilidade do modelo cinético obtido, analisando sua aplicabilidade no processo.
- Pré-tratamento dos dados espectroscópicos a fim de minimizar os possíveis erros sistemáticos e ruídos.
- Desenvolver um modelo de PLS capaz de prever as concentrações de celobiose, glicose e xilose de bagaço pré-tratado hidrotermicamente a partir de espectros de infravermelho próximo
- Avaliar a precisão do modelo através de técnicas de validação cruzada

3 Revisão Bibliográfica

3.1 Bagaço de cana de açúcar como matéria prima para etanol 2G

O Brasil destaca-se no cenário global como principal produtor de cana de açúcar, tendo produzido, em 2023, 677,6 milhões de toneladas (CONAB, 2023). Nesse cenário, a indústria sucroalcooleira desempenha um papel crucial, não apenas pela significativa produção de açúcar e biocombustíveis, mas também pela geração de grandes quantidades de subprodutos, destacando-se principalmente o bagaço, obtido após a moagem. São produzidos 140 kg de bagaço para cada tonelada de cana processada, sendo que essa biomassa é predominantemente queimada para atender às demandas energéticas (PIPPO et al., 2010; PEREIRA et al., 2015).

É fundamental compreender que, no contexto específico da indústria sucroalcooleira, o bagaço de cana de açúcar constitui uma forma valiosa de biomassa lignocelulósica. Sua relevância vai além da produção de energia, tendo em vista sua composição complexa. O conceito de biomassa é entendido como o grupo de materiais orgânicos não fósseis de origem vegetal, animal ou microbiana (DIAS et al., 2012). Sob este contexto estão os materiais lignocelulósicos, formados pela parede celular de organismos vegetais, estes são compostos principalmente por celulose (25-50%), hemicelulose (20-30%) e lignina (10-25%), além de pequenas quantidades de cinzas, pectinas e extrativos (YOUSUF; PIROZZI; SANNINO, 2020; SUN et al., 2016). Em geral, a celulose apresenta-se em microfibras cristalinas, envoltas em uma matriz amorfa de hemicelulose e lignina, conforme a figura 1:

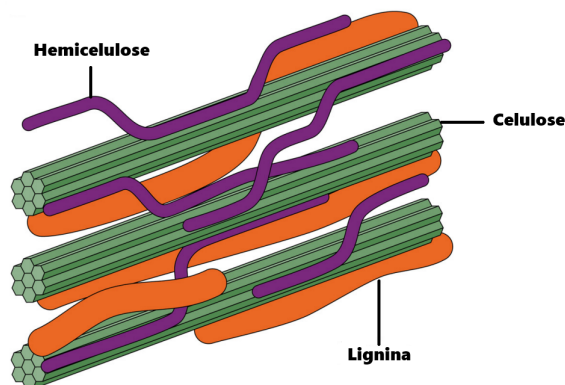


Figura 1 – Estrutura da biomassa lignocelulósica (RYTIOJA et al., 2014).

Em termos gerais, os procedimentos adotados nos processos de produção de etanol a partir de biomassa compreendem uma fase inicial de recebimento e condicionamento da matéria-prima, cuja natureza varia conforme a biomassa empregada. Em sequência, são conduzidas operações voltadas para a conversão do material lignocelulósico em açúcares

fermentescíveis. Essas operações englobam as fases de pré-tratamento e de hidrólise, respectivamente. As etapas finais do processo correspondem à fermentação dos açúcares à destilação do etanol.

3.1.1 Celulose

A celulose consiste em um polissacarídeo linear composto por moléculas de glicose unidas através de ligações glicosídicas. A estrutura da celulose é caracterizada pela presença de hidroxilas ao longo de suas cadeias, possibilitando a ocorrência de interações inter ou intracadeia por meio de ligações de hidrogênio, além das forças de van der Waals. Essa característica, aliada à alta simetria das cadeias, culmina na organização do polissacarídeo na forma de fibras, resultando na formação de zonas cristalinas (ZHANG; LYND, 2004). Essa organização única confere à celulose propriedades fundamentais que a tornam uma biomassa essencial, não apenas em termos estruturais, mas também como matéria-prima valiosa em diversas aplicações industriais e biotecnológicas.

3.1.2 Hemicelulose

A hemicelulose é composta por um grupo de polissacarídeos de cadeia irregular, de maneira que este não apresenta regiões cristalinas como as da celulose. Sua estrutura é composta por hexoses (galactose, glicose e manose), pentoses (arabinose e xilose) além de ácidos hexenurônicos (EDLUND; ALBERTSSON, 2008)

3.1.3 Lignina

A lignina, consiste em um polímero amorfo de natureza aromática (BEZERRA; RAGAUSKAS, 2016). Esta é composta principalmente pelos álcoois p-cumarílico, coniferílico e sinapílico que organizam-se ao redor das regiões cristalinas de celulose, formando uma barreira contra agentes externos (MAULIDIYAH et al., 2017). Tal barreira, por sua vez, é capaz de adsorver celulases, o que dificulta a hidrólise (ZHANG; WU; XIE, 2017).

3.2 Pré-tratamento

O processo de transformação de bagaço em bioetanol passa inicialmente por uma etapa de pré-tratamento das fibras, necessária devido a alta recalcitrância à hidrólise. Dessa maneira, a etapa do pré-tratamento da biomassa é fundamental na degradação de seus componentes para conversão em bioenergia (MANKAR et al., 2021). A fim de otimizar a capacidade das etapas subsequentes na conversão bioenergética (hidrólise e fermentação), o pré-tratamento visa à alteração das características da biomassa de maneira a minimizar as barreiras físicas e químicas impostas pela estrutura da lignocelulose. Dentre

as ações promovidas, estão principalmente: a remoção seletiva da lignina e da hemicelulose e/ou a dissolução da celulose em fase líquida diretamente (ZHAO et al., 2022). Além destes, pré-tratamentos podem ter como objetivo a redução do tamanho de partículas, solubilização de frações de hemicelulose, aumento de área superficial ou redução do grau de cristalinidade e polimerização da celulose (SILVEIRA et al., 2015).

Os diferentes pré-tratamentos podem ser classificados em três principais abordagens: a física, a química e a biológica. Além disso, determinados processos podem apresentar aspectos combinados dessas classes (SILVEIRA et al., 2015). A escolha do processo varia conforme a matéria-prima e produto final desejado, de maneira que não há uma única escolha evidente, sendo observadas vantagens e desvantagens para cada opção. Um resumo dos diferentes pré-tratamentos e suas características pode ser analisado na tabela 1.

Para o presente trabalho, foi estudada a hidrólise de um bagaço de cana pré-tratado hidrotermicamente. Este encaixa-se como método térmico, consistindo em submeter a biomassa à água em estado líquido a altas pressões por períodos de tempo de 1 até 10 minutos em temperaturas entre 180 e 240°C. Este processo, além de enfraquecer a estrutura da lignocelulose, causa a liberação de inibidores como furfural e 5-hidroximetilfurfural (PAULA, 2020).

3.3 Hidrólise de biomassa

A etapa da hidrólise da biomassa lignocelulósica visa à quebra das cadeias de celulose e hemicelulose em hexoses e pentoses, que por sua vez constituem os substratos na etapa da fermentação para a produção de etanol (PRUNESCU; SIN, 2013). Dentro do contexto de viabilidade econômica do etanol de segunda geração, a hidrólise deve ocorrer sob alta carga de sólidos (mínimo de 15 a 20% m/m), de maneira a apresentar uma alta concentração de hidrolisados (8% m/m em média) a fim de permitir a operação eficiente das etapas subsequentes (BATTISTA et al., 2018).

Para a hidrólise, há duas abordagens: a química e a enzimática. A primeira consiste no uso de soluções ácidas inorgânicas, principalmente ácido sulfúrico, sobre a biomassa pré-tratada (EL-ZAWAWY et al., 2011). Esta alternativa acarreta em um sério problema de tratamento de resíduos, além de manutenção do reator por corrosão deste (SU et al., 2020). A hidrólise enzimática, por sua vez, baseia-se na aplicação de coquetéis de enzimas, cada qual reponsável por uma porção do substrato lignocelulósico. Apesar de apresentarem um maior custo em relação aos ácidos inorgânicos, a via enzimática apresenta menor produção de subprodutos tóxicos e menor risco de corrosão do reator. Sendo assim, uma alternativa mais interessante no sentido da minimização de impacto ambiental, assim como no próprio rendimento de etanol (LIMAYEM; RICKE, 2012).

Tabela 1 – Métodos de pré-tratamento (adaptado de (ZHAO et al., 2022)).

Método		Vantagens	Desvantagens
Físico	Mecânico	sem produção de inibidores, operação simples	Custo e consumo energético altos
	Ultrassom	Baixo consumo energético, sem formação de inibidores	Baixa seletividade
	Térmico	Baixa poluição ambiental	Formação de inibidores, Altas temperaturas
Químico	Ácido	Baixo tempo de reação	Alta toxicidade e corrosividade, formação de inibidores, alto custo
	Alcalino	Baixa produção de inibidores	Poluição ambiental, longa duração, alto consumo energético
	Oxidativo	Baixa poluição ambiental, Alta seletividade	Formação de compostos tóxicos e inibidores
	Organosolv	Baixo tempo de reação	Alta toxicidade e corrosividade, Alta poluição ambiental
Biológico	Fúngico	Condições amenas de operação, Baixo consumo energético e poluição ambiental	Longa duração, Baixa recuperação de celulose
	Bacterial	Baixo consumo energético, baixa produção de compostos tóxicos, pouca poluição ambiental	Altos tempo de incubação e pré-tratamento
	Termite	Condições amenas de operação, baixo consumo energético e impacto ambiental	Longa duração
Físico-químico	Explosão à vapor	Baixo custo, baixa poluição ambiental,	Alto gasto energético e altas temperaturas
	Calor alcalino	Baixo custo capital	Altas pressões e temperatura
	Expansão de fibra de amônia	Condições amenas de operação, sem produção de compostos inibidores	Altas pressões, temperaturas, consumo energético e poluição ambiental
	Extrusão	Operação simples, baixa poluição ambiental	

3.3.1 Hidrólise enzimática

Para a hidrólise da fração de celulose do material lingo-celulósico, as principais enzimas utilizadas são as celulases. Tais enzimas são classificadas: as endoglicanases, as celobiohidrolases, β -glicosidases e monooxigenases (SANCHEZ, 2009). Estas enzimas trabalham em conjunto de modo que as endoglicanases atuam nas regiões internas e amorfas da cadeia de celulose formando oligômeros. Na sequência, as celobiohidrolases são responsáveis pela degradação dos oligômeros e liberação de celobiose. A β -glicosidase entra então para hidrolisar a celobiose, produzindo por fim glicose como produto. Já para as monooxigenases, seu principal meio de atuação é possivelmente o trabalho em sinergia com as outras enzimas produzindo uma maior quantidade de extremidades livres para hidrólise posterior (TAHA et al., 2016).

Para realização da hidrólise da hemicelulose, em comparação com a celulose, é necessário um coquetel mais complexo de enzimas, devido à sua natureza mais heterogênea. Para a sua degradação em pentoses, hexoses e ácidos urônicos, a hemicelulose deve ser submetida a uma combinação de xilanases, esterases e arabinofuranosidases em proporções específicas (SAHA, 2003; TAHA et al., 2016).

Por fim, a complexidade estrutural da lignina requer o trabalho de um coquetel específico de enzimas para o enfraquecimento desta, permitindo que as enzimas hidrolíticas tenham acesso às cadeias de polissacarídeos. Para tal, utilizam-se principalmente lignina-peroxidases, manganês-peroxidases e laccases (PEREZ et al., 2002; TAHA et al., 2016).

3.3.2 Estado-da-arte e desafios da hidrólise enzimática

Apesar de vantajosa, a hidrólise enzimática enfrenta desafios significativos para sua aplicação em larga escala. Dentre eles, o equilíbrio entre o rendimento da conversão de biomassa em monossacarídeos (maximizada a baixas cargas de sólidos) e a alta concentração de hidrolisado necessária (obtida sob maior carga de sólidos), que acarreta na dificuldade de obter-se um produto viável economicamente. Além disso, a natureza heterogênea da biomassa pré-tratada resulta em uma cinética lenta, assim como um alto gasto energético na mistura e agitação do substrato (FURLONG, 2019; OLIVIERI et al., 2021).

3.4 Sistema de reatores

Com base nos desafios apresentados na seção anterior, um sistema de reatores desenvolvido durante o trabalho de doutorado de Vitor Badiale Furlong foi utilizado neste projeto. Este sistema consiste na utilização de um reator horizontal de batelada alimentada com alta fração de sólidos seguido por um reator de tanque agitado, para a fração líquida, conforme a figura 2. Uma das principais características propostas consiste na separação

das fases líquida e sólida, permitindo tanto menores gastos energéticos quanto ganhos na cinética da reação (FURLONG, 2019).

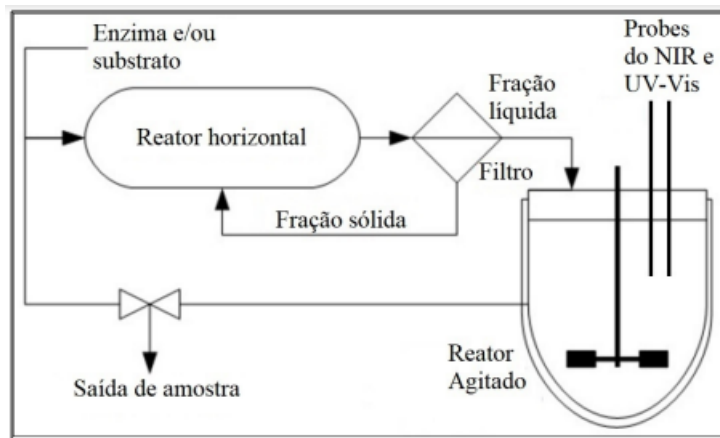


Figura 2 – Representação esquemática do sistema de reatores utilizado (FURLONG, 2019).

O menor custo energético do processo no reator origina do desacoplamento dos mecanismos de agitação para cada respectiva fase. Isso se dá a partir da suposição de que, em reatores horizontais tradicionais, a agitação do líquido é responsável pelo maior gasto de energia, apesar de não necessitar da mesma intensidade de agitação (JIN et al., 2010; LIU; ZHANG; BAO, 2015; FURLONG, 2019).

Outra vantagem da separação é a maior facilidade na retirada de hidrolisados. Carboidratos fermentescíveis (glicose e xilose) podem ser removidos do sistema, minimizando assim efeitos inibitórios causados por acúmulo de produto. Além disso, a fração sólida possibilita a fabricação de outros produtos de alto valor agregado, como compostos de nanocelulose (FARINAS; MARCONCINI; MATTOSO, 2018).

Na prática, a separação de fases é obtida através de uma malha de aço inoxidável que reveste as paredes do reator de sólidos, permitindo a passagem de líquido, que por sua vez é recuperado em uma rampa e redirecionado para um reator de tanque agitado, conforme a figura 3. Esta configuração permite a separação de fases sem a necessidade de energia adicional

3.5 Modelagem matemática

3.5.1 Modelo cinético semi-mecanístico

O uso de modelos cinéticos para projetar biorreatores impulsionados pela hidrólise enzimática, especialmente na sacarificação de biomassa lignocelulósica, é complexo principalmente devido a fatores como heterogeneidade do substrato e atuação conjunta de um grupo de enzimas (OLIVIERI et al., 2021). Embora vários modelos tenham sido

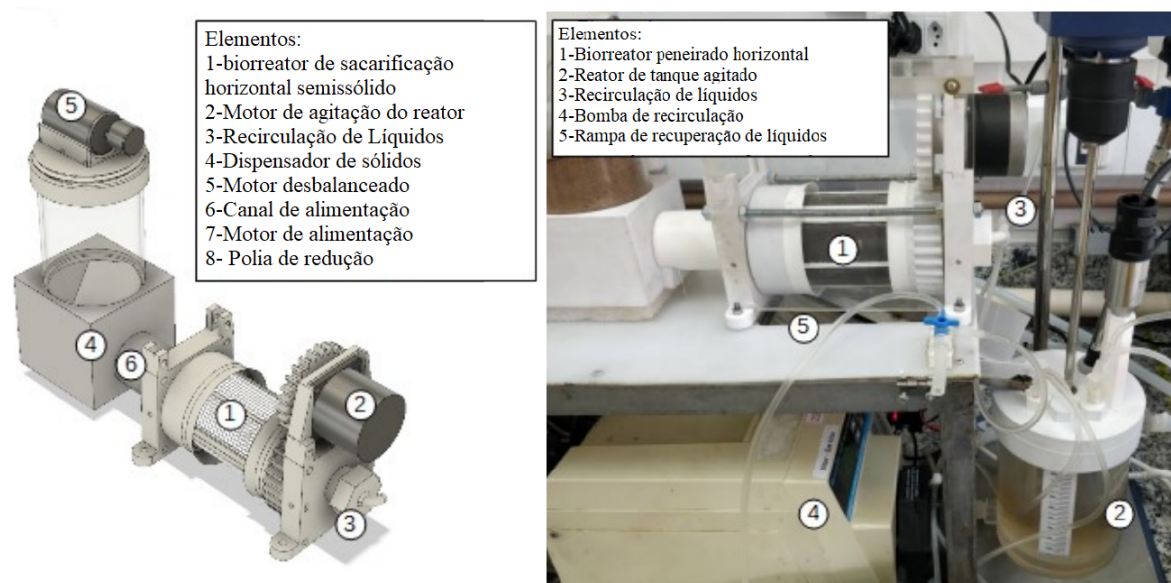


Figura 3 – Projeto do reator de sólidos (esquerda) e protótipo do sistema utilizado (direita) (FURLONG, 2019).

propostos para entender e descrever a hidrólise em operações de alto teor de sólidos, este muitas vezes não capturam completamente o comportamento dinâmico do sistema ao longo de períodos prolongados, como em operações de batelada alimentada. Adicionar complexidades a esses modelos, seja por modificações estruturais ou funcionais, introduz mais parâmetros. Quando derivados dos mesmos dados empíricos, esses parâmetros podem se correlacionar, perdendo seus significados físicos intrínsecos. Apesar desses obstáculos, há uma variada gama de modelos para o processo, podendo estes ser classificados em: empíricos, mecanísticos e semi-mecanísticos.

Dentre estes, modelos semi-mecanísticos diferem de mecanísticos no grau de detalhamento empregado, utilizando menores quantidades de parâmetros e equações, assim como um maior número de simplificações para a representação do sistema (NETO, 2016). A abordagem semi-mecanística, portanto, por apresentar estrutura simples e ser capaz de fornecer valores sem grande esforço computacional, será utilizada neste trabalho.

Os modelos semi-mecanísticos mais amplamente utilizados para hidrólise enzimática de biomassas lignocelulósicas são os modelos pseudo-homogêneos de Michaelis-Menten (MM). A natureza heterogênea da reação de despolimerização, no entanto, invalida algumas das hipóteses de transporte de massa utilizadas na cinética clássica de Michaelis-Menten, baseada em sistemas homogêneos de reação (KADAM; RYDHOLM; MCMILLAN, 2004). A condição de substrato em excesso em relação à enzima, por exemplo, é descartada uma vez que a fração de ligações glicosídicas disponíveis não atinge os níveis necessários para estabelecer excesso. Além disso, a degradação da celulose é descrita como um processo unidimensional, no qual as celobiohidrolases percorrem a molécula em busca de sítios ativos (LYND et al., 2002).

Sendo assim, entre as abordagens mais comuns na modelagem das velocidades de reação da hidrólise estão: a adsorção da enzima solúvel ao substrato sólido e a inibição competitiva de produto (PRATTO et al., 2015; PROCENTESE et al., 2020). Ao considerar a adsorção enzimática sob o substrato sólido ($[S]$) supõe-se que, em estado estacionário, a concentração de enzima ($[E]$) que apresenta excesso. Sendo assim, a equação de velocidade de reação para este modelo modificado de Michaelis-Menten (MMM) pode ser descrita conforme a equação 3.1 (BAILEY, 1989; CARRILLO et al., 2005):

$$\alpha_i = \frac{k_i \cdot [E] \cdot [S_i]}{Km_i + [E]} \quad (3.1)$$

Nesta, α ($g.L^{-1}.s^{-1}$) denota a velocidade de reação, Km_i ($g.L^{-1}$) é a constante de MM, representando neste caso metade da constante de saturação enzimática, k_i ($g.L^{-1}.s^{-1}$) é a constante cinética. Em adição a estas alterações, a utilização de inibição competitiva de produto ($[P]$) é um fator importante na modelagem, modificando portanto a equação 3.1 ao introduzir uma constante Kp_i ($g.L^{-1}$) de inibição conforme a equação 3.2 (CARVALHO et al., 2013).

$$\alpha_i = \frac{k_i \cdot [E] \cdot [S]}{Km_i \cdot (1 + \frac{[P_i]}{Kp_i}) + [E]} \quad (3.2)$$

A utilização de modelos semi-mecanísticos, apesar de apresentar um equilíbrio entre complexidade e capacidade descritiva, pode não ser capaz de descrever completamente a mudança no comportamento do sistema durante todo o processo. A adição de complexidade ao modelo, reconhecendo outros efeitos na sua formulação matemática exige mais parâmetros, o que pode acarretar em uma alta correlação entre estes e, como resultado, o modelo perde significância física.

Uma alternativa proposta consiste em utilizar diferentes modelos dentro do mesmo processo. Liu *et al.* propôs que a hidrólise ocorre sob duas dinâmicas separadas: a liquefação, quando a concentração de sólidos na suspensão é muito alta, e a sacarificação, quando a maior parte do substrato foi liquefeito (LIU; ZHANG; BAO, 2015). Estes mecanismos, no entanto, podem ocorrer de maneira simultânea, como em uma batelada alimentada, na qual sólidos são alimentados ao longo do processo. Por conta disso, apenas uma classe de modelo pode ser insuficiente para descrever a cinética completa da reação, sendo necessários diferentes métodos de inteligência computacional para descrever o sistema em condições sob as quais ambas as dinâmicas ocorrem (FURLONG, 2019).

3.5.2 Lógica fuzzy

Sob essa circunstância de uso da inteligência computacional para descrever o comportamento do sistema com múltiplos mecanismos reacionais, a lógica *fuzzy* emerge

como uma possível ferramenta crucial. Na lógica *fuzzy* os valores veritativos de uma variável podem ser parcialmente verdadeiros, empregando regras de inferência para modelar um determinado sistema (NOVÁK; PERFILEVA; MOCKOR, 2012). Em um sistema desse tipo, os elementos possuem graus de pertinência, caracterizados por uma função, que podem assumir quaisquer valores reais entre zero e um. Nestes sistemas, o processo de inferência é guiado por um conjunto de regras *IF-THEN* ('se-então') que estabelecem as relações entre as variáveis de entrada e saída. Cada regra *fuzzy* é composta por duas partes: a antecedente, na qual as condições para a ativação da regra são descritas, e a consequente, descrevendo a conclusão da regra (MEHRAN, 2008).

Dentre os principais métodos de modelagem *fuzzy* destaca-se o modelo proposto por Takagi e Sugeno (TS) (TAKAGI; SUGENO, 1985). Este modelo é delineado por meio de regras *fuzzy*, para as quais a saída do modelo é definida como uma função das variáveis de entrada, com base nos valores das funções de pertinência dos conjuntos dos antecedentes. Sua principal característica consiste na capacidade de expressar a dinâmica local de cada regra através de um modelo de sistema linear. Sendo assim, a configuração global do modelo resultante é alcançada através da combinação dos modelos lineares.

Dessa maneira, é possível representar diferentes as diferentes dinâmicas do processo ocorrendo simultaneamente. Ao descrever modelos individuais para cada dinâmica predominante, uma regra *fuzzy* pode ser utilizada para pesar entre os modelos individuais em regiões nas quais ambas ocorrem com diferentes intensidades (AL-HADITHI; JIMÉNEZ; MATÍA, 2012).

3.6 Monitoramento espectroscópico e calibração multivariada

No contexto da hidrólise de bagaço, os espectros de absorbância derivados da biomassa abrangem uma variedade de compostos, manifestados por distintas bandas de absorção. Portanto, é factível empregar a calibração multivariada, estabelecendo relações entre os dados de absorbância das amostras e um ou mais conjuntos de dados, a fim de quantificar os compostos presentes em cada amostra (ANDRADE et al., 2003).

3.6.1 Lei de Lambert-Beer

A base física para o cálculo de concentrações na calibração multivariada consiste na lei de Lambert-Beer. Esta enuncia que, em condições ideais, isto é, sem espalhamento de luz e em baixa concentração de analitos, a absorbância de um raio de luz de comprimento de onda λ é proporcional à concentração de um analito, conforme a equação 3.3 (ROGER; MALLET; MARINI, 2022):

$$A_0(\lambda) = -\log\left(\frac{I(\lambda)}{I_0(\lambda)}\right) = \epsilon(\lambda)LC \quad (3.3)$$

Onde $A_0(\lambda)$ representa a absorbância do analito quando a onda incidente possui comprimento λ , $I(\lambda)$ e $I_o(\lambda)$ correspondem à intensidade da luz transmitida no mesmo comprimento de onda, L representa o caminho óptico, ϵ é a absortividade molar do analito, sendo a concentração do analito denotada por C .

As condições ideais, no entanto, raramente ocorrem. Sob condições usuais, uma variedade de fenômenos impede a aplicação da equação 3.3. Além dos ruídos normalmente presentes, o espalhamento da luz é causa dos principais problemas: o alongamento do caminho óptico, e a perda de fótons para o meio. Ambos fenômenos introduzem termos que afetam a lei de Lambert-Beer, estes termos são multiplicativos e aditivos, respectivamente. Sendo assim, para manter-se algum nível de precisão e confiabilidade em dados espectrométricos, é crucial um pré-tratamento adequado dos dados (ROGER; MALLET; MARINI, 2022). Além disso, a linearidade entre a absorbância e concentração de analito só é válida sob certas faixas de concentração (SKOOG; HOLLER; CROUCH, 2017; RIBEIRO, 2007). Por fim, a sobreposição espectral causada pela presença de múltiplos componentes resulta em um alargamento de bandas no espectro de absorbância, dificultando a aplicação de métodos que considerem apenas um analito (CORRER et al., 2005).

3.6.2 Espectroscopia do infravermelho próximo

A espectroscopia do infravermelho próximo (NIR) consiste em um método baseado na interação da radiação eletromagnética com a matéria, mais especificamente na faixa dos 700 aos 2500 nm, produzindo respostas detectáveis (SKVARIL; KYPRIANIDIS; DAHLQUIST, 2017). A interação medida é baseada na transição dos elétrons entre níveis de energia quando expostos a uma onda eletromagnética, sendo quantificada a razão entre a radiação incidente e a radiação absorvida. As transições ocorrentes dependem da estrutura eletrônica do material, e portanto de sua composição (PASQUINI, 2003).

Essa interação decorre nas mudanças em níveis de energia vibracionais. Tais mudanças produzem respostas na região do espectro NIR principalmente em ligações moleculares não simétricas como O-H, C-H, C-O, S-H, N-H e C=O (CEN; HE, 2007; ROGER; MALLET; MARINI, 2022). Desse modo, a espectroscopia NIR apresenta-se como uma ferramenta importante para a quimiometria em processos envolvendo biomassa lignocelulósica e seus constituintes (SKVARIL; KYPRIANIDIS; DAHLQUIST, 2017).

4 Materiais e métodos

4.1 Coleta de dados dos ensaios

A presente pesquisa baseia-se na análise e monitoramento da hidrólise enzimática de bagaço de cana de açúcar, utilizando dados provenientes de ensaios realizados anteriormente pelo grupo de pesquisa no sistema de reatores descrito na seção 3.4. Além dos ensaios documentados em sua tese, Furlong realizou um total de seis ensaios adicionais sob diferentes condições experimentais. Este trabalho gerou uma grande quantidade de dados brutos, incluindo espectros de infravermelho próximo (NIR), dados de cromatografia líquida de alta eficiência (HPLC), anotações de laboratório e outras informações.

Este trabalho utilizou esses dados para as avaliações cinéticas e monitoramento espectroscópico. Os dados foram inicialmente reunidos, organizados e tratados, verificando-se a consistência destes e removendo *outliers*. As anotações de laboratório foram analisadas para obtenção dos balanços de massa e outras informações relevantes, culminando na representação dos dados em função do tempo.

4.1.1 Condições experimentais

Nos seis ensaios em estudo, três perfis de alimentação foram utilizados: batelada, batelada com alimentação de enzimas e batelada alimentada de bagaço e enzimas, todos com 48 horas de duração. Cada um utiliza-se de 20 gramas de biomassa pré-tratada hidrotérmicamente sob as mesmas condições detalhadas na tese, sendo sua composição mássica determinada: $60,1 \pm 0,2\%$ de celulose, $4,76 \pm 0,5\%$ de hemicelulose, $27,3 \pm 0,3\%$ de lignina e $1,1 \pm 0,1\%$ de cinzas. O coquetel enzimático utilizado também foi o mesmo: o CELIC CTEC 2, doado pela Novozymes América Latina, operado sob um tampão de citrato com pH 4,8 e força iônica de 50mM.

Além da composição e quantidade do substrato, outras condições experimentais foram mantidas em relação ao trabalho original: a carga enzimática ($3,7 mg_{enz}/g_{sólidos}$), a carga de sólidos (15%). Nos ensaios com alimentação de enzima, assim como na batelada alimentada, metade da carga enzimática foi adicionada no começo da reação e o restante foi alimentado ao longo das 36 primeiras horas de ensaio com uma vazão volumétrica de $0,21 mL/min$ e $0,23 mL/min$ respectivamente. A batelada alimentada de bagaço, por sua vez, foi iniciada com 5g de bagaço e os 15g foram adicionados a uma vazão volumétrica de $7 mg/min$ sendo interrompidos juntamente com a alimentação do complexo enzimático após 36 horas de ensaio. A partir dessas informações, foram estimadas as concentrações iniciais de celulose e hemicelulose para cada ensaio conforme a tabela 2:

Tabela 2 – Condições iniciais estimadas dos ensaios.

Ensaio	Modo de operação	massa seca inicial de substrato (g)	Volume inicial no reator (mL)	Celulose inicial calculada (g/L)	Hemicelulose inicial calculada (g/L)
1	Batelada	20	513,32	23,37	1,87
2		20	613,32	19,56	1,57
3	Batelada com alimentação de enzima	20	613,32	19,56	1,57
4		20	613,32	19,48	1,57
5	Batelada	5	528,33	5,69	0,45
6	alimentada	5	528,33	5,44	0,45

4.1.2 Aparato analítico

Os dados de cromatogramas foram obtidos em um HPLC Shimadzu SCL-10A® com detector de índice de refração Shimadzu RID10-A®, uma coluna Aminex HPX-87H® Bio-rad e ácido sulfúrico 5mM a 0,6mL/min como fase móvel, com amostragens feitas a cada 2 horas (SLUITER et al., 2006). Já para os espectros de infravermelho próximo, foi utilizada uma sonda de refletância Falcata 6 acoplada a um Frontier NIR Standard da Perkin Elmer®, registrando varreduras a cada 6 minutos em uma faixa de número de onda de 10000cm⁻¹ a 1200cm⁻¹ com uma resolução de 4cm⁻¹ (FURLONG, 2019).

4.2 Modelo Cinético

4.2.1 Ajuste dos parâmetros cinéticos

Durante a estimação de parâmetros, busca-se que o modelo descreva os dados experimentais com precisão, minimizando o erro de previsão. Este é geralmente abordado como um problema de otimização, com o critério de mínimos quadrados comumente utilizado, conforme a equação 4.1 (KRISTENSEN, 2002).

$$\min_{\theta} F = \sum_{l=1}^L W_l \sum_{i=1}^{I_l} w^2 \sum_{n=1}^{N_{li}} (Y_{lin} - \bar{y}_{lin})^2 \quad (4.1)$$

Para a resolução deste problema, uma vasta gama de métodos existem. Para a obtenção dos parâmetros cinéticos apresentados neste trabalho foi utilizada uma abordagem meta-heurística através do algoritmo de recozimento simulado (*Simulated Annealing*, SA). A escolha dessa se baseia na alta dimensionalidade do espaço de busca, que consiste em dezenas de parâmetros, de maneira que seu emprego desempenha melhor no ajuste de modelos complexos quando comparados a métodos mais diretos, como Levenberg–Marquardt (EFTAXIAS et al., 2002).

Inspirado no processo físico de têmpera em metal, o SA simula o resfriamento lento de um sistema. O algoritmo funciona mantendo um estado atual e iterativamente realizando perturbações nesse estado de acordo com uma distribuição probabilística. Se uma perturbação leva a uma solução melhor, ela é aceita automaticamente. Caso contrário,

a perturbação é aceita com uma probabilidade que diminui ao longo do tempo, permitindo que o algoritmo escape de mínimos locais. Para a obtenção dos intervalos de confiança do ajuste, utiliza-se o método de Monte-Carlo, uma vez que o cálculo da covariância resulta frequentemente em intervalos contendo valores nulos (KRAUSCH et al., 2019).

4.2.2 Desenvolvimento do modelo cinético

Para a descrição das reações do processo, foram consideradas as hidrólises da celulose e da hemicelulose do bagaço. No caso da celulose, utilizou-se de duas vias: a degradação direta em glicose e a intermediada pela celobiose. Já para a hemicelulose foi considerada apenas sua sacarificação em xilose. Além disso, inativação enzimática de primeira ordem foi também assumida. As reações consideradas estão esquematizadas abaixo conforme a figura 4 e denotadas na forma matricial em 4.2 (DOCHAIN; BASTIN, 1990).

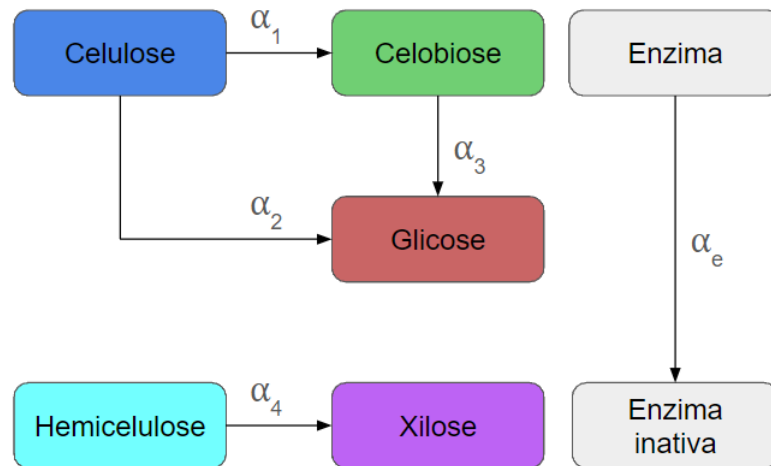


Figura 4 – Esquematização das reações de hidrólise consideradas.

$$\frac{d}{dt} \begin{bmatrix} Cl \\ Cb \\ Gl \\ He \\ Xy \\ En \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ \gamma_{Cl-Cb} & 0 & -1 & 0 & 0 \\ 0 & \gamma_{Cl-Gl} & \gamma_{Cb-Gl} & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & \gamma_{He-Xy} & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_e \end{bmatrix} \quad (4.2)$$

O lado esquerdo das equações compreende as concentrações de celulose ([Cl]), celobiose ([Cb]), glicose ([Gl]), hemicelulose ([He]), xilose ([Xy]) e atividade enzimática ([En]). A matriz 6x5 do lado direito da equação representa a matriz pseudo-estequiométrica, e o vetor (α_i) é composto pelas taxas de reação das reações de 1 a 4, conforme previamente descrito. Os valores de γ são as relações de massa pseudo-estequiométricas entre substratos e produtos para cada reação. Os valores utilizados foram: $\gamma_{Cl-Cb} = 1.056$, $\gamma_{Cl-Gl} = 1.111$,

$\gamma_{Cb-Gl} 1.056$, $\gamma_{He-Xy} = 0.841$ (KADAM; RYDHOLM; MCMILLAN, 2004; YAO et al., 2015).

Inicialmente, considerou-se para as reações 1,2 e 4, um modelo de Michaelis-menten modificado e com inibição de produto, conforme a equação 3.2 apresentada anteriormente. Para o caso da reação 3, de conversão da celobiose em glicose, o substrato foi considerado solúvel e, portanto resultando em uma velocidade de reação dada por 4.3:

$$\alpha_2 = \frac{k_3 \cdot [E] \cdot [Cb]}{Km_3 \cdot \left(1 + \frac{[Gl]}{Kp_i}\right) + [Cb]} \quad (4.3)$$

Este modelo de MM com inibição por produto (MM-IP), conforme ainda será discutido na seção de resultados, foi capaz de ajustar satisfatoriamente o comportamento dos ensaios de hidrólise. No entanto, o mesmo não foi capaz de convergir para um único conjunto de parâmetros cinéticos que explicasse simultaneamente as diferentes condições de alimentação de bagaço. Esta dificuldade sugere que há diferentes mecanismos de reação que ocorrem no sistema. Sendo assim, fez-se necessária a investigação de estratégias para descrever o comportamento da hidrólise sob as diferentes condições de alimentação do processo.

4.2.3 Abordagem *fuzzy* cinética

Visando abordar a problemática apresentada na seção anterior, uma alternativa consiste na utilização de dois conjuntos de parâmetros cinéticos ajustados para diferentes concentrações de substratos, empregando uma função de pertinência para interpolar entre os dois modelos. Tal abordagem mostrou-se eficaz na literatura para a obtenção de um único modelo semi-mecanístico simples capaz de descrever o comportamento da hidrólise sob diferentes concentrações de sólidos reativos (SR) em reatores de tanque agitado. Isso se dá a partir da ideia de que o processo de hidrólise passa por diferentes estágios de liquefação da biomassa. Além disso, quando a biomassa é alimentada ao reator, dois mecanismos podem coexistir: um conduzido pelo ataque enzimático do sólido pelas enzimas, e o outro refletindo o fato de que parte do substrato presente no reator já foi pelo menos parcialmente hidrolisada. Portanto, apenas uma simples classe de modelo pode não ser suficiente para descrever a cinética da reação (FURLONG et al., 2019).

A coexistência destes mecanismos portanto, pode explicar a dificuldade em ajustar um único modelo para todos os dados analisados. Dessa maneira, foram investigadas e empregadas técnicas computacionais para distinguir entre as regiões em que cada mecanismo predomina. Neste contexto, a lógica *fuzzy* apresenta-se como uma poderosa ferramenta, uma vez que esta permite a combinação de modelos em casos em que ambos os mecanismos ocorrem em proporções de maneira que nenhum deles prevalece (BEZDEK; PAL, 1994).

Como observado nas condições dos experimentos analisados, há uma diferença considerável na concentração de sólidos reativos (celulose e celobiose) entre as bateladas e bateladas alimentadas. Isso levanta, portanto, a hipótese de que a aplicação de uma função de pertinência *fuzzy* permitiria a descrição de ambos os comportamentos pelo modelo também neste sistema de reatores.

Com base nisso, o modelo *fuzzy cinético* inicialmente proposto por Furlong e discutido neste trabalho é composto de, essencialmente, dois modelos cinéticos modificados de Michaelis-menten com inibição competitiva por produto: um para alta concentração de sólidos reativos (AC) e outro para a baixa concentração dos mesmos (BC), cada qual com sua respectiva velocidade de reação (α^{AC} e α^{BC}) (FURLONG et al., 2019). Para a interpolação entre esses dois modelos, uma função de pertinência linear Takagi-Sugeno foi empregada, conforme a equação 4.4 (TAKAGI; SUGENO, 1985).

$$GP = \left\{ \begin{array}{ll} 0 & SR \leq ACS_{inf} \\ \frac{SR - ACS_{inf}}{ACS_{sup} - ACS_{inf}} & se \quad ACS_{inf} \leq SR \leq ACS_{sup} \\ 1 & SR \geq ACS_{sup} \end{array} \right\} \quad (4.4)$$

Nessa função, o grau de pertinência (GP) é calculado com base na concentração de sólidos reativos presente no sistema em determinado momento. Os valores de ACS_{sup} e ACS_{inf} denotam respectivamente os limites superior e inferior a partir dos quais o modelo de ACS passa a ser considerado. A partir do valor de GP é então utilizado para obter-se as velocidades finais para cada reação (α^{fuzzy}) conforme a equação 4.5.

$$\alpha^{fuzzy} = GP \cdot \alpha^{AC} + (1 - GP) \cdot \alpha^{BC} \quad (4.5)$$

Os parâmetros cinéticos do modelo utilizado foram ajustados originalmente para um reator de tanque agitado e utilizando um pré-tratamento por explosão a vapor, ambas condições diferem das presentes neste trabalho. Portanto, é previsível que estes valores não fossem capazes de descrever a hidrólise nos ensaios realizados. Dessa maneira, os dados de HPLC das bateladas foram utilizados para obtenção dos parâmetros de ACS, enquanto a batelada alimentada originou o modelo BCS. Em sequência, as informações de todos os seis experimentos foram submetidas a um novo ajuste para estimar os limites de transição da função de pertinência.

4.2.4 Identificabilidade dos parâmetros

Alcançar um conjunto de parâmetros identificáveis depende de duas condições fundamentais: primeiramente, os dados experimentais devem responder às variações individuais dos parâmetros, e, em segundo lugar, não deve haver correlação entre os parâmetros (RUBIO, 2010). A primeira condição pode ser abordada através do cálculo de uma medida

média de sensibilidade, para cada parâmetro (θ_j), facilitando a avaliação de sua importância individual. É importante observar que a remoção de parâmetros não significativos é essencial, uma vez a presença destes indica uma complexidade maior do que a necessária no modelo (KRISTENSEN, 2002). A matriz de sensibilidade é construída a partir da variação da saída do modelo (y_i) a cada parâmetro (θ_j) na solução ótima (θ) de parâmetros, conforme a equação 4.6:

$$s_{ij} = \left. \frac{dy_i}{d\theta_j} \right|_{\hat{\theta}} \quad (4.6)$$

Para que as diferentes ordens de grandeza de cada parâmetro e saída do modelo sejam normalizadas, é utilizada a forma adimensional descrita em 4.7.

$$s_{ij}^{nd} = \frac{\Delta\theta_j}{Sc_i} \frac{dy_i}{d\theta_j} \quad (4.7)$$

Onde $\Delta\theta_j$ define uma medida *a priori* da amplitude do parâmetro θ_j e Sc_i é um fator de escala com a mesma dimensão observada para as saídas do modelo. Neste estudo, os valores utilizados para o fator de escala e de amplitude foram, respectivamente, o valor médio de saída do modelo para cada substância e o próprio valor estimado dos parâmetros (BRUN; REICHERT; KÜNSCH, 2001). A medida de sensibilidade média (δ) é definida conforme a equação 4.8:

$$\delta_j = \sqrt{\frac{1}{m} \sum_{i=1}^m (s_{ij}^{nd})^2} \quad (4.8)$$

Subsequentemente, deve-se garantir que não haja correlação entre os parâmetros. Essa correlação, no entanto, frequentemente ocorre, a exemplo do observado neste trabalho. Para lidar com essas interações em modelos de maior dimensionalidade, utiliza-se a colinearidade das sensibilidades paramétricas (GABOR; VILLAVARDE; BANGA, 2017). Para calcular-se a colinearidade, primeiro normaliza-se a matriz S conforme a equação 4.9:

$$\tilde{s}_{ij} = \frac{s_{ij}^{nd}}{\|s_j^{nd}\|} \quad (4.9)$$

Onde $\|s_j^{nd}\|$ é a norma euclidiana da j-ésima coluna da matriz adimensional de sensibilidade.

O índice de Colinearidade (γ) é calculado para um subconjunto de parâmetros K, calculado com a combinação $C_k^{n_p} \in 2, 3, \dots, n_p$. O índice mede o grau de dependência linear entre as colunas da submatriz \tilde{S}_k , contendo as colunas que correspondem aos parâmetros do subconjunto K. O índice γ_k é definido como:

$$\gamma_k = \frac{1}{\sqrt{\lambda_k}} \quad (4.10)$$

Onde λ_k representa o menor autovalor do produto $\tilde{S}_k^T \tilde{S}_k$. Caso as colunas de \tilde{S}_k sejam linearmente dependentes, o valor de γ_k será grande e o subconjunto é dado como não identificável. O limite para que o subconjunto seja considerado não identificável na

literatura, e também utilizado neste estudo foi de 20 (GABOR; VILLAVERDE; BANGA, 2017).

Calcular a colinearidade de todos os subconjuntos de parâmetros possíveis permite definir quantos destes são identificáveis, a inferência de quantos parâmetros podem ser estimados a partir dos dados experimentais, além de avaliar a colinearidade entre três ou mais parâmetros. As análises de sensibilidade e identificabilidade no presente projeto foram implementadas em *Python* 3.11 com auxílio da biblioteca *NumPy* (HARRIS et al., 2020).

4.3 Modelagem espectroscópica

Os espectros NIR correspondentes aos mesmos tempos de reação dos dados cromatográficos utilizados foram utilizados nas regressões PLS para o ajuste, assim como para a validação cruzada. Os dados espectrais completos, retirados a cada 6 minutos conforme a seção 4.1.2, por sua vez, serviram para observar a capacidade preditiva da regressão otimizada.

Para garantir a qualidade dos dados espectrais brutos, os quais apresentam alta quantidade de ruído e *offset*, foi avaliada a região de comprimentos entre 4500 a 10000 cm^{-1} aplicando os pré-tratamentos de alisamento por média móvel com uma janela de 13 pontos, assim como filtro de primeira derivada de Savitzky–Golay (SG) com uma janela de 15 pontos e um polinômio de segunda ordem. Além disso, foi avaliada a utilização de Algoritmo genético para a seleção de variáveis, utilizando a minimização do $RMSE_{CV}$ como função objetivo, selecionando janelas de 10 comprimentos de onda e empregando 100 gerações e 50% de convergência de *fitness* da população como condições de parada. A implementação do ajuste PLS foi realizada através do *software* MATLAB com auxílio da ferramenta *PLS Toolbox* (INC., 2023).

4.3.1 Calibração multivariada

A calibração multivariada (CM) caracteriza um conjunto de técnicas de quantificação simultânea de múltiplos analitos. Baseando-se na lei de Lambert-Beer, a maioria dos métodos de CM permite a determinação de concentração de múltiplos compostos a partir do espectro de absorção de cada amostra (ANDRADE et al., 2003). No presente trabalho é abordado para interpretação dos dados espectrométricos, o método dos mínimos quadrados parciais (*Partial Least Squares*, PLS).

Para a formulação do problema a ser ajustado, parte-se do modelo inverso à lei de Lambert-Beer pelo método de regressão linear múltipla (*Multiple Linear Regression*,

MLR):

$$\mathbf{C}_{(m \times l)} = \mathbf{A}_{(m \times n)} \cdot \mathbf{P}_{(n \times l)} + \mathbf{E}_{(m \times l)} \quad (4.11)$$

Na qual \mathbf{A} compõe a matriz de absorvância, composta por m amostras e n comprimentos de onda, \mathbf{C} descreve as concentrações de l analitos em m amostras \mathbf{E} compreende os erros de calibração das concentrações não ajustados no modelo e, por fim, \mathbf{P} denota uma matriz de coeficientes. Desta maneira, esta última pode ser ajustada através da solução clássica de mínimos quadrados, utilizada para verificar a relação de linearidade entre a absorvância e concentração:

$$\mathbf{P}_{\text{opt}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{C} \quad (4.12)$$

Este método, no entanto, apresenta um problema relacionado à matriz a ser invertida na equação 4.11 ($\mathbf{A}^T \mathbf{A}$, de dimensão $n \times n$). A matriz em questão apresenta um *rank* que não excede m , isto é, a quantidade de comprimentos de onda utilizados em \mathbf{A} não pode exceder o número de amostras, o que apresenta uma impraticidade no método. Além disso, mesmo que o número de amostras exceda a quantidade de comprimentos de onda, problemas de co-linearidade podem aparecer, resultando em uma matriz mal condicionada (RIBEIRO, 2007).

Para contornar tal problemática, as duas principais estratégias utilizadas envolvem: a seleção de comprimentos de onda específicos para reduzir a singularidade de $\mathbf{A}^T \mathbf{A}$ ou a transformação do espectro em um novo sistema de coordenadas, reduzindo assim o número de entradas (também denominadas de variáveis latentes) no processo. No tocante a redução de dimensionalidade, destacam-se a regressão de componentes principais (PCR) e os mínimos quadrados parciais (PLS).

No método de PCR, gera-se uma matriz de projeções \mathbf{B} , que substitui a matriz de absorvância a fim de reduzir suas dimensões. A matriz \mathbf{B} é calculada de maneira a maximizar a covariância dos dados ao longo de seus eixos ortogonais. Portanto, pode-se selecionar apenas as direções de maior variância enquanto descartam-se direções de menor variância. O método PLS, por sua vez, diferencia-se do PCR por utilizar também os dados da matriz \mathbf{C} para a obtenção dos novos eixos coordenados (NELLES, 2001).

4.3.1.1 Validação cruzada

Para garantir o desempenho dos modelos obtidos através das regressões frente a dados não utilizados no ajuste lança-se mão de validação cruzada. Nesta, uma fração dos dados de ajuste é retirada do conjunto para utilização em testes ao longo dos passos de calibração. Dessa maneira, o modelo realiza uma estimativa dos dados retirados e é

calculado o erro associado a essa predição (NELLES, 2001). Este erro é calculado através da raiz quadrada do erro médio (*Root-mean-square error of cross-validation, RMSE_{CV}*), sendo este calculado para a validação cruzada através da equação 4.13.

$$RMSE_{CV}^{(i)} = \sqrt{\frac{1}{m} \sum_{k=1}^m (c_k^{(i)} - \hat{c}_k^{(i)})^2} \quad (4.13)$$

Nessa equação $c_k^{(i)} - \hat{c}_k^{(i)}$ representa respectivamente os valores de referência e preditos para as concentrações do i -ésimo analito na k -ésima amostra.

O emprego dessa técnica permite determinar o número ótimo de regressores para representar os dados de absorvância e concentração de maneira a minimizar o erro associado a tal validação. Esse ótimo é atingido quando o $RMSE_{CV}$ começa a se afastar da tendência de acompanhar o erro médio do ajuste ($RMSE_{cal}$). Esse afastamento é interpretado como o ponto a partir do qual a complexidade do modelo torna-se maior do que o necessário no sistema, e o ruído das amostras passa a ser considerado no reconhecimento dos padrões, configurando assim, um sobreajuste (*overfitting*). Esse comportamento está exemplificado na figura 5.

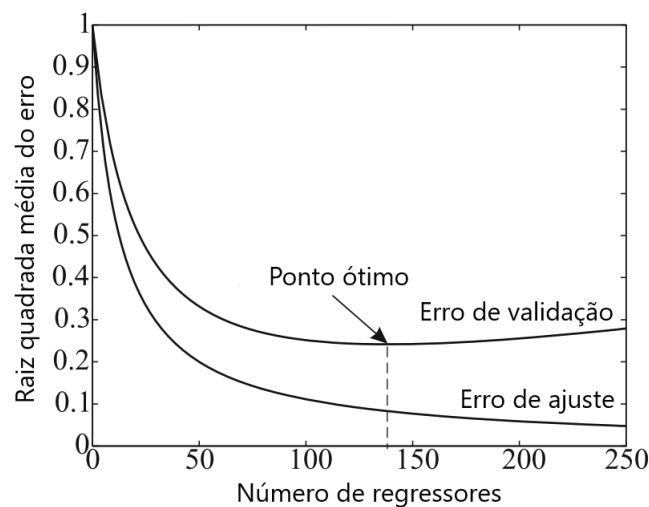


Figura 5 – Exemplificação do comportamento entre erro de validação/ajuste (Adaptado de (NELLES, 2001)).

4.3.2 Pré-tratamento dos dados espectrométricos

Dados provenientes de espectroscopia são altamente sujeitos a ruídos aleatórios e erros sistemáticos. Sendo assim, geralmente faz-se necessária a aplicação de pré-tratamentos para a minimização desses erros, melhorando portanto a inferência de modelos obtidos a partir dessas informações (BRERETON, 2007). Dentre a vasta gama de opções existentes de pré-tratamentos, no presente trabalhos destacam-se o filtro de Savitzky–Golay, o

alisamento por média móvel e o uso de otimização por algoritmo genético para a seleção de comprimentos de onda.

4.3.2.1 Alisamento por média móvel

A técnica de alisamento de dados espectrais por média móvel é comumente utilizada para a redução de ruído em sinais. Ela consiste na aplicação de uma janela deslizante sobre o sinal original seguida do cálculo da média aritmética dos valores dentro dessa janela em cada respectiva posição, conforme a equação 4.14.

$$y_i = \frac{1}{n} \sum_{j=i-k}^{i+k} x_j \quad (4.14)$$

Onde y_i é o valor na posição i do sinal suavizado, n é o comprimento da janela móvel, x_j são os valores de espectro original na posição j e k é o ponto central da janela.

4.3.2.2 Filtro de Savitzky–Golay com primeira derivada

O filtro de Savitzky-Golay (SG) com primeira derivada consiste em uma técnica de suavização de dados projetada para estimar a primeira derivada de um sinal enquanto preserva suas características mais relevantes. Ela é uma extensão do filtro SG convencional (SAVITZKY; GOLAY, 1964).

A abordagem desse filtro envolve o ajuste de um polinômio de ordem m a um conjunto de pontos em uma janela de tamanho n . Neste caso, ajusta-se o polinômio à primeira derivada dos dados.

A utilização da primeira derivada permite a correção de efeitos de deslocamento da linha de base em uma quantidade constante ao longo do espectro (BRERETON, 2007).

4.3.2.3 Seleção de variáveis por algoritmo genético

Em análise quimiométrica, os algoritmos genéticos (*Genetic Algorithm*, GA) são empregados para selecionar um subconjunto relevante de variáveis de um conjunto de dados multivariados. Essas variáveis compreendem os comprimentos de onda no espectro. Esse processo simplifica o modelo ao reduzir a quantidade de comprimentos de onda utilizados no ajuste (HASEGAWA; MIYASHITA; FUNATSU, 1997). Em um GA, cada solução candidata é denominada de indivíduo, e cada indivíduo apresenta um conjunto de genes, cada qual corresponde a um comprimento de onda, codificado como valores binários indicando a inclusão ou não da variável na seleção. Posteriormente, uma população inicial de soluções é gerada aleatoriamente. Cada solução é avaliada através de uma função de aptidão (*fitness*) que mensura o desempenho do modelo quimiométrico construído usando apenas as variáveis selecionadas. Os indivíduos mais aptos são selecionados para

reprodução, e então ocorre o crossover, onde partes dos cromossomos dos pais são trocadas para gerar descendentes. Adicionalmente, mutações aleatórias podem ocorrer para manter a diversidade genética. Os descendentes substituem os indivíduos menos aptos na população, garantindo que apenas as soluções mais aptas persistam. O processo é repetido por várias gerações ou até que um critério de parada seja atingido, como um número máximo de iterações ou uma convergência de *fitness* pela maior parte da população (EIBEN; SMITH, 2015; CARR, 2014).

5 Resultados e discussão

5.1 Modelo cinético

5.1.1 Ajuste inicial

Com os dados organizados, procedeu-se à primeira etapa do estudo para a otimização de um modelo cinético capaz de descrever o comportamento da hidrólise sob o novo sistema. Inicialmente, ajustou-se um modelo MM conforme o apresentado na seção 4.2.2, sem utilizar a regra *fuzzy*. Para tal, utilizou-se as informações das bateladas (ensaios 1 e 2) e bateladas alimentadas (ensaios 5 e 6). Os ensaios com alimentação apenas de enzima (3 e 4) foram utilizados como dados de validação, a fim de avaliar a capacidade de extrapolação dos modelos ajustados.

O ajuste inicial do modelo cinético, sem a implementação da regra *fuzzy*, apresentou um desempenho relativamente satisfatório na descrição do perfil de formação de glicose. No tocante aos demais produtos, o modelo teve dificuldades em prever o comportamento da celobiose em ambas as configurações de alimentação, já para a xilose, sua velocidade de formação foi drasticamente superestimada. Os perfis de formação de produtos sob esse ajuste inicial estão dispostos nas figuras 6 e 7. Alterações nas estimativas iniciais, nesse modelo, não foram capazes de solucionar os problemas apresentados em relação à xilose e celobiose.

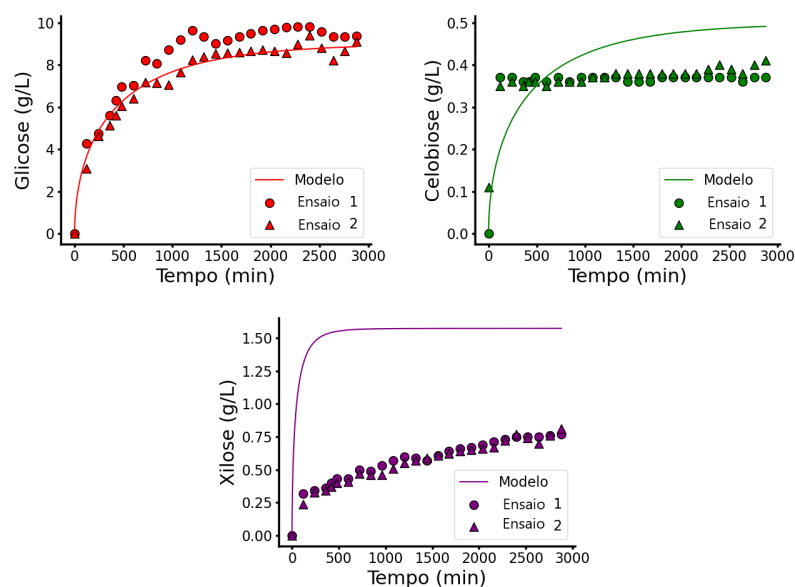


Figura 6 – Comparação entre o modelo MM-IP e dados experimentais no ensaio de batelada.

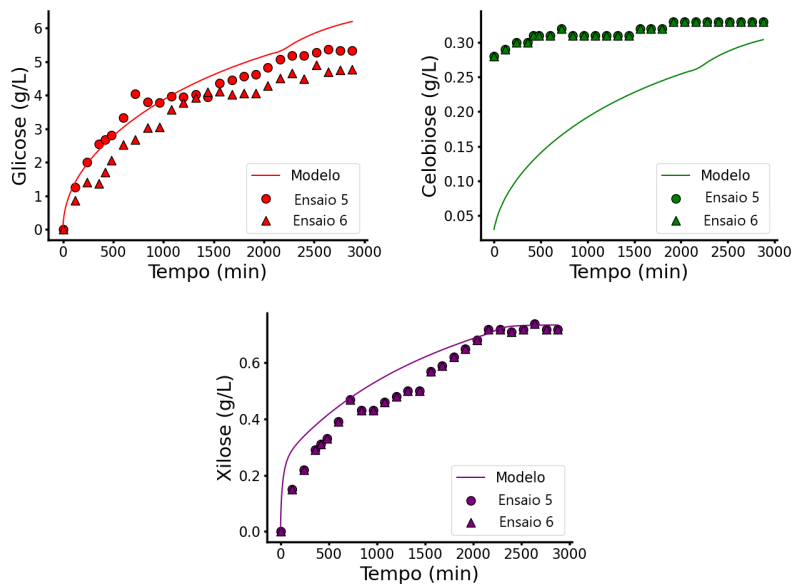


Figura 7 – Comparação entre o modelo MM-IP e dados experimentais no ensaio de batelada alimentada.

Dada a dificuldade na obtenção do modelo, foi introduzida a função de pertinência *fuzzy* linear apresentada na seção 4.2.3. O modelo, agora dividido em dois grupos de parâmetros cinéticos foi então ajustado juntamente com os limites da equação 4.4. Neste novo modelo, foram ajustados os limites superior e inferior de $8,25\text{g/L}$ e $7,96\text{g/L}$ respectivamente, além disso. Apesar de significativamente mais complexo, o novo modelo foi capaz de prever o comportamento da curva de formação da xilose em todos os 4 ensaios de maneira satisfatória, restando apenas a dificuldade em acompanhar o perfil da celobiose, conforme as figuras 8 e 9.

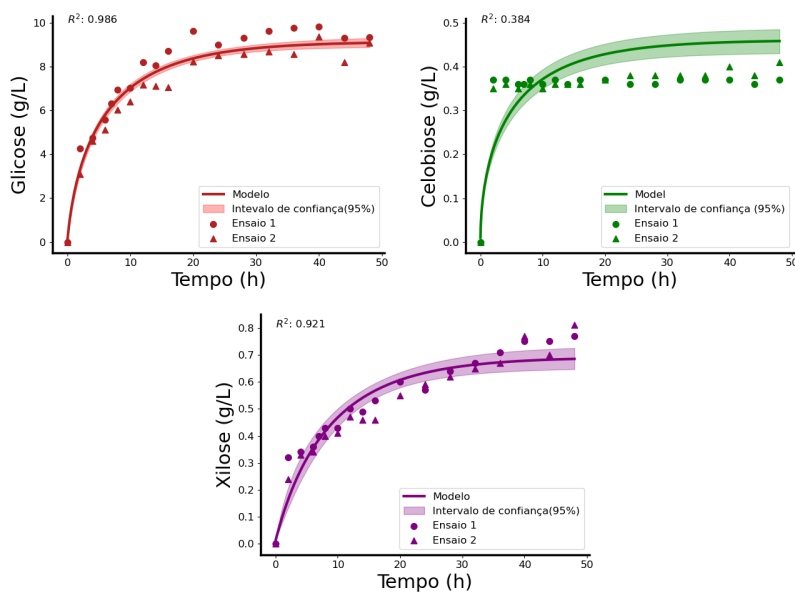


Figura 8 – Comparação entre o modelo MM-IP *fuzzy* e dados experimentais no ensaio de batelada.

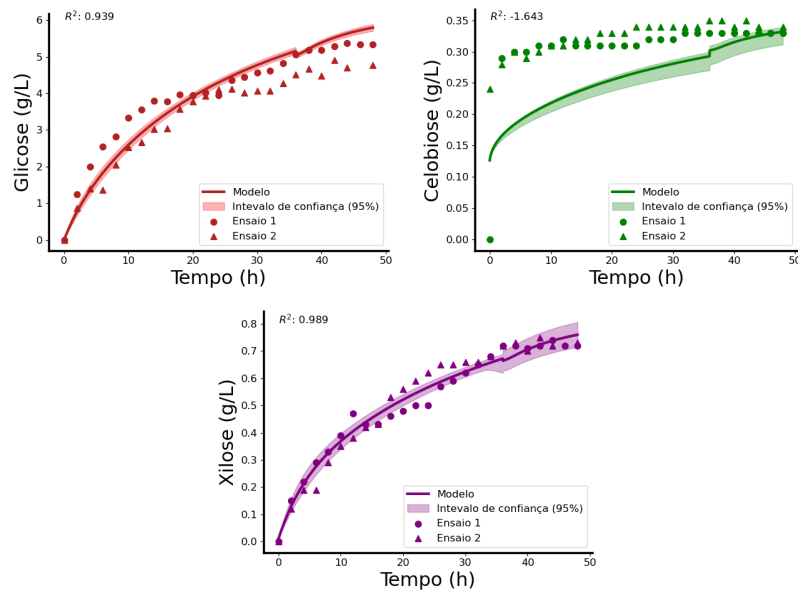


Figura 9 – Comparação entre o modelo MM-IP *fuzzy* dados experimentais no ensaio de batelada alimentada.

A abordagem *fuzzy* cinética proposta em outro trabalho de *Furlong*, portanto, mostra-se mais eficaz no ajuste da hidrólise para este sistema de reatores quando comparado a um simples modelo de Michaelis-Menten com inibição por formação de produto. Os parâmetros obtidos nesse ajuste, no entanto, estão contidos em grandes intervalos de confiança, conforme observável na tabela 7 do apêndice A, de maneira a indicar uma alta correlação entre eles. Tal situação é indesejada, uma vez que um determinado parâmetro pode assumir qualquer valor dentro de seu intervalo e isso pode ser compensado por mudanças apropriadas nos parâmetros correlacionados, prejudicando assim o significado físico dos valores obtidos. A alta correlação é confirmada pela matriz apresentada na figura 10, na qual os valores dos parâmetros são numerados conforme a reação ao qual pertencem, e o sufixo "h" (*high*) denota que este pertence ao modelo ACS e o sufixo "l" (*low*) indica pertencimento ao conjunto BCS.

Nota-se que há uma alta correlação entre diversos pares de parâmetros. Apesar do bom ajuste, a alta correlação paramétrica observada indica que nem todos os parâmetros podem ser estimados unicamente a partir dos dados experimentais disponíveis. Sendo assim, foi realizada a análise de identificabilidade, conforme delineado na seção 4.2.4. É importante ressaltar que essa análise é local, uma vez que as derivadas são calculadas em torno da solução ótima obtida no ajuste.

Deve-se notar também que essas análises permitem apenas a reestruturação do modelo de maneira a melhor utilizar a informações disponíveis. Para que atinja-se um conjunto totalmente identificável a partir de dados insuficientes há duas abordagens possíveis: a obtenção de novos dados experimentais e a fixação de determinados parâmetros a partir de dados da literatura. Este último, no entanto, não representa uma opção confiável,

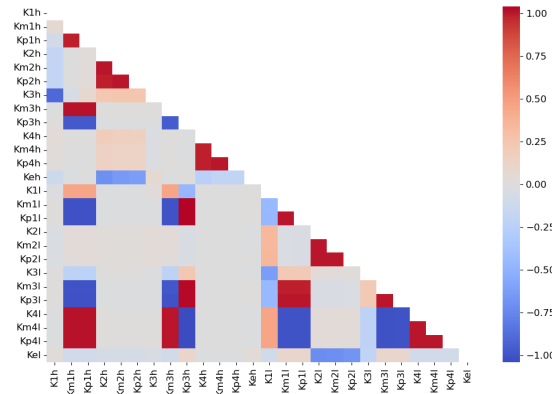


Figura 10 – Matriz de correlação para o modelo MM-IP *fuzzy*

uma vez que não há valores presentes na literatura para hidrólise enzimática com separação de fases, dado que trata-se de uma arquitetura de reatores não-convencional desenvolvida internamente. As análises realizadas a seguir auxiliam apenas na escolha de parâmetros a serem excluídos do modelo e no projeto de experimentos futuros.

5.1.2 Análise de identificabilidade

A análise de identificabilidade foi então realizada, iniciando-se com a análise de sensibilidade e utilizando como medida principal a medida de sensibilidade média (δ), demonstrada na figura 11. A partir deste cálculo, infere-se que o modelo apresenta sensibilidade alta apenas aos parâmetros $Kp3_h$, $Kp3_l$, $Kp2_h$ e $Kp2_l$, uma pequena sensibilidade a $k1_h$, $k2_h$, $Km2_h$, $k3_h$, $k1_l$, $k2_l$, $Km2_l$, $k3_l$, $k4_l$, $Kp4_l$ e praticamente nenhuma sensibilidade aos demais. Isso sugere que poucos valores neste modelo são identificáveis, o que corrobora com a alta correlação observada anteriormente. As constantes cinéticas não sensíveis, no entanto, não podem ser diretamente excluídas do modelo.

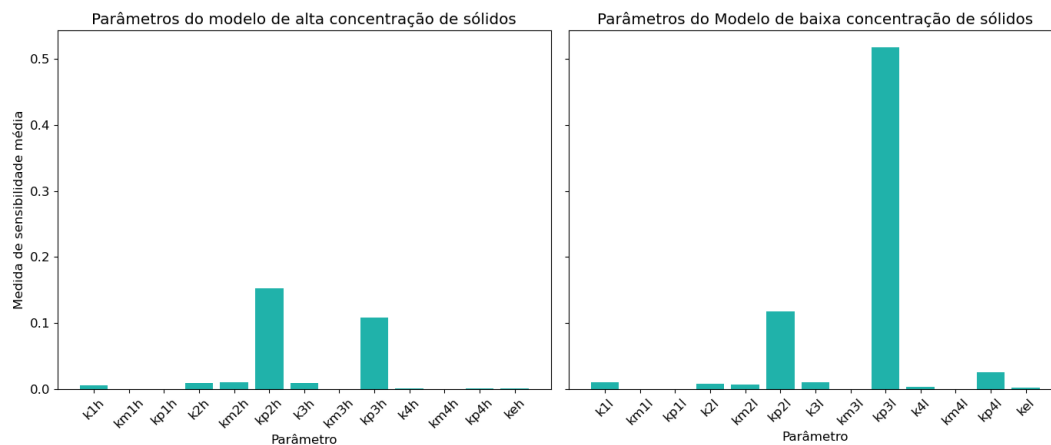


Figura 11 – Valores de sensibilidade média para os conjuntos de parâmetros ACS (esquerda) e BCS (direita) no modelo *fuzzy* linear.

A segunda etapa da análise de identificabilidade consiste na estimativa dos índices de colinearidade (γ). Estes foram calculados para todos os possíveis subconjuntos de parâmetros, sendo considerados identificáveis apenas aqueles que apresentaram um índice inferior a 20. A Tabela 3 agrupa os subconjuntos conforme o tamanho destes e indica quantos destes, entre todas as combinações possíveis, são identificáveis. Esta análise permite inferir a quantidade máxima de parâmetros identificáveis a partir dos dados experimentais,

Tabela 3 – Número total e combinações identificáveis de acordo com o número de parâmetros considerados no modelo MM-IP *fuzzy*.

Número de Parâmetros	Total de combinações possíveis	combinações identificáveis ($\gamma < 20$)	Percentual de subconjuntos identificáveis
2	325	296	91,076%
3	2600	1734	66,692%
4	14950	5929	39,658%
5	65780	11569	17,587%
6	230230	11687	5,076%
7	657800	3779	0,574%
8	1562275	83	0,005%
9	3124550	0	0,00%
10-26	47543227	0	0,00%

Com base nesses cálculos, um total de 35077 subconjuntos de parâmetros foram considerados identificáveis, representando cerca de 0,05% dos conjuntos possíveis. A partir da tabela 3, infere-se que no máximo 8 parâmetros podem ser identificados a partir dos dados disponíveis, dentre os 26 propostos no modelo. A baixa identificabilidade indica que a complexidade do modelo não é contemplada pelos dados experimentais, sendo necessárias mais investigações para que o modelo seja capaz de descrever o comportamento da hidrólise no sistema de reatores. A partir da análise de identificabilidade do modelo, mostraram-se necessárias alterações no modelo, que serão discutidas a seguir.

5.1.3 Reidentificação paramétrica

A alta sensibilidade apresentada pelos valores das constantes de inibição das reações envolvendo formação de glicose (2 e 3) indica que, quando estes variam sem que haja acomodação em parâmetros correlacionados, uma resposta drasticamente diferente na curva do modelo é gerada como consequência. Sendo assim, propõe-se que o modelo seja ajustado conforme a cinética de Michaelis-Menten pseudohomogênea sem considerar efeitos inibitórios. Essa alteração permite a redução da quantidade de parâmetros, além de permitir a investigação da necessidade de tais efeitos para a descrição dos experimentos realizados no reator. Sendo assim, as equações de taxa de reação foram remodeladas, retirando-se o termo inibitório e reduzindo o número de constantes a serem estimadas.

Além disso, duas outras adaptações foram feitas: a divisão das funções de pertinência e a alteração da regra de transição entre os modelos. A primeira visa investigar se ambas

as curvas de formação de glicose e xilose possuem as mesmas condições para alternar entre mecanismos de reação, com base no comportamento da xilose no modelo sem função *fuzzy*, além de isolar os parâmetros cinéticos envolvendo as reações paralelas. Isso foi realizado a partir da introdução de dois graus de pertinência: um para governar a transição de modelos para as reações envolvendo glicose como produto, e outro para a xilose. A segunda visa uma transição mais suave entre os modelos, substituindo a função Takagi-Sugeno utilizada na literatura por uma função sigmóide. Dessa forma, a equação 4.4 é reescrita como:

$$GP = \begin{cases} 0 & \text{se } SR \leq ACS_{lb} \\ \frac{1}{1+e^{-\alpha \cdot (SR-SR_0)}} & \text{se } ACS_{lb} \leq SR \leq ACS_{ub} \\ 1 & \text{se } SR \geq ACS_{ub} \end{cases} \quad (5.1)$$

Onde:

$$SR_0 = \frac{ACS_{inf} + ACS_{sup}}{2} \quad (5.2)$$

$$\alpha = -\frac{\ln\left(\frac{1}{0.99} - 1\right)}{ACS_{sup} - ACS_{inf}}$$

Os dados experimentais foram ajustados ao novo modelo, obtendo-se um novo conjunto de parâmetros cinéticos, dispostos na tabela 8 no apêndice A. O comportamento do novo modelo em relação aos dados de ajuste apresentaram melhoras, conforme observável nas figuras 12 e 13 abaixo. Os limites inferiores e superiores das funções de pertinência obtidos para esse novo modelo foram respectivamente de 0,14g/L e 9,81g/L para a função de pertinência da xilose e 3,43g/L e 9,28g/L para os demais.

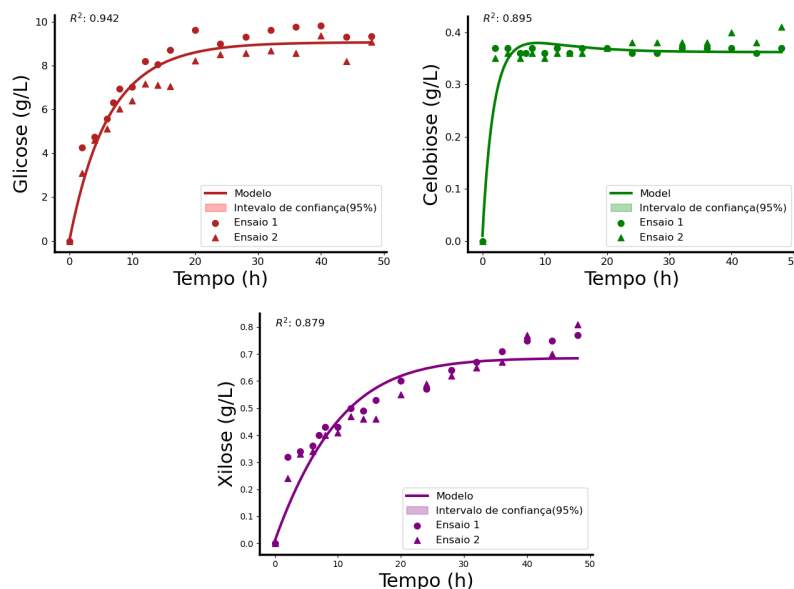


Figura 12 – Comparação entre o modelo MM *fuzzy* sigmoide e dados experimentais no ensaio de batelada.

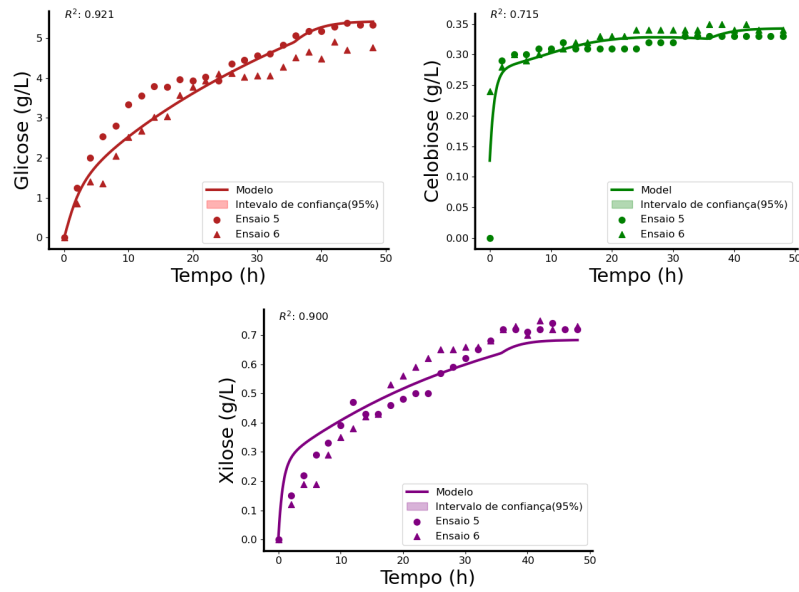


Figura 13 – Comparação entre o modelo MM *fuzzy* sigmoide e dados experimentais no ensaio de batelada alimentada.

Sob esse novo modelo, o ajuste das curvas de formação de glicose mantiveram-se satisfatórias enquanto a celobiose apresentou uma melhora significativa. Dessa maneira, as alterações realizadas demonstram de imediato uma melhor capacidade de compreender os dados dos ensaios utilizados no ajuste. A matriz de correlação paramétrica para esse ajuste, disposta na figura 14, indica uma melhora, apesar de ainda apresentar alta correlação entre os pares $k2_h - Km2_h$, $k4_h - Km4_h$, $k1_l - Km1_l$, $k1_h - Km1_h$ e $k3_l - Km3_l$.

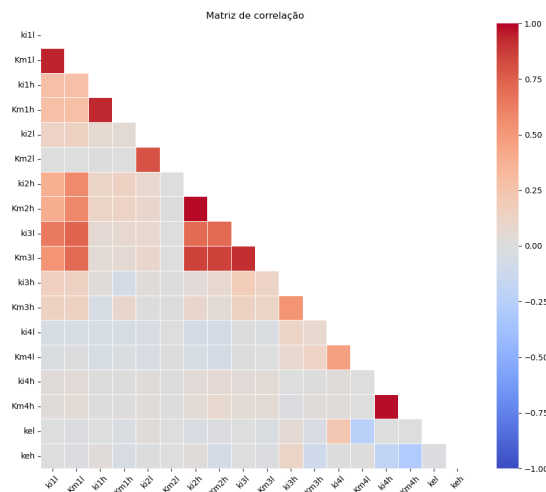


Figura 14 – Matriz de correlação para o modelo MM *fuzzy* sigmoide

A análise indica uma alta sensibilidade média nos parâmetros $k3_h$ e $Km3_h$, enquanto $Km1_h$, $k2_h$, $Km2_h$, $Km2_l$, $k3_l$, $Km3_l$, e $k4_l$ demonstram valor menor de δ . Em contrapartida, os restantes mostram-se insignificantes em termos de sensibilidade média.

Este novo modelo demonstra uma melhora mínima na distribuição da sensibilidade entre os parâmetros, portanto não há grandes ganhos em sensibilidade paramétrica neste reajuste.

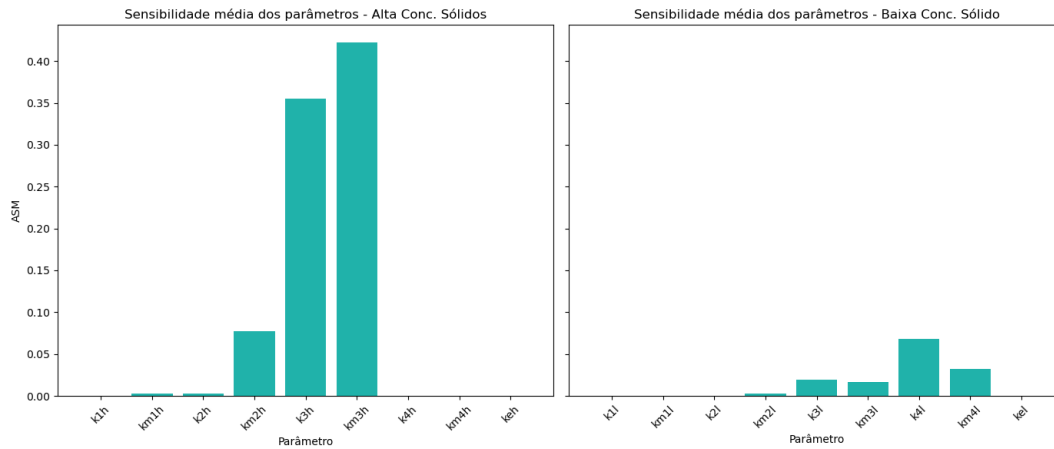


Figura 15 – Valores de sensibilidade média para os conjunto de parâmetros ACS (esquerda) e BCS (direita) no modelo *fuzzy* sigmoide.

Os índices de colinearidade para este modelo também indicam uma pequena melhora em termos de identificabilidade dos subconjuntos de parâmetros. As quantidades e o percentual de subconjuntos identificáveis para esse modelo estão dispostos na tabela 4

Tabela 4 – Número total e combinações identificáveis de acordo com o número de parâmetros considerados no modelo MM com *fuzzy* sigmoide.

Número de Parâmetros	Total de combinações possíveis	combinações identificáveis ($\gamma < 20$)	Percentual de subconjuntos identificáveis
2	153	126	82,353%
3	816	547	67,03%
4	3060	1632	53,33%
5	8568	3569	41,65%
6	18564	5584	30,08%
7	31824	6170	19,39%
8	43758	4922	11,25%
9	48620	2800	5,76%
10	43758	1106	2,53%
11	31824	292	0,92%
12	18564	44	0,24%
13	8568	3	0,03%
14-18	4048	0	0,00%

Os cálculos de colinearidade para este novo modelo indicam 26795 subconjuntos identificáveis, representando 10% dos possíveis subconjuntos. É possível inferir então, conforme a tabela 4, que 13 parâmetros podem ser identificados a partir dos dados experimentais nesse novo modelo. Essa informação aponta que, ainda, a complexidade do modelo não é contemplada pelos dados experimentais, mesmo indicando uma maior parte dos parâmetros como identificáveis, sendo necessários ajustes posteriores para garantir a

identificabilidade do modelo. Dentre os três subconjuntos compostos por 13 constantes cinéticas, o de menor colinearidade identificado foi: $Km1_l$, $k1_h$, $Km1_h$, $k2_l$, $Km2_l$, $Km3_l$, $k3_h$, $Km3_h$, $k4_l$, $Km4_l$, $k4_h$, $Km4_h$, ke_h .

Através destes cálculos infere-se as alterações na modelagem cinética resultaram em uma melhor identificabilidade e menor correlação entre os parâmetros ajustados. No entanto, o conjunto total de parâmetros ainda não pode ser estimado exclusivamente a partir dos dados experimentais disponíveis, conforme indicam as análises de sensibilidade e colinearidade realizadas. Deve-se ressaltar que, uma prática comum para a obtenção de um conjunto identificável é a fixação de constantes com base em valores retirados de trabalhos semelhantes. No entanto, por tratar-se de uma arquitetura de reatores não convencional desenvolvida internamente, a literatura carece de informações. Sendo assim, a maneira mais adequada de refinar o modelo cinético para este conjunto consiste na obtenção de mais informações experimentais.

5.1.4 Validação do modelo

Os dados cromatográficos dos ensaios de batelada com alimentação apenas de enzima (3 e 4) foram então utilizados a fim de avaliar a capacidade de extrapolação do modelo para mais condições de alimentação. O desempenho do modelo nestes experimentos está apresentada na figura 16 abaixo.

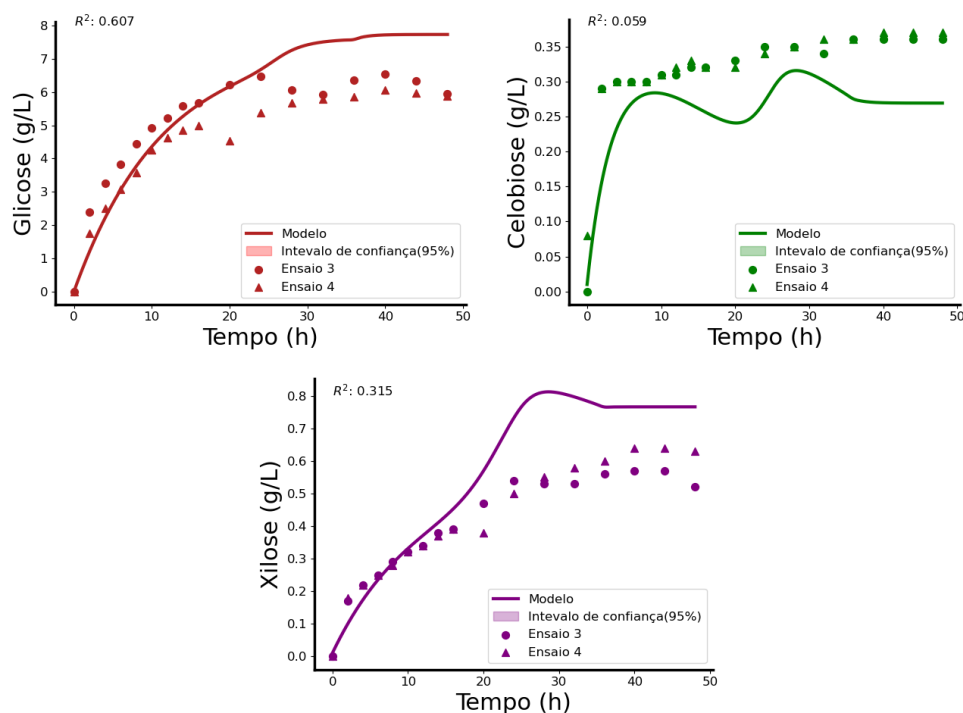


Figura 16 – Comparação entre o modelo MM *fuzzy sigmoide* e dados experimentais de validação.

Ao observar comportamento da curva do modelo em relação aos experimentos de validação, nota-se que após cerca de 20 horas de ensaio há um desvio de previsão considerável do modelo em relação aos dados experimentais para a glicose, enquanto para os outros carboidratos, a velocidade de reação sobe rapidamente e depois desacelera. Entende-se que os rápidos aumentos e diminuições na concentração dos produtos nesse período da reação não representam um comportamento compatível com a cinética esperada de um sistema de hidrólise lignocelulósico. Este momento da reação, coincide com o início da transição entre os grupos de parâmetros ACS e de BCS, conforme observável na figura 17.

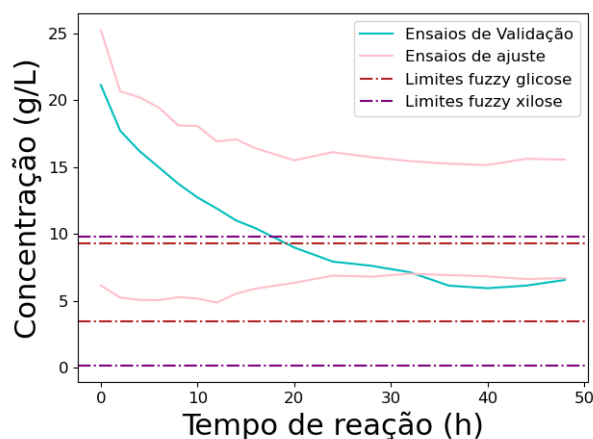


Figura 17 – Concentração calculada de sólidos reativos ao longo dos experimentos de validação.

Nota-se que os desvios ocorrem a partir do momento em que o modelo adentra a região de transição *fuzzy*. Portanto infere-se que o modelo ajustado, apesar de demonstrar capacidade descritiva processo para diferentes condições de alimentação de bagaço, ainda requer mais informações de calibração para uma estimativa precisa dos limites de pertinência.

5.2 Monitoramento espectroscópico

Os dados espectroscópicos utilizados neste estudo foram adquiridos por meio da coleta de espectrogramas do NIR em intervalos de seis minutos, para todos os ensaios experimentais. Para permitir o ajuste da regressão PLS, foram identificadas as amostras para as quais haviam dados do HPLC. Nessa etapa, descartaram-se alguns espectros com alto nível de ruído, assim como os dados para os quais não foram detectados picos de absorvância. Essa verificação inicial levou à desconsideração do ensaio 1, por conta da ausência dos picos, e do ensaio 6, por apresentar a maior parte das amostras como ruidosas demais. A figura 18 ilustra as amostras brutas, assim como as selecionadas para ajustes subsequentes.

Para a validação cruzada, 10 % dos dados foram selecionados aleatoriamente para teste, enquanto os 90% restantes foram utilizados na calibração. O número de regressores para cada modelo foi então selecionado a partir do menor $RMSE_{CV}$, a partir do qual há um afastamento entre este e o erro de calibração.

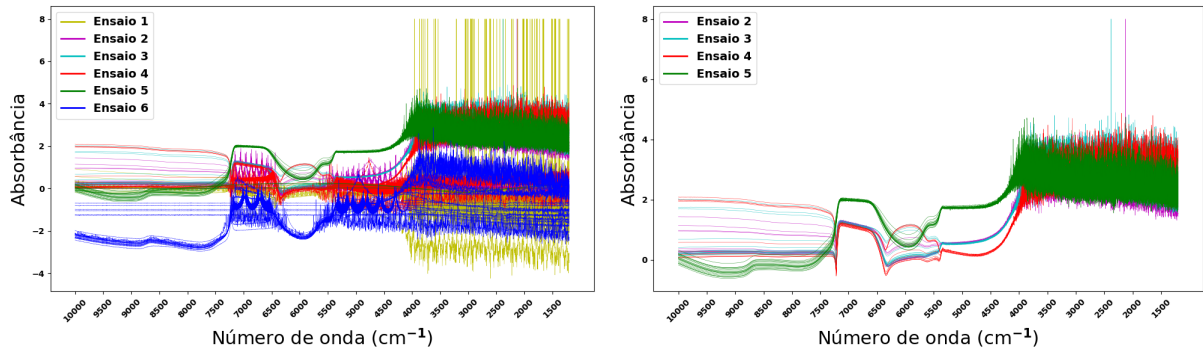


Figura 18 – Espectros NIR brutos (esquerda) e amostras selecionadas (direita).

Com base nas amostras selecionadas, nota-se que a região entre os números de onda 1200 e 4500 cm^{-1} é dominada por ruídos, além disso, também foi observado um deslocamento vertical na linha de base entre os experimentos. Portanto, visando à redução de ruídos espectrais, a região mencionada foi desconsiderada para análises futuras e um alisamento por média móvel, com janela de 11 pontos. Ademais, para contornar problemas relacionados a linha de base, outro alisamento aplicado foi um filtro de Savitzky-Golay com primeira derivada. Os espectros resultantes após o corte e uso dos filtros estão dispostos na figura 19

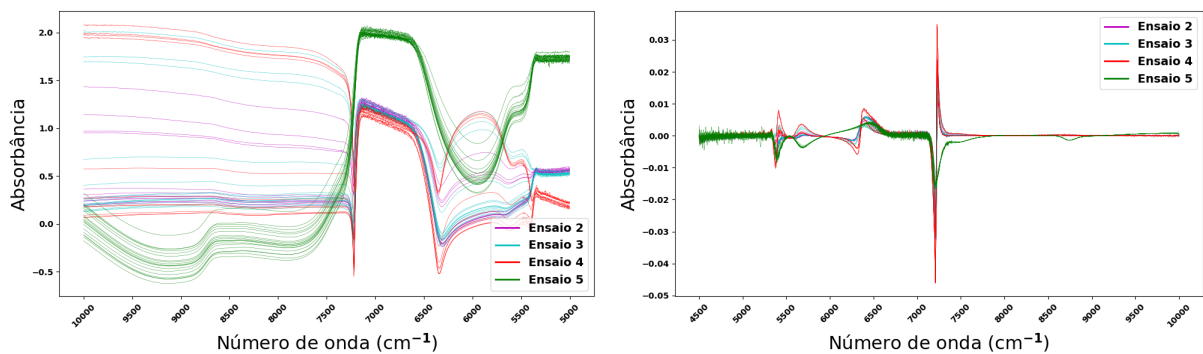


Figura 19 – Espectros após o corte (esquerda) e após os filtros (direita).

Sendo assim, ajustaram-se 4 modelos PLS a fim de avaliar a efetividade dos pré-tratamentos propostos, observando principalmente o erro de validação cruzada ($RMSE_{CV}$) para cada modelo. Os modelos ajustados estão dispostos na tabela 5

Nota-se que os modelos 2 e 4 apresentaram os menores valores de validação cruzada. Dentre os métodos investigados, constata-se que o corte de comprimento de onda é predominantemente responsável pela redução do erro, demonstrando maior eficácia. Os pré-tratamentos destes dois modelos foram então utilizados para a etapa subsequente.

Tabela 5 – Modelos PLS ajustados

Modelo	1			2			3			4		
	Celobiose	Glicose	Xilose	Celobiose	Glicose	Xilose	Celobiose	Glicose	Xilose	Celobiose	Glicose	Xilose
Número de regressores	4	4	4	5	5	5	4	4	4	4	4	4
$RMSE_{CV}$ (g/L)	0,025	1,170	0,121	0,016	0,900	0,077	0,015	0,931	0,088	0,014	0,895	0,085
$RMSE_{cal}$ (g/L)	0,012	0,679	0,094	0,014	0,803	0,071	0,011	0,663	0,060	0,011	0,682	0,063
Faixa de número de onda (cm^{-1})	10000 - 1200			10000 - 4500			10000 - 4500			10000 - 4500		
Pré-tratamento (em ordem)	x			x			1ªderiv. Savitzky-Golay			Média móvel, 1ªderiv. Savitzky-Golay		

Após selecionados os pré-tratamentos, dois novos modelos foram ajustados com a utilização de um algoritmo genético para seleção de faixas de comprimentos de onda, conforme os parâmetros discriminados na seção 4.3. Os respectivos resultados da seleção de variáveis para o pré-tratamento apenas com o corte (Modelo 5) e para o modelo com filtros de Média móvel e derivada SG (Modelo 6) estão dispostos na figura 20. A tabela 6 abaixo representa os regressores e os erros associados a estes ajustes.



Figura 20 – Variáveis selecionadas para os modelos 5 e 6.

Tabela 6 – Modelos PLS ajustados com seleção de variáveis

Modelo	5			6		
	Celobiose	Glicose	Xilose	Celobiose	Glicose	Xilose
Número de regressores	8	8	8	5	5	5
$RMSE_{CV}$ (g/L)	0,012	0,767	0,078	0,012	0,650	0,072
$RMSE_{cal}$ (g/L)	0,010	0,607	0,0688	0,009	0,33	0,046
Faixa de número de onda (cm^{-1})	10000 - 4500			10000 - 4500		
Pré-tratamento (em ordem)	x			Média móvel, 1ªderivada SG		

Ambos os modelos exibem melhorias significativas na precisão da calibração após a seleção de variáveis. O pré-processamento realizado com média móvel e filtro de derivada SG resultou no menor erro de validação cruzada observado entre todas as abordagens, apesar da disparidade entre os valores de $RMSE_{CV}$ e $RMSE_{cal}$ ao ajustar para a concentração de glicose. Para os outros analitos, não se observaram diferenças significativas nos erros

entre as duas abordagens, o que torna difícil afirmar a eficácia dos métodos de suavização empregados. No entanto, o modelo 6, baseado nos dados pré-processados e selecionados pelo GA, alcançou resultados comparáveis utilizando uma estrutura mais simples em comparação com o modelo 5. Portanto, optou-se por utilizar o modelo 6 nas etapas subsequentes.

Associados ao ajuste PLS do modelo 6, foram traçados os gráficos de concentrações preditas *versus* as concentrações de referência para todos os analitos, representados na figura 21. A medida de R^2 em relação à bissetriz na figura quantifica a precisão do modelo em relação aos dados de calibração. Em geral, nota-se que as predições estão satisfatórias, com uma baixa dispersão em torno da bissetriz.

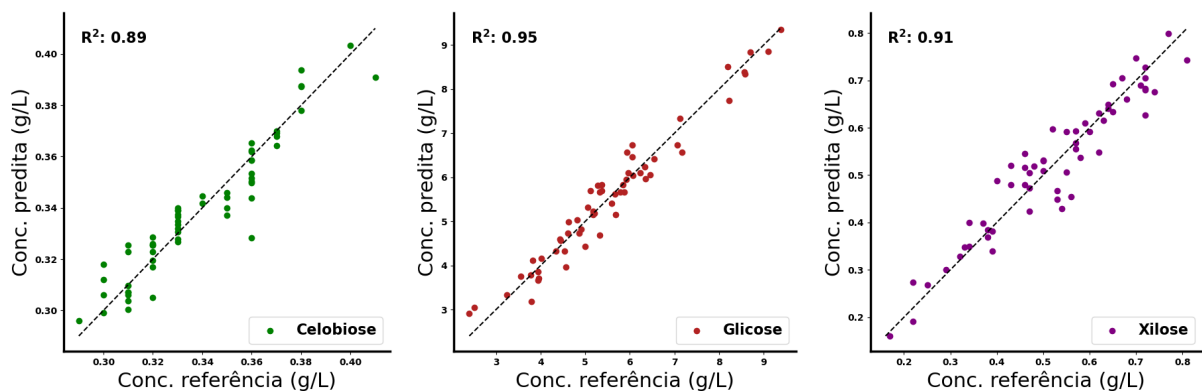


Figura 21 – Gráficos das concentrações de referência (HPLC) x concentrações preditas para a validação cruzada do modelo 6.

5.2.1 Curva de predição de concentrações do modelo PLS

Por fim, a curva de inferência do modelo 6 para os analitos foi traçada utilizando todos os espectros disponíveis nos ensaios, e os valores preditos estão dispostos nas figuras 22, 23, 24 e 25 abaixo.

Analisando o comportamento da curva de previsão, nota-se que o modelo foi capaz de acompanhar bem a concentração de celobiose nos ensaios de batelada e batelada com alimentação de enzima (2, 3 e 4). Já para a glicose e xilose nesses ensaios, o modelo apresentou dificuldades na predição das concentrações nas primeiras horas. O ensaio de batelada alimentada de bagaço (5), apesar de demonstrar um comportamento semelhante, apresentou um desvio muito grande na previsão de todos os carboidratos nas 10 primeiras horas de ensaio.

O modelo desenvolvido neste trabalho então, teve um bom desempenho na previsão dos carboidratos monitorados sob os modos de operação de batelada e batelada com alimentação de enzima, com exceção das 6 horas iniciais do processo. Já para a batelada com alimentação de bagaço, os valores previstos só atingem uma qualidade satisfatória após

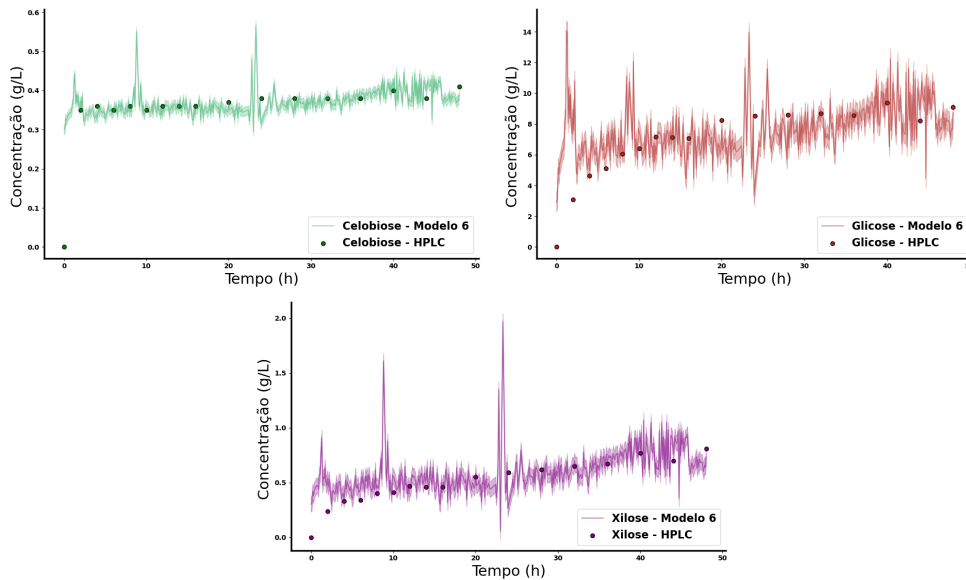


Figura 22 – Predição das concentrações de celobiose, glicose e xilose pelo modelo 6 para o ensaio 2 (Batelada).

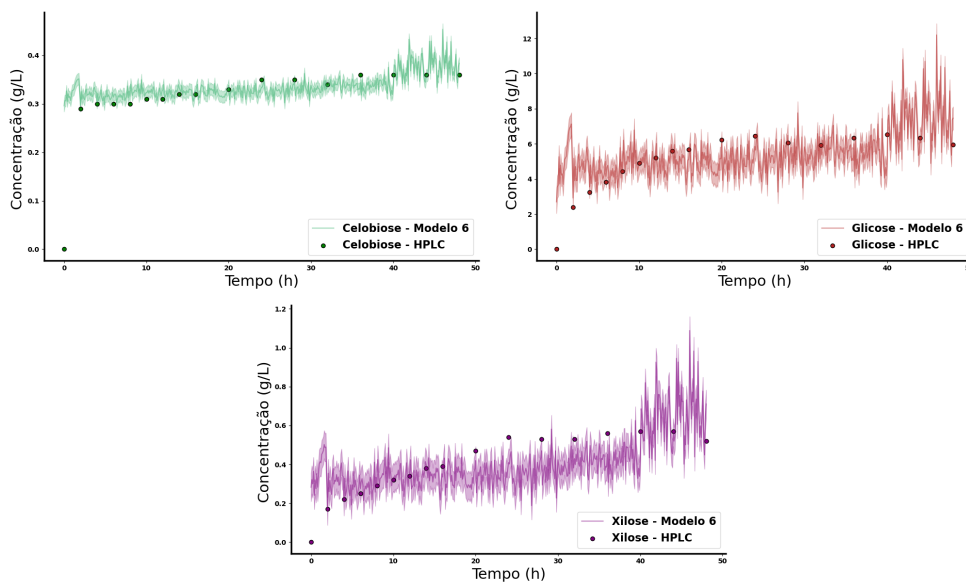


Figura 23 – Predição das concentrações de celobiose, glicose e xilose pelo modelo 6 para o ensaio 3 (Batelada alimentada de ezima).

as 12 primeiras horas de ensaio. Dessa maneira, recomenda-se que experimentos futuros, principalmente com alimentação de bagaço e enfoque nas horas iniciais do processo, sejam realizados.

Além disso, os pré-tratamentos aplicados aos espectros também podem ser otimizados. Uma vez que não foi possível, neste estudo, avaliar uma ampla gama de métodos. Ainda que o resultado apresentado seja satisfatório, uma análise mais profunda do impacto de outros pré-tratamentos na inferência de produtos da hidrólise pode revelar opções que resultem em monitoramentos espectroscópicos de melhor qualidade.

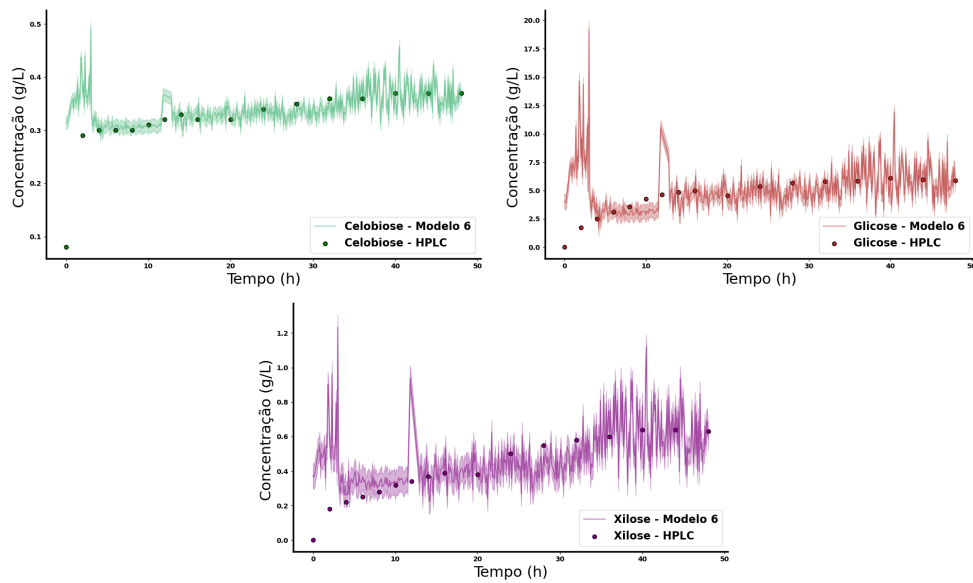


Figura 24 – Predição das concentrações de celobiose, glicose e xilose pelo modelo 6 para o ensaio 4 (Batelada alimentada de enzima).

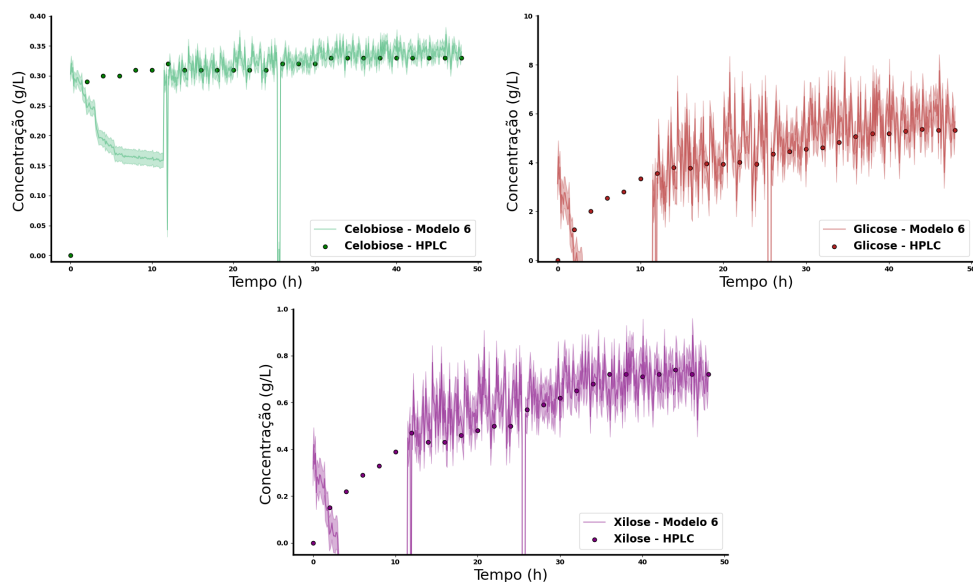


Figura 25 – Predição das concentrações de celobiose, glicose e xilose pelo modelo 6 para o ensaio 5 (Batelada alimentada de bagaço).

6 Conclusões e próximos passos

Neste trabalho, foram avaliados diferentes conjuntos de dados experimentais de HPLC e NIR. As informações disponíveis foram sistematicamente analisadas e avaliadas para duas finalidades: o ajuste de um modelo cinético para a modelagem da hidrólise enzimática do bagaço de cana pré-tratados e a obtenção de um modelo PLS para a previsão das concentrações a partir do espectro do infravermelho próximo, ambos em um sistema de reatores com separação de fases. Por mais que trate-se de um sistema reacional complexo, foi possível desenvolver modelos capazes de descrever o processo para os experimentos. De maneira geral, são necessárias mais informações de validação para que seja possível verificar a capacidade preditiva desses modelos.

Na avaliação cinética, desenvolveu-se um modelo capaz de descrever a curva de formação de produtos da hidrólise sob diferentes modos de alimentação através de uma abordagem *fuzzy* adaptada da literatura. Os parâmetros ajustados apresentaram um bom desempenho para as bateladas e bateladas alimentadas na faixa de concentração de sólidos estudadas. O ajuste dos ensaios de batelada com alimentação de enzima, assim como as análises de identificabilidade realizadas, indicam que alterações no modelo são necessárias para que este seja capaz de incorporar diferentes modos de operação. Ademais, ressalta-se que futuros experimentos são necessários para investigar o comportamento do processo com diferentes concentrações de sólidos reativos, uma vez que as informações disponíveis foram insuficientes para uma compreensão abrangente dos efeitos dessa variável na hidrólise.

Na modelagem espectroscópica, foi implementado um modelo de Multicalibração baseado na lei de Lambert-Beer, para a previsão de concentrações da glicose, celobiose e xilose a partir de espectros da região do infravermelho próximo. O modelo desenvolvido empregou técnicas de pré-tratamento do espectro e seleção de variáveis por algoritmo genético, a fim de diminuir ruídos e reter as informações relevantes para a regressão. Os cálculos realizados resultaram em um modelo satisfatório para as bateladas e bateladas com alimentação de enzima. Para as bateladas alimentadas ainda são necessários dados experimentais mais consistentes e mais informações, principalmente nas horas iniciais, para uma previsão de qualidade sob essas condições. Além disso, uma análise mais profunda e abrangente dos pré-tratamentos espectrais mostra-se necessária, uma vez que não foi possível constatar a efetividade dos pré-tratamentos aplicados na redução dos erros de validação cruzada.

Em virtude disso, o modelo ainda deve ser refinado e reavaliado, e novos experimentos devem ser realizados. Para trabalhos futuros, recomenda-se o refino dos pré-tratamentos espectrométricos e a avaliação de outros meios de seleção de variáveis em conjunto com a regressão PLS.

Referências

- ABAS, N.; KALAIR, A.; KHAN, N. Review of fossil fuels and future energy technologies. *Futures*, Elsevier BV, v. 69, p. 31–49, may 2015. Citado na página 19.
- AL-HADITHI, B. M.; JIMÉNEZ, A.; MATÍA, F. A new approach to fuzzy estimation of takagi–sugeno model and its applications to optimal control for nonlinear systems. *Applied Soft Computing*, Elsevier, v. 12, n. 1, p. 280–290, 2012. Citado na página 31.
- ANDRADE, S. R. et al. Comparison of multivariate calibration methods to determine simultaneously mebendazole–cambendazole and mebendazole–thiabendazole in pharmaceutical preparations by uv–visible spectrophotometry. *Journal of Pharmaceutical and Biomedical Analysis*, Elsevier BV, v. 33, n. 4, p. 655–665, nov. 2003. ISSN 0731-7085. Citado 2 vezes nas páginas 31 e 39.
- BAILEY, C. J. Enzyme kinetics of cellulose hydrolysis. *Biochemical Journal*, Portland Press Ltd., v. 262, n. 3, p. 1001–1001, set. 1989. ISSN 1470-8728. Citado na página 30.
- BATTISTA, F. et al. Enzymatic hydrolysis at high dry matter content: The influence of the substrates’ physical properties and of loading strategies on mixing and energetic consumption. *Bioresource Technology*, Elsevier BV, v. 250, p. 191–196, fev. 2018. ISSN 0960-8524. Citado na página 25.
- BERTRAND, E.; DUSSAP, C.-G. First generation bioethanol: Fundamentals—definition, history, global production, evolution. In: _____. *Liquid Biofuels: Bioethanol*. Cham: Springer International Publishing, 2022. p. 1–12. ISBN 978-3-031-01241-9. Disponível em: <https://doi.org/10.1007/978-3-031-01241-9_1>. Citado na página 19.
- BEZDEK, J. C.; PAL, S. K. Fuzzy models for pattern recognition. 1 1994. Disponível em: <<https://www.osti.gov/biblio/81602>>. Citado na página 36.
- BEZERRA, T. L.; RAGAUSKAS, A. J. A review of sugarcane bagasse for second-generation bioethanol and biopower production. *Biofuels, Bioproducts and Biorefining*, Wiley, v. 10, n. 5, p. 634–647, jul. 2016. ISSN 1932-1031. Citado na página 24.
- BRERETON, R. G. *Applied chemometrics for scientists*. [S.l.]: John Wiley and Sons, 2007. Citado 2 vezes nas páginas 41 e 42.
- BRUN, R.; REICHERT, P.; KÜNSCH, H. R. Practical identifiability analysis of large environmental simulation models. *Water Resources Research*, American Geophysical Union (AGU), v. 37, n. 4, p. 1015–1030, apr 2001. Citado na página 38.
- CARR, J. An introduction to genetic algorithms. *Senior Project*, v. 1, n. 40, p. 7, 2014. Citado na página 43.
- CARRILLO, F. et al. Effect of alkali pretreatment on cellulase hydrolysis of wheat straw: Kinetic study. *Process Biochemistry*, Elsevier BV, v. 40, n. 10, p. 3360–3364, out. 2005. ISSN 1359-5113. Citado na página 30.

- CARVALHO, M. L. et al. Kinetic study of the enzymatic hydrolysis of sugarcane bagasse. *Brazilian Journal of Chemical Engineering*, FapUNIFESP (SciELO), v. 30, n. 3, p. 437–447, sep 2013. Citado 2 vezes nas páginas 19 e 30.
- CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science and Technology*, Elsevier BV, v. 18, n. 2, p. 72–83, feb 2007. Citado na página 32.
- CHANDEL, A. K. et al. The path forward for lignocellulose biorefineries: Bottlenecks, solutions, and perspective on commercialization. *Bioresource Technology*, v. 264, p. 370–381, 2018. ISSN 0960-8524. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0960852418307831>>. Citado na página 19.
- CHENEY, E. S.; HAWKES, M. W. The future of hydrocarbons: Hubbert’s peak or a plateau. *GSA Today*, v. 17, n. 6, p. 69–70, 2007. Citado na página 19.
- CONAB, C. N. de A. Acompanhamento da safra brasileira de cana-de-açúcar. v. 11, n. 3, nov. 2023. Disponível em: <<https://www.conab.gov.br/info-agro/safras/cana>>. Citado na página 23.
- CORRER, C. J. et al. Determinação de ácido kójico em produtos farmacêuticos por espectroscopia uv-vis e processo de calibração multivariada. *acta farmacéutica bonaerense*, COLEGIO DE FARMACEUTICOS, v. 24, n. 3, p. 416, 2005. Citado na página 32.
- DIAS, J. d. S. et al. Produção de briquetes e péletes a partir de resíduos agrícolas, agroindustriais e florestais. Brasília, DF: Embrapa Agroenergia, 2012., 2012. Citado na página 23.
- DOCHAIN, D.; BASTIN, G. Adaptive control of fedbatch bioreactors. *Chemical Engineering Communications*, Taylor & Francis, v. 87, n. 1, p. 67–85, 1990. Citado na página 35.
- EDLUND, U.; ALBERTSSON, A.-C. A microspheric system: Hemicellulose-based hydrogels. *Journal of Bioactive and Compatible Polymers*, SAGE Publications, v. 23, n. 2, p. 171–186, mar. 2008. ISSN 1530-8030. Citado na página 24.
- EFTAXIAS, A. et al. Nonlinear kinetic parameter estimation using simulated annealing. *Computers and Chemical Engineering*, Elsevier BV, v. 26, n. 12, p. 1725–1733, dez. 2002. ISSN 0098-1354. Citado na página 34.
- EIBEN, A. E.; SMITH, J. E. *Introduction to evolutionary computing*. [S.l.]: Springer, 2015. Citado na página 43.
- EL-ZAWAWY, W. K. et al. Acid and enzyme hydrolysis to convert pretreated lignocellulosic materials into glucose for ethanol production. *Carbohydrate Polymers*, Elsevier BV, v. 84, n. 3, p. 865–871, mar. 2011. ISSN 0144-8617. Citado na página 25.
- FARINAS, C. S.; MARCONCINI, J. M.; MATTOSO, L. H. C. Enzymatic conversion of sugarcane lignocellulosic biomass as a platform for the production of ethanol, enzymes and nanocellulose. *Journal of Renewable Materials*, Computers, Materials and Continua (Tech Science Press), v. 6, n. 2, p. 203–216, mar. 2018. ISSN 2164-6325. Citado na página 28.

FURLONG, V. B. *Design and automation of a new reactor for enzymatic hydrolysis of lignocellulosic biomass at high solids loading*. Tese (phdthesis) — Universidade Federal de São Carlos, 2019. Citado 7 vezes nas páginas 13, 20, 27, 28, 29, 30 e 34.

FURLONG, V. B. et al. Fuzzy-enhanced modeling of lignocellulosic biomass enzymatic saccharification. *Energies*, MDPI, v. 12, n. 11, p. 2110, 2019. Citado 2 vezes nas páginas 36 e 37.

GABOR, A.; VILLAVERDE, A. F.; BANGA, J. R. Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems. *BMC Systems Biology*, Springer Science and Business Media LLC, v. 11, n. 1, may 2017. Citado 2 vezes nas páginas 38 e 39.

HARRIS, C. R. et al. Array programming with NumPy. *Nature*, Springer Science and Business Media LLC, v. 585, n. 7825, p. 357–362, set. 2020. Disponível em: <<https://doi.org/10.1038/s41586-020-2649-2>>. Citado na página 39.

HASEGAWA, K.; MIYASHITA, Y.; FUNATSU, K. Ga strategy for variable selection in qsar studies: ga-based pls analysis of calcium channel antagonists. *Journal of Chemical Information and Computer Sciences*, American Chemical Society (ACS), v. 37, n. 2, p. 306–310, mar. 1997. ISSN 1520-5142. Citado na página 42.

INC., E. R. *PLS Toolbox 9.3*. 2023. Disponível em: <<https://eigenvector.com/software/pls-toolbox/>>. Citado na página 39.

JIN, J. et al. Studies on the performance of a rotating drum bioreactor for bioleaching processes — oxygen transfer, solids distribution and power consumption. *Hydrometallurgy*, Elsevier BV, v. 103, n. 1–4, p. 30–34, jun. 2010. ISSN 0304-386X. Citado na página 28.

KADAM, K.; RYDHOLM, E.; MCMILLAN, J. Development and validation of a kinetic model for enzymatic saccharification of lignocellulosic biomass. *Biotechnology Progress*, Wiley, v. 20, n. 3, p. 698–705, jun. 2004. ISSN 8756-7938. Citado 2 vezes nas páginas 29 e 36.

KRAUSCH, N. et al. Monte carlo simulations for the analysis of non-linear parameter confidence intervals in optimal experimental design. *Frontiers in Bioengineering and Biotechnology*, Frontiers Media SA, v. 7, maio 2019. ISSN 2296-4185. Citado na página 35.

KRISTENSEN, N. *Fed-Batch Process Modelling for State Estimation and Optimal Control*. Tese (Doutorado) — Technical University of Denmark., 2002. Citado 2 vezes nas páginas 34 e 38.

LIMAYEM, A.; RICKE, S. C. Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. *Progress in Energy and Combustion Science*, Elsevier BV, v. 38, n. 4, p. 449–467, ago. 2012. ISSN 0360-1285. Citado 2 vezes nas páginas 19 e 25.

LIU, K.; ZHANG, J.; BAO, J. Two stage hydrolysis of corn stover at high solids content for mixing power saving and scale-up applications. *Bioresource Technology*, Elsevier BV, v. 196, p. 716–720, nov. 2015. ISSN 0960-8524. Citado 2 vezes nas páginas 28 e 30.

- LYND, L. R. et al. Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, American Society for Microbiology, v. 66, n. 3, p. 506–577, set. 2002. ISSN 1098-5557. Citado na página 29.
- MANKAR, A. R. et al. Pretreatment of lignocellulosic biomass: A review on recent advances. *Bioresource Technology*, Elsevier BV, v. 334, p. 125235, ago. 2021. ISSN 0960-8524. Citado na página 24.
- MAULIDIYAH, M. et al. Lignin degradation of oil palm empty fruit bunches using tio2 photocatalyst as antifungal of fusarium oxysporum. *Oriental journal of chemistry*, An International Open Free Access, Peer Reviewed Research Journal, v. 33, n. 6, p. 3101–3106, 2017. Citado na página 24.
- MEHRAN, K. Takagi-sugeno fuzzy modeling for process control. *Industrial Automation, Robotics and Artificial Intelligence (EEE8005)*, Newcastle University Newcastle upon Tyne, UK, v. 262, p. 1–31, 2008. Citado na página 31.
- NELLES, O. *Nonlinear System Identification*. 2nd. ed. [S.l.]: Springer Berlin Heidelberg, 2001. ISBN 978-3-030-47438-6. Citado 3 vezes nas páginas 13, 40 e 41.
- NETO, J. M. *Modelagem matemática do processo de hidrólise enzimática do bagaço de cana-de-açúcar*. Tese (phdthesis) — Universidade Estadual de Campinas, 2016. Citado na página 29.
- NOVÁK, V.; PERFILIEVA, I.; MOCKOR, J. *Mathematical principles of fuzzy logic*. [S.l.]: Springer Science & Business Media, 2012. v. 517. Citado na página 31.
- OLIVIERI, G. et al. Bioreactor and bioprocess design issues in enzymatic hydrolysis of lignocellulosic biomass. *Catalysts*, MDPI AG, v. 11, n. 6, p. 680, may 2021. Citado 2 vezes nas páginas 27 e 28.
- PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, FapUNIFESP (SciELO), v. 14, n. 2, p. 198–219, apr 2003. Citado na página 32.
- PATTANAIK, L. et al. Biofuels from agricultural wastes. Elsevier, p. 103–142, 2019. Citado na página 19.
- PAULA, A. R. de. *Monitoramento da hidrólise enzimática de bagaço de cana-de-açúcar por métodos espectrofotométricos*. Dissertação (mathesis) — Universidade Federal de São Carlos, 2020. Disponível em: <<https://repositorio.ufscar.br/handle/ufscar/12786>>. Citado na página 25.
- PEREIRA, S. C. et al. 2g ethanol from the whole sugarcane lignocellulosic biomass. *Biotechnology for Biofuels*, Springer Science and Business Media LLC, v. 8, n. 1, mar. 2015. ISSN 1754-6834. Citado na página 23.
- PEREZ, J. et al. Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *International Microbiology*, Springer Science and Business Media LLC, v. 5, n. 2, p. 53–63, apr 2002. Citado na página 27.

- PIPPO, W. A. et al. Energy recovery from sugarcane-trash in the light of 2nd generation biofuels. part 1: Current situation and environmental aspects. *Waste and Biomass Valorization*, Springer Science and Business Media LLC, v. 2, n. 1, p. 1–16, nov. 2010. ISSN 1877-265X. Citado na página 23.
- PRATTO, B. et al. Enzymatic hydrolysis of pretreated sugarcane straw: Kinetic study and semi-mechanistic modeling. *Applied Biochemistry and Biotechnology*, Springer Science and Business Media LLC, v. 178, n. 7, p. 1430–1444, dez. 2015. ISSN 1559-0291. Citado na página 30.
- PROCENTESE, A. et al. Kinetic characterization of enzymatic hydrolysis of apple pomace as feedstock for a sugar-based biorefinery. *Energies*, MDPI AG, v. 13, n. 5, p. 1051, fev. 2020. ISSN 1996-1073. Citado na página 30.
- PRUNESCU, R. M.; SIN, G. Dynamic modeling and validation of a lignocellulosic enzymatic hydrolysis process – a demonstration scale study. *Bioresource Technology*, Elsevier BV, v. 150, p. 393–403, dez. 2013. ISSN 0960-8524. Disponível em: <<https://doi.org/10.1016/j.biortech.2013.10.029>>. Citado na página 25.
- RIBEIRO, M. P. d. A. *Design and automation of a new reactor for enzymatic hydrolysis of lignocellulosic biomass at high solids loading*. Tese (phdthesis) — Universidade Federal de São Carlos, 2007. Disponível em: <<https://repositorio.ufscar.br/handle/ufscar/3849>>. Citado 2 vezes nas páginas 32 e 40.
- ROGER, J.-M.; MALLET, A.; MARINI, F. Preprocessing NIR spectra for aquaphotomics. *Molecules*, MDPI AG, v. 27, n. 20, p. 6795, oct 2022. Citado 2 vezes nas páginas 31 e 32.
- RUBIO, O. A. P. Integration of bioreactor and membrane separation processes: A model based approach: Reverse electro-enhanced dialysis process for lactic acid fermentation. Technical University of Denmark, 2010. Citado na página 37.
- RYTIOJA, J. et al. Plant-polysaccharide-degrading enzymes from basidiomycetes. *Microbiology and Molecular Biology Reviews*, American Society for Microbiology, v. 78, n. 4, p. 614–649, dez. 2014. ISSN 1098-5557. Citado 2 vezes nas páginas 13 e 23.
- SAGMEISTER, P. et al. Soft sensor assisted dynamic bioprocess control: Efficient tools for bioprocess development. *Chemical Engineering Science*, Elsevier BV, v. 96, p. 190–198, jun. 2013. ISSN 0009-2509. Citado na página 20.
- SAHA, B. C. Hemicellulose bioconversion. *Journal of Industrial Microbiology and Biotechnology*, Oxford University Press (OUP), v. 30, n. 5, p. 279–291, may 2003. Citado na página 27.
- SANCHEZ, C. Lignocellulosic residues: Biodegradation and bioconversion by fungi. *Biotechnology Advances*, Elsevier BV, v. 27, n. 2, p. 185–194, mar. 2009. ISSN 0734-9750. Citado na página 27.
- 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. Citado na página 42.
- SILVEIRA, M. H. L. et al. Current pretreatment technologies for the development of cellulosic ethanol and biorefineries. *ChemSusChem*, Wiley, v. 8, n. 20, p. 3366–3390, set. 2015. ISSN 1864-564X. Citado na página 25.

- SKOOG, D. A.; HOLLER, F. J.; CROUCH, S. R. *Principles of instrumental analysis*. [S.l.]: Cengage learning, 2017. Citado na página 32.
- SKVARIL, J.; KYPRIANIDIS, K. G.; DAHLQUIST, E. Applications of near-infrared spectroscopy (NIRS) in biomass energy conversion processes: A review. *Applied Spectroscopy Reviews*, Informa UK Limited, v. 52, n. 8, p. 675–728, jan 2017. Citado na página 32.
- SLUITER, A. et al. Determination of sugars, byproducts, and degradation products in liquid fraction process samples. *Golden: National Renewable Energy Laboratory*, v. 11, p. 65–71, 2006. Citado na página 34.
- SOUZA, A. P. de et al. Composition and structure of sugarcane cell wall polysaccharides: Implications for second-generation bioethanol production. *BioEnergy Research*, Springer Science and Business Media LLC, v. 6, n. 2, p. 564–579, nov 2012. Citado na página 19.
- SU, T. et al. Lignocellulosic biomass for bioethanol: Recent advances, technology trends, and barriers to industrial development. *Current Opinion in Green and Sustainable Chemistry*, Elsevier BV, v. 24, p. 56–60, aug 2020. Citado na página 25.
- SUAREZ, C. A. G. et al. Modeling the kinetics of complex systems: enzymatic hydrolysis of lignocellulosic substrates. *Applied biochemistry and biotechnology*, Springer, v. 173, p. 1083–1096, 2014. Citado na página 19.
- SUN, S. et al. The role of pretreatment in improving the enzymatic hydrolysis of lignocellulosic materials. *Bioresource Technology*, Elsevier BV, v. 199, p. 49–58, jan. 2016. ISSN 0960-8524. Citado na página 23.
- TAHA, M. et al. Commercial feasibility of lignocellulose biodegradation: possibilities and challenges. *Current Opinion in Biotechnology*, Elsevier BV, v. 38, p. 190–197, apr 2016. Citado na página 27.
- TAKAGI, T.; SUGENO, M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, Institute of Electrical and Electronics Engineers (IEEE), SMC-15, n. 1, p. 116–132, jan. 1985. ISSN 2168-2909. Citado 2 vezes nas páginas 31 e 37.
- YAO, S. et al. Efficient extraction of bagasse hemicelluloses and characterization of solid remainder. *Bioresource Technology*, Elsevier, v. 185, p. 21–27, 2015. Citado na página 36.
- YOUSUF, A.; PIROZZI, D.; SANNINO, F. Fundamentals of lignocellulosic biomass. Elsevier, p. 1–15, 2020. Citado na página 23.
- ZHANG, H.; WU, S.; XIE, J. Evaluation of the effects of isolated lignin on enzymatic hydrolysis of cellulose. *Enzyme and Microbial Technology*, Elsevier BV, v. 101, p. 44–50, jun. 2017. ISSN 0141-0229. Citado na página 24.
- ZHANG, Y. P.; LYND, L. R. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnology and Bioengineering*, Wiley, v. 88, n. 7, p. 797–824, nov. 2004. ISSN 1097-0290. Citado na página 24.
- ZHAO, L. et al. Advances in pretreatment of lignocellulosic biomass for bioenergy production: Challenges and perspectives. *Bioresource Technology*, Elsevier BV, v. 343, p. 126123, jan. 2022. ISSN 0960-8524. Citado 3 vezes nas páginas 15, 25 e 26.

APÊNDICE A – Parâmetros dos modelos *fuzzy*-cinéticos ajustados.

Tabela 7 – Parâmetros ajustados para o modelo MM-IP *fuzzy* linear

Parâmetro	Intervalo de confiança (2,5%)	Valor estimado do parâmetro	Intervalo de confiança (97,5%)	Unidade
Modelo de Alta concentração de sólidos (ACS)				
$k1_h$	5.338×10^{-4}	5.103×10^{-3}	2.099×10^{-2}	min^{-1}
$km1_h$	5.6669×10^{-6}	8.3294×10^{-6}	2.0407×10^{-6}	$g.L^{-1}$
$kp1_h$	1.6275×10^{-7}	5.8485×10^{-7}	5.0382×10^{-6}	$g.L^{-1}$
$k2_h$	4.5454×10^{-3}	7.5176×10^{-3}	1.8969×10^{-2}	min^{-1}
$km2_h$	7.1014×10^{-3}	7.7735×10^{-3}	1.1059×10^{-2}	$g.L^{-1}$
$kp2_h$	5.4883×10^{-2}	1.1814×10^{-1}	2.3286×10^{-1}	$g.L^{-1}$
$k3_h$	2.9391×10^{-3}	7.2923×10^{-3}	1.3880×10^{-2}	min^{-1}
$km3_h$	4.6021×10^{-9}	4.6021×10^{-9}	2.2763×10^{-5}	$g.L^{-1}$
$kp3_h$	6.9280×10^{-2}	1.5437×10^{-1}	1.1822×10^2	$g.L^{-1}$
$k4_h$	1.0340×10^{-3}	1.2927×10^{-3}	1.7294×10^{-3}	min^{-1}
$km4_h$	5.3173×10^{-5}	5.8554×10^{-5}	8.8062×10^{-5}	$g.L^{-1}$
$kp4_h$	6.0836×10^{-4}	7.0667×10^{-4}	1.0558×10^{-3}	$g.L^{-1}$
ke_h	9.8474×10^{-4}	1.1080×10^{-3}	1.2493×10^{-3}	min^{-1}
Modelo de Baixa concentração de sólidos (BCS)				
$k1_l$	3.1183×10^{-3}	1.2079×10^{-2}	8.5294×10^{-2}	min^{-1}
$km1_l$	5.5898×10^{-6}	7.9129×10^{-6}	1.6276×10^{-5}	$g.L^{-1}$
$kp1_l$	1.3468×10^{-7}	7.0973×10^{-7}	2.6224×10^{-6}	$g.L^{-1}$
$k2_l$	5.7129×10^{-3}	7.6699×10^{-3}	1.1940×10^{-2}	min^{-1}
$km2_l$	6.3293×10^{-3}	7.2204×10^{-3}	1.3101×10^{-2}	$g.L^{-1}$
$kp2_l$	9.0190×10^{-2}	1.1280×10^{-1}	2.2167×10^{-1}	$g.L^{-1}$
$k3_l$	2.5185×10^{-3}	9.3844×10^{-3}	1.8527×10^{-2}	min^{-1}
$km3_l$	5.0062×10^{-11}	9.5159×10^{-11}	4.0705×10^{-8}	$g.L^{-1}$
$kp3_l$	3.5319×10^{-1}	6.0241×10^{-1}	1.6536×10^2	$g.L^{-1}$
$k4_l$	3.7171×10^{-3}	4.2020×10^{-3}	5.3731×10^{-3}	min^{-1}
$km4_l$	2.1478×10^{-8}	2.2714×10^{-5}	1.7489×10^{-4}	$g.L^{-1}$
$kp4_l$	8.1662×10^{-3}	2.5652×10^{-2}	2.4317×10^{-1}	$g.L^{-1}$
ke_l	1.5902×10^{-3}	1.9403×10^{-3}	2.3554×10^{-3}	min^{-1}

Tabela 8 – Parâmetros ajustados para o modelo MM *fuzzy* sigmoide

Parâmetro	Intervalo de confiança (2,5%)	Valor estimado do parâmetro	Intervalo de confiança (97,5%)	Unidade
Modelo de Alta concentração de sólidos (ACS)				
$k1_h$	5.338×10^{-4}	5.103×10^{-3}	2.099×10^{-2}	min^{-1}
$km1_h$	5.6669×10^{-6}	8.3294×10^{-6}	2.0407×10^{-6}	$g.L^{-1}$
$k2_h$	4.5454×10^{-3}	7.5176×10^{-3}	1.8969×10^{-2}	min^{-1}
$km2_h$	7.1014×10^{-3}	7.7735×10^{-3}	1.1059×10^{-2}	$g.L^{-1}$
$k3_h$	2.9391×10^{-3}	7.2923×10^{-3}	1.3880×10^{-2}	min^{-1}
$km3_h$	4.6021×10^{-9}	4.6021×10^{-9}	2.2763×10^{-5}	$g.L^{-1}$
$k4_h$	1.0340×10^{-3}	1.2927×10^{-3}	1.7294×10^{-3}	min^{-1}
$km4_h$	5.3173×10^{-5}	5.8554×10^{-5}	8.8062×10^{-5}	$g.L^{-1}$
ke_h	9.8474×10^{-4}	1.1080×10^{-3}	1.2493×10^{-3}	min^{-1}
Modelo de Baixa concentração de sólidos (BCS)				
$k1_l$	3.1183×10^{-3}	1.2079×10^{-2}	8.5294×10^{-2}	min^{-1}
$km1_l$	5.5898×10^{-6}	7.9129×10^{-6}	1.6276×10^{-5}	$g.L^{-1}$
$k2_l$	5.7129×10^{-3}	7.6699×10^{-3}	1.1940×10^{-2}	min^{-1}
$km2_l$	6.3293×10^{-3}	7.2204×10^{-3}	1.3101×10^{-2}	$g.L^{-1}$
$k3_l$	2.5185×10^{-3}	9.3844×10^{-3}	1.8527×10^{-2}	min^{-1}
$km3_l$	5.0062×10^{-11}	9.5159×10^{-11}	4.0705×10^{-8}	$g.L^{-1}$
$k4_l$	3.7171×10^{-3}	4.2020×10^{-3}	5.3731×10^{-3}	min^{-1}
$km4_l$	2.1478×10^{-8}	2.2714×10^{-5}	1.7489×10^{-4}	$g.L^{-1}$
ke_l	1.5902×10^{-3}	1.9403×10^{-3}	2.3554×10^{-3}	min^{-1}