



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA FUNDAMENTAL
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

BRUNO HENRIQUE DA SILVA MELO

**MONITORAMENTO DA PRODUÇÃO DE AÇÚCAR EMPREGANDO
ESPECTRÔMETRO NIR PORTÁTIL**

Recife

2021

BRUNO HENRIQUE DA SILVA MELO

**MONITORAMENTO DA PRODUÇÃO DE AÇÚCAR EMPREGANDO
ESPECTRÔMETRO NIR PORTÁTIL**

Dissertação apresentada ao Programa de Pós-Graduação
em Química da Universidade Federal de Pernambuco
como requisito parcial à obtenção do título de Mestre em
Química.

Área de concentração: Química Analítica

Orientadora: Prof.^a Dr.^a Maria Fernanda Pimentel Avelar

Coorientadora: Dr.^a Rafaella de Figueiredo Sales

Recife

2021

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

M528m Melo, Bruno Henrique da Silva
 Monitoramento da produção de açúcar empregando espectrômetro NIR
 portátil / Bruno Henrique da Silva Melo. – 2021.
 80 f.: il., fig., tab.

 Orientadora: Maria Fernanda Pimentel Avelar.
 Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN,
 Química, Recife, 2021.

 Inclui referências e apêndice.

 1. Química analítica. 2. Regressão não-linear. I. Avelar, Maria Fernanda
 Pimentel (orientadora). II. Título.

543

CDD (23. ed.)

UFPE - CCEN 2022-59

BRUNO HENRIQUE DA SILVA MELO

MONITORAMENTO DA PRODUÇÃO DE AÇÚCAR EMPREGANDO ESPECTRÔMETRO NIR PORTÁTIL.

Dissertação apresentada ao Programa de Pós-Graduação no Departamento de Química Fundamental da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Química.

Aprovada em: 16/03/2021

BANCA EXAMINADORA

Profa. Maria Fernanda Pimentel Avelar (Orientadora)

Departamento de Engenharia Química
Universidade Federal de Pernambuco

Profa. Claudete Fernandes Pereira

Departamento de Química Fundamental
Universidade Federal de Pernambuco

Prof. Neirivaldo Cavalcante da Silva

Faculdade de Química
Universidade Federal do Pará

AGRADECIMENTOS

Agradeço a Deus por tantas bênçãos em minha vida, pela proteção e pessoas incríveis que tem colocado em meu caminho.

À toda minha família, em especial aos meus pais, Emy e Benjamim, meus irmãos Bárbara e Breno e cunhada Amanda, pelo apoio incondicional em todos esses anos longe de casa. Foram e são minha fortaleza em todos os momentos, não seria nada sem vocês.

À minha orientadora, Prof.^a Dr.^a Maria Fernanda Pimentel, a quem sou extremamente agradecido, desde o momento que aceitou um aluno até então desconhecido, mas confiou nas referências da Prof.^a Dr.^a Simone Simões e Dr. Everaldo Medeiros. A professora Fernanda é um exemplo de profissional, mulher, mãe e cientista, que inspira imensamente a todos os seus “filhos científicos”. Agradeço o acolhimento, confiança, paciência e valiosos conselhos e ensinamentos.

À minha coorientadora, Dr.^a Rafaella Sales, pelo imenso apoio, valiosas discussões, críticas, sugestões, dedicação e paciência. Poder contar com você e sua amizade foram essenciais para o desenvolvimento deste trabalho, assim como, crescimento pessoal e profissional.

Aos gestores e profissionais da usina onde este trabalho foi realizado, principalmente pela disponibilidade em ensinar e ajudar em qualquer necessidade, o que foi fundamental no entendimento do processo e rotina de trabalho. A interação Universidade/Indústria foi de grande enriquecimento pessoal e profissional.

Aos amigos do Laboratório de Combustíveis (LAC) em especial Aline, Amanda, Erklaylle, Fernando e Flávio, pela parceria, momentos de descontração, por compartilharem das dores e alegrias da pós-graduação e pela amizade sincera construída esses anos.

À Carolina, Rafaella e Neirivaldo pelas aulas maravilhosas nas reuniões de grupo, por todo o conhecimento transmitido, vocês são exemplo de dedicação e profissionalismo.

Agradeço pelo suporte dos profissionais do LAC, em especial Giseli e Vanessa e a todos que contribuíram com o desenvolvimento deste trabalho direta ou indiretamente.

Aos amigos que torcem pelo meu sucesso e felicidade, vocês transbordam minha vida de amor e carinho.

À coordenação e secretaria do Programa de Pós-Graduação em Química da UFPE por todo apoio, dedicação e zelo por seus alunos.

Ao CNPQ, pela concessão da bolsa de mestrado, à CAPES, ao INCTAA, NUQAAPE e FACEPE por todos os auxílios concedidos, à UFPE pelo suporte institucional.

RESUMO

O uso da espectroscopia no infravermelho próximo (NIRS) para identificar e quantificar diferentes tipos de analitos apresenta vantagens em comparação às metodologias convencionais: é uma técnica rápida, não destrutiva e não invasiva. Por essa razão, vem sendo amplamente empregada no setor industrial, pois permite agilidade na tomada de decisões diante de urgências na cadeia produtiva. No setor sucroalcooleiro em particular, destacam-se a determinação do teor de sólidos solúveis (brix) e sacarose aparente (pol), parâmetros importantes no monitoramento da produção de açúcar. Lidar com amostras reais em uma indústria pode ser desafiador, uma vez que efeitos inerentes ao processo estão comumente presentes nos dados, como os acarretados por variações no tipo de amostra, concentração de analitos, viscosidade, entre outros, que podem afetar a qualidade dos espectros e dificultar o tratamento de dados. Diferentes técnicas de calibração multivariada podem ser empregadas para lidar com estes efeitos, geralmente associados a desvios da linearidade. Nesse contexto, o presente trabalho direcionou-se a avaliar o desempenho de um espectrômetro NIR portátil e comparar as técnicas de regressão por mínimos quadrados parciais (*Partial Least Squares* - PLS) e máquinas de vetores de suporte (*Support Vector Machines* - SVM) na construção de modelos para quantificação de brix e pol de diversos substratos da fabricação de açúcar de uma usina situada no estado de Pernambuco, visando maior eficiência no controle do processo, rapidez na obtenção de resultados e controle de perdas. Para isso, medidas de transmitância e reflectância difusa, assim como as análises de referência, foram realizadas nas amostras provenientes dos processos de recepção de matéria-prima, extração e tratamento do caldo, evaporação, cristalização e cozimento, centrifugação e refinamento de açúcar VHP (*Very High Polarization*). As amostras foram coletadas em 2 safras (2019/2020 e 2020/2021), totalizando 1507, e ainda submetidas à análise pelo método de referência para brix, e 769 dessas, quantificadas também para pol. Através da análise espectral e comparação entre os modelos, foi constatada a presença de efeitos não-lineares, sobretudo nas amostras medidas por reflectância difusa, por serem mais concentradas e apresentarem maior variabilidade de tipo de amostra. Dentre os modelos obtidos, aqueles construídos com SVM demonstraram maior capacidade em lidar com não-linearidades, e menores valores de RMSEP (*Root Mean Square Error of Prediction*) foram observados: 0,59 e 0,69% (m/m) para brix e pol por transmitância, e 1,44 e 2,44% (m/m) para brix e pol por reflectância difusa, respectivamente. O emprego de espectrômetros portáteis e calibração multivariada no setor sucroalcooleiro mostra uma melhor relação custo/benefício comparado aos espectrômetros de bancada, que são em geral mais caros, indicando a viabilidade de se utilizar a estratégia de modelagem sugerida para o monitoramento dos parâmetros de qualidade do processo.

Palavras-chave: tecnologia analítica de processo; calibração multivariada; regressão não-linear; monitoramento de processo; cana-de-açúcar.

ABSTRACT

The use of near infrared spectroscopy (NIRS) to identify and quantify different types of analyte presents advantages in comparison to the conventional methodologies: it is a fast, non-destructive and non-invasive technique. For this reason, it has been widely employed in the industry, allowing a faster decision-making process in the face of emergencies in the production. In the sugarcane industry, in particular, the soluble solids content (brix) and apparent sucrose content (pol) are important parameters needed to be determined in the monitoring of sugar production. Dealing with real industrial samples can present challenging aspects because of the inherent effects within the process that are commonly present in the data, such as variations in type of sample, analyte concentration and viscosity. These effects can affect the quality of the spectra, hindering data processing. Different multivariate calibration techniques can be used to deal with these effects, generally associated with deviations from linearity. In this context, this work aimed to evaluate the performance of a portable NIR spectrometer in the development of calibration models to quantify brix and pol of various substrates of sugar production in a sugarcane mill located in the state of Pernambuco. A comparison between two regression techniques was also evaluated: partial least squares regression (Partial Least Squares - PLS) and support vector machines (Support Vector Machines - SVM). The objective of the work was to provide higher efficiency in process control, by obtaining faster results and controlling losses in the process. To this end, transmittance and reflectance measurements were acquired from the samples collected from the processes of cane reception, juice extraction and treatment, evaporation, crystallization, boiling, centrifugation and VHP (*Very High Polarization*) sugar refining. A total of 1507 samples were collected from two harvesting seasons (2019/2020 and 2020/2021). All of these were analysed by the reference method to quantify brix; 769 were also submitted to pol determination. Through spectral analysis and model comparison, it was possible to verify the presence of non-linear effects in the data, in particular for samples analysed in the reflectance mode, as they are more concentrated and present greater variability in type of sample. Among the models obtained, the SVM models demonstrated a greater ability to deal with nonlinearities. Lower values of RMSEP (Root Mean Square Error of Prediction) were then obtained for this approach: 0.59 and 0.69% (m/m) for brix and pol by transmittance, and 1.44 and 2.44% (m/m) for brix and pol by reflectance, respectively. The use of portable spectrometers and multivariate calibration in the sugar-alcohol sector demonstrated a higher benefit–cost ratio when compared to benchtop spectrometers, which are in general more expensive. This study confirms then the feasibility of using the modelling strategy suggested in this work for monitoring the quality parameters of the sugarcane production process.

Keywords: process analytical technology; multivariate calibration; non-linear regression; process monitoring; sugarcane.

LISTA DE FIGURAS

Figura 1 –	Esquema da moenda de quatro rolos.....	17
Figura 2 –	Esquema do conjunto ou tandem de moagem e embebição composta.....	18
Figura 3 –	Evaporador de quádruplo efeito.....	20
Figura 4 –	Sistema de cozimento de 3 massas.....	21
Figura 5 –	Amostras selecionadas utilizando o algoritmo <i>Onion</i> com diferentes valores de fração de loop.....	28
Figura 6 –	Ilustração do princípio da regressão por SVM.....	31
Figura 7 –	Espectrômetro portátil acoplado ao acessório de transmitância.....	37
Figura 8 –	Procedimento experimental para a determinação da sacarose aparente do Mel B.....	41
Figura 9 –	Espectros NIR pré-processados com SNV adquiridos utilizando cubetas de quartzo de diferentes caminhos ópticos.....	44
Figura 10 –	Gráficos dos valores preditos vs. referência com inclusão de amostras intermediárias.....	46
Figura 11 –	Espectros brutos das amostras coletadas no modo transmitância.....	47
Figura 12 –	Espectros de transmitância pré-processados.....	48
Figura 13 –	Espectros brutos das amostras coletadas no modo reflectância difusa.....	50
Figura 14 –	Espectros pré-processados das amostras coletadas no modo reflectância difusa	51
Figura 15 –	Espectros médios pré-processados com SNV para as amostras coletadas no modo reflectância difusa + espectro médio pré-processado com SNV para 20 amostras de açúcar refinado.....	52
Figura 16 –	Gráficos dos valores preditos vs. referência para os conjuntos de calibração e de predição; e os VIPs escores dos modelos PLS.....	56
Figura 17 –	Superfícies de resposta da primeira varredura para ajuste de parâmetros SVM para o modelo preditivo de brix por transmitância.....	59
Figura 18 –	Superfícies de resposta da segunda varredura para ajuste de parâmetros SVM para o modelo preditivo de brix por transmitância.....	60
Figura 19 –	Gráficos dos valores preditos vs. referência para os conjuntos de calibração e de predição dos modelos PLS (à esquerda) e SVM (à direita)	63

LISTA DE TABELAS

Tabela 1 –	Informações acerca das amostras coletadas na usina e analisadas por espectroscopia NIR.....	36
Tabela 2 –	Fatores de diluição para as amostras do processo de fabricação de açúcar.....	38
Tabela 3 –	Modelos desenvolvidos para monitoramento de parâmetros de qualidade.....	43
Tabela 4 –	Resultados para as etapas de calibração e de predição dos modelos PLS preditivos de brix para avaliação de inclusão de amostras intermediárias.....	46
Tabela 5 –	Resultados para as etapas de calibração e de predição dos modelos de regressão PLS para diferentes pré-processamentos.....	53
Tabela 6 –	Resultados para as etapas de calibração e de predição dos modelos de regressão SVM para avaliação de parâmetros de ajuste.....	61
Tabela 7 –	Métricas de desempenho dos modelos otimizados para brix e pol e comparação com os resultados de modelos obtidos por outros autores.....	67

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
ANN	Redes Neurais Artificiais (<i>Artificial Neural Networks</i>)
ATR	Açúcar Total Recuperável
FDA	Administração de Alimentos e Medicamentos (<i>Food and Drug Administration</i>)
IBGE	Instituto Brasileiro de Geografia e Estatística
iSPA	Algoritmo das Projeções Sucessivas para Seleção de Intervalos (<i>Successive Projections Algorithm for Interval selection</i>)
LVF	Filtro Linear Variável (<i>Linear Variable Filter</i>)
NIR	Infravermelho Próximo (<i>Near Infrared</i>)
MSC	Correção de Espalhamento Multiplicativo (<i>Multiplicative Scatter Correction</i>)
PAT	Tecnologia Analítica de Processo (<i>Process Analytical Technology</i>)
PCTS	Pagamento de Cana por Teor de Sacarose
PLS	Mínimos Quadrados Parciais (<i>Partial Least Squares</i>)
RBF	Função de Base Radial (<i>Radial Basis Function</i>)
RMSEC	Raiz do Erro Quadrático Médio de Calibração (<i>Root Mean Square Error of Calibration</i>)
RMSECV	Raiz do Erro Quadrático Médio de Validação Cruzada (<i>Root Mean Square Error of Cross-Validation</i>)
RMSEP	Raiz do Erro Quadrático Médio de Predição (<i>Root Mean Square Error of Prediction</i>)
SNV	Padronização Normal de Sinal (<i>Standard Normal Variate</i>)
SVM	Máquinas de Vetores de Suporte (<i>Support Vector Machines</i>)
VHP	Polarização Muito Alta (<i>Very High Polarization</i>)
VIP	Importância das Variáveis na Projeção (<i>Variable Importance in Projection</i>)
VL	Variáveis Latentes

SUMÁRIO

1	INTRODUÇÃO	12
2	OBJETIVOS	14
2.1	OBJETIVO GERAL	14
2.2	OBJETIVOS ESPECÍFICOS	14
3	FUNDAMENTAÇÃO TEÓRICA	15
3.1	PRODUÇÃO DE AÇÚCAR.....	15
3.1.1	Recepção da matéria-prima	15
3.1.1.1	Determinação do teor de sólidos solúveis totais	15
3.1.1.2	Determinação do teor de sacarose aparente	16
3.1.2	Preparo da cana para extração	17
3.1.3	Extração do caldo	17
3.1.4	Tratamento do caldo	18
3.1.5	Evaporação do caldo	19
3.1.6	Cozimento e cristalização	20
3.1.7	Refinamento de açúcar VHP	22
3.2	ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO (NIR).....	22
3.3	QUIMIOMETRIA.....	24
3.3.1	Pré-processamento de dados espectroscópicos	25
3.3.2	Calibração Multivariada	27
3.3.2.1	Método de seleção de amostras	27
3.3.2.2	Regressão por mínimos quadrados parciais	28
3.3.2.3	Regressão por máquinas de vetores de suporte	30
3.4	MONITORAMENTO DE PROCESSOS	33
3.4.1	Espectroscopia no infravermelho próximo aplicada ao monitoramento de processos no setor sucroalcooleiro	33
4	METODOLOGIA	36
4.1	COLETA DE AMOSTRAS	36
4.2	AQUISIÇÃO ESPECTRAL	37
4.3	ANÁLISES DE REFERÊNCIA.....	38
4.3.1	Determinação do teor de sólidos solúveis totais.....	38
4.3.2	Determinação do teor de sacarose aparente	39
4.3.2.1	Pol do caldo extraído.....	39

4.3.2.2	Pol dos licores, mel e massa refinada.....	40
4.3.2.3	Pol das demais amostras.....	40
4.4	ANÁLISE DE DADOS E DESENVOLVIMENTO DE MODELOS DE CALIBRAÇÃO	41
4.4.1	Desenvolvimento de modelos para definição do modo de aquisição espectral para amostras intermediárias do processo	41
4.4.2	Desenvolvimento dos modelos de calibração	42
5	RESULTADOS E DISCUSSÃO	44
5.1	SELEÇÃO DO CAMINHO ÓPTICO PARA MEDIDAS DE TRANSMITÂNCIA .	44
5.2	DEFINIÇÃO DA METODOLOGIA DE AQUISIÇÃO ESPECTRAL PARA AMOSTRAS INTERMEDIÁRIAS DO PROCESSO	45
5.3	DESENVOLVIMENTO E OTIMIZAÇÃO DOS MODELOS DE CALIBRAÇÃO .	47
5.3.1	Avaliação dos espectros de transmitância.....	47
5.3.2	Avaliação dos espectros de reflectância difusa	49
5.3.3	Avaliação de modelos PLS desenvolvidos com diferentes pré-processamentos ..	53
5.3.4	Avaliação de modelos SVM	58
6	CONCLUSÃO	68
7	PERSPECTIVAS FUTURAS.....	69
	REFERÊNCIAS	70
	APÊNDICE A - SUPERFÍCIES DE RESPOSTA DOS PARÂMETROS DE AJUSTE SVM PARA OS MODELOS PREDITIVOS DE POL POR TRANSMITÂNCIA E BRIX E POL POR REFLECTÂNCIA DIFUSA	75

1 INTRODUÇÃO

O Brasil atualmente é o maior produtor mundial de cana-de-açúcar e um dos principais exportadores de açúcar e álcool. A cultura desta matéria-prima é a terceira maior no país, tendo participação importante no PIB, além de estar difundida em todas as suas regiões. O estado de Pernambuco encontra-se como o segundo maior produtor do Norte e Nordeste, segundo dados da safra 2019/2020 (IBGE, 2020). Por ser uma das principais atividades econômicas da agricultura do estado, o setor demanda pesquisa de novas tecnologias para melhorias em processos e produtos.

No que diz respeito ao controle de qualidade no processamento da cana-de-açúcar, parâmetros chave, como o teor de sólidos solúveis totais (brix) e o teor de sacarose aparente (pol), são frequentemente determinados e, tradicionalmente, essas determinações são realizadas em laboratório utilizando técnicas *off-line*, que demandam preparo apropriado para a maioria das amostras, necessitando de etapas prévias de diluição e emprego de reagentes químicos nocivos. Além do mais, essa abordagem fornece uma informação atrasada sobre o processo, o que pode causar ineficiência em termos de seu monitoramento e controle (MEHROTRA; SIESLER, 2003).

A aplicação de métodos que proporcionam informações confiáveis sobre o desempenho das diversas etapas e operações do processo em tempo real pode oferecer uma solução efetiva para aprimorar o monitoramento, o controle, o entendimento e a tomada de decisões na indústria. Neste contexto, a espectroscopia no infravermelho próximo (NIR) pode ser uma importante aliada devido à sua natureza não destrutiva, rapidez e possibilidade de análises em tempo real, através do uso de analisadores *in-line*, *on-line* ou *in-situ* (PASQUINI, 2018). Quando associada à quimiometria, a espectroscopia NIR torna possível o desenvolvimento de modelos de calibração multivariada, que podem ser empregados para acompanhamento de parâmetros de qualidade ao longo do processo de produção de açúcar, conforme descrito por Simpson e Naidoo (2010) e Tange *et al.* (2015), que obtiveram resultados satisfatórios no que diz respeito ao uso de instrumentos NIR de bancada para a determinação de brix e pol de amostras coletadas em algumas etapas do processo.

O desenvolvimento tecnológico dos últimos anos tem contribuído para a miniaturização dos espectrômetros, tornando-os mais compactos e robustos. O custo fundamentalmente mais baixo e o pequeno formato desses instrumentos os tornam promissores para aplicações em monitoramento e controle de processos, podendo trazer benefícios para a indústria (GALAVARNA *et al.*, 2018; PUIG-BERTOTTO; COELLO; MASPOCH, 2019).

No setor sucroalcooleiro, a utilização de espectrômetros NIR portáteis é focada principalmente nas aplicações em campo. Vários trabalhos descritos na literatura demonstraram resultados satisfatórios quanto à quantificação de parâmetros de qualidade, como os teores de fibra, pol e brix, medidos diretamente no colmo de cana-de-açúcar ou no caldo extraído do mesmo (MARAPHUM *et al.*, 2018; PHUPHAPHUD *et al.*, 2019; TAIRA; UENO; SAENGPRACHATANARUG; *et al.*, 2013). Entretanto, até o momento, não há relatos na literatura de trabalhos que utilizam um NIR portátil visando o monitoramento do processo de fabricação de açúcar em suas diversas etapas, o que confere caráter inédito ao presente trabalho.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Avaliar o desempenho de um espectrômetro portátil no infravermelho próximo, associado a técnicas quimiométricas, para determinação dos principais parâmetros de qualidade necessários ao monitoramento do processo de fabricação de açúcar de uma usina sucroalcooleira, visando maior eficiência no controle do processo, rapidez na obtenção de resultados e controle de perdas.

2.2 OBJETIVOS ESPECÍFICOS

- Coletar amostras das diversas etapas da fabricação de açúcar: Recepção de matéria-prima; Extração e Tratamento do caldo; Evaporação; Cristalização e Cozimento; Centrifugação; e Refinamento do açúcar VHP (*Very High Polarization*);
- Adquirir espectros no infravermelho próximo de todas as amostras utilizando um espectrômetro portátil ultracompacto;
- Determinar os parâmetros de qualidade, brix e pol das amostras empregando os métodos de referência;
- Desenvolver modelos de regressão por mínimos quadrados parciais (PLS) para quantificação de brix, com o intuito de definir a metodologia de aquisição espectral mais adequada para as amostras intermediárias do processo (caldos concentrados do 1º e do 2º efeito da evaporação);
- Utilizar técnicas de regressão por PLS e máquinas de vetores de suporte (SVM) para construção e otimização dos modelos de calibração, e avaliá-los quanto à quantificação mais eficiente para brix e pol nas etapas do processo de fabricação do açúcar.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 PRODUÇÃO DE AÇÚCAR

A produção de açúcar sob a forma de sacarose cristalizada a partir da cana-de-açúcar é realizada pelas usinas, cujo processo tecnológico é formado por uma sequência de operações industriais que se resumem a:

- Extração do caldo dos colmos de cana-de-açúcar;
- Tratamento do caldo extraído;
- Concentração do caldo clarificado por evaporação e cristalização da sacarose.

Essas operações, assim como outras etapas importantes do processo, estão descritas a seguir.

3.1.1 Recepção da matéria-prima

Antes do seu processamento, a cana-de-açúcar é analisada pelo laboratório de PCTS (Pagamento de Cana por Teor de Sacarose) para determinação do ATR (Açúcar Total Recuperável), parâmetro estimado a partir dos teores de brix, pol e fibras que define o valor a ser pago aos fornecedores da cana, assim como para atestar a qualidade da matéria-prima que entra na fábrica. Nesta etapa, alíquotas são coletadas dos caminhões transportadores por meio de sondas e logo após desintegradas e homogeneizadas, gerando o bagaço.

Para as análises, $500 \pm 0,5$ g do bagaço são pesados em balança semi-analítica e enviados para a prensa hidráulica, onde o caldo é extraído com pressão mínima e constante de 24,5 Mpa, correspondente à 250 kgf/cm² sobre a amostra de bagaço durante 1 minuto (CONSECANA, 2015). Após extração, o caldo extraído é analisado quanto a brix e pol, e o bagaço úmido é pesado para estimativa do teor de fibras.

3.1.1.1 Determinação do teor de sólidos solúveis totais

A determinação do brix (% m/m) é realizada em refratômetro digital, de leitura automática, com correção de temperatura e resolução máxima de 0,1% (m/m), devendo o valor

final ser expresso a 20°C. A leitura é realizada após filtração de parte do caldo extraído em papel de filtro qualitativo, utilizando algumas gotas do filtrado para tal.

3.1.1.2 Determinação do teor de sacarose aparente

A sacarose aparente é medida através de técnica de polarização, definida como a porcentagem em massa de sacarose aparente contida em uma solução açucarada, determinada pelo desvio da luz polarizada ao atravessar essa solução (INSTITUTO ADOLFO LUTZ, 2008).

Os açúcares são substâncias quirais (opticamente ativos), nos quais a molécula tem um átomo de carbono assimétrico, ou seja, está ligada a quatro grupos diferentes, provocando a rotação (desvio angular) do plano de uma luz polarizada. A análise da polarização de açúcar é realizada em um equipamento chamado sacarímetro, um polarímetro construído com uma escala que permite a leitura em graus sacarimétricos ao invés de graus angulares. A escala sacarimétrica, ou escala internacional de açúcar, é construída de forma que uma solução de 26,000 g de sacarose pura, dissolvida em água até um volume final de 100,000 mL, a 20 °C, apresente o valor 100 °Z na escala, sendo essa solução denominada solução normal. O ponto zero da escala sacarimétrica é o conferido à água pura (LOPES, 2011). Não deve ser esquecido que a rotação óptica de uma solução é a soma algébrica da rotação de todos os componentes opticamente ativos presentes na solução, e não da sacarose somente. Dessa forma, para as soluções reais, que não são compostas somente de sacarose, deve-se considerar a pol como sendo o teor de sacarose aparente.

Como o caldo extraído de cana-de-açúcar é composto por uma série de constituintes, ele não apresenta uma transmitância óptica que permita a passagem do raio de luz polarizada do equipamento analisador, por isso, é realizada uma etapa anterior de clarificação, para remoção de coloides e impurezas que conferem cor ao caldo, sendo o sub-acetato de chumbo ($\text{Pb}(\text{C}_2\text{H}_3\text{O}_2)_2$) o clarificante químico mais empregado nas usinas. O mesmo, segundo a CONSECANA-PE (Conselho de Produtores de Cana-de-açúcar, Açúcar e Etanol do Estado de Pernambuco), é adicionado na relação mínima de 1,5 g / 100 mL de caldo extraído, podendo ser utilizado o agitador eletromagnético por 30 segundos e/ou mixer para homogeneização do clarificante, seguido de filtração em papel de filtro qualitativo. A leitura do filtrado é realizada em sacarímetro digital automático a 589 nm com tubo polarimétrico de 200 mm com fluxo contínuo e temperatura corrigida a 20 °C. Posteriormente, a leitura em °Z é convertida em pol (%m/m).

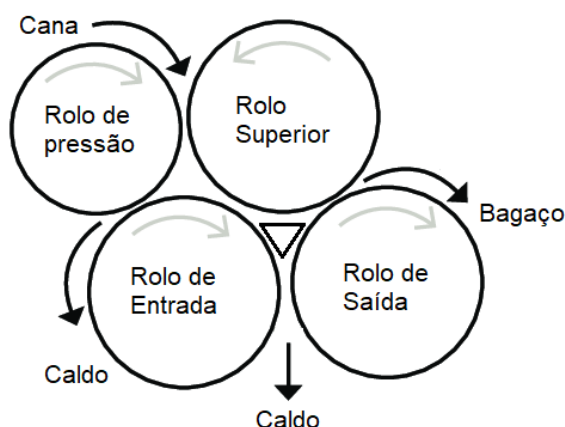
3.1.2 Preparo da cana para extração

Atestada a qualidade da matéria-prima, a cana segue para descarregamento em pátio de armazenagem, ou diretamente na mesa alimentadora, onde é lavada por jatos de água quente ou fria para remoção de impurezas, como terra, areia e detritos. Em seguida, a cana é picada e desfibrada, com o objetivo de romper a sua estrutura, facilitando a extração do caldo na moagem e assim aumentando sua eficiência. Paralelamente, o colchão de cana formado é nivelado, a fim de garantir uma alimentação contínua e uniformizada na etapa seguinte.

3.1.3 Extração do caldo

A extração do caldo da cana é uma operação realizada, na maioria das usinas brasileiras, pelo processo de moagem, que consiste em fazer o colchão de cana-de-açúcar ser comprimido entre pares de cilindros. Uma moenda (ou terno) é classicamente constituída por três rolos e recebe duas pressões, sendo a primeira entre o rolo superior e o rolo de entrada e uma segunda compressão entre o rolo superior e o rolo de saída. Possui também um rolo de pressão que é responsável por uma melhor alimentação e uma bagaceira com a finalidade de conduzir o bagaço, parcialmente esmagado, para a próxima compressão (HUGOT, 1977; PAYNE, 1989), como pode ser observado na Figura 1.

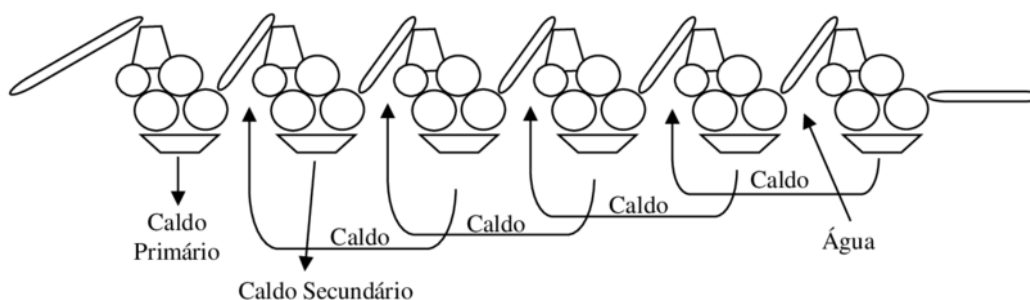
Figura 1- Esquema da moenda de quatro rolos.



Fonte: Adaptado de Zanardi e Junior (2016)

Nas usinas, encontram-se geralmente, de 4 a 6 moendas acopladas em série, formando o chamado conjunto ou tandem de moagem, como exemplificado na Figura 2. Após a passagem do bagaço por todos os ternos, pode-se considerar que a sacarose contida na cana foi extraída até seu limite econômico.

Figura 2 – Esquema do conjunto ou tandem de moagem e embebição composta.



Fonte: Adaptado de Zanconato e Bukvic (2018)

No contrafluxo da cana, na última moenda, é adicionada água quente ao colchão de cana, a fim de extrair o máximo de sacarose retida no bagaço e aumentar o rendimento industrial. O caldo diluído obtido nessa unidade retorna para a moenda anterior pela mesma razão, e o ciclo é repetido até chegar na segunda moenda, conforme observado na Figura 2. Este procedimento é chamado de embebição composta (FILHO; PICCIRILLI, 2012). A mistura dos caldos primário (obtido da primeira moenda) e secundário (obtido do processo de embebição) formam o caldo misto que segue para etapa de tratamento, enquanto o bagaço resultante da extração é enviado às caldeiras para ser utilizado como combustível. O calor gerado aquece a água contida em tubulações, formando o vapor que movimenta a fábrica e torna as usinas autossuficientes em energia (PAYNE, 1989)

3.1.4 Tratamento do caldo

O caldo possui uma composição bastante complexa, onde estão presentes, além da sacarose, outros carboidratos como glicose e frutose, compostos orgânicos como ácidos, aminoácidos, proteínas e gomas, além de diversos componentes inorgânicos (HUGOT, 1977). Os materiais presentes no caldo se apresentam tanto sob a forma solubilizada, como sob a forma de suspensão coloidal. O mesmo também contém impurezas como bagacilhos (resíduos do

bagaço) e areia, denominados material grosseiro. O objetivo do tratamento do caldo da cana é a máxima remoção dessas impurezas, utilizando métodos físicos e químicos.

Como tratamento físico para a remoção de material grosseiro em suspensão, a filtração com peneira rotativa é a mais comum e aceita pelas usinas em geral. A capacidade de retenção desse equipamento pode chegar a 83%. Após filtração, o caldo filtrado segue para a adição do leite de cal (Ca(OH)_2) e ácido fosfórico (H_3PO_4), que tem como propósito a elevação do pH para valores próximos de 7,0, com o propósito de inibir a inversão da sacarose e desencadear a formação do fosfato de cálcio, que precipita e arrasta consigo impurezas presentes no caldo (FILHO; PICCIRILLI, 2012; LOPES, 2011).

O caldo caleado segue para trocadores de calor onde atinge temperatura média de 105 °C; a troca térmica é realizada com vapor de água saturado em geral proveniente das caixas de evaporação. As principais finalidades desta elevação de temperatura são descritas por Filho e Piccirilli (2012):

- Acelerar e facilitar a coagulação e floculação de coloides e não-açúcares;
- Aumentar a eficiência da decantação;
- Eliminar microrganismos por esterilização;
- Liberar gases presentes no caldo.

A posterior decantação do caldo ocorre nos decantadores que operam em regime contínuo. As partículas em suspensão depositam-se no fundo do decantador pela ação da gravidade e com o auxílio de raspadores, que giram a velocidades constantes e baixas, direcionando as impurezas (lodo) ao fundo, de onde é conduzido à filtragem. No topo do decantador é coletado o caldo clarificado (ou decantado), que é enviado ao setor de evaporação para concentração (HUGOT, 1977).

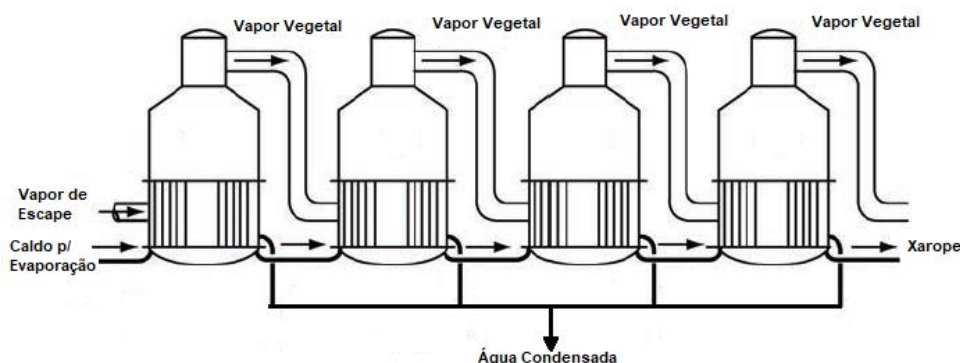
3.1.5 Evaporação do caldo

A evaporação é uma operação unitária que tem por objetivo retirar a água presente no caldo, resultando num produto líquido de maior concentração, o xarope, com brix variando entre 65 e 70% (m/m).

A concentração do caldo ocorre em uma série de evaporadores de funcionamento subsequente, conhecidos como evaporadores de múltiplo efeito, denominados conforme o

número de efeitos (caixas) de evaporação, geralmente de quádruplo ou quántuplo efeito, ou seja, quatro ou cinco caixas, respectivamente. Nesta etapa, é necessário injetar vapor apenas no primeiro efeito de evaporação, pois a própria água evaporada irá aquecer o caldo nas caixas seguintes. O vapor que alimenta a primeira caixa é denominado vapor de escape (proveniente da turbina de acionamento de uma moenda ou de outro equipamento). Enquanto o vapor gerado pela ebulição do caldo da primeira caixa, que é aproveitado para provocar ebulição na caixa seguinte, é denominado vapor vegetal (PAYNE, 1989). Um esquema para um evaporador de quádruplo efeito é mostrado na Figura 3.

Figura 3 – Evaporador de quádruplo efeito.



Fonte: Adaptado de Lopes (2011)

Este procedimento sequencial só pode ser obtido com a diferença decrescente de pressão existente entre as caixas, mantida por um gerador de vácuo ligado ao último efeito, ou seja, o decaimento da temperatura é compensado pela diminuição da pressão ao longo das caixas.

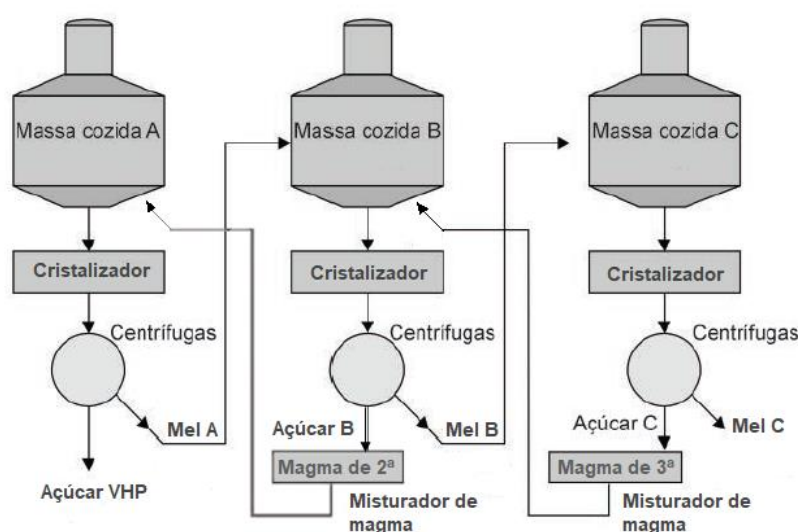
O caldo inicia o processo com uma temperatura de 120 a 125 °C e brix entre 14 e 16% (m/m), chegando ao fim do último efeito com brix de 65 a 70% (m/m) e com ponto de ebulição reduzido ao patamar de 60 °C por meio da baixa pressão do sistema, entre 58 e 62 cm Hg (FILHO; PICCIRILLI, 2012).

3.1.6 Cozimento e cristalização

O cozimento é a segunda etapa de concentração do caldo, onde ocorrerá a formação e crescimento dos cristais de sacarose. O xarope é aquecido em evaporadores de simples efeito, denominados cozedores, similares aos evaporadores de múltiplo efeito, mas operam totalmente independentes uns dos outros, pois a densidade do produto não permite a sua transferência de um cozedor a outro (LOPES, 2011). Através da baixa pressão e de trocas térmicas com vapor

nos cozedores, a água evapora do xarope para tornar a solução sobressaturada, que possui elevada viscosidade. Neste ponto, a chamada semente é adicionada, formada por uma suspensão de açúcar em álcool na proporção de 1:2 (m/v), que vai facilitar a nucleação e crescimento dos cristais de açúcar no mel, gerando assim a massa cozida. Quando os cristais atingem tamanho ideal, a massa cozida é descarregada nos cristalizadores, dotados de sistema de resfriamento que finalizam o processo de cristalização e crescimento dos cristais. Estes equipamentos servem de “pulmão” para as centrífugas, que são alimentadas pela massa, e separam os cristais de açúcar do mel por centrifugação. Posteriormente, os cristais são lavados com jatos e/ou vapor de água para remoção da película de mel que envolve o cristal de açúcar. Para uma máxima recuperação de sacarose, o mel obtido na centrifugação, passa por mais duas etapas de cristalização, visando seu total esgotamento, ou seja, a máxima recuperação de sacarose economicamente viável (FILHO; PICCIRILLI, 2012; HUGOT, 1977). Este processo completo é denominado cozimento de 3 massas e é representado na Figura 4.

Figura 4 – Sistema de cozimento de 3 massas.



Fonte: Adaptado de Lopes (2011)

As massas e méis mencionados anteriormente, agora possuem 3 diferentes nomenclaturas, massas e méis A, B e C, respectivamente, estes diferem em teor de pureza aparente, que decresce de A a C. Nesse processo, o cozedor C recebe a semente de açúcar em álcool mencionada anteriormente e, ao final do processo de cristalização e centrifugação, produz o açúcar C de baixa pureza e cor elevada. Este por sua vez, é misturado com água condensada formando uma pasta chamada de magma de terceira, que serve de semente no

cozedor B, que após os respectivos processos de cristalização e centrifugação, produz o açúcar B, de cristais maiores e de cor mais clara que o C, entretanto a pureza ainda não é ideal. Então o processo se repete, sendo o açúcar B misturado com água condensada para formar o magma de segunda, que é a semente utilizada no cozedor A, que após cristalização e centrifugação vai gerar o açúcar VHP, com cristais bem definidos, elevada pureza (entre 99 e 99,49 °Z) e de cor amarelo claro.

Ao final do processo, o mel C, também chamado de melaço, é enviado para a produção de álcool, pois a recuperação de sacarose nesse subproduto é economicamente inviável devido à sua baixa pureza. Já o açúcar VHP obtido, é seco e destinado para exportação ou refinaria para se obter o açúcar refinado.

3.1.7 Refinamento de açúcar VHP

O açúcar VHP é reprocessado com o objetivo de se obter o açúcar refinado com alta polarização (acima de 99,85 °Z) e, conseqüentemente, mais claro. Para isso, o açúcar VHP é dissolvido em água condensada, formando o licor que tem brix em torno de 60% (m/m), e é adicionado de descolorantes para remoção de impurezas. Essas, por sua vez, são removidas por processos de filtração e flotação. O licor filtrado gerado, como no processo de obtenção do açúcar VHP, segue para as etapas de cozimento e cristalização em sistema de apenas 1 massa (devido à sua alta pureza) e por fim, é centrifugado, separando os cristais de açúcar da massa refinada. O mel refinado obtido na centrifugação retorna ao processo de clarificação para elevação da pureza. O processo finaliza com a secagem do açúcar refinado, que posteriormente é ensacado e armazenado até expedição.

3.2 ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO (NIR)

A espectroscopia no infravermelho é uma espectroscopia vibracional que se baseia na energia dos fótons absorvida pelas moléculas para gerar mudanças nos seus estados vibracionais e rotacionais. As moléculas são formadas por ligações químicas, que se mantêm em constante vibração e deslocamento em uma frequência que dependerá da força de ligação e das massas dos átomos envolvidos. A maior parte das moléculas encontram-se em estados fundamentais de energia vibracional. Quando um feixe de radiação eletromagnética na região do infravermelho próximo incide sobre uma molécula, esse feixe não apresenta energia suficiente para promover transições eletrônicas, no entanto, a energia absorvida será suficiente para

provocar mudanças no nível vibracional e/ou rotacional das moléculas, levando ao aumento da amplitude do movimento (PASQUINI, 2003).

As vibrações moleculares podem se apresentar na forma de estiramentos ou deformações angulares. Quando há uma mudança na posição relativa dos átomos da molécula, ou seja, na distância interatômica ao longo do eixo de ligação, ocorrem as vibrações de estiramento, que podem ser simétricas ou assimétricas. Já as deformações angulares ocorrem quando há mudança em relação ao ângulo de ligação entre duas ligações (SKOOG; HOLLER; CROUCH, 2018). É importante ressaltar que para que ocorram mudanças nos níveis de vibração da molécula, é necessário mais do que a incidência de um feixe com energia adequada. É preciso também que o campo elétrico oscilante da radiação interaja com o campo elétrico resultante da variação do momento dipolo que ocorre durante o movimento vibracional. A magnitude da variação do momento dipolo está relacionada com a intensidade da banda de absorção (PASQUINI, 2003).

As absorções na região do NIR estão associadas, principalmente, aos sobretons e às bandas de combinação das vibrações fundamentais das ligações C—H, N—H, O—H e S—H. Assim, as principais aplicações dessa técnica estão relacionadas às determinações quantitativas e qualitativas de espécies que apresentam grupos funcionais contendo essas ligações (SKOOG; HOLLER; CROUCH, 2018).

A espectroscopia vibracional na região do NIR estende-se da região mais próxima do visível a 750 nm até 2500 nm (em número de onda: 4.000 a 12.500 cm^{-1}). (SKOOG; HOLLER; CROUCH, 2018). Os espectros obtidos nessa região apresentam bandas largas, com sobreposições de sobretons e bandas de combinação, algo que costumava trazer dificuldades com relação à especificidade e ao uso da técnica. No entanto, com o avanço dos métodos estatísticos e a aplicação da quimiometria, a espectroscopia NIR vem sendo cada vez mais utilizada e aplicada, inclusive, em processos industriais. Dentre as diversas vantagens do uso dessa técnica para essa finalidade, pode-se citar a possibilidade de realizar análises rápidas, não-destrutivas e diretas, ou seja, sem a necessidade de preparo de amostras (PASQUINI, 2003).

O crescente desenvolvimento tecnológico proporcionou a miniaturização dos espectrômetros, inclusive NIR, algo que torna muito prático sua aplicação em contexto industrial, através de análises na linha de produção ou em campo (GALAVARNA *et al.*, 2018). Outra vantagem desses aparelhos compactos e robustos é o seu baixo custo; em contrapartida, há redução no desempenho óptico e na janela espectral comparado aos equipamentos de bancada. Entretanto, Puig-Bertotto; Coelho e Maspocho (2019), ao avaliar a quantificação de

dois princípios ativos em comprimidos, afirmaram que a precisão dos resultados obtida com um espectrômetro portátil (MicroNIR na faixa espectral de 950 a 1650 nm), apesar de ser menor em comparação a um NIR de bancada, é elevada o suficiente para ser utilizada em diversas situações práticas. Os autores completam que o fato de o MicroNIR registrar principalmente a faixa espectral dos sobretons das ligações, não é um fator limitante, visto que há informações nesta faixa para resolver a maioria dos problemas analíticos. Diversos trabalhos do nosso grupo têm também evidenciado a eficiência dos equipamentos portáteis para diferentes aplicações, tais como: monitoramento da produção de biodiesel (SALES *et al.*, 2019); avaliação da qualidade de combustíveis (SILVA *et al.*, 2019) e de frutas (MALEGORI *et al.*, 2017; MARQUES *et al.*, 2016); e detecção de sangue em diferentes substratos (PEREIRA *et al.*, 2017).

3.3 QUIMIOMETRIA

O termo “Quimiometria” foi usado pela primeira vez por Svante Wold, em 1972, com o intuito de se referir a seu grupo de pesquisa, pois aplicava matemática a problemas de química. Alguns anos depois, em 1975, Bruce Kowalski definiu a Quimiometria como um novo ramo da química a ser utilizado para extrair informações químicas úteis de dados brutos. Naquela época, o uso de computadores ainda não era popularizado, e as técnicas quimiométricas estavam sendo prioritariamente utilizadas para reconhecimento de padrões (BROWN, 2017). Ambos os autores não imaginariam a pluralidade de técnicas que são atualmente utilizadas e suas infinitudes de aplicações.

Com o avanço das tecnologias e o aumento da disponibilidade de equipamentos analíticos cada vez mais avançados, o acesso a um maior número de informações é algo cada vez mais iminente. Dessa forma, a demanda pelo uso de métodos estatísticos avançados capazes de tratar dados multivariados de origem química vem aumentando progressivamente (BRERETON *et al.*, 2017). Dentre as principais aplicações da quimiometria na área de Química Analítica, uma das que mais se destaca é o desenvolvimento de modelos de calibração multivariada (FERREIRA; TOBYN, 2015). Usualmente, a aplicação das técnicas quimiométricas para solução de problemas analíticos, empregando dados espectrais, envolve quatro etapas: a avaliação dos dados espectrais e o seu pré-tratamento, análise exploratória dos dados, a construção de modelos e a validação desses modelos construídos.

3.3.1 Pré-processamento de dados espectroscópicos

A maioria dos dados espectrais devem ser submetidos a algum tipo de pré-processamento dependendo da instrumentação utilizada para adquirir os dados. Essa é uma etapa essencial para construção de modelos quimiométricos e seu principal objetivo é a remoção de informações não relacionadas ao objetivo da modelagem, como a presença de ruídos e variação de linha de base (RINNAN; BERG; ENGELSEN, 2009). Essa etapa é de extrema importância e deve ser realizada com bastante cuidado, geralmente, em um processo de tentativa e erro, pois ela pode acabar distorcendo informações de interesse analítico ou amplificando sinais de ruídos (FERREIRA, 2015). De modo que é essencial ter um conhecimento prévio dos dados.

Dentre as principais técnicas de pré-processamento utilizadas para tratamento de espectros no infravermelho próximo estão a correção de espalhamento multiplicativo (MSC) e a padronização normal de sinal (SNV), que são usadas para corrigir variações de linha de base, decorrentes de efeitos aditivos e multiplicativos devido ao espalhamento da radiação. Os pré-tratamentos mencionados são aplicados simultaneamente a todas as variáveis das amostras, de modo que todo o espectro é transformado de acordo com o método selecionado (RINNAN; BERG; ENGELSEN, 2009). Além disso, a derivada também é bastante empregada para correção dos efeitos decorrentes da dispersão da radiação, sendo capaz de remover os efeitos aditivos e multiplicativos dos espectros.

A Padronização Normal de Sinal se baseia em uma centralização na média do próprio espectro, seguido de uma normalização pelo desvio padrão dos valores de intensidade espectral. O espectro corrigido ($x_{i(corr)}$) pode ser obtido através da Equação 1, em que $a_{0,i}$ é o valor médio de intensidade do espectro e $a_{1,i}$ é o desvio-padrão dos valores de intensidade do espectro que será corrigido ($x_{i,org}$) (RINNAN; BERG; ENGELSEN, 2009).

$$x_{i(corr)} = \frac{x_{i,org} - a_{0,i}}{a_{1,i}} \quad (1)$$

A Correção de Espalhamento Multiplicativo é uma ferramenta que, como o nome já diz, tem o objetivo corrigir o espalhamento da radiação. Ela se baseia na estimativa dos coeficientes de regressão linear ($b_{0,i}$) e angular ($b_{1,i}$), calculados conforme pode ser visto na Equação 2, em que x_i é um dos espectros originais, x_{ref} é um espectro de referência que, geralmente,

corresponde à média dos espectros e $\mathbf{e}_i$ é um vetor de resíduos (RINNAN; BERG; ENGELSEN, 2009).

$$\mathbf{x}_i = b_{0,i} + b_{1,i} \cdot \mathbf{x}_{ref} + \mathbf{e}_i \quad (2)$$

A partir do cálculo dos coeficientes de correção, pode-se calcular o espectro corrigido $\mathbf{x}_{i(corr)}$ conforme a Equação 3.

$$\mathbf{x}_{i(corr)} = \frac{\mathbf{x}_i - b_{0,i}}{b_{1,i}} = \mathbf{x}_{ref} + \frac{\mathbf{e}_i}{b_{1,i}} \quad (3)$$

Geralmente, o formato dos espectros corrigidos utilizando o MSC assemelha-se bastante com os espectros corrigidos com SNV, mas sua geometria no espaço espectral não é a mesma (FEARN *et al.*, 2009).

Outra técnica bastante utilizada no tratamento de espectros é a derivada. Essa técnica é muito útil para corrigir os efeitos de dispersão da radiação, removendo variações de linha de base, e pode auxiliar na separação de bandas sobrepostas, aumentando a resolução dos dados. Enquanto a primeira derivada é empregada para corrigir efeitos aditivos, a segunda derivada é útil para a correção de efeitos multiplicativos (OTTO, 2017; RINNAN; BERG; ENGELSEN, 2009; BEEBE; PELL; SEASHOLTZ, 1998).

Como os sinais são intensificados ao utilizar essa técnica, uma consequência de seu uso é o aumento dos ruídos espectrais. Dessa forma, dificilmente ela é empregada sozinha, sendo frequentemente associada a um método de suavização. Nesse sentido, existem diferentes métodos que podem ser aplicados, como o Norris-William e o Savitzky-Golay, sendo esse último o mais utilizado. Esse método ajusta uma função polinomial para uma janela de pontos, e o ponto central da janela é substituído pelo valor obtido após a derivada da função polinomial. Essa janela é então deslocada até percorrer todo o espectro, permitindo a sua suavização. O grau do polinômio a ser ajustado e a janela de pontos são parâmetros decididos pelo quimiometrista, e devem ser escolhidos com atenção (RINNAN; BERG; ENGELSEN, 2009; BEEBE; PELL; SEASHOLTZ, 1998).

3.3.2 Calibração Multivariada

Na química analítica, a Calibração Multivariada é uma técnica de modelagem utilizada, em geral, para relacionar uma medida instrumental a uma propriedade de interesse que se deseja quantificar, sendo uma das linhas da quimiometria bastante aplicada no monitoramento de processos (FERREIRA; TOBYN, 2015). Neste tópico, são abordados alguns temas da calibração multivariada que foram aplicados no desenvolvimento deste trabalho: o método *Onion* utilizado para a separação de um conjunto de amostras em calibração e predição e os métodos de regressão PLS e SVM.

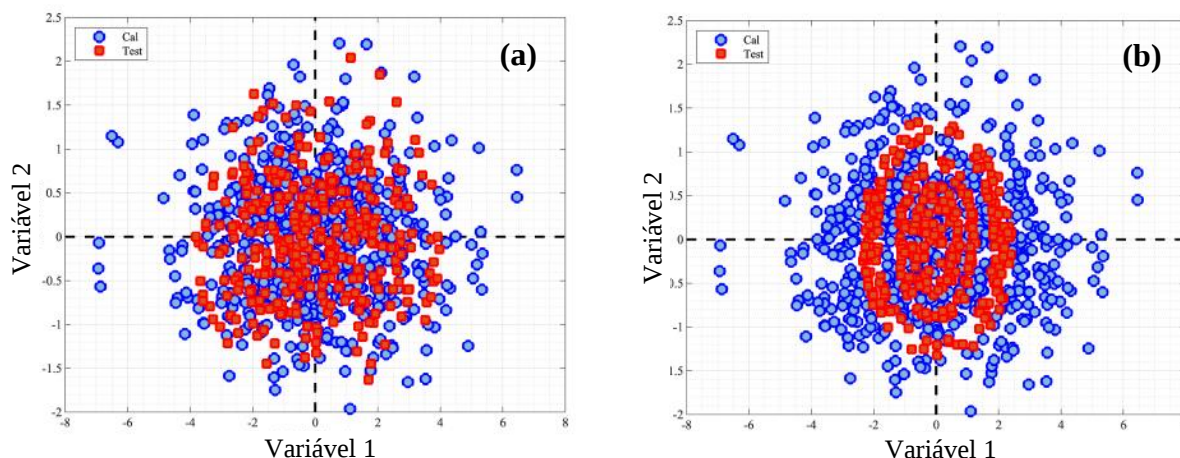
3.3.2.1 Método de seleção de amostras

Com a finalidade de construir e validar modelos de calibração multivariada, o conjunto de dados original é repartido em dois subconjuntos: calibração e predição (ou validação externa). Dentre as metodologias disponíveis para essa separação, o algoritmo *Onion*, que utiliza a distância Euclidiana para a seleção das amostras, foi empregado no presente trabalho. Como o nome sugere, o algoritmo seleciona amostras em camadas, semelhante à estrutura de uma cebola. A primeira camada, mais externa, é alocada para o conjunto de calibração e o número de amostras nesta camada é ditado pela fração de loop, conforme Equação 4:

$$\text{N}^\circ \text{ de amostras na camada} = (\text{fração de loop}) \times (\text{fração C}) \times (\text{N}^\circ \text{ Amostras}) \quad (4)$$

Em que a “fração C” equivale à fração de amostras no conjunto de calibração, e “N° amostras” à quantidade total de amostras do conjunto de dados. A segunda camada tem o mesmo número de amostras, e é alocada no conjunto de predição. Esse ciclo se repete até todas as amostras terem sido alocadas. Uma pequena fração de loop indica que as amostras do conjunto de predição estão mais próximas do exterior do espaço dos dados (GALLAGHER; SULLIVAN, 2020). Na Figura 5 são mostrados diferentes frações de loop na seleção de amostras pelo algoritmo *Onion*.

Figura 5 - Amostras selecionadas utilizando o algoritmo *Onion* com diferentes valores de fração de loop.



Valores de fração de loop iguais a 0,1 (a) e 0,5 (b).

Fonte: Adaptado de Gallagher e Sullivan (2020)

Um valor menor de fração de loop, como mostrado na Figura 5a, acarreta em um menor número de amostras por camada, o que sugere uma variabilidade dos dados semelhante para os conjuntos de calibração (em azul) e predição (em vermelho), pois as camadas estão mais próximas. Enquanto um maior valor de fração de loop, deixa o conjunto de predição mais interno ao espaço dos dados, sendo possível observar as “camadas da cebola”, que ficam em evidência devido à maior quantidade de amostras por camada (GALLAGHER; SULLIVAN, 2020).

3.3.2.2 Regressão por mínimos quadrados parciais

A regressão por mínimos quadrados parciais (PLS) é um dos métodos mais utilizados na construção de modelos de calibração multivariada, e tem como principal objetivo estabelecer um modelo linear entre os dados coletados de um sistema (neste trabalho representados pelos espectros NIR), organizados em uma matriz $\mathbf{X}$, e um parâmetro de interesse desse sistema, que pode ser de natureza química ou física, representado por um vetor $\mathbf{y}$. A relação entre os dois conjuntos é promovida por meio de fatores, denominados variáveis latentes (VL), (FERREIRA, 2015). Essas, por sua vez, são calculadas na primeira etapa de desenvolvimento do modelo PLS, a calibração, de modo a maximizar a covariância entre a matriz $\mathbf{X}$ e o vetor $\mathbf{y}$. Enquanto na segunda etapa (predição), esse modelo é usado para a predição da propriedade de interesse em novas amostras, as quais não foram empregadas na primeira etapa.

Segundo Olivieri (2018), a equação geral para um modelo linear entre a matriz $\mathbf{X}$ e o vetor com a propriedade de interesse do analito ($\mathbf{y}_n$), para n amostras utilizadas na etapa de calibração, é dada pela Equação 5, em que $\mathbf{e}$ é o vetor de erros e $\mathbf{b}_n$, o vetor dos coeficientes de regressão.

$$\mathbf{y}_n = \mathbf{X}\mathbf{b}_n + \mathbf{e} \quad (5)$$

É bastante comum as variáveis em $\mathbf{X}$ apresentarem correlação, daí o papel das variáveis latentes em representar a matriz de dados originais e reter a principal fonte da variabilidade em $\mathbf{X}$ e $\mathbf{y}_n$. Com isso, $\mathbf{X}$ é substituída por $\mathbf{T}_A$, que corresponde à matriz de escores, considerada uma versão comprimida de $\mathbf{X}$, em que A representa o número de variáveis latentes utilizado na modelagem. A Equação 5 pode ser reescrita de acordo com a Equação 6:

$$\mathbf{y}_n = \mathbf{T}_A\mathbf{v}_n + \mathbf{e} \quad (6)$$

Sendo $\mathbf{v}_n$ o vetor dos coeficientes de regressão definidos no espaço das A variáveis latentes, em que o mesmo é calculado através da Equação 7, completando a fase de calibração.

$$\mathbf{v}_n = \mathbf{T}_A^+ \mathbf{y}_n \quad (7)$$

Em que $\mathbf{T}_A^+$ é a inversa generalizada de $\mathbf{T}_A$. Antes de seguir para a etapa de predição, vale compreender o modo em que a matriz $\mathbf{T}_A$ é calculada, conforme a Equação 8:

$$\mathbf{T}_A = \mathbf{X}^T \mathbf{W}_A (\mathbf{P}_A^T \mathbf{W}_A)^{-1} \quad (8)$$

Sendo $\mathbf{W}_A$ a matriz de pesos (*weight loadings*) e $\mathbf{P}_A$ é a matriz de *loadings*, ambas definidas para as A variáveis latentes. É justamente o emprego dessas duas matrizes no cálculo de $\mathbf{T}_A$ que faz com que as variáveis latentes sejam estimadas não somente com base na matriz de dados originais, mas também com base no vetor da propriedade de interesse do analito.

Na etapa de predição, o vetor dos coeficientes de regressão do modelo desenvolvido na etapa de calibração ($\mathbf{v}_n$) é empregado para estimar o parâmetro de interesse das amostras do conjunto de predição. Antes disso, no entanto, uma etapa preliminar é requerida, a qual envolve

o cálculo dos escores ($\mathbf{t}_A$) para as amostras de predição, a partir da resposta instrumental obtida para essas amostras ($\mathbf{x}$), de acordo com a Equação 9:

$$\mathbf{t}_A = (\mathbf{W}_A^T \mathbf{P}_A)^{-1} \mathbf{W}_A^T \mathbf{x} \quad (9)$$

Por fim, calculado o vetor $\mathbf{t}_A$, a predição do parâmetro de interesse ($\hat{y}$) da amostra pode ser estimada pela Equação 10:

$$\hat{y} = \mathbf{v}_n^T \mathbf{t}_A \quad (10)$$

3.3.2.3 Regressão por máquinas de vetores de suporte

Como dito anteriormente, a regressão por PLS assume uma relação aproximadamente linear entre o parâmetro de interesse e os dados espectrais, por exemplo, e em geral, os modelos gerados são de fácil construção e interpretação. Entretanto, uma relação linear nem sempre é obtida, e, apesar de o PLS conseguir modelar dados que apresentem certo grau de não-linearidade, se a presença de efeitos não-lineares for significativa, pode gerar modelos com baixo desempenho (OLIVIERI, 2018; TANGE *et al.*, 2017).

Os efeitos não-lineares estão comumente associados com: elevada concentração do analito, que ocasiona desvios da Lei de Beer; variação na composição da matriz amostral (devido às interações químicas, principalmente interações de ligações de hidrogênio, do analito com outros componentes na amostra, que afetam a forma como a radiação interage com ela); e outros fatores intrínsecos das amostras, como temperatura, pH, viscosidade e tamanho de partícula. Esses efeitos podem gerar mudanças na posição das bandas de absorção dos espectros, alargamento de bandas e espalhamento da radiação (BERTRAN *et al.*, 1999; MILLER, 1993).

Uma alternativa para lidar com esses efeitos consiste na aplicação de modelagens que lidam com dados não-lineares, em que a regressão por máquinas de vetores de suporte (SVM) tem destaque (BALABIN; LOMAKINA, 2011; PIERNA *et al.*, 2011; TANGE *et al.*, 2017). Essa técnica tem como base transformar o problema não linear em um problema linear, de mais fácil resolução, mapeando os dados originais em um espaço de dimensão superior. O mapeamento pode ser feito usando a função kernel *Radial Basis Function* (Função de Base Radial – RBF), que é mais empregada em dados não-lineares e, por requerer apenas um

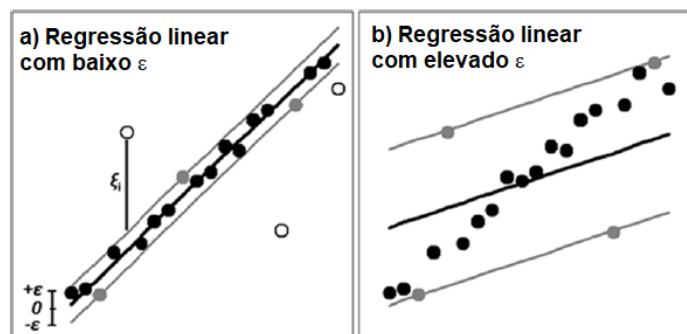
parâmetro a ser otimizado, implica em menor esforço computacional (BRERETON; LLOYD, 2010).

Para facilitar o entendimento sobre o método, pode-se partir da regressão linear por SVM, que, segundo Schölkopf e Smola (2002), é expressa pela Equação 11:

$$f(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + b + e \quad (11)$$

Em que $\mathbf{x}$ é um vetor de dados coletados de um sistema (como o espectro de uma única amostra); $\mathbf{w}$ é o vetor de regressão; b é o *offset* e e , o erro de predição. A ideia básica da regressão por SVM é minimizar o erro de predição encontrando uma função $f(\mathbf{x})$ que apresente um desvio de no máximo ε (*epsilon*) em relação ao valor da propriedade de interesse obtido para cada observação (SCHÖLKOPF; SMOLA, 2002). Como pode ser visualizado na Figura 6a, ε define a margem de um “tubo” (representada pelas linhas cinzas) ao redor do vetor de regressão $\mathbf{w}$ (linha preta). É possível perceber que essa ideia é satisfeita para as amostras representadas pelos círculos pretos e cinzas, visto que estão dentro ou sobre a margem delimitada por ε , respectivamente.

Figura 6 – Ilustração do princípio da regressão por SVM.



Círculos de cor cinza representam os vetores de suporte (VS) na margem ε , círculos brancos representam os VS fora da margem ε e os círculos pretos não são vetores de suporte.

Fonte: Adaptado de Tange *et al.* (2015)

Por vezes, a inclusão de todas as amostras dentro da margem, através do aumento de ε , não é benéfica, visto que pode aumentar o erro de predição para a maioria dessas, como observado na Figura 6b. As observações fora do tubo, representadas pelos círculos brancos na Figura 6a, contribuem para a chamada função de perda ε -insensível, em que somente amostras com erros de predição maiores que ε contribuem para a perda. Essas contribuições são

chamadas de variáveis de folga (ξ), definidas como a magnitude do erro de predição maior que ε (TANGE et al., 2015).

Outra característica fundamental dos modelos SVM é minimizar a norma do vetor $\mathbf{w}$, buscando o menor coeficiente de regressão para uma dada margem ε (BRERETON; LLOYD, 2010). Agora é possível compreender que as amostras cinzas e brancas, que estão sobre e fora da margem do tubo definida por ε , respectivamente, sustentam o modelo, pois elas otimizam o coeficiente de regressão, sendo denominadas vetores de suporte (VS). Enquanto as amostras em preto não contribuem para a definição de ε , tampouco para a função de perda, por isso, após a estimativa do modelo, podem ser removidas sem consequências ao modelo (BRERETON; LLOYD, 2010; TANGE et al., 2015).

O objetivo da regressão por SVM pode ser generalizado para múltiplas dimensões e é matematicamente definido pela Equação 12:

$$\text{minimizar } \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (12)$$

Em que $\mathbf{w}$ é o vetor de regressão da Equação (11), ξ_i e ξ_i^* são as variáveis de folga positivas e negativas para uma amostra $\mathbf{x}_i$, e n o número de observações. O novo parâmetro C (custo) é o erro de penalidade e determina o quão importante é garantir que todas as amostras estejam dentro das margens definidas por ε . Assim como ε , este é um parâmetro definido pelo usuário, e, em dados com presença de *outliers*, deve ser analisado com cautela, pois valores altos para C tendem a incluir possíveis *outliers* no modelo SVM, sobreajustando os dados. Por outro lado, um valor de C muito baixo pode gerar falta de ajuste no modelo (BRERETON; LLOYD, 2010).

Quando o objetivo na Equação 12 é resolvido no espaço de dimensão superior, a solução final da modelagem por SVM é expressa pela Equação 13:

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i + \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (13)$$

A solução é expressa em termos dos pesos (α_i e α_i^*) dos vetores de suporte $\mathbf{x}_i$ e da função Kernel K . Como mencionado anteriormente, neste trabalho, foi empregada a função Kernel de base radial (RBF) mostrada na Equação 14:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (14)$$

Em que $\frac{1}{2\sigma^2} = \gamma$ (*gamma*), sendo este o parâmetro que determina a flexibilidade da solução. Em geral, altos valores de γ induzem à elevada flexibilidade e a uma solução mais complexa, enquanto o contrário leva a uma solução mais suave (TANGE *et al.*, 2015).

Por fim, os parâmetros da regressão por SVM, ε e C , e o parâmetro para a função de kernel, γ , devem ser definidos. Como um parâmetro influencia o outro, devem ser otimizados simultaneamente. A otimização pode ser realizada através de pesquisa num dado intervalo de valores para os 3 parâmetros, avaliando o desempenho da validação cruzada (BRERETON; LLOYD, 2010; TANGE *et al.*, 2015).

3.4 MONITORAMENTO DE PROCESSOS

A ideia de monitoramento de processo está intimamente conectada às Tecnologias Analíticas de Processo (PAT), que consistem, dentre outros aspectos, na utilização de diferentes técnicas analíticas, geralmente associadas à aquisição e análise de dados multivariados, a fim de obter uma maior compreensão, monitoramento contínuo e melhoria do processo (CHANDA *et al.*, 2015; FDA, 2004). Além disso, empregar técnicas analíticas que oferecem alto desempenho como: precisão, exatidão, rapidez e robustez são mais atrativas para PAT. A espectroscopia no infravermelho próximo associada à quimiometria, por exemplo, possuem tais características ideais para fins de monitoramento e controle de processos (GALAVARNA *et al.*, 2018; PASQUINI, 2003).

3.4.1 Espectroscopia no infravermelho próximo aplicada ao monitoramento de processos no setor sucroalcooleiro

A crescente demanda por produtos de maior qualidade implica em consumidores cada vez mais exigentes e, conseqüentemente, na busca, pela indústria, de métodos cada vez mais aprimorados para garantir a qualidade de seus produtos e aumentar a sua competitividade em diferentes ramos do mercado. Nesse sentido, as técnicas analíticas como os métodos espectroscópicos são cada vez mais procuradas para aplicações industriais, como vem ocorrendo nos setores agrícolas (CORTÉS *et al.*, 2019). No setor sucroalcooleiro, a

espectroscopia NIR, por exemplo, tem destaque tanto no campo, quanto na indústria para monitoramento e controle de processos.

Como mencionado anteriormente, a cana-de-açúcar é analisada no laboratório PCTS para estimativa do ATR a partir das medidas de brix e pol realizadas no caldo extraído da cana. As metodologias convencionais empregadas nessas análises são laboriosas e demandam o emprego de produtos químicos tóxicos. A metodologia para determinação desses parâmetros por espectroscopia NIR possui homologação junto à ABNT (NBR 16770 e 16771 de 2019) e CONSECANA, cabendo aos gestores das usinas brasileiras, determinar qual metodologia utilizar. No Japão, por exemplo, a espectroscopia NIR é o método oficial para determinar os preços da cana em todas as usinas de açúcar bruto (TAIRA; UENO; FURUKAWA; *et al.*, 2013). Neste segmento, vários trabalhos têm incentivado o setor, mostrando as vantagens no emprego desta tecnologia. Através de análises rápidas, não-destrutivas e com um mínimo preparo de amostras, os espectros NIR de caldo extraído de cana foram adquiridos em espectrômetros de bancada e, a partir deles, os modelos de calibração multivariada foram construídos para quantificar brix, pol e açúcares redutores (AR) (CHEN *et al.*, 2014; SOROL *et al.*, 2010; VALDERRAMA; BRAGA; POPPI, 2007).

Para fins de monitoramento do processo produtivo, poucos trabalhos foram desenvolvidos no setor. Como exemplo, Simpson e Naidoo (2010) relataram o emprego de um espectrômetro NIR de bancada com faixa espectral de 800 a 2500 nm na construção de modelos PLS para quantificação de brix, pol e sacarose em amostras do processo de fabricação de açúcar: caldo clarificado; xarope; méis A e B; e massas cozidas B e C. Foram desenvolvidos modelos locais para cada amostra a partir de medidas de transmitância das amostras diluídas (exceto caldo clarificado) e em geral, resultados satisfatórios foram obtidos.

Numa abordagem um pouco diferente, Tange *et al.* (2015) avaliaram modelos locais e globais para a quantificação de brix e pol de 4 tipos de amostras do processo de fabricação de açúcar: caldo primário; xarope; massa cozida; e melaço. Foram construídos e comparados modelos de regressão PLS e SVM a partir de medidas de transfectância em um espectrômetro Vis-NIR com faixa espectral de 400 – 2498 nm e equipado com um refletor de ouro. Os espectros foram adquiridos logo após a coleta das amostras na usina, sem diluição prévia. De modo geral, os modelos SVM se sobressaíram aos PLS, o que foi atribuído aos efeitos não-lineares presentes nas amostras (variações de temperatura, pH e viscosidade, assim como elevadas concentrações em alguns tipos de amostras).

Por fim, se tratando do emprego de aparelhos portáteis no setor sucroalcooleiro, as aplicações são voltadas para o campo. O emprego de espectrômetros Vis-NIR portáteis foi

avaliado na construção de modelos de calibração multivariada para quantificação de parâmetros de qualidade como brix, pol e teor de fibras. A aquisição espectral foi realizada diretamente no colmo (caule) de cana-de-açúcar (ou em sua seção transversal, conforme (NAWI; CHEN; JENSEN, 2013)), visando avaliar a maturidade da cana e determinar o início da colheita (CORRÊDO et al., 2020; MARAPHUM et al., 2018; NAWI et al., 2013; PHUPHAPHUD et al., 2019; TAIRA; UENO; SAENGPRACHATANARUG; et al., 2013).

4 METODOLOGIA

4.1 COLETA DE AMOSTRAS

Os dados utilizados no presente trabalho foram adquiridos em uma usina de açúcar e álcool situada no estado de Pernambuco – Brasil. Amostras foram coletadas em diferentes etapas ao longo do processo de fabricação de açúcar em duas safras distintas, 2019/2020 e 2020/2021. A amostragem realizada seguiu o mesmo procedimento de rotina adotado pelo laboratório industrial da usina, onde em espaços de tempo já estabelecidos, os coletores levam ao laboratório amostras de todo o processo, para determinação de parâmetros de qualidade. Para o desenvolvimento deste trabalho, parte do que foi coletado também foi analisado por espectroscopia NIR. Na Tabela 1, é apresentado um resumo das etapas de coleta, tipos e quantidades das diferentes amostras coletadas e analisadas.

Tabela 1 – Informações acerca das amostras coletadas na usina e analisadas por espectroscopia NIR.

Etapas do Processo	Tipo de Amostra	Nº de Amostras	Amostras Coletadas
Recepção da matéria-prima		329	Caldo extraído
Extração	Caldo	133	Caldo primário, residual e misto
Tratamento do caldo		293	Caldo filtrado, decantado e para evaporação
Evaporação	Caldo concentrado	138	1º ao 5º Efeito de evaporação
Tanque de xarope	Xarope	50	Xarope
Cozimento e Cristalização	Massa Cozida	212	Massas A, B, C e refinada
Centrifugação	Mel	169	Méis A, B, C e refinado
	Magma	57	Magmas de 2ª e 3ª
Refinamento de Açúcar VHP	Licor	126	Licores dissolvido, misto, filtrado e flotado
		1507	TOTAL

Fonte: O autor (2021)

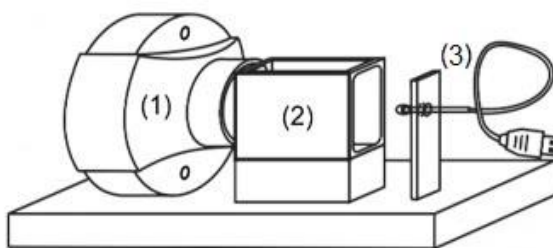
4.2 AQUISIÇÃO ESPECTRAL

Os espectros NIR foram adquiridos utilizando um espectrômetro portátil (MicroNIR™ 1700, Viavi Solutions, EUA), que opera na faixa de comprimento de onda de 950 a 1650 nm e pode fazer medidas em reflectância difusa ou transmitância. Este instrumento tem um filtro linear variável (LVF) e um arranjo de detector InGaAs de 128 pixel, que permite a determinação dos valores de absorbância de forma simultânea para todos os comprimentos de onda. Sua resolução óptica é menor do que 1,25% (geralmente, 1%) do comprimento de onda central (por exemplo, para 1000 nm, a resolução seria menor do que 12,5 nm). Para análise das amostras, cada espectro foi obtido como uma média de 50 varreduras com tempo de integração de 10 ms, registrando dados a cada 6,2 nm aproximadamente.

Considerando a variabilidade do tipo de amostra demonstrada na Tabela 1, os dois modos de leitura do espectrômetro portátil foram utilizados. Amostra líquidas (como os caldos) tiveram seus espectros registrados por transmitância, enquanto as amostras mais densas (caldos concentrados, xaropes, méis, licores, massas e magmas), geradas através das etapas de concentração do açúcar, tiveram registro espectral empregando reflectância difusa. Alguns dos caldos concentrados (1º e 2º efeito de evaporação), aqui tratados como amostras intermediárias, foram testados em ambas as configurações, para determinar qual a melhor metodologia de aquisição espectral. Em todos os casos, a amostra foi analisada diretamente, sem necessidade de prévia diluição.

Para as medidas no modo transmitância, utilizou-se um acessório semelhante ao descrito por Paiva *et al.*, (2015), mostrado na Figura 7, que permite usar diferentes caminhos ópticos que variam com o tamanho das cubetas de quartzo selecionadas. Neste estudo, foram testadas as cubetas com 2, 5 e 10 mm de caminho óptico. Para essas medidas, o 100% de transmitância foi adquirido usando uma cubeta limpa e vazia e o 0% de transmitância foi obtido bloqueando-se a passagem da luz entre a fonte externa do acessório (mini lâmpada de tungstênio) e o espectrômetro.

Figura 7 – Espectrômetro portátil acoplado ao acessório de transmitância.



(1) MicroNIR; (2) Compartimento de cubeta; (3) Mini lâmpada de tungstênio.

Fonte: Adaptado de Paiva *et al.*, (2015)

Para as medidas em reflectância difusa, uma pequena placa de Petri de vidro contendo aproximadamente 2 mL de cada amostra foi empregada na aquisição espectral. Para tais medidas, o espectro de 100% de reflectância foi obtido usando um tarugo de Teflon, nome comercial do Politetrafluoretileno.

4.3 ANÁLISES DE REFERÊNCIA

As amostras também foram analisadas através dos métodos de referência descritos por Caldas (2012) para quantificar os teores de sólidos solúveis totais (brix) e sacarose aparente (pol), conforme apresentado a seguir. Do total de amostras coletadas (1507), todas foram analisadas para determinação do brix, enquanto 769 foram submetidas ao método para quantificação de pol.

4.3.1 Determinação do teor de sólidos solúveis totais

Alguns dos substratos do processo de fabricação de açúcar necessitam de diluição para facilitar a leitura da amostra, pois alguns são sólidos e outros têm elevada viscosidade. Na Tabela 2, são apresentados os fatores de diluição para os substratos do processo.

Tabela 2 – Fatores de diluição para as amostras do processo de fabricação de açúcar.

Amostra do Processo	Fator de Diluição	%(m/m)
Caldos	-	-
Caldo concentrado	-	-
Xarope	1:5	100/500
Massas A, B e C	1:5	200/1000
Mel A	1:6	100/600
Méis B e C	1:8	100/800
Magmas	1:5	100/500
Licores	-	-
Mel Refinado	-	-
Massa Refinada	1:2	100/200

Fonte: O autor (2021)

A metodologia de análise seguiu, em resumo, as seguintes instruções:

- Homogeneizar a amostra original;
- Diluir em água destilada conforme Tabela 2;
- Homogeneizar em agitador mecânico;
- Colocar algumas gotas da solução preparada no refratômetro e fazer a leitura;
- Calcular o Brix da amostra multiplicando a leitura pelo fator de diluição.

A leitura do brix foi realizada em refratômetro digital automático (RDA 9000, Acatec, Brasil).

4.3.2 Determinação do teor de sacarose aparente

4.3.2.1 Pol do caldo extraído

O procedimento experimental para o caldo extraído da etapa de recepção da matéria-prima é resumido por:

- Separar cerca de 200 mL do caldo extraído por pressão hidráulica;
- Adicionar acetato básico de chumbo em pó, em quantidade suficiente para clarificação da solução;
- Homogeneizar em agitador mecânico e filtrar em papel de filtro qualitativo;
- Realizar a leitura em sacarímetro;
- Calcular a pol através da equação (15) (CONSECANA, 2015).

$$\text{Pol}(\%) = \text{LPb}(0,2605 - 0,0009882 \times \text{Brix}) \quad (15)$$

Na equação 15, LPb é a leitura sacarimétrica equivalente à de acetato básico de chumbo como substância clarificante.

4.3.2.2 Pol dos licores, mel e massa refinada

Por se tratarem de amostras com elevada pureza, dispensam a utilização de clarificante químico e seguem procedimento similar à obtenção da solução normal de sacarose descrita no item 3.1.1.2:

- Pesar 26g da amostra diretamente em balão volumétrico de 100 mL (para licores e mel refinado), ou em béquer (para a massa refinada), seguida de diluição em água destilada e transferência para o balão volumétrico;
- Completar o volume do balão com água destilada e homogeneizar a solução;
- Filtrar com papel de filtro qualitativo;
- Realizar a leitura em sacarímetro, onde a leitura observada é equivalente ao teor de pol.

4.3.2.3 Pol das demais amostras

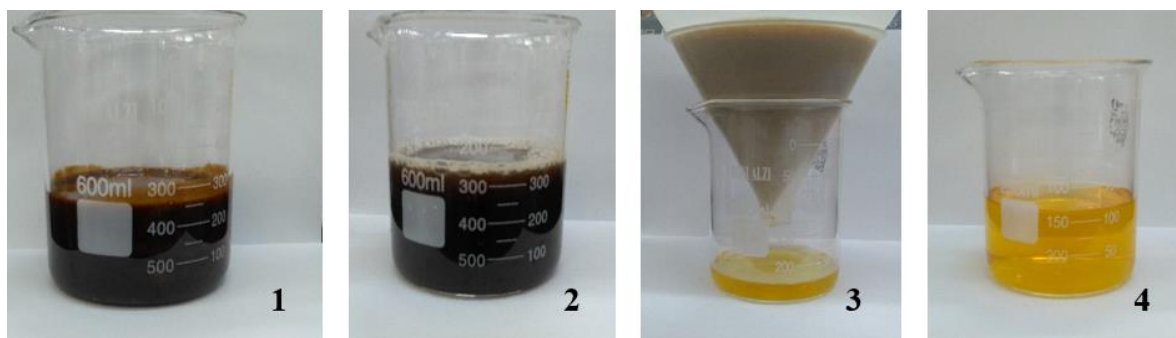
O procedimento experimental para as amostras de caldos (exceto caldo extraído), xaropes, méis A, B e C, massas A, B e C e magmas é resumido por:

- Separar cerca de 200 mL da amostra diluída usada para determinação do brix;
- Adicionar acetato básico de chumbo em pó, em quantidade suficiente para clarificação da solução (a qual depende do tipo de amostra analisada);
- Homogeneizar em agitador mecânico e filtrar em papel de filtro qualitativo; na Figura 8 é mostrada a sequência do procedimento para o mel B;
- Realizar a leitura em sacarímetro;
- Calcular a pol através da equação (16) (CALDAS, 2012).

$$\text{Pol(\%)} = L_s \times F_{\text{pol}} \times F_d \quad (16)$$

Na equação 2, L_s é a leitura sacarimétrica corrigida a 20 °C; F_{pol} é o fator de polarização descrito em (CALDAS, 2012); e F_d , o fator de diluição (quando houver).

Figura 8 – Procedimento experimental para a determinação da sacarose aparente do Mel B.



(1) Amostra de mel B; (2) Amostra diluída; (3) Filtração após adição do clarificante; (4) Amostra clarificada.

Fonte: O autor (2021)

A leitura da pol foi realizada em sacarímetro automático (Autopol 880, Rudolph Research Analytical, EUA).

4.4 ANÁLISE DE DADOS E DESENVOLVIMENTO DE MODELOS DE CALIBRAÇÃO

4.4.1 Desenvolvimento de modelos para definição do modo de aquisição espectral para amostras intermediárias do processo

Com o intuito de determinar qual dos dois modos de aquisição espectral seria mais adequado para a análise das amostras intermediárias do processo (caldos concentrados do 1º e do 2º efeito da evaporação), modelos PLS para a quantificação de brix foram construídos com a ausência e a presença de 26 amostras intermediárias, e comparados entre si. Para os modelos de pol, não foi necessário este estudo, pois as amostras intermediárias consideradas são apenas monitoradas através do brix na rotina de trabalho da usina.

Os quatro modelos desenvolvidos nesta etapa foram gerados fixando-se a 1ª derivada com filtro Savitzky-Golay (com janela de 5 pontos e polinômio de 2ª ordem) como pré-processamento. Para os modelos desenvolvidos com as amostras intermediárias, a inclusão dessas amostras aos conjuntos de dados obtidos por reflectância e por transmitância foi realizada antes da separação dos conjuntos de calibração e predição. A obtenção desses conjuntos foi feita através do algoritmo *Onion*, separando-se as amostras na proporção de 70% para a calibração e 30% para a predição, aproximadamente. Esse método, descrito em Olsson; Gottfries; Wold, (2004), está implementado no PLS Toolbox (Eigenvector Research Inc., EUA), software utilizado no desenvolvimento deste trabalho.

4.4.2 Desenvolvimento dos modelos de calibração

Após definido o modo de aquisição espectral mais adequado para as amostras intermediárias e, conseqüentemente, estabelecidas as faixas de concentração dos modelos, partiu-se para o desenvolvimento e otimização dos modelos de regressão. Para isso, além de 1ª derivada com filtro Savitzky-Golay, outros pré-processamentos, como SNV e MSC, também foram avaliados em modelos PLS, para determinar o mais adequado aos dados. Os dados pré-processados e as medidas de referência foram ainda centrados na média antes da construção dos modelos. Além disso, para elaboração e validação dos modelos, as amostras coletadas durante as duas safras foram divididas em conjuntos de calibração e predição, na proporção de 70 e 30%, respectivamente. Para este propósito, foi empregado o método *Onion*.

Definido o melhor pré-processamento com modelos PLS, outra abordagem foi avaliada no desenvolvimento dos modelos de calibração: a regressão por SVM. Apesar dos modelos PLS serem mais difundidos e abordados nas indústrias, não são, em geral, adequados para dados com presença de não-linearidades (desvios da Lei de Beer). Nesses casos, a aplicação da regressão por SVM pode ser mais apropriada (SEXTON *et al.*, 2018; TANGE *et al.*, 2015).

Os modelos SVM foram construídos através da função de Kernel, RBF (*radial basis function*). Para sua otimização, três parâmetros (custo - C, epsilon - ϵ e gamma - γ) foram avaliados a fim de selecionar o modelo com melhor desempenho em termos da raiz do erro quadrático médio da validação cruzada (RMSECV). A detecção e remoção de *outliers* foi realizada utilizando-se os gráficos dos resíduos (Q residual x T² Hotelling) gerados nos modelos PLS. Portanto, outliers detectados nesta abordagem de modelagem também foram removidos dos modelos SVM. A validação cruzada *venetian blinds* foi aplicada para ambas as abordagens de modelagem e o desempenho dos modelos de calibração foi avaliado usando os conjuntos de predição, com base nos valores da raiz do erro quadrático médio da predição (RMSEP) e de *bias*, que avalia a presença de erro sistemático no modelo.

Dessa forma, variando entre modo de aquisição espectral (transmitância e reflectância difusa), modelagem abordada (regressão por PLS ou SVM) e parâmetro de qualidade (Brix e Pol), ao final das avaliações, um total de 8 modelos foram desenvolvidos, como mostrado na Tabela 3. O número de amostras apresentado para os conjuntos de calibração e de predição não inclui os *outliers* eliminados durante o desenvolvimento dos modelos.

Tabela 3 – Modelos desenvolvidos para monitoramento de parâmetros de qualidade.

Método de Análise	Parâmetro de Qualidade	Amostras de calibração	Amostras de predição	Modelagem
Transmitância	Brix	547	234	PLS SVM
	Pol	324	138	PLS SVM
Reflectância Difusa	Brix	508	218	PLS SVM
	Pol	215	92	PLS SVM

Fonte: O autor (2021)

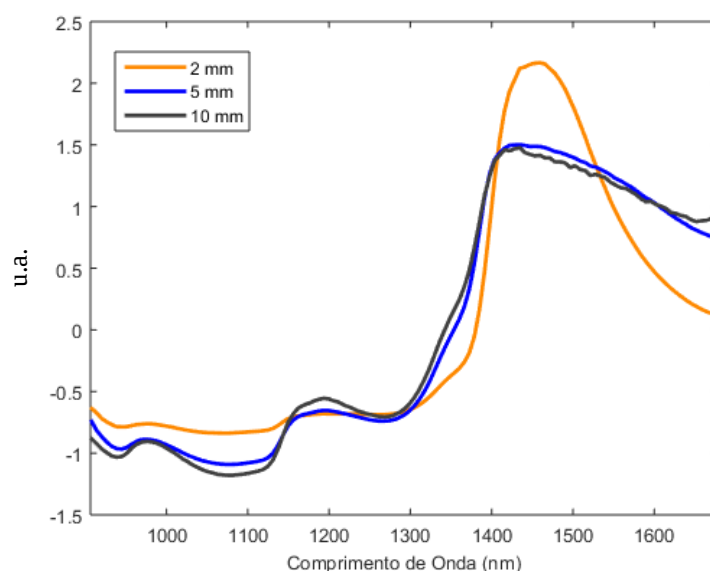
A análise e tratamento dos dados foi realizada no software MATLAB® (R2010a 7.10.0.499, MathWorks, EUA). Os modelos foram desenvolvidos e avaliados empregando o PLS Toolbox (Eigenvector Research Inc., EUA), que funciona em ambiente Matlab.

5 RESULTADOS E DISCUSSÃO

5.1 SELEÇÃO DO CAMINHO ÓPTICO PARA MEDIDAS DE TRANSMITÂNCIA

Inicialmente, algumas amostras do caldo extraído da cana-de-açúcar foram analisadas em diferentes caminhos ópticos para determinar a metodologia ideal de aquisição espectral. Para isso, utilizou-se o acessório de transmitância com cubetas de quartzo de 2, 5 e 10 mm de caminho óptico. Na Figura 9, são apresentados os espectros pré-processados com SNV de uma amostra de caldo, adquiridos utilizando as três cubetas.

Figura 9 - Espectros NIR pré-processados com SNV adquiridos utilizando cubetas de quartzo de diferentes caminhos ópticos.



Fonte: O autor (2021)

Nota-se na Figura 9 que o espectro obtido com a cubeta de 10 mm (em preto) apresentou sinais de absorção saturados a partir da região de 1400 nm, aproximadamente. A absorção nesta faixa é devida, principalmente, ao primeiro sobretom do grupamento O-H (WORKMAN; WEYER, 2007). Assim, isso pode ter ocorrido devido ao excesso de água presente no caldo. Comportamento semelhante foi observado no espectro obtido com a cubeta de 5 mm (em azul). Enquanto o espectro registrado com a cubeta de 2 mm (em laranja), apresentou a banda de absorção entre 1400 e 1600 nm com melhor definição. Com base nessas informações, o caminho óptico de 2 mm foi selecionado para aquisição espectral de caldos e caldos concentrados do processo no modo transmitância.

5.2 DEFINIÇÃO DA METODOLOGIA DE AQUISIÇÃO ESPECTRAL PARA AMOSTRAS INTERMEDIÁRIAS DO PROCESSO

Com a variabilidade do tipo de amostra entre líquidos e sólidos no processo de fabricação do açúcar, ficou claro que as duas abordagens de aquisição espectral pelo MicroNIR seriam necessárias, por transmitância e reflectância difusa. Dessa forma, uma modelagem global que atendesse a todo o processo produtivo não poderia ser considerada nas condições em que este trabalho foi realizado. No entanto, com o intuito de simplificar a aplicação dos modelos no monitoramento do processo, optou-se por incluir em um mesmo modelo, amostras de diferentes etapas, cujos espectros tinham sido adquiridos da mesma forma. Assim, foi preciso definir em um estudo preliminar, a faixa de concentração de cada um dos modelos, avaliando-se qual a melhor abordagem a ser utilizada na análise de amostras intermediárias do processo, em relação ao brix.

O processo produtivo evolui de caldo a açúcar e uma das principais operações unitárias envolvida na obtenção do produto final é a evaporação da água presente, que vai tornando a amostra mais concentrada e viscosa. Consequentemente, a evaporação do caldo foi a etapa que determinou a faixa de concentração para cada metodologia empregada. Considerando que os caldos concentrados do 3º efeito de evaporação, com valores variando entre 45 e 64% (m/m), já possuem viscosidade suficiente para dificultar a aquisição espectral em cubeta de quartzo no modo transmitância, foi estipulado que, desse ponto do processo em diante, as medidas seriam feitas no modo reflectância. Enquanto as amostras do 1º e 2º efeitos de evaporação (amostras intermediárias), que apresentaram teores menores do que 45% (m/m), foram avaliadas pelas duas abordagens.

Para essa avaliação, modelos PLS foram construídos conforme descrito no item 2.4.1. Vale ressaltar que a inclusão das 26 amostras intermediárias foi realizada antes da separação dos conjuntos de calibração e predição, se certificando logo após, que a distribuição dessas amostras foi homogênea nas duas modelagens, a fim de garantir uma comparação justa. Os resultados das etapas de calibração e de predição dos modelos desenvolvidos com a presença e a ausência dessas amostras intermediárias estão apresentados na Tabela 4.

Tabela 4 - Resultados para as etapas de calibração e de predição dos modelos PLS preditivos de brix para avaliação de inclusão de amostras intermediárias.

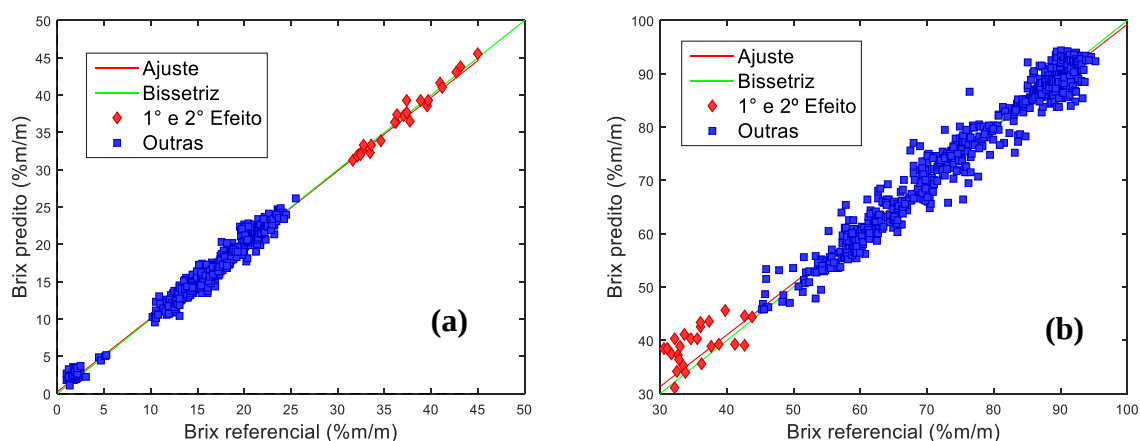
Modo de Aquisição Espectral	1° e 2° Efeitos	Faixa de Concentração (%m/m)	VL	RMSEC (%m/m)	RMSECV (%m/m)	R ² cv	RMSEP (%m/m)	R ² p
Transmitância	Presentes	1,02 - 45	4	0,71	0,72	0,98	0,74	0,99
	Ausentes	1,02 - 25,5	4	0,64	0,65	0,98	0,72	0,98
Reflectância Difusa	Presentes	30,6 - 95,3	4	2,58	2,66	0,97	2,62	0,97
	Ausentes	45,3 - 95,3	4	2,25	2,33	0,97	2,36	0,97

Fonte: O autor (2021)

No que diz respeito aos modelos desenvolvidos utilizando a reflectância difusa como metodologia de aquisição espectral, os valores de RMSEP obtidos com e sem a inclusão das amostras intermediárias apresentaram diferença estatisticamente significativa, com base em um teste F a um nível de 95% de confiança. Enquanto não foi observada diferença estatística nos valores de RMSEP para os modelos obtidos com transmitância, com a inclusão das amostras intermediárias e, consequente, aumento da faixa de concentração desses modelos.

A Figura 10 apresenta os gráficos de valores de referência *versus* preditos para os modelos com presença das amostras de 1° e 2° efeito de evaporação.

Figura 10 - Gráficos dos valores preditos vs. referência com inclusão de amostras intermediárias.



Amostras obtidas no modo transmitância (a) e no modo reflectância difusa (b).

Fonte: O autor (2021)

É possível notar que as amostras intermediárias se ajustaram melhor ao modelo obtido com a técnica de transmitância. Com base nesses resultados, optou-se pela inclusão das amostras de 1° e 2° efeito de evaporação no modelo gerado através de medidas de transmitância.

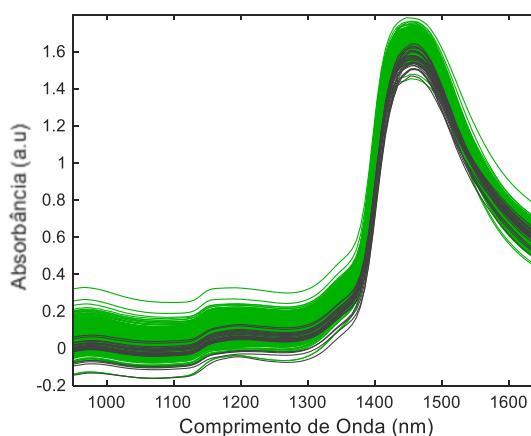
5.3 DESENVOLVIMENTO E OTIMIZAÇÃO DOS MODELOS DE CALIBRAÇÃO

Após a definição da faixa de concentração dos modelos para a quantificação de brix, seguiu-se para a etapa de otimização desses modelos e construção dos modelos para determinação da pol. Nessa etapa, diferentes pré-processamentos foram avaliados na construção dos modelos de calibração. Para guiar na escolha das técnicas de pré-processamento, foi feita uma avaliação dos espectros obtidos por transmitância e por reflectância, as quais estão apresentadas a seguir. Essa avaliação permitiu, inclusive, fazer uma análise comparativa dos espectros das amostras coletadas em diferentes etapas do processo e incluídas no mesmo modelo.

5.3.1 Avaliação dos espectros de transmitância

A Figura 11 apresenta os espectros brutos de 781 amostras adquiridas no modo transmitância.

Figura 11 – Espectros brutos das amostras coletadas no modo transmitância.



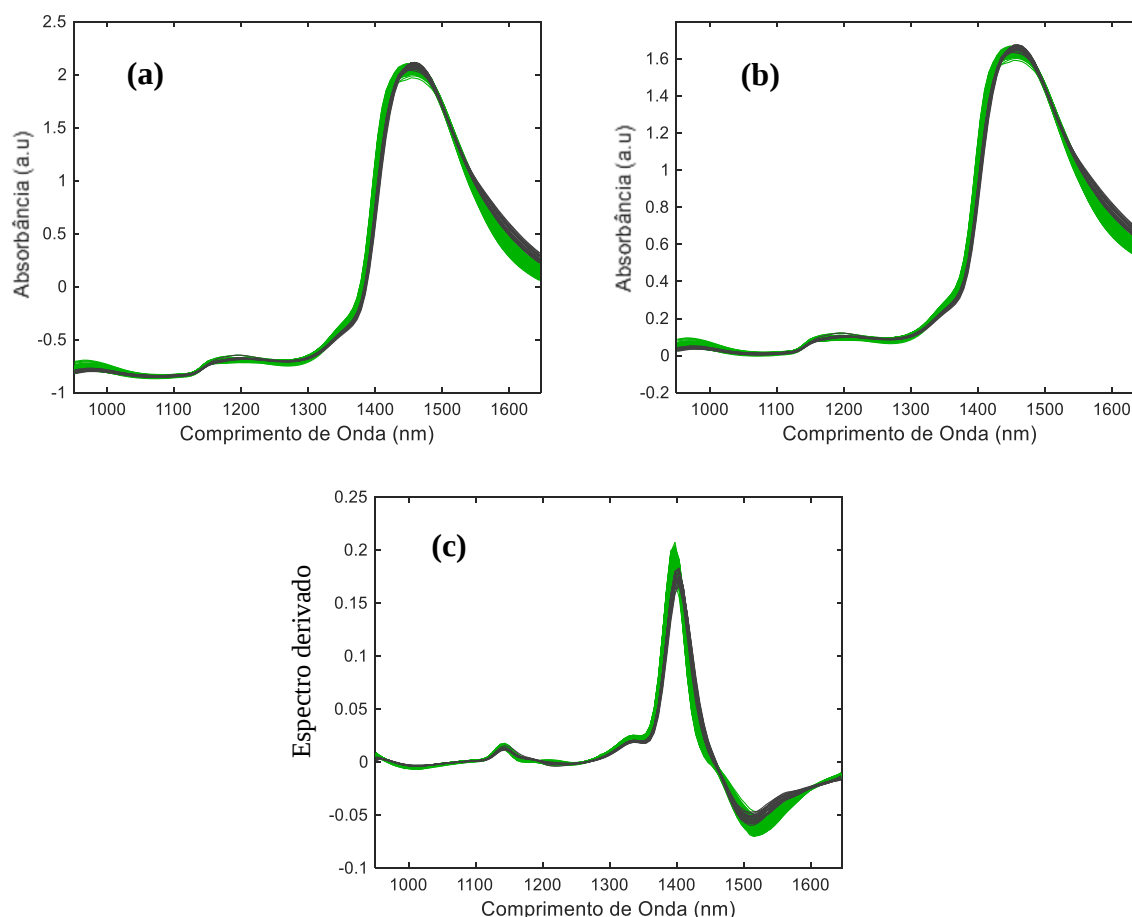
Caldos (em verde) e caldos concentrados – 1º e 2º efeito – (em preto)

Fonte: O autor (2021)

Nos diferentes espectros, nota-se que a principal fonte de variação entre as amostras é advinda de espalhamento da radiação (gerando diferenças na linha de base), causado por fenômenos físicos possivelmente associados à suspensão de sólidos insolúveis inerentes ao caldo de cana-de-açúcar. Visando corrigir esses efeitos indesejáveis e visualizar a fonte de variação entre as amostras proveniente da informação química, alguns pré-processamentos podem ser

empregados. Na Figura 12, são mostrados os espectros de transmitância pré-processados com SNV, MSC e 1ª derivada com polinômio de 2ª ordem e janela de 5 pontos.

Figura 12 – Espectros de transmitância pré-processados.



(a) SNV; (b) MSC e (c) 1ª derivada com polinômio de 2ª ordem e janela de 5 pontos

Caldos (em verde) e caldos concentrados – 1º e 2º efeito – (em preto)

Fonte: O autor (2021)

Nestes exemplos é possível notar uma leve diferença entre os espectros de caldo comparados aos espectros de caldo concentrado, principalmente nos espectros pré-processados com 1ª derivada na região entre 1400 e 1600 nm, onde ocorre absorção do primeiro sobretom do grupamento O-H (WORKMAN; WEYER, 2007), ou seja, diferenciam-se em relação a quantidade de água e açúcares presente nas amostras.

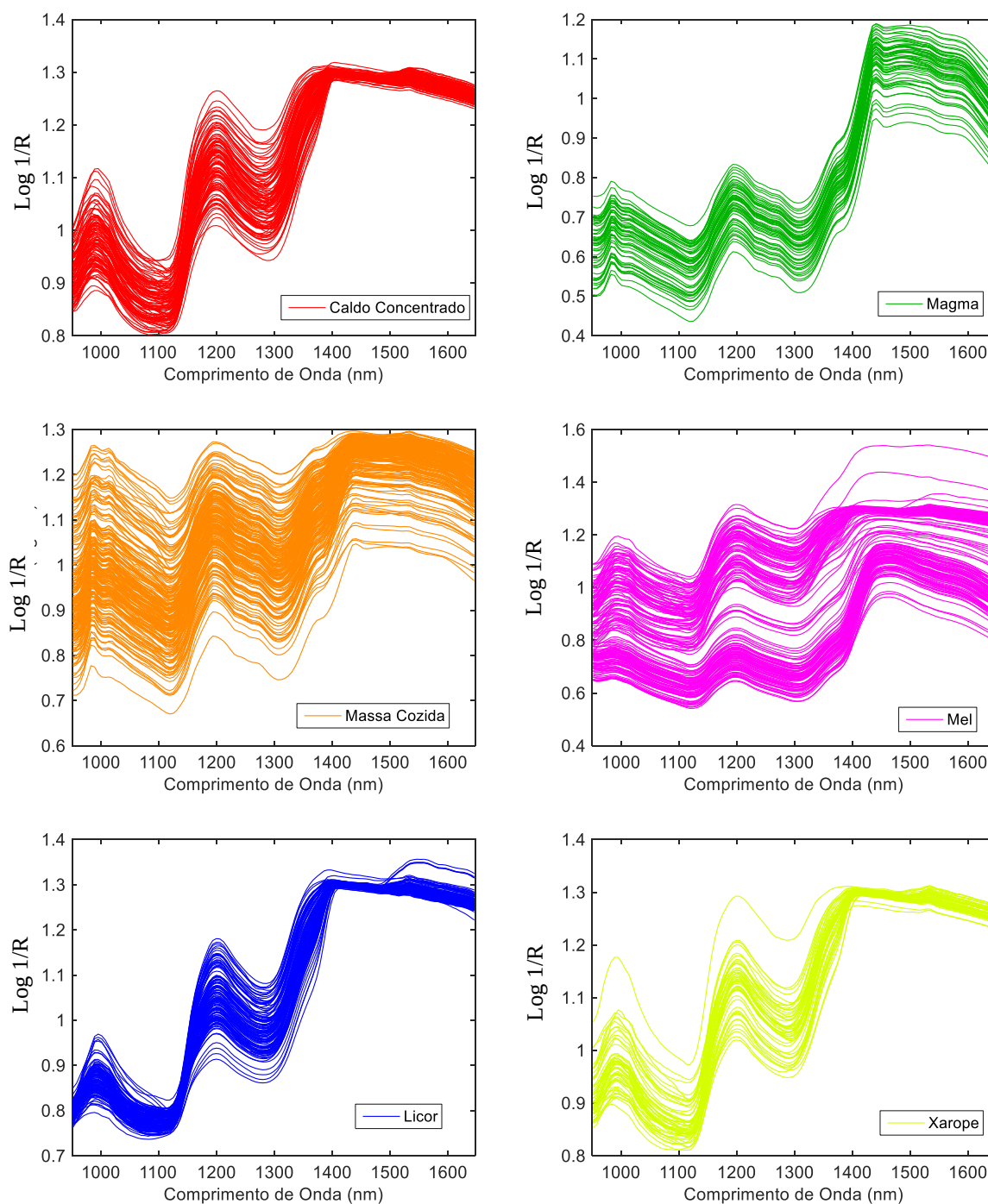
Os três pré-processamentos avaliados corrigiram os defeitos observados nos espectros, efeitos aditivos de linha de base. Os resultados para SNV e MSC são semelhantes e facilitam a interpretação dos resultados, por preservarem a forma original do espectro, enquanto a derivada,

apesar da interpretação ser mais complexa, também é um método eficaz para correção de linha de base dos espectros.

5.3.2 Avaliação dos espectros de reflectância difusa

A Figura 13 apresenta os espectros brutos de 726 amostras adquiridas no modo reflectância difusa nas diversas etapas do processo de fabricação de açúcar, coloridos e separados de acordo com o tipo de amostra para melhor visualização.

Figura 13 – Espectros brutos das amostras coletadas no modo reflectância difusa.



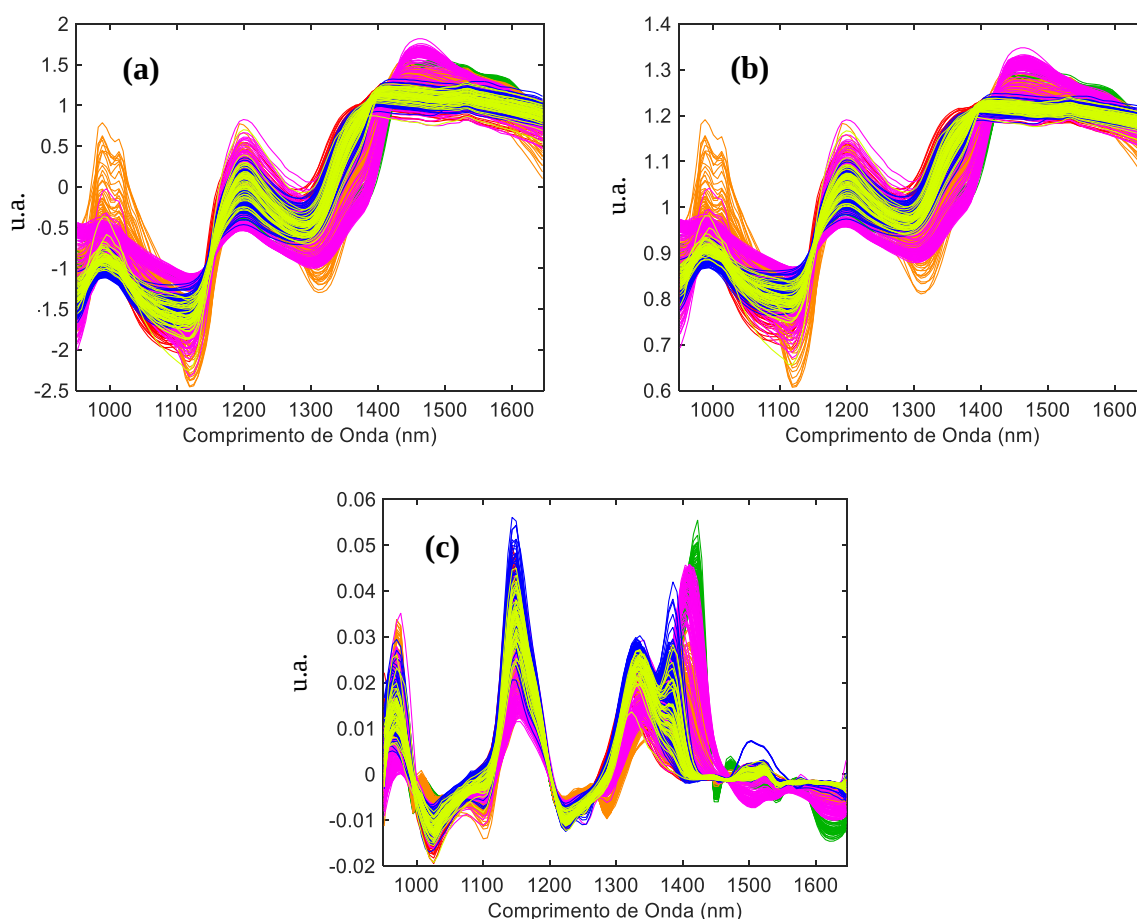
Fonte: O autor (2021)

Diferentemente do comportamento observado nos espectros adquiridos por transmitância, os espectros das amostras analisadas por reflectância difusa mudam de perfil conforme o tipo de amostra. Isso ocorre, principalmente, devido à constituição física e química destas, pois variam de, xarope (viscoso, porém líquido) à magma (de predominância sólida). Além dos diferentes perfis espectrais, também se observa um elevado efeito de espalhamento

da radiação, que é inerente da técnica de reflectância difusa, mas é acentuado por características físicas das amostras, a exemplo das massas, formadas por cristais de açúcar embebidos em mel. Os diferentes tamanhos dos cristais geram espalhamento da radiação e, consequentemente, deslocamento da linha de base.

Para corrigir os efeitos aditivos e multiplicativos de linha base e visualizar a fonte de variação entre as amostras proveniente principalmente da informação química, alguns pré-processamentos foram empregados. Na Figura 14, são mostrados os espectros de reflectância difusa pré-processados com SNV, MSC e 1ª derivada com polinômio de 2ª ordem e janela de 5 pontos. Apesar de ter sido testada também a 1ª derivada com polinômio de 2ª ordem e janela de 7 pontos, os espectros não foram mostrados pois, visualmente, apresentam perfis semelhantes aos pré-processados com janela de 5 pontos.

Figura 14 - Espectros pré-processados das amostras coletadas no modo reflectância difusa.



Caldo concentrado (em vermelho); Xarope (em amarelo); Massa cozida (em laranja); Méis (em rosa); Magma (em verde) e licores (em azul).

(a) SNV; (b) MSC e (c) 1ª derivada com polinômio de 2ª ordem e janela de 5 pontos

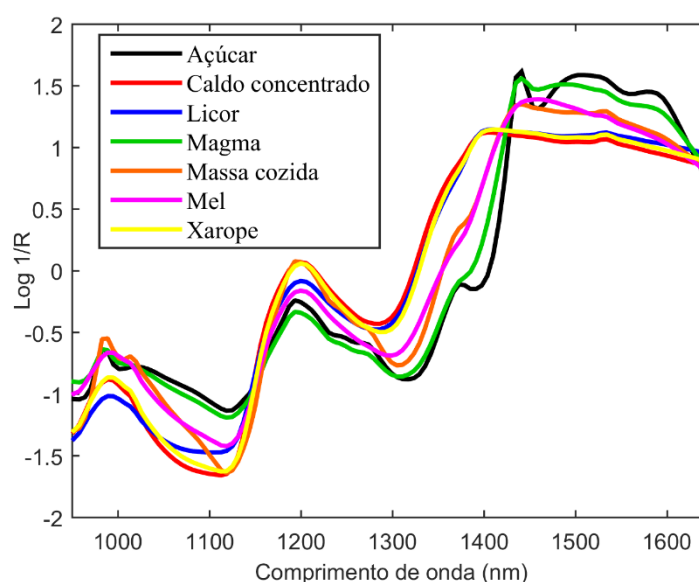
Fonte: O autor (2021)

Avaliando os espectros pré-processados, é possível notar que as principais fontes de variação entre as amostras do processo ocorrem nas regiões em torno de 1000, 1200 e 1400 nm, correspondendo ao 2º sobretom da ligação O-H, 2º sobretom de estiramento da ligação C-H e 1º sobretom da ligação O-H, respectivamente, regiões atribuídas principalmente à presença de açúcares e água (regiões de 1000 e 1400 nm) nas amostras. (OSBORNE, B.G.; FEARN, T.; HINDLE, P.H, 1993).

De modo geral, houve redução dos efeitos de espalhamento da radiação nos pré-processamentos aplicados. Entretanto, as correções por SNV e MSC, apesar de manterem o perfil espectral e facilitarem a comparação entre os diversos tipos de amostras, foram menos eficientes na redução desses efeitos quando comparados à derivada. De fato, segundo Ferreira (2015), em espectros de reflectância difusa onde efeitos de espalhamento da luz são naturais, a primeira derivada é em geral o método mais apropriado para correção.

Para melhor visualização e comparação entre os diversos tipos de amostra, a Figura 15 apresenta os espectros médios de reflectância difusa, pré-processados com SNV, para as diversas etapas da fabricação, além do espectro médio também pré-processado com SNV de 20 amostras de açúcar refinado, todas de elevada pureza, com pol maior que 99,85%.

Figura 15 – Espectros médios pré-processados com SNV para as amostras coletadas no modo reflectância difusa + espectro médio pré-processado com SNV para 20 amostras de açúcar refinado.



Fonte: O autor (2021)

É possível observar que à medida que o processo avança para amostras mais sólidas, o perfil espectral destas assemelham-se ao espectro do produto final, o açúcar. A semelhança é

notória com os espectros de massa cozida e, principalmente, com os espectros de magma. Essas amostras são obtidas após separar o açúcar do mel, na etapa de centrifugação, sendo assim, é o perfil espectral mais próximo ao açúcar puro, diferindo principalmente em umidade.

5.3.3 Avaliação de modelos PLS desenvolvidos com diferentes pré-processamentos

Após a avaliação dos espectros, seguiu-se para a etapa de otimização dos modelos PLS, onde as métricas de desempenho dos modelos gerados com cada pré-processamento (PP) foram analisadas para determinar o melhor tratamento a ser utilizado no desenvolvimento dos modelos. Os seguintes métodos foram testados: SNV, MSC e primeira derivada com filtro Savitzky-Golay (ajustando-se um polinômio de 2ª ordem com janela de 5 (Drv./5 pts.) e 7 (Drv./7 pts.) pontos). Os resultados das etapas de calibração e de predição dos modelos desenvolvidos estão apresentados na Tabela 5.

Tabela 5 - Resultados para as etapas de calibração e de predição dos modelos de regressão PLS para diferentes pré-processamentos.

Modo de análise	PQ	PP	Faixa (%m/m)	VL	RMSEC (%m/m)	RMSECV (%m/m)	R ² cv	RMSEP (%m/m)	R ² p
Transmitância	Brix	SNV	1,0-45	3	0,76	0,79	0,98	0,88	0,98
		MSC		3	0,78	0,79	0,98	0,88	0,98
		Drv./5 pts.		4	0,71	0,72	0,98	0,74	0,99
		Drv./7 pts.		4	0,72	0,72	0,98	0,74	0,99
	Pol	SNV	1,1-20,6	3	0,85	0,88	0,93	0,91	0,93
		MSC		3	0,85	0,88	0,93	0,91	0,93
		Drv./5 pts.		4	0,74	0,76	0,95	0,77	0,95
		Drv./7 pts.		4	0,75	0,77	0,95	0,76	0,95
Reflectância Difusa	Brix	SNV	45,3-95,3	6	2,69	2,77	0,96	2,84	0,96
		MSC		5	2,87	2,95	0,95	3,04	0,95
		Drv./5 pts.		4	2,25	2,33	0,97	2,36	0,97
		Drv./7 pts.		4	2,32	2,38	0,97	2,41	0,97
	Pol	SNV	36,2-90,0	4	4,45	4,65	0,92	4,28	0,94
		MSC		4	4,85	5,16	0,9	4,69	0,92
		Drv./5 pts.		5	2,86	2,99	0,97	2,98	0,97
		Drv./7 pts.		5	2,81	2,92	0,97	2,95	0,97

* PQ, parâmetro de qualidade; PP, pré-processamento; VL, variáveis latentes.

Fonte: Própria (2021)

Por meio das métricas de desempenho apresentadas na Tabela 5, foi possível analisar a técnica de pré-processamento mais adequada para o desenvolvimento dos modelos. Isso foi feito de modo a selecionar, se possível, um único pré-processamento para os modelos PLS, com o intuito de facilitar a interpretação e aplicação dos modelos no monitoramento do processo.

Para as 4 modelagens avaliadas (brix e pol por transmitância e por reflectância difusa), os valores de RMSEP para os modelos PLS desenvolvidos com SNV e MSC apresentaram diferença estatisticamente significativas em relação aos modelos obtidos com a 1ª derivada, com base em teste F a 95% de confiança. Assim, apesar de os modelos com derivada terem precisado, em geral, de mais variáveis latentes, esse pré-processamento foi escolhido por gerar modelos com RMSEP significativamente menor. Entre as derivadas, para todos os modelos, a alteração da janela não apontou diferença estatística para os valores de RMSEP (teste F, 95% de confiança), nem provocou alteração do número de variáveis latentes requeridas na construção do modelo. Dessa forma, a 1ª derivada com polinômio de 2ª ordem e janela de 5 pontos foi escolhida, e utilizada na discussão dos resultados dos modelos apresentada a seguir. A escolha desse pré-processamento ratificou ainda o comportamento observado nos espectros, principalmente para a reflectância difusa, em que a aplicação da derivada foi mais eficiente na correção de efeitos de espalhamento da radiação.

Para a quantificação de brix por transmitância, após a eliminação de *outliers*, 547 amostras de caldos e caldos concentrados com teores menores do que 45% (m/m) foram utilizadas na etapa de calibração, e 234 amostras desse tipo, para a etapa de predição. Os resultados obtidos para esse modelo foram considerados satisfatórios tanto para a calibração (RMSECV de 0,72% (m/m)) quanto para a predição (RMSEP de 0,74% (m/m)).

Após a investigação e exclusão de *outliers*, o modelo para quantificar pol por transmitância foi construído utilizando-se 324 amostras de caldo no conjunto de calibração e 138 amostras para o conjunto de predição. Após a avaliação desses modelos, resultados satisfatórios também foram obtidos tanto para a etapa de calibração (RMSECV = 0,76% (m/m)) quanto para a de predição (RMSEP = 0,77% (m/m)).

Como comparação, Sorol *et al.* (2010) obtiveram valores de RMSEP de 0,52% (m/m) quando utilizaram espectros de transmitância obtidos por um espectrômetro Vis-NIR de bancada, com faixa espectral de 400 a 2498 nm, e de 0,28% (m/m), aplicando algoritmo genético como seleção de variáveis, para estimar o brix em caldos extraídos de cana-de-açúcar por modelos PLS. A faixa de concentração de brix utilizada na construção dos modelos, de 11,76 a 23,15% (m/m), foi menor do que a empregada no presente trabalho. Valores de RMSEP de 0,28% (m/m) e 0,42% (m/m) foram encontrados por Valderrama, Braga & Poppi (2007) de um conjunto de predição formado por 378 amostras de caldo extraído de cana-de-açúcar, para modelos PLS preditivos de

brix e pol, respectivamente. Os modelos foram construídos através de medidas de transmitância em um espectrômetro NIR de bancada com faixa espectral de 1100 a 2500 nm e faixas de concentração de brix e pol de 8 a 26% (m/m) e de 4 a 23% (m/m), aproximada e respectivamente.

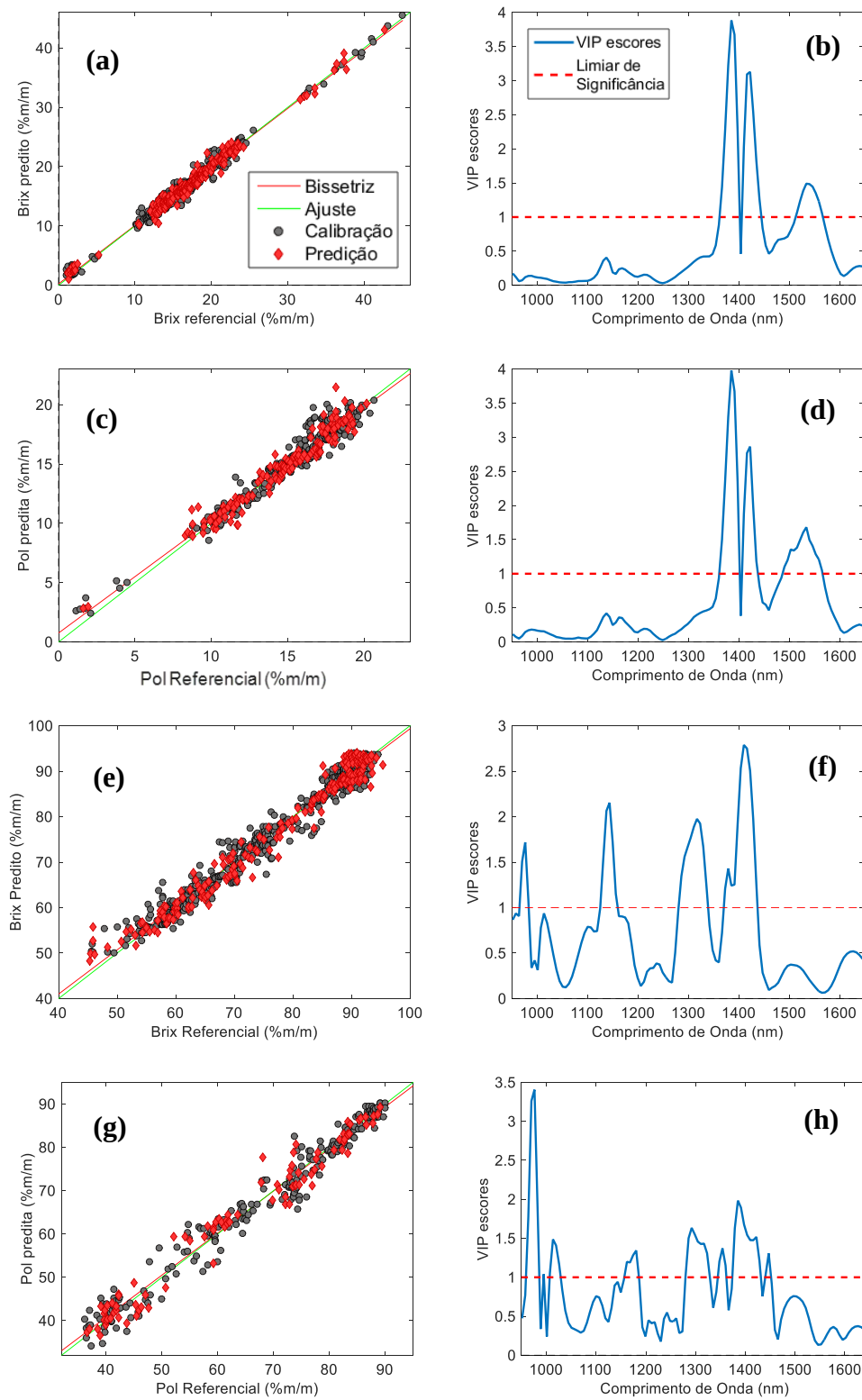
Em relação ao modelo desenvolvido para os dados adquiridos em reflectância difusa, na quantificação de brix, foram utilizadas amostras de caldos concentrados com teores maiores do que 45% (m/m), xaropes, méis, magmas, massas cozidas e licores da etapa de refinamento. Após verificação e exclusão de amostras anômalas, gerou-se o modelo formado por 508 amostras no conjunto de calibração e 218 no conjunto de predição, com resultados satisfatórios para a calibração (RMSECV = 2,33% (m/m)), bem como para a predição (RMSEP = 2,36% (m/m)).

Quanto ao modelo para determinação de pol por reflectância difusa, amostras de xaropes, méis, magmas, massas cozidas e licores foram empregadas na construção do mesmo, que apresentou resultados aceitáveis para as etapas de calibração (RMSECV de 2,99% (m/m)) e de predição (RMSEP de 2,98% (m/m)), utilizando para tal, 215 amostras na calibração e 92 amostras na predição, após a retirada de *outliers*.

Tange *et al.* (2015) desenvolveram modelos PLS globais e locais para quantificar os teores de brix e pol de 4 diferentes tipos de amostras do processo de fabricação do açúcar (caldo, xarope, massa cozida e mel) utilizando espectros de transreflectância obtidos por um espectrômetro Vis-NIR de bancada com faixa espectral de 400 a 2498 nm. As medidas foram realizadas diretamente nas amostras sem necessidade de diluição. Os valores de RMSEP para os modelos globais de brix e pol foram de 1,15 e 1,54% (m/m), respectivamente, considerando a faixa de concentração de 15 a 95% (m/m) para brix e de 13 a 78% (m/m), para pol, aproximadamente.

Adicionalmente, em relação aos modelos PLS selecionados, é importante ressaltar que os mesmos não apresentaram *bias* estatisticamente significativo, de acordo com um teste t, em um nível de confiança de 95% (*bias* de predição para os modelos de brix e pol por transmitância e por reflectância difusa, respectivamente: -0,001, 0,01, -0,08 e 0,12). Os gráficos de valores de referência *versus* preditos desses modelos para as amostras de calibração (em preto) e as amostras do conjunto de predição (em vermelho) estão apresentados na Figura 16 (a, c, e, g). Além disso, os gráficos de importância das variáveis (VIP) para esses modelos estão apresentados na Figura 16 (b, d, f, h). Valores que apresentarem VIPs escores acima do limiar de significância, linha tracejada em vermelho, são considerados importantes para a construção do modelo.

Figura 16 - Gráficos dos valores preditos vs. referência para os conjuntos de calibração e de predição; e os VIPs escores dos modelos PLS.



Modelos para quantificar brix (a e b) e pol (c e d) por transmitância; e brix (e e f) e pol (g e h) por reflectância difusa.

Fonte: O autor (2021)

Como observado na Figura 16 (a, c, e, g), de modo geral, pode-se perceber que as amostras de calibração e de predição estão bem distribuídas entre si devido ao método empregado para a seleção de amostras (*Onion method*). No gráfico de valores preditos versus referência do modelo de brix por reflectância, é possível perceber uma certa tendência de comportamento não linear. As possíveis não-linearidades presentes nos dados, as quais podem limitar o desempenho do modelo PLS, serão discutidas a seguir, quando os modelos SVM forem avaliados.

Na Figura 16 (b e d), é possível observar que as variáveis mais importantes para a quantificação do brix e pol por transmitância estão localizadas principalmente nas regiões em torno de 1400 e 1550 nm, atribuídas ao 1^a sobretom da ligação O-H, presente nas moléculas de sacarose e água das amostras. A similaridade entre esses dois gráficos pode estar relacionada com a elevada pureza aparente das amostras, calculada pela Equação 17, que para os caldos, em média, foi superior a 80%. Logo, os valores de brix e pol são próximos entre si e indicativos da sacarose presente.

$$\frac{pol}{brix} \times 100 \quad (17)$$

As faixas de comprimento de onda consideradas importantes para o desenvolvimento dos modelos de transmitância deste trabalho, também ficaram entre as selecionadas por Sorol *et al.* (2010), através de diferentes métodos de seleção de variáveis empregados para o desenvolvimento de modelos PLS preditivos de brix em caldos extraídos de cana-de-açúcar.

Para a quantificação do brix por reflectância difusa, além da região em 1400 nm, as regiões em torno de 950, 1150 e 1300 nm também foram importantes, representando o 2^o sobretom da ligação O-H, 2^o sobretom de estiramento da ligação C-H e regiões de combinação do grupo C-H, respectivamente. Todas essas são ligações presentes, principalmente, na estrutura da molécula de sacarose. Ao contrário do comportamento observado para os VIP escores dos modelos de transmitância, as variáveis consideradas importantes na Figura 16 (f e h), apesar de serem as mesmas regiões para brix e pol, diferem quanto à intensidade de importância. A pureza aparente mais uma vez pode estar envolvida neste comportamento, uma vez que, para as amostras de reflectância difusa, apresenta elevada variabilidade (de 42% para mel C a 99,9 % para massa refinada por exemplo). Com isso, os teores de brix e pol não são similares para determinadas amostras, o que pode ter ocasionado as diferenças observadas nos gráficos.

Como comparação, Ramirez-Morales *et al.* (2016), através do método de seleção de variáveis univariado baseado em teste t, obtiveram como resultado diversos comprimentos de onda selecionados das faixas de 1350 a 1888 nm para o brix, e 950 a 1050 nm e 1150 a 1888 nm, dentre outras variáveis no visível, para a pol. Essas regiões foram consideradas importantes para a

construção de modelos SVM destinados a quantificação de brix e pol de amostras diversas do processo de fabricação de açúcar. Enquanto Almeida *et al.* (2018), utilizando iSPA como método de seleção de variáveis, tiveram as seguintes faixas selecionadas: de 800 a 1500 nm, aproximadamente, para brix, e de 600 a 800 nm e 1250 a 1450 nm, aproximadamente, para pol. Estas foram utilizadas na construção de modelos kernel-PLS para quantificar esses parâmetros de qualidade da cana-de-açúcar para diferentes tipos de amostras do processo. Ambos os trabalhos foram desenvolvidos através do conjunto de dados descrito em Tange *et al.* (2015), composto por cerca de 1800 espectros de transfectância Vis-NIR varrendo a faixa de 400 a 2498 nm e valores de referência para brix e pol de 4 tipos de amostra do processo de fabricação de açúcar (caldo, xarope, mel e massa cozida).

5.3.4 Avaliação de modelos SVM

Apesar dos resultados satisfatórios obtidos, buscando otimizar ainda mais os modelos gerados, optou-se por testar a regressão por SVM, em que os novos modelos foram construídos utilizando-se apenas a técnica de pré-processamento selecionada na etapa anterior, assim como, as mesmas amostras consideradas na modelagem PLS. Essa estratégia foi adotada devido à maior dificuldade que se tem de avaliar *outliers* na modelagem por SVM.

Esta técnica foi considerada devido à possível presença de efeitos não-lineares minimizando o desempenho dos modelos PLS, advindos de parâmetros físicos e químicos que modificam a interação entre a radiação NIR e a amostra. Algumas características do conjunto de dados utilizado nesse trabalho que podem estar associadas aos efeitos não-lineares, de acordo com (BERTRAN *et al.*, 1999; MILLER, 1993), são: elevada concentração de brix e pol que, como discutido anteriormente, podem causar desvios na Lei de Beer; tamanho de partícula, que causa efeitos de espalhamento da radiação; e fatores externos como temperatura.

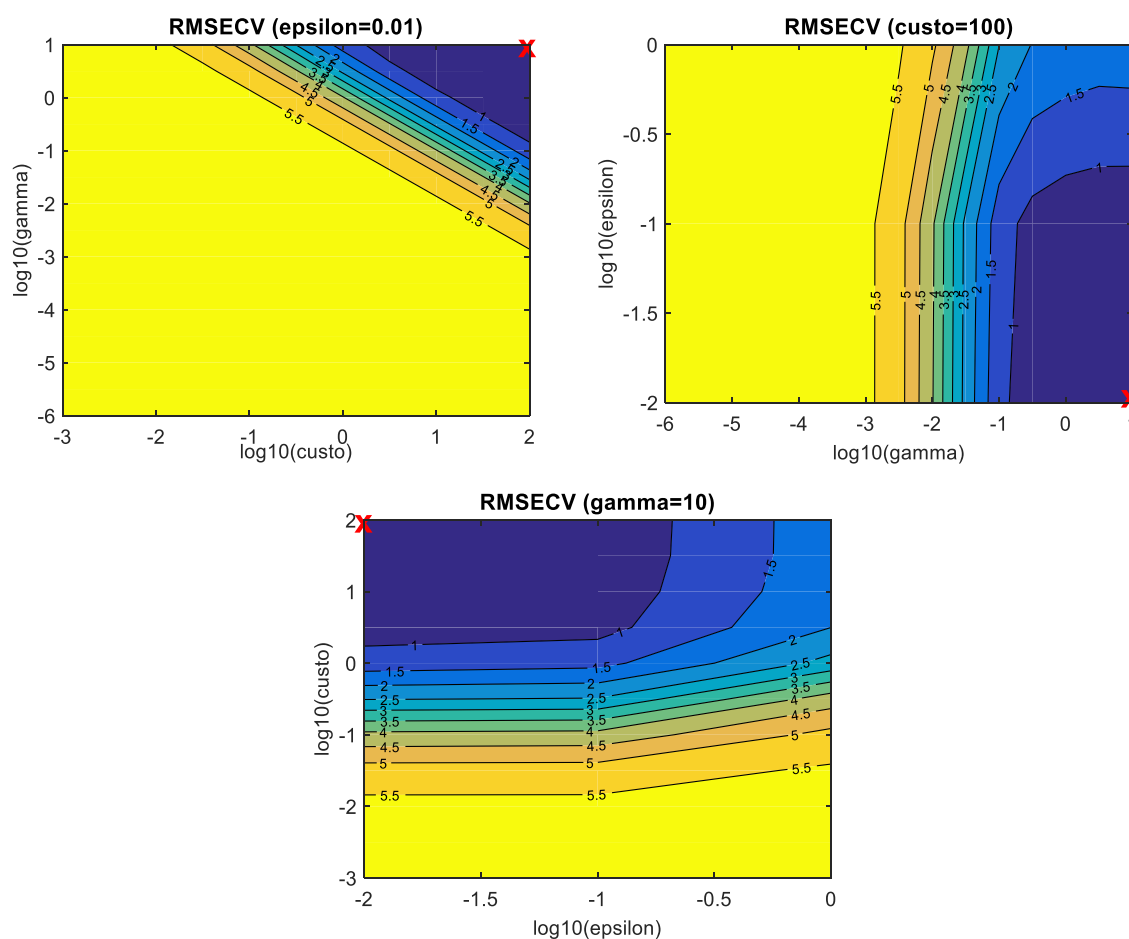
Uma das formas de tratar dados não-lineares, é usando técnicas que possam lidar com esses efeitos. Ao comparar diferentes técnicas de regressão, PLS, ANN e SVM, na quantificação da pol de diferentes substratos do processo de fabricação de açúcar, Tange *et al.* (2017) pontuaram algumas vantagens de se utilizar SVM em aplicações em que se têm dados reais, dentre as quais, pode-se citar o fato de que os resultados com SVM obtiveram, de modo geral, maior acurácia e uma elevada capacidade de generalização, mesmo quando empregado um reduzido conjunto de calibração.

Ao construir esses modelos no *software* PLS Toolbox, é realizada uma varredura na grade de parâmetros de ajustes (custo - C, epsilon - ϵ e gamma - γ), usando validação cruzada para selecionar os valores dos parâmetros que fornecem o menor RMSECV, e construir o

modelo usando esses valores. Se um ponto ótimo não é encontrado na primeira varredura, a faixa de parâmetros de ajuste pode ser redefinida (através da análise dos gráficos de superfície de resposta que auxiliam na determinação dessas novas faixas), e o modelo é recalculado até sua otimização. A fim de evitar o sobreajuste no conjunto de calibração dos modelos, o conjunto de predição é continuamente avaliado a cada varredura.

Os gráficos de superfície de resposta da primeira varredura para ajuste de parâmetros SVM no modelo preditivo de brix por transmitância são apresentados na Figura 17.

Figura 17 – Superfícies de resposta da primeira varredura para ajuste de parâmetros SVM para o modelo preditivo de brix por transmitância.

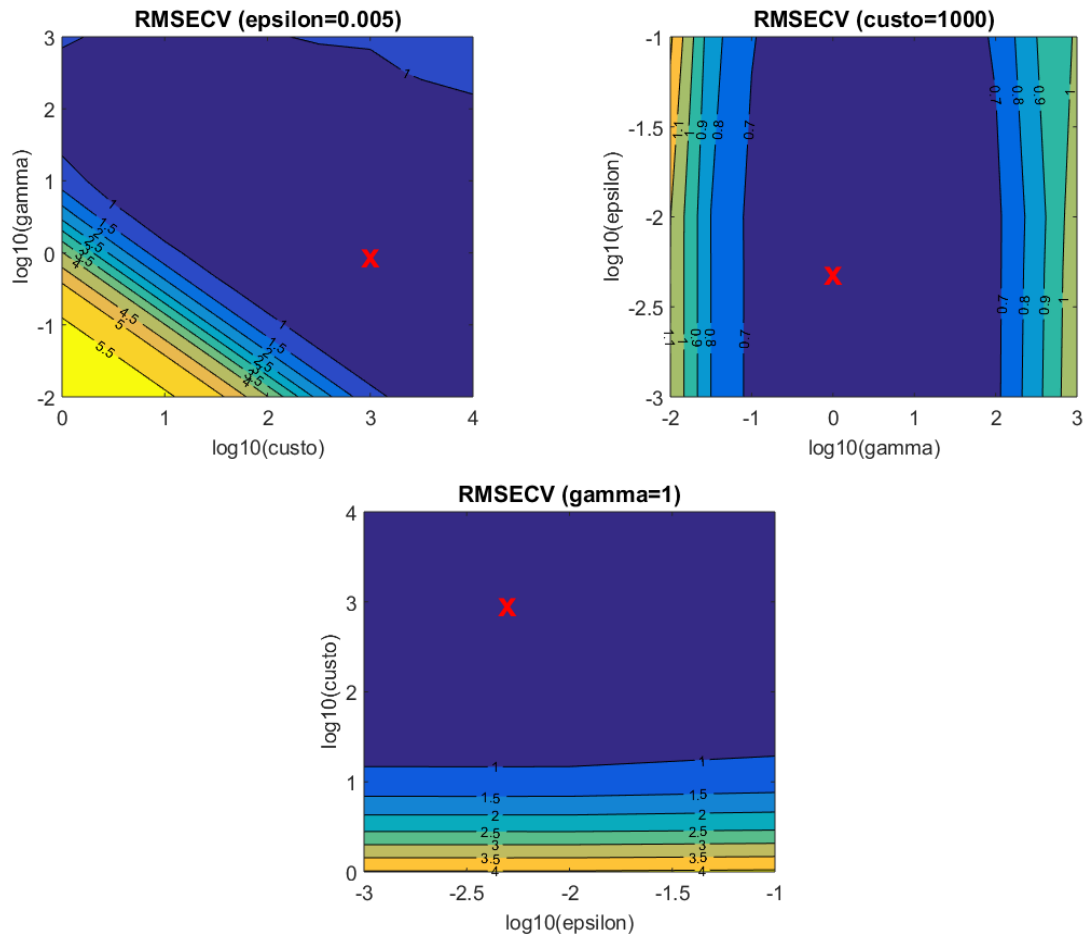


Fonte: O autor (2021)

A região otimizada é demarcada pelo X vermelho em cada gráfico bidimensional, representando os parâmetros otimizados para a construção do modelo ($\epsilon = 0,01$; $C = 100$; $\gamma = 10$). É possível observar que nos 3 gráficos, o ponto ótimo está no limite da faixa de cada parâmetro, indicando que valores maiores para os parâmetros custo e gamma, e menores para epsilon, podem gerar um modelo com RMSECV menor. Assim, a faixa para esses parâmetros

foi redefinida, e uma nova varredura foi realizada. Os gráficos com os parâmetros otimizados são mostrados na Figura 18.

Figura 18 – Superfícies de resposta da segunda varredura para ajuste de parâmetros SVM para o modelo preditivo de brix por transmitância.



Fonte: O autor (2021)

Com a nova varredura, o ponto ótimo de RMSECV para o modelo SVM preditivo de brix por transmitância foi determinado através dos parâmetros do modelo ($\epsilon = 0,005$; $C = 1000$; $\gamma = 1$). Novas varreduras não são necessárias, já que o valor otimizado se encontra dentro das faixas consideradas. Para os outros 3 modelos avaliados (pol por transmitância e brix e pol por reflectância difusa), também foram necessárias à execução de duas varreduras para otimização dos parâmetros SVM, as superfícies de resposta geradas encontram-se no APÊNDICE A.

As superfícies de resposta são indicativas do modelo SVM ideal, mas como dito anteriormente, o conjunto de predição deve ser analisado paralelamente a fim de evitar sobreajustes nos modelos. Para isso, as métricas de desempenho são avaliadas para determinar se o resultado gerado é satisfatório, definindo assim, o modelo ideal. Os resultados das etapas

de calibração e de predição dos modelos SVM desenvolvidos para cada varredura estão apresentados na Tabela 6.

Tabela 6 - Resultados para as etapas de calibração e de predição dos modelos de regressão SVM para avaliação de parâmetros de ajuste.

MA	PQ	V	Ajuste SVM			VS	RMSEC (%m/m)	RMSECV (%m/m)	R ² cv	RMSEP (%m/m)	R ² p
			ϵ	C	γ						
Transmitância	Brix	1 ^a	0,01	100	10	465	0,59	0,61	0,99	0,59	0,99
		2 ^a	0,005	1000	1	512	0,60	0,61	0,99	0,59	0,99
	Pol	1 ^a	0,1	100	10	192	0,69	0,72	0,95	0,73	0,96
		2 ^a	0,1	316,6	31,6	184	0,62	0,69	0,96	0,69	0,96
Reflectância Difusa	Brix	1 ^a	0,01	100	10	455	1,70	1,81	0,98	1,78	0,98
		2 ^a	0,005	316,6	31,6	478	1,35	1,64	0,98	1,44	0,99
	Pol	1 ^a	0,1	100	10	96	2,20	2,43	0,98	2,44	0,98
		2 ^a	0,1	10000	3,16	106	1,74	2,36	0,98	2,49	0,98

MA, modo de análise; PQ, parâmetro de qualidade; V, varredura; VS, vetores de suporte.

Fonte: O autor (2021)

A partir das métricas de desempenho apresentadas na Tabela 6, foi possível analisar os efeitos da alteração dos parâmetros de ajuste nos modelos gerados em cada varredura.

Na avaliação do primeiro conjunto de dados (brix por transmitância), é observado que a mudança nos parâmetros de ajuste praticamente não altera o valor de RMSEP do modelo. Como essa alteração não causou mudanças significativas no RMSEP (teste F, 95% de confiança) e ainda aumentou o número de vetores de suportes necessários para modelar o conjunto de calibração, então, por parcimônia, a 1^a varredura foi selecionada.

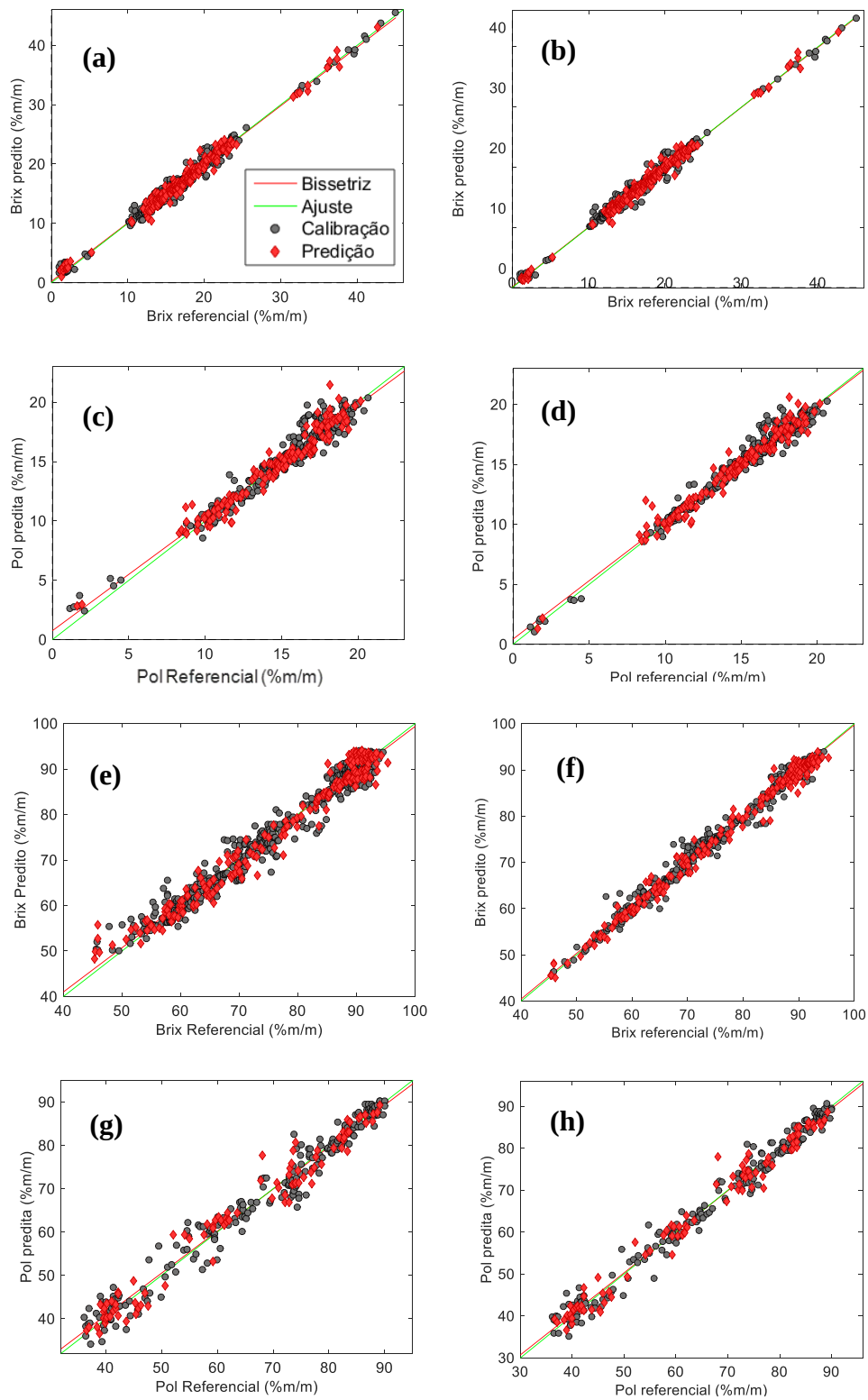
Entre a 1^a e 2^a varredura dos modelos preditivos de pol por transmitância, C e γ foram alterados. A combinação dos parâmetros otimizados requereu menos vetores de suporte para explicar o conjunto de calibração, por esse motivo, apesar da 2^a varredura não ser estatisticamente diferente da 1^a por meio de teste F a 95% de confiança, foi selecionada para o modelo SVM preditivo de pol.

Diminuição no parâmetro ϵ e acréscimo nos valores de C e γ foram observados ao passar da 1^a para a 2^a varredura nos modelos preditivos de brix por reflectância difusa. A otimização desses parâmetros diminuiu o RMSEP do modelo sem sobreajustar os dados de calibração. Sendo essa diminuição estatisticamente significativa (teste F com 95% de confiança), a 2^a varredura foi selecionada.

A região otimizada da validação cruzada para a pol por reflectância difusa foi obtida após aumento do custo e diminuição de gamma. Entretanto, com a segunda otimização, o erro de calibração diminuiu acentuadamente, enquanto o erro de predição aumentou, apesar de não ser estatisticamente diferente (teste F, 95% de confiança) do valor de RMSEP da primeira varredura, indicando assim sobreajuste dos dados e gerando um modelo com menor capacidade preditiva para amostras externas. Dessa forma, a 1ª varredura é mais satisfatória para o modelo SVM preditivo de pol por reflectância difusa.

Os modelos SVM selecionados para cada parâmetro não apresentaram *bias* estatisticamente significativo, de acordo com um teste t a 95% de confiança (*bias* de predição para os modelos de brix e pol por transmitância e por reflectância difusa, respectivamente: 0,015, 0,05, -0,08, -0,13). A título de comparação, os gráficos de valores de referência *versus* preditos dos modelos PLS (a, c, e, g) e SVM (b, d, f, h), estão apresentados na Figura 19.

Figura 19 - Gráficos dos valores preditos vs. referência para os conjuntos de calibração e de predição dos modelos PLS (à esquerda) e SVM (à direita).



Modelos para quantificar brix (a e b) e pol (c e d) por transmitância; e brix (e e f) e pol (g e h) por reflectância difusa.

Fonte: O autor (2021)

Através de teste F a 95% de confiança, a diferença entre os valores de RMSEP foi estatisticamente significativa entre os modelos PLS e SVM para cada parâmetro avaliado, corroborando a presença dos efeitos não-lineares sugeridos anteriormente.

Analisando os gráficos da Figura 19 e a Tabela 6, é possível observar que os modelos de brix (Figura 19 e, f) e pol (Figura 19 g, h) por reflectância difusa se ajustaram melhor de acordo com a técnica SVM do que os modelos por transmitância (Figura 19 a, b, c, d), sugerindo que esses são mais susceptíveis a efeitos não-lineares, o que de fato é explicado pelas condições do processo. Os modelos por reflectância difusa possuem as amostras mais concentradas, com elevados teores de brix e pol e são mais susceptíveis a desvios da Lei de Beer, ou seja, a relação entre absorvância e concentração nestas condições desvia da linearidade (MILLER, 1993). Além do mais, diferenças de viscosidade e tamanho de partícula presentes nas diversas amostras incluídas nesses modelos, como os méis, massas e magmas é mais acentuada. Por outro lado, os modelos obtidos por transmitância são formados, em geral, por caldos, que possuem concentrações mais baixas, sendo assim, o desvio de linearidade observado é menor.

Os resultados apresentados foram comparados aos obtidos por outros autores. Tange *et al.* (2015), por exemplo, para avaliar a presença e causas de efeitos não-lineares em quatro substratos do processo de fabricação do açúcar bruto, construíram modelos PLS e SVM locais e globais para quantificação de brix e pol das amostras do processo (caldo, xarope, melaço e massa cozida). Os modelos foram construídos utilizando-se um espectrômetro Vis-NIR com faixa espectral de 400 a 2498 nm através de medidas de transflectância. Como também foi observado no presente trabalho, nesse estudo foi mostrado que efeitos não-lineares, como elevada concentração, pH, viscosidade, tamanho de partícula, temperatura e composição amostral influenciam todas as etapas do processo, principalmente para melaço e massas. O baixo desvio da linearidade observado para os caldos é justificado pela reduzida interação intermolecular, por ser um tipo de amostra com baixa concentração de sólidos. A abordagem SVM se sobressaiu em comparação ao PLS na modelagem dos efeitos não-lineares e resultados satisfatórios de RMSEP foram obtidos nos modelos globais de brix e pol, 0,59 e 0,64% (m/m), respectivamente.

Utilizando a mesma biblioteca de dados do trabalho anterior, Almeida *et al.* (2018) propuseram um novo método de seleção de variáveis para calibração multivariada não-linear, o iSPA-Kernel-PLS, o qual é comparado ao Kernel-PLS utilizando o espectro total. Os autores obtiveram valores de RMSEP de 1,0 e 1,1% (m/m) para brix e pol, respectivamente, utilizando o novo método, e concluindo que o mesmo possuiu capacidade preditiva superior ao modelo gerado por Kernel-PLS.

Também utilizando a biblioteca de dados de Tange *et al.* (2015), Ramirez-Morales *et al.* (2016) avaliaram um método de seleção de variáveis univariado baseado em teste t, na quantificação de brix e pol através de modelagem SVM, em que os modelos globais apresentaram resultados de RMSEP mais baixos que os descritos no trabalho original, de 0,305 e 0,485% (m/m) para brix e pol respectivamente.

Por fim, para facilitar uma análise comparativa, os resultados obtidos no presente trabalho para os modelos PLS e SVM, desenvolvidos com 1ª derivada (polinômio de 2ª ordem e janela de 5 pontos), estão apresentados na Tabela 7. Adicionalmente, são apresentados também alguns resultados obtidos por outros autores.

Para brix, o equipamento portátil avaliado no presente trabalho, empregando dados de transmitância e modelos SVM, apresentou resultados comparáveis aos apresentados por Tange *et al.* (2015) e Ramirez-Morales *et al.* (2016), que utilizaram um equipamento de bancada. Entretanto, é importante ressaltar que as faixas de brix são diferentes. Para uma faixa de brix mais elevada foi necessária a aquisição espectral por reflectância difusa empregando o equipamento portátil. Neste caso, os valores de RMSEP são mais elevados que os apresentados por Tange *et al.* (2015), porém semelhantes aos valores de RMSEP obtidos por Almeida *et al.* (2018).

No caso do parâmetro pol, observa-se que para as amostras menos concentradas cujos espectros foram obtidos por transmitância empregando o equipamento portátil, os valores de RMSEP são também semelhantes aos reportados por Tange *et al.* (2015) e Ramirez-Morales *et al.* (2016), e menores que os de Almeida *et al.* (2018), que usaram i-SPA-Kernel-PLS. Para uma faixa de pol mais elevada, os modelos SVM construídos com os espectros adquiridos por reflectância apresentaram RMSEP maiores que os reportados nos outros trabalhos. Entretanto, considerando que o valor de RMSEP de 2,44% (m/m) representa 3,9% em valores relativos à média da faixa de pol, pode ser considerado um desempenho bastante satisfatório. No caso de Tange *et al.* (2015), Almeida *et al.* (2018) e Ramirez-Morales *et al.* (2016) os valores de RMSEP relativos são 1,41%, 2,42% e 1,07% respectivamente.

Apesar do desempenho de predição inferior observado para o modelo de pol por reflectância difusa, os resultados obtidos pelo presente estudo são muito promissores para o monitoramento do processo de fabricação de açúcar. É importante destacar a heterogeneidade e complexidade das amostras industriais incluídas neste trabalho, que foram adquiridas ao longo de duas safras, de diversas etapas da cadeia produtiva, abrangendo todas as variações de amostras discutidas acima, ressaltando que, o emprego de um equipamento portátil na construção desses modelos, no ponto de vista comercial e industrial, traz uma ótima relação

custo/benefício para o setor, pois, um espectrômetro NIR portátil de baixo custo associado a uma metodologia rápida e não-destrutiva na quantificação de brix e pol, demonstrou no presente estudo, resultados equiparáveis aos obtidos por instrumentos de bancada, que em geral possuem alto custo de aquisição e manutenção. Sendo assim, uma proposta bastante viável para o setor sucroalcooleiro.

Tabela 7 – Métricas de desempenho dos modelos otimizados para brix e pol e comparação com os resultados de modelos obtidos por outros autores.

Autor	Instrumento	Faixa Espectral (nm)	Modo de Análise	PQ	Faixa de Concentração (%m/m)	Modelo	RMSEP (%m/m)
Presente trabalho	Equipamento portátil MicroNIR™	950-1650	Transmitância	Brix	1,0 - 45	PLS	0,74
						SVM	0,59
				Pol	1,1 - 20,6	PLS	0,77
						SVM	0,69
			Reflectância Difusa	Brix	45,3 - 95,3	PLS	2,36
						SVM	1,44
				Pol	36,2 - 90,0	PLS	2,98
						SVM	2,44
Tange <i>et al.</i> (2015)	Equipamento de bancada NIRS™ DS2500 (FOSS AB) equipado com refletor de ouro (0,5 mm)	400 - 2498	Transflectância	Brix	15 - 95	SVM	0,59
Pol				13 - 78	0,64		
Almeida <i>et al.</i> (2018)		Seleção de Variáveis entre 400 - 2498		Brix	15 - 95	iSPA-Kernel-PLS	1,0
				Pol	13 - 78		1,1
Ramirez-Morales <i>et al.</i> (2016)				Brix	15 - 95	SVM	0,305
				Pol	13 - 78		0,485

*Parâmetro de Qualidade

Fonte: O autor (2021)

6 CONCLUSÃO

O presente trabalho avaliou o desempenho de um espectrômetro portátil na região do NIR (MicroNIR) para o monitoramento de parâmetros de qualidade em uma usina sucroalcooleira, através da quantificação do teor de sólidos solúveis (brix) e sacarose aparente (pol), em diversas etapas do processo de fabricação de açúcar utilizando diferentes estratégias de calibração multivariada. Estudos preliminares determinaram o melhor método de aquisição espectral para as amostras do processo produtivo e os modelos desenvolvidos permitiram avaliar a presença de efeitos não-lineares ao comparar regressões PLS e SVM.

Os modelos SVM construídos apresentaram melhor desempenho na quantificação dos parâmetros avaliados quando comparados aos modelos PLS. Assim, valores de RMSEP significativamente menores (95% de confiança) foram observados para os modelos SVM: 0,59 e 0,69% (m/m) para brix e pol por transmitância, e 1,44 e 2,44% (m/m) para brix e pol por reflectância difusa, respectivamente. Tais resultados evidenciaram os desvios de linearidade sugeridos, assim como mostraram a capacidade da regressão por SVM em modelar as não-linearidades dos espectros NIR.

A introdução desses métodos pode beneficiar todo o processo da indústria sucroalcooleira, tendo em vista que após a construção dos modelos de calibração, a previsão de novas amostras é rápida e prática, reduzindo o tempo na obtenção de resultados, assim como da utilização de substâncias nocivas ao meio ambiente e seres humanos, para assim, buscar maior eficiência no monitoramento e controle do processo.

O emprego de espectrômetros portáteis para o controle de processos no setor sucroalcooleiro apesar de ser pouco discutido na literatura, apresenta em geral, desempenho satisfatório quando comparado aos equipamentos de bancada, como mostrado no presente estudo principalmente para os modelos por transmitância, e junto ao seu baixo custo de aquisição, somam uma boa relação custo x benefício para o setor.

7 PERSPECTIVAS FUTURAS

Como perspectivas para continuidade do trabalho propõe-se:

- Alimentar os modelos com mais espectros provenientes das amostras de outras safras, aumentando a variabilidade dos dados e buscando melhorar os resultados;
- Avaliar as demandas e dificuldades industriais nos diversos setores da fabricação de açúcar e de álcool (Extração; Tratamento de caldo; Evaporação; Cozimento e Cristalização; Refinaria de açúcar; Fermentação; Destilaria; Expedição), para traçar novas estratégias de monitoramento e controle de processos;
- Avaliar o desempenho de outro equipamento portátil (Scanner -NeoSpectra), que opera em uma faixa espectral mais ampla (1350 a 2500 nm) e pode fornecer informações importantes a respeito das diferenças entre as amostras do processo;
- Estudar e construir, através de ferramentas quimiométricas, sensores virtuais para determinação de propriedades de qualidade, tanto com base nos dados espectroscópicos coletados, quanto nas variáveis de processo (como temperatura, pressão, vazão, etc);
- Estender os modelos de monitoramento e controle de qualidade a outros parâmetros, como teor de açúcar redutores, pureza da cana e concentração de etanol;
- Desenvolver cartas de controle estatístico de processos multivariados para monitorar o andamento e evolução do processo produtivo;

REFERÊNCIAS

- ALMEIDA, V. E.; DE ARAÚJO GOMES, A.; DE SOUSA FERNANDES, D. D.; GOICOECHEA, H. C.; GALVÃO, R. K. H.; ARAÚJO, M. C. U. Vis-NIR spectrometric determination of Brix and sucrose in sugar production samples using kernel partial least squares with interval selection based on the successive projections algorithm. **Talanta**, v. 181, p. 38–43, 2018.
- BALABIN, R. M.; LOMAKINA, E. I. Support vector machine regression (SVR/LS-SVM) - An alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. **Analyst**, v. 136, n. 8, p. 1703–1712, 2011.
- BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. **Chemometrics: A Practical Guide**. 1. ed. New York - NY: Wiley-Interscience, 1998.
- BERTRAN, E.; BLANCO, M.; MASPOCH, S.; ORTIZ, M. C.; SÁNCHEZ, M. S.; SARABIA, L. A. Handling intrinsic non-linearity in near-infrared reflectance spectroscopy. **Chemometrics and Intelligent Laboratory Systems**, v. 49, n. 2, p. 215–224, 1999.
- BRERETON, R. G.; JANSEN, J.; LOPES, J.; MARINI, F.; POMERANTSEV, A.; RODIONOVA, O.; ROGER, J. M.; WALCZAK, B.; TAULER, R. Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools. **Analytical and Bioanalytical Chemistry**, v. 409, n. 25, p. 5891–5899, 2017.
- BRERETON, R. G.; LLOYD, G. R. Support Vector Machines for classification and regression. **Analyst**, v. 135, n. 2, p. 230–267, 2010.
- BROWN, S. D. The chemometrics revolution re-examined. **Journal of Chemometrics**, v. 31, n. 1, p. 1–23, 2017.
- CALDAS, C. **Novo Manual para Laboratórios Sucroalcooleiros**. 1. ed. Piracicaba: STAB - Sociedade dos Técnicos Açucareiros e Alcooleiros do Brasil, 2012.
- CHANDA, A.; DALY, A. M.; FOLEY, D. A.; LAPACK, M. A.; MUKHERJEE, S.; ORR, J. D.; REID, G. L.; THOMPSON, D. R.; WARD, H. W. Industry perspectives on process analytical technology: Tools and applications in API development. **Organic Process Research and Development**, v. 19, n. 1, p. 63–83, 2015.
- CHEN, H.-Z.; WEN, J.-B.; CHEN, J.-C.; LI, L.-H.; HUO, Y.-J. Near-infrared spectroscopy modeling optimization for quantitative determination of sugar brix in initial-pressure juice. **International Journal of Technical Research and Applications**, v. 2, n. 6, p. 69–72, 2014.
- CONSECANA. **Manual de Instruções**. 6. ed. Piracicaba: Conselho dos Produtores de Cana-de-Açúcar, Açúcar e Etanol do Estado de São Paulo, 2015.
- CORRÊDO, L. de P.; CANATA, T. F.; MALDANER, L. F.; DE LIMA, J. de J. A.; MOLIN, J. P. Sugarcane Harvester for In-field Data Collection: State of the Art, Its Applicability and Future Perspectives. **Sugar Tech**, , p. 1–14, 2020.

CORTÉS, V.; BLASCO, J.; ALEIXOS, N.; CUBERO, S.; TALENS, P. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. **Trends in Food Science and Technology**, v. 85, p. 138–148, 2019.

FDA. Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Development, manufacturing, and Quality Assurance. Pharmaceutical CGMPs, 2004. Disponível em <<https://www.fda.gov/media/71012/download>>. Acesso em: 25 Jan, 2021.

FEARN, T.; RICCIOLI, C.; GARRIDO-VARO, A.; GUERRERO-GINEL, J. E. On the geometry of SNV and MSC. **Chemometrics and Intelligent Laboratory Systems**, v. 96, n. 1, p. 22–26, 2009.

FERREIRA, A. P.; TOBYN, M. Multivariate analysis in the pharmaceutical industry: Enabling process understanding and improvement in the PAT and QbD era. **Pharmaceutical Development and Technology**, v. 20, n. 5, p. 513–527, 2015.

FERREIRA, M. M. C. **Quimiometria - Conceitos, Métodos e Aplicações**. 1. ed. Campinas - SP: Editora Unicamp, 2015.

FILHO, G. Z.; PICCIRILLI, J. P. **O Processo de Fabricação do Açúcar e Alcool: Desde a Lavoura da Cana até o Produto Acabado**. 1. ed. Santa Cruz do Rio Pardo - SP: Editora Viena, 2012.

GALAVARNA, R.; RIBESSI, R. L.; ROHWEDDER, J. J. R.; PASTRE, J. C. Coupling Continuous Flow Microreactors to MicroNIR Spectroscopy: Ultracompact Device for Facile In-Line Reaction Monitoring. **Organic Process Research and Development**, v. 22, n. 7, p. 780–788, 2018.

GALLAGHER, N. B.; SULLIVAN, D. O. Selection of Representative Learning and Test Sets Using the Onion Method. **Eigenvector**, Manson - WA, 2020.

HUGOT, E. **Manual da Engenharia Açucareira**. 1. ed. São Paulo - SP: Mestre Jou, 1977.
INSTITUTO ADOLFO LUTZ. **Métodos físico-químicos para análise de alimentos**. 4. ed. São Paulo: Instituto Adolfo Lutz, 2008.

LOPES, C. H. **Tecnologia de Produção de Açúcar (Coleção UAB-UFSCar)**. 1. ed. São Carlos: EdUFSCar, 2011.

MALEGORI, C.; NASCIMENTO MARQUES, E. J.; DE FREITAS, S. T.; PIMENTEL, M. F.; PASQUINI, C.; CASIRAGHI, E. Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms. **Talanta**, v. 165, p. 112–116, 2017.

MARAPHUM, K.; CHUAN-UDOM, S.; SAENGPRACHATANARUG, K.; WONGPICHET, S.; POSOM, J.; PHUPHAPHUD, A.; TAIRA, E. Effect of waxy material and measurement position of a sugarcane stalk on the rapid determination of Pol value using a portable near infrared instrument. **Journal of Near Infrared Spectroscopy**, v. 26, n. 5, p. 287–296, 2018.

MARQUES, E. J. N.; DE FREITAS, S. T.; PIMENTEL, M. F.; PASQUINI, C. Rapid and non-destructive determination of quality parameters in the “Tommy Atkins” mango using a novel

- handheld near infrared spectrometer. **Food Chemistry**, v. 197, p. 1207–1214, 2016.
- MEHROTRA, R.; SIESLER, H. W. Application of Mid Infrared/Near Infrared Spectroscopy in Sugar Industry. **Applied Spectroscopy Reviews**, v. 38, n. 3, p. 307–354, 2003.
- MILLER, C. E. Sources of Non-Linearity in near Infrared Methods. **NIR news**, v. 4, n. 6, p. 3–5, 1993.
- NAWI, N. M.; CHEN, G.; JENSEN, T. Visible and Shortwave near Infrared Spectroscopy for Predicting Sugar Content of Sugarcane Based on a Cross-Sectional Scanning Method. **Journal of Near Infrared Spectroscopy**, v. 21, n. 4, p. 289–297, 2013.
- NAWI, N. M.; CHEN, G.; JENSEN, T.; MEHDIZADEH, S. A. Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared. **Biosystems Engineering**, v. 115, n. 2, p. 154–161, 2013.
- OLIVIERI, A. C. **Introduction to Multivariate Calibration: A Practical Approach**. 1. ed. Springer International Publishing, 2018.
- OLSSON, I. M.; GOTTFRIES, J.; WOLD, S. D-optimal onion designs in statistical molecular design. **Chemometrics and Intelligent Laboratory Systems**, v. 73, n. 1, p. 37–46, 2004.
- OSBORNE, B.G.; FEARN, T.; HINDLE, P. H. **Practical NIR Spectroscopy: With Applications in Food and Beverage Analysis**. 2. ed. Harlow: Prentice Hall, 1993.
- OTTO, M. **Chemometrics: Statistics and Computer Application in Analytical Chemistry**. 3. ed. Weinheim, Germany: Wiley-VCH Verlag GmbH &Co, 2017.
- PAIVA, E. M.; ROHWEDDER, J. J. R.; PASQUINI, C.; PIMENTEL, M. F.; PEREIRA, C. F. Quantification of biodiesel and adulteration with vegetable oils in diesel/biodiesel blends using portable near-infrared spectrometer. **Fuel**, v. 160, p. 57–63, 2015.
- PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, p. 8–36, 2018.
- PASQUINI, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198–219, 2003.
- PAYNE, J. H. **Operações unitárias na produção de açúcar da cana**. 1. ed. São Paulo - SP: Nobel Stab, 1989.
- PEREIRA, J. F. Q.; SILVA, C. S.; VIEIRA, M. J. L.; PIMENTEL, M. F.; BRAZ, A.; HONORATO, R. S. Evaluation and identification of blood stains with handheld NIR spectrometer. **Microchemical Journal**, v. 133, p. 561–566, 2017.
- PHUPHAPHUD, A.; SAENGPRACHATANARUG, K.; POSOM, J.; MARAPHUM, K.; TAIRA, E. Prediction of the fibre content of sugarcane stalk by direct scanning using visible-shortwave near infrared spectroscopy. **Vibrational Spectroscopy**, v. 101, p. 71–80, 2019.
- PIERNA, J. A. F.; LECLER, B.; CONZEN, J. P.; NIEMOELLER, A.; BAETEN, V.; DARDENNE, P. Comparison of various chemometric approaches for large near infrared

spectroscopic data of feed and feed products. **Analytica Chimica Acta**, v. 705, n. 1–2, p. 30–34, 2011.

PUIG-BERTOTTO, J.; COELLO, J.; MASPOCH, S. Evaluation of a handheld near-infrared spectrophotometer for quantitative determination of two APIs in a solid pharmaceutical preparation. **Analytical Methods**, v. 11, n. 3, p. 327–335, 2019.

RAMÍREZ-MORALES, I.; RIVERO, D.; FERNÁNDEZ-BLANCO, E.; PAZOS, A. Optimization of NIR calibration models for multiple processes in the sugar industry. **Chemometrics and Intelligent Laboratory Systems**, v. 159, p. 45–57, 2016.

RINNAN, Å.; BERG, F. van den; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC - Trends in Analytical Chemistry**, v. 28, n. 10, p. 1201–1222, 2009.

SALES, R.; DA SILVA, N. C.; DA SILVA, J. P.; FRANÇA, H. H.; PIMENTEL, M. F.; STRAGEVITCH, L. Handheld near-infrared spectrometer for on-line monitoring of biodiesel production in a continuous process. **Fuel**, v. 254, 2019.

SCHÖLKOPF, B.; SMOLA, A. J. **Learning with Kernels – Support Vector Machines, Regularization, Optimization, and Beyond**. Cambridge - MA: MIT Press, 2002.

SEXTON, J.; EVERINGHAM, Y.; DONALD, D.; STAUNTON, S.; WHITE, R. A comparison of non-linear regression methods for improved on-line near infrared spectroscopic analysis of a sugarcane quality measure. **Journal of Near Infrared Spectroscopy**, v. 26, n. 5, p. 297–310, 2018.

SILVA, N. C.; MASSA, A. R. C. G.; DOMINGOS, D.; AMIGO, J. M.; REBOUÇAS, M. V.; PASQUINI, C.; PIMENTEL, M. F. NIR-based octane rating simulator for use in gasoline compounding processes. **Fuel**, v. 243, p. 381–389, 2019.

SIMPSON, R.; NAIDOO, Y. Progress in improving laboratory efficiencies using near infrared spectroscopy (NIRS). **Proc. Int. Soc. Sugar Cane Technol.**, v. 27, p. 1–9, 2010.

SKOOG, D. A.; HOLLER, F. J.; CROUCH, S. R. **Principles of Instrumental Analysis**. 7. ed. Boston - MA: Cengage Learning, 2018.

SOROL, N.; ARANCIBIA, E.; BORTOLATO, S. A.; OLIVIERI, A. C. Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice. A test field for variable selection methods. **Chemometrics and Intelligent Laboratory Systems**, v. 102, n. 2, p. 100–109, 2010.

TAIRA, E.; UENO, M.; FURUKAWA, N.; TASAKI, A.; KOMAKI, Y.; NAGAI, J.; SAENGPRACHATANARUG, K. Networking system employing near infrared spectroscopy for sugarcane payment in Japan. **Journal of Near Infrared Spectroscopy**, v. 21, n. 6, p. 477–483, 2013.

TAIRA, E.; UENO, M.; SAENGPRACHATANARUG, K.; KAWAMITSU, Y. Direct sugar content analysis for whole stalk sugarcane using a portable near infrared instrument. **Journal of Near Infrared Spectroscopy**, v. 21, n. 4, p. 281–287, 2013.

TANGE, R.; RASMUSSEN, M. A.; TAIRA, E.; BRO, R. Application of Support Vector Regression for Simultaneous Modelling of near Infrared Spectra from Multiple Process Steps. **Journal of Near Infrared Spectroscopy**, v. 23, n. 2, p. 75–84, 2015.

TANGE, R.; RASMUSSEN, M. A.; TAIRA, E.; BRO, R. Benchmarking support vector regression against partial least squares regression and artificial neural network: Effect of sample size on model performance. **Journal of Near Infrared Spectroscopy**, v. 25, n. 6, p. 381–390, 2017.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Validation of multivariate calibration models in the determination of sugar cane quality parameters by near infrared spectroscopy. **Journal of the Brazilian Chemical Society**, v. 18, n. 2, p. 259–266, 2007.

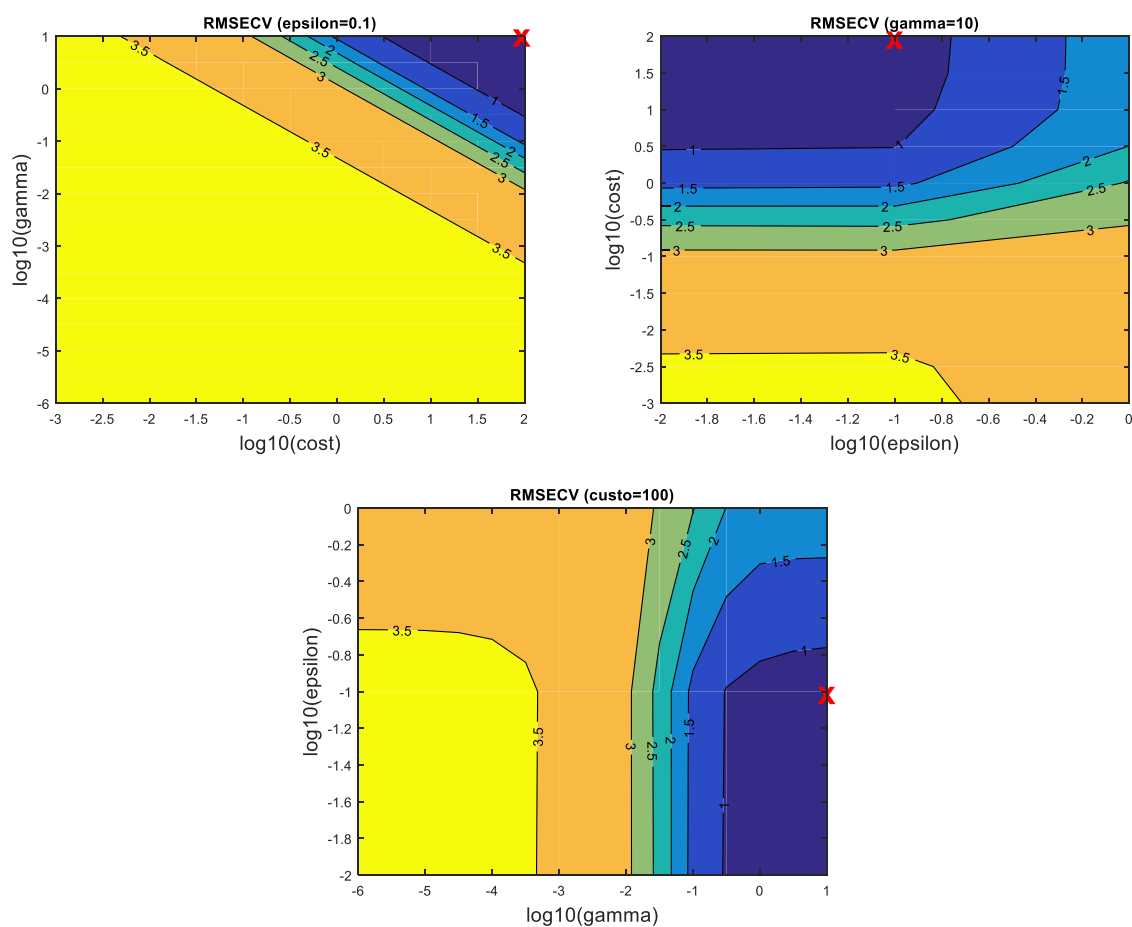
WORKMAN, J.; WEYER, L. **Practical Guide to Interpretive Near-Infrared Spectroscopy**. 1. ed. New York: CRC Press: Taylor & Francis Group, 2007.

ZANARDI, M. S.; JUNIOR, E. F. C. Tecnologia e perspectiva da produção de etanol no Brasil. **Revista Liberato**, v. 17, n. 27, p. 20–34, 2016.

ZANCONATO, A. G.; BUKVIC, G. Estudo de caso: Viabilidade econômica da utilização de um robô aplicador de chapisco nos frisos dos rolos de moenda. *In*: ANDRADE, D. F. (org.). **Gestão da Produção em Foco - Volume 3**. 1. ed. Belo Horizonte - MG: Poisson, 2018. v. 3, p. 39–47.

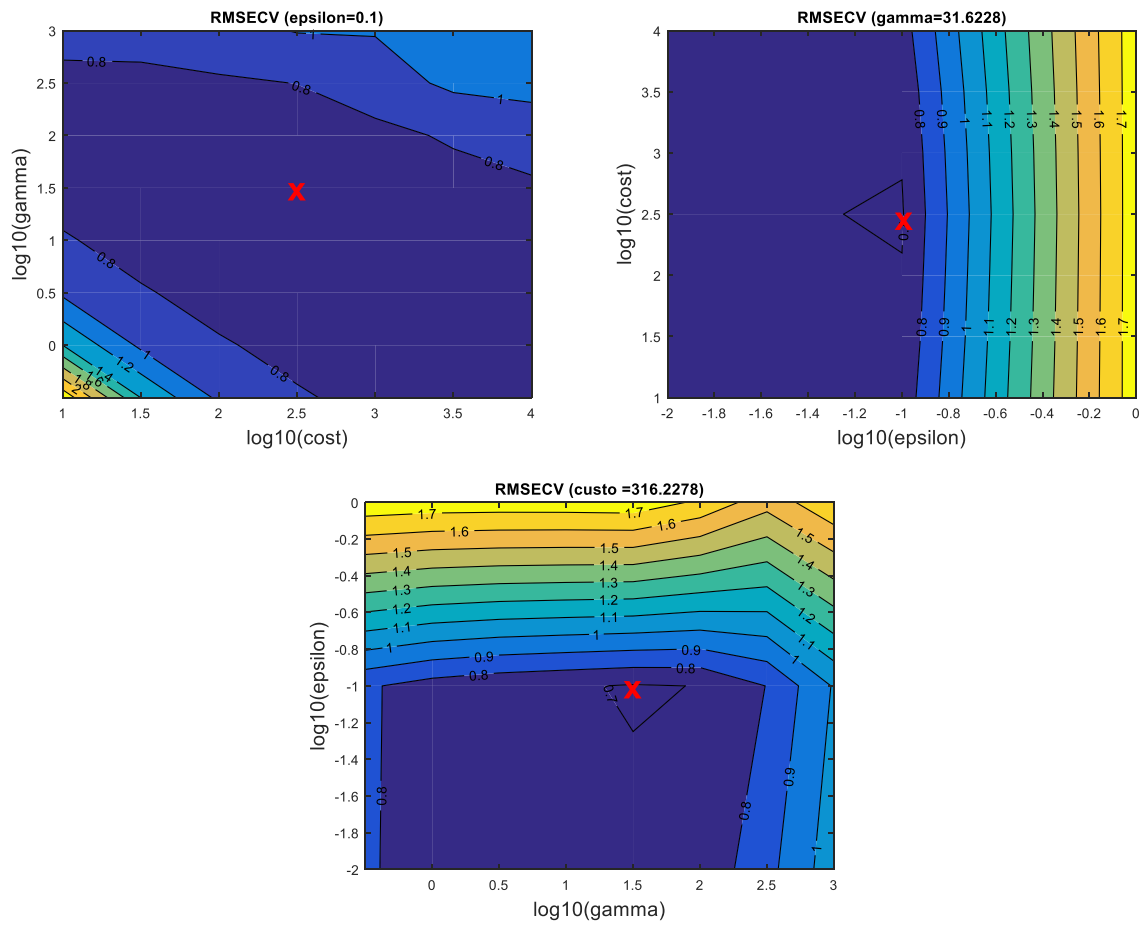
APÊNDICE A - SUPERFÍCIES DE RESPOSTA DOS PARÂMETROS DE AJUSTE SVM PARA OS MODELOS PREDITIVOS DE POL POR TRANSMITÂNCIA E BRIX E POL POR REFLECTÂNCIA DIFUSA

Figura A1 – Superfícies de resposta da primeira varredura para ajuste de parâmetros SVM para o modelo preditivo de pol por transmitância.



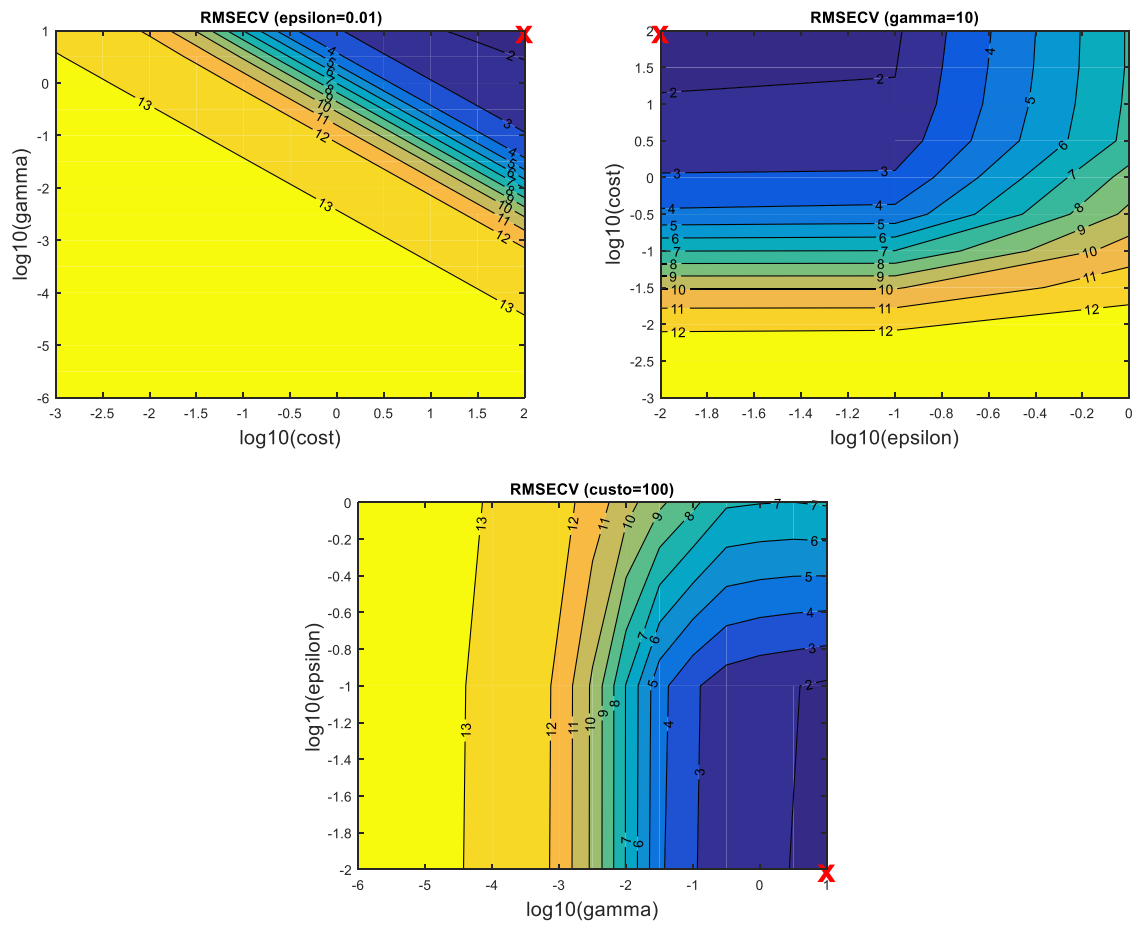
Fonte: O autor (2021)

Figura A2 – Superfícies de resposta da segunda varredura para ajuste de parâmetros SVM para o modelo preditivo de pol por transmitância.



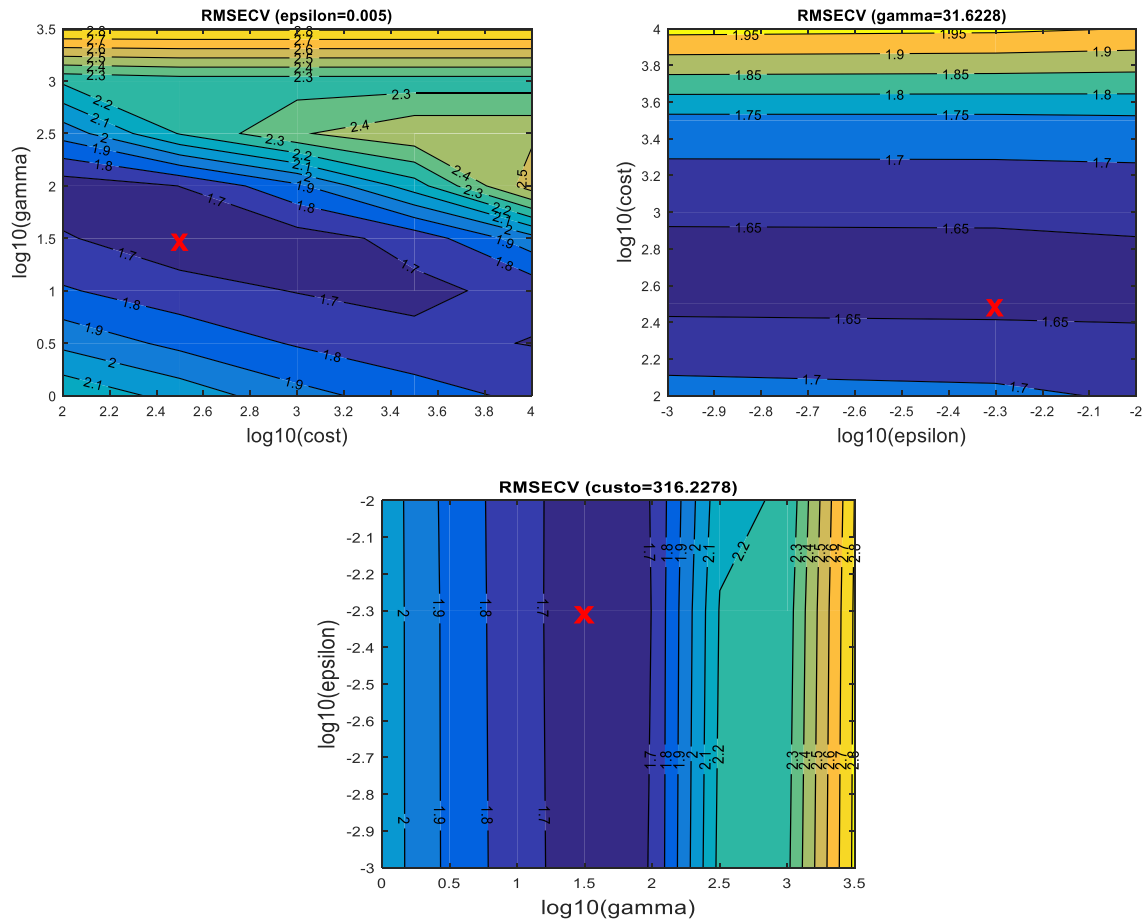
Fonte: O autor (2021)

Figura A3 – Superfícies de resposta da primeira varredura para ajuste de parâmetros SVM para o modelo preditivo de brix por reflectância difusa.



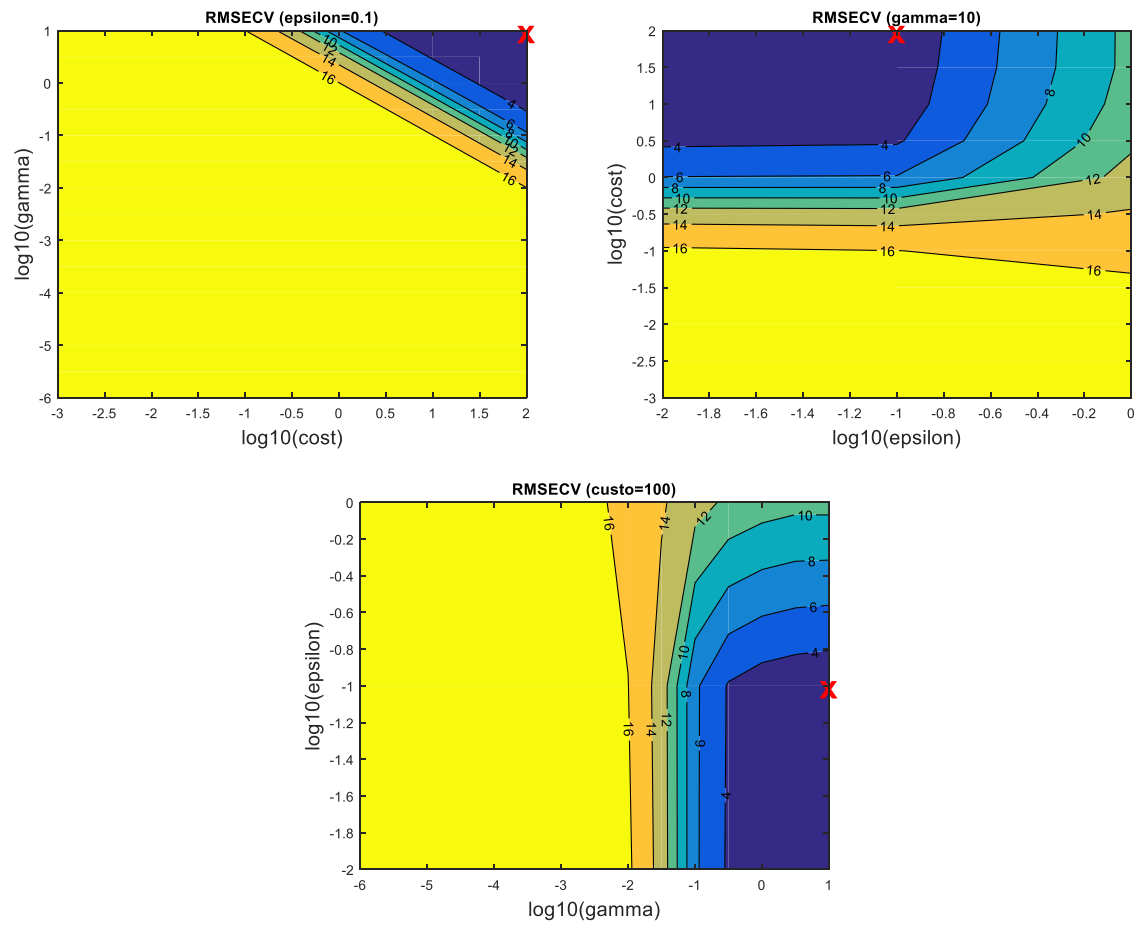
Fonte: O autor (2021)

Figura A4 – Superfícies de resposta da segunda varredura para ajuste de parâmetros SVM para o modelo preditivo de brix por reflectância difusa.



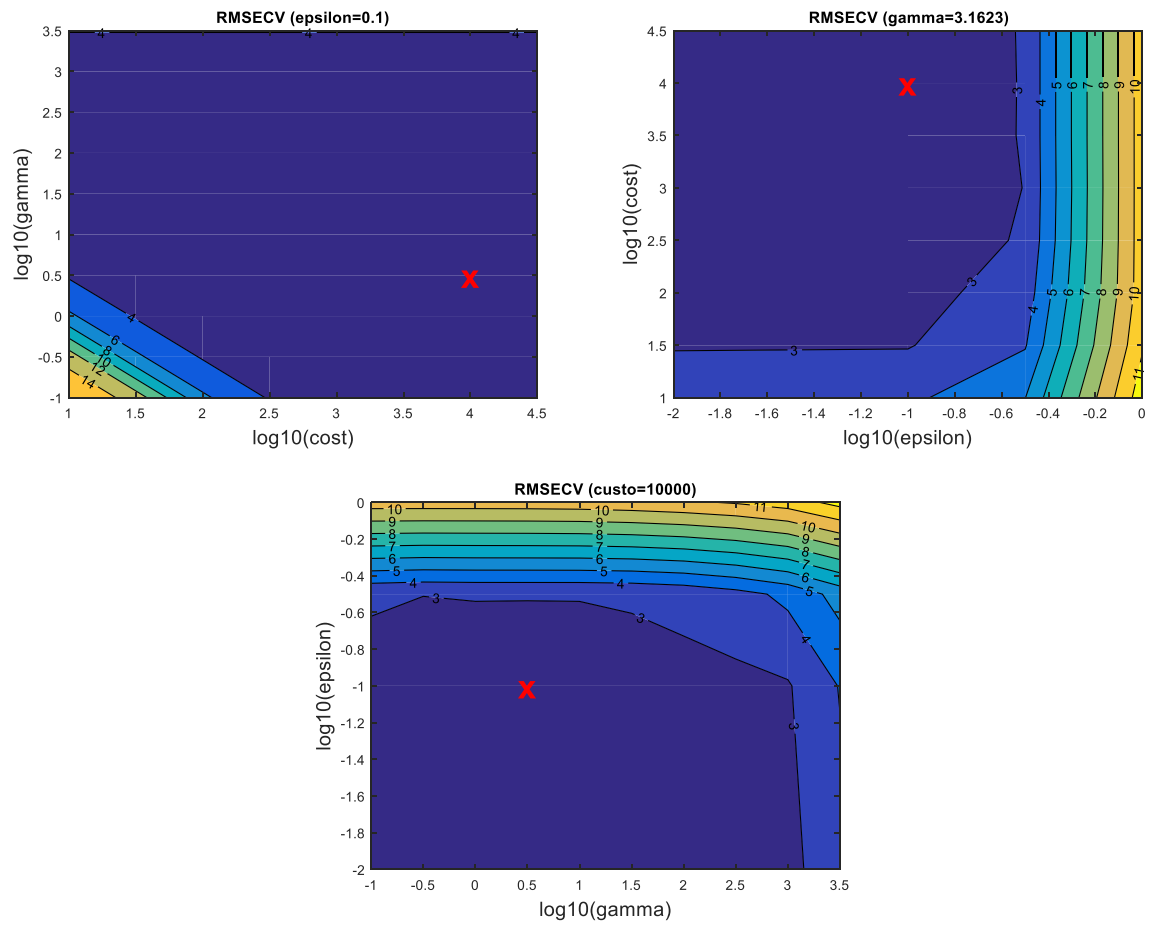
Fonte: O autor (2021)

Figura A5 – Superfícies de resposta da primeira varredura para ajuste de parâmetros SVM para o modelo preditivo de pol por reflectância difusa.



Fonte: O autor (2021)

Figura A6 – Superfícies de resposta da segunda varredura para ajuste de parâmetros SVM para o modelo preditivo de brix por reflectância difusa.



Fonte: O autor (2021)