

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

**UTILIZAÇÃO DE TÉCNICAS ESPECTROANALÍTICAS PARA
AVALIAÇÃO DE AÇÚCAR DE CANA-DE-AÇÚCAR**

Douglas Romano Beletti*

Dissertação apresentada como parte dos requisitos para obtenção do título de MESTRE PROFISSIONAL EM QUÍMICA, área de concentração: QUÍMICA TECNOLÓGICA.

Orientador (a): Prof. Dr. Edenir Rodrigues Pereira Filho

*** CETEC Equipamentos para Laboratórios**

**São Carlos - SP
2021**



UNIVERSIDADE FEDERAL DE SÃO CARLOS

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

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Douglas Romano Beletti, realizada em 16/06/2021.

Comissão Julgadora:

Prof. Dr. Edenir Rodrigues Pereira Filho (UFSCar)

Prof. Dr. Eduardo Poggi e Borges (Fermentec Ltda.)

Prof. Dr. Felipe Manfroi Fortunato (Laboratório Sinergia)

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

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

Dedico este trabalho àqueles que já se foram.

Ao meu avô, **SALIM**, por todo seu amor. Minha avó, **CÉLIA**, por toda sua força. E ao meu melhor amigo, **TIAGO**, por toda sua amizade.

Hoje pela manhã eu agradei. Acordei, levantei, me preparei para ir trabalhar e agradei. Pois em tudo que os meus braços alcançavam Ele me guiou e deu forças para que eu fizesse, e aquilo que meus braços não alcançavam Ele fez por mim.

AGRADECIMENTOS

Primeiramente agradeço à Deus, por ter me dado saúde, sabedoria e disposição sempre guiando meus passos, fazendo que todo esse trabalho fosse possível.

Aos meus pais, Sérgio e Cristiane, por todo seu amor e dedicação e por sempre terem acreditado que a educação seria uma das coisas mais preciosas que poderiam nos dar, sem eles eu não teria chego até aqui.

Aos meus irmãos, Letícia e Leonardo, que são parte de mim. Meus avós, Humberto e Dalva que são meus exemplos de vida, e familiares próximos que sempre estiveram presentes no meu dia a dia, por todo apoio e incentivo durante essa jornada.

À minha noiva, Maísa, pelo amor, carinho, companheirismo e por estar ao meu lado nos momentos mais difíceis. Também sua família pelo suporte.

À CETEC, por seus diretores Dorival e Wagner que acreditaram no meu potencial e me deram todas as oportunidades de crescimento pessoal e profissional, além do suporte e incentivo para realizar esse trabalho. Aos gerentes Anderson e Alex pelo suporte e incentivo. E também ao Vinícius pelo compartilhamento dos conhecimentos e companheirismo diário no setor de Aplicação.

Ao Prof. Edenir Rodrigues Pereira Filho, pela orientação, paciência e amizade. Que em um momento tão difícil como uma pandemia teve a resiliência e dedicação de manter todos seus alunos produtivos, a empatia para entender os momentos difíceis, e a sensibilidade para nos levar em frente permitindo que esse trabalho fosse concluído.

Ao grupo GAIA, Departamento de Química e ao programa de pós-graduação em Química da Universidade Federal de São Carlos, através de seus alunos, professores e funcionários que deram suporte, estrutura, propiciaram a realização deste Mestrado Profissional, fazendo com que a elaboração de projetos em conjunto com as empresas fosse possíveis.

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

Por fim, a todos os amigos e pessoas que de alguma forma colaboraram com este projeto e meu crescimento pessoal e profissional.

Muito obrigado.

LISTA DE ABREVIATURAS

ANN	<i>Artificial Neural Networks</i>
ICUMSA	<i>International Commission for Uniform Methods of Sugar Analysis</i>
INTCPT	Intercept
LIBS	<i>Laser Induced-Breakdown Spectroscopy</i>
LV	<i>Latent Variables</i>
Max	Valor Máximo
Min	Valor Mínimo
MLR	<i>Multiple Linear Regression</i>
N	Número de Amostras
NIRS	<i>Near Infrared Spectroscopy</i>
PCA	<i>Principal Component Analysis</i>
PLS	<i>Partial Least Squares</i>
PP	Pontos Pretos
RH	<i>Relative Humidity</i>
RMSEC	<i>Root Mean Square Error of Calibration</i>
RMSECV	<i>Root Mean Square Error of Cross-Validation</i>
RMSEP	<i>Root Mean Square Error of Prediction</i>
SEL	<i>Standard Error of Laboratory</i>
RSQ	<i>R-Squared (Coeficiente de Determinação)</i>
RSQVal	<i>R-Squared Validation</i>
SD	<i>Standard Deviation</i>
SNV	<i>Standard Normal Variate</i>
VHP	<i>Very High Polarization</i>
VVHP	<i>Very Very High Polarization</i>

ÍNDICE DE TABELAS

Tabela 1. Parâmetros de Qualidade do Açúcar do Grupo II (Vigente a partir de 1º de julho de 2019).....	11
Tabela 2. Descrições para quatro tipos de métodos de cross-validação com representação gráfica.....	18
Tabela 3. Doehlert design usado para determinar as melhores condições operacionais do LIBS para os experimentos.....	28
Tabela 4. Descrição dos 12 modos de normalizações aplicados.....	30
Tabela 5. Condições experimentais definidas para elaboração dos experimentos LIBS.....	31
Tabela 6. Avaliação dos modelos PLS para as 12 normalizações no parâmetro Areia% (N=61).....	32
Tabela 7. Avaliação dos modelos PLS para as 12 normalizações no parâmetro Açúcar% (N=61).....	33
Tabela 8. Avaliação dos modelos PLS para as 12 normalizações no parâmetro PP% (N=61).....	33
Tabela 9. Descrição dos critérios de seleção de variáveis para otimização dos modelos.....	36
Tabela 10. Avaliação dos modelos PLS para seleção de variáveis no parâmetro Açúcar%.....	37
Tabela 11. Avaliação dos modelos PLS para seleção de variáveis no parâmetro Areia%.....	37
Tabela 12. Avaliação dos modelos PLS para seleção de variáveis no parâmetro PP (pp/100g).....	37
Tabela 13. Avaliação dos modelos PLS e mPLS para o parâmetro Areia%.....	38
Tabela 14. Avaliação dos modelos PLS e mPLS para o parâmetro Açúcar%.....	38
Tabela 15. Avaliação dos modelos PLS e mPLS para o parâmetro PP(pp/100g).....	39
Tabela 16. Comparação dos modelos PLS obtidos através da fusão de dados com individuais.....	39
Tabela 17. Comparação dos modelos utilizando grupos de calibração e validação.....	40
Tabela 18. Resultados previstos para novas misturas.....	41
Tabela 19. Resultados de Areia (%) previstos nos modelos NIR para diferentes tipos de açúcar.....	42
Tabela 20. p-valores obtidos através do teste de Wilcoxon para modelos com seleção de variáveis.....	45
Tabela 21. p-valores obtidos através do teste de Wilcoxon para diferentes modelos a partir de algoritmos diferentes.....	45
Tabela 22. Quantidade de amostras por tipo de açúcar.....	48
Tabela 23. Avaliação para o parâmetro Cor ICUMSA (UI).....	51
Tabela 24. Avaliação para o parâmetro Polarização (°Z).....	51

Tabela 25. Avaliação para o parâmetro Cinzas (% m/m)	52
Tabela 26. Parâmetros de calibração para modelos mPLS.	53
Tabela 27. Parâmetros de calibração para modelos ANN.	54
Tabela 28. Estatísticas dos grupos de calibração e validação de Açúcar VHP	56
Tabela 29. Desempenho de calibração e validação para Açúcar VHP	56
Tabela 30. Estatísticas dos grupos de calibração e validação de Açúcar Cristal	58
Tabela 31. Desempenho de calibração e validação para Açúcar Cristal	58
Tabela 32. Estatísticas dos grupos de calibração e validação de Açúcar Ref. Amorfo	60
Tabela 33. Desempenho de calibração e validação para Açúcar Refinado Amorfo ..	60
Tabela 34. Estatísticas dos grupos de calibração e validação de Açúcar Refinado Granulado.....	62
Tabela 35. Desempenho de calibração e validação para Açúcar Refinado Granulado	62
Tabela 36. Estatísticas dos grupos de calibração e validação para dados Globais ..	64
Tabela 37. Desempenho de calibração e validação para dados globais de Açúcar..	64
Tabela 38. Desempenho dos modelos globais nos tipos de Açúcar individuais	66

ÍNDICE DE FIGURAS

Figura 1. Fluxograma de produtos e subprodutos da cadeia da cana-de-açúcar.....	3
Figura 2. Açúcar bruto (esquerda) e areia (direita).....	4
Figura 3. Esquema típico da instrumentação LIBS	12
Figura 4. Espectro eletromagnético com faixas de comprimento de onda do UV até MIR.....	13
Figura 5. Esquema instrumentação NIRS	14
Figura 6. Representação gráfica das matrizes e vetores usados na PLS.	20
Figura 7. ANN simplificado onde cada círculo representa um neurônio; neste exemplo, três neurônios ocultos são usados e apenas quatro entradas (representando variáveis espectrais) com uma saída.	22
Figura 8. Função de transferência não-linear – função tangente hiperbólica sigmoidal.	23
Figura 9. Diagrama de mistura ternária nas concentrações planejadas.....	27
Figura 10. (Esquerda) LIBS modelo J200 Applied Spectra (Direita) Pastilhas preparadas para leitura no instrumento LIBS.	28
Figura 11. Amostra de açúcar na célula de leitura Small Cup.....	29
Figura 12. Espectros das amostras de açúcar obtido pelo LIBS.....	31
Figura 13. Espectros das amostras de açúcar obtido pelo NIRS.	31
Figura 14. Valores reais versus os previstos para as 3 normalizações escolhidas nos parâmetros Areia (%), Açúcar (%) e Pontos Pretos (PP/100g).	34
Figura 15. Intensidade dos vetores de regressão obtidos através dos modelos calculados nas normalizações 1 (LV=3) e 6 (LV=2).	35
Figura 16. Regressão da calibração (círculos azuis) e validação (triângulos vermelhos) para os modelos otimizados na técnica LIBS, NIRS e Fusão de dados nos parâmetros Areia (%), Açúcar (%) e Pontos Pretos (PP/100g).	41
Figura 17. Gráficos de barras com valores de RMSEP comparativos para os parâmetros Areia (%), Açúcar (%) e PP (pp/100g) nos modelos obtidos via LIBS, NIRS e Fusão de dados.	43
Figura 18. Gráfico de dispersão Cor ICUMS versus Areia a partir dos resultados das amostras de açúcar VHP da Tabela 19.....	44
Figura 19. Distribuição regional dos instrumentos para coleta de açúcar na região centro-sudeste do Brasil.....	47
Figura 20. (Esquerda) NIRS modelo DS2500 da fabricante FOSS. (Direita) Amostra de açúcar na célula de leitura Large Cup.....	53
Figura 21. Espectros de açúcar utilizando pré-tratamento de SNV (<i>only scale</i>) e <i>detrend</i> com derivativos em 1, 16, 16, 1.	54
Figura 22. Espectros de Açúcar obtidos pelo NIRS DS2500	55

Figura 23. Regressão para os modelos mPLS, calibração (círculo azul) e validação (triângulo laranja) para parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol no açúcar VHP.	57
Figura 24. Regressão para os modelos ANN calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol no açúcar VHP.	57
Figura 25. Regressão para os modelos mPLS, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol no açúcar Cristal.	59
Figura 26. Regressão para os modelos ANN, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol no açúcar Cristal.	59
Figura 27. Regressão para os modelos mPLS, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol no açúcar Refinado Amorfo.	61
Figura 28. Regressão para os modelos ANN, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cor no açúcar Refinado Amorfo.	61
Figura 29. Regressão para os modelos mPLS, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol no açúcar Refinado Granulado.	63
Figura 30. Regressão para os modelos ANN, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cor e (B) Umidade no açúcar Refinado Granulado.	63
Figura 31. Regressão para os modelos mPLS, , calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol para dados globais de açúcar.	65
Figura 32. Regressão para os modelos ANN, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol para dados globais de açúcar.	65
Figura 33. Gráfico de barras comparativo dos valores de RMSEP dos modelos para Cor ICUMSA.	67
Figura 34. Gráfico de barras comparativo dos valores de RMSEP dos modelos para Cinzas.	67
Figura 35. Gráfico de barras comparativo dos valores de RMSEP dos modelos para Umidade.	68
Figura 36. Gráfico de barras comparativo dos valores de RMSEP dos modelos para Pol.	68
Figura 37. Gráfico de barras comparativo dos valores de RMSEP entre os modelos, SEL e R obtidos pelo ensaio de proficiência para o parâmetro Cor ICUMSA.	69
Figura 38. Gráfico de barras comparativo dos valores de RMSEP entre os modelos, SEL e R obtidos pelo ensaio de proficiência para o parâmetro Cinzas.	69
Figura 39. Gráfico de barras comparativo dos valores de RMSEP entre os modelos, SEL e R obtidos pelo ensaio de proficiência para o parâmetro Pol.	70

Figura 40. Gráfico de barras comparativo dos valores de RMSEP entre os modelos e literatura para Cor ICUMSA.....	71
Figura 41. Gráfico de barras comparativo dos valores de RMSEP entre os modelos e literatura para Cinzas.	71
Figura 42. Gráfico de barras comparativo dos valores de RMSEP entre os modelos e literatura para Umidade.....	72
Figura 43. Gráfico de barras comparativo dos valores de RMSEP entre os modelos e literatura para Pol.	72
Figura 44. Gráfico de dispersão laboratório versus previsto para o parâmetro cinzas.	73
Figura 45. Gráficos de dispersão laboratório versus previsto para Cor ICUMSA (UI) utilizando os dados globais nos modelos mPLS (esquerda) e ANN (direita).	73
Figura 46. (a) PC1 e PC2 para Açúcar Cristal (Verde), Açúcar VHP (Amarelo), Açúcar Granulado (Azul) e Açúcar Amorfo (Roxo) (b) Loading #1 da PCA	74

RESUMO

UTILIZAÇÃO DE TÉCNICAS ESPECTROANALÍTICAS PARA AVALIAÇÃO DE AÇÚCAR DE CANA-DE-AÇÚCAR

O Brasil é o maior produtor de cana-de-açúcar no mundo, e sua indústria consolidada também o coloca em posição de maior produtor e exportador do açúcar. Com um processo bastante variável e também com grandes volumes transportados, o controle de adulterações e análise dos seus parâmetros físico-químicos se torna etapa crucial em diversos pontos da sua cadeia de suprimento, estabelecendo os padrões de qualidade e a demanda fim do mesmo. Dada a importância do controle, no âmbito laboratorial, busca-se novas técnicas e aplicações que possibilitem análises rápidas, não destrutivas, sem emprego de reagentes químicos e de fácil operação, possibilitando assim tomadas rápidas de ação que conduzam a melhorias na eficiência dos processos e diminuição das perdas. Desse modo, o objetivo deste trabalho é a avaliação e desenvolvimento de novas aplicações para análise do açúcar utilizando as técnicas de Espectrometria de Emissão Óptica com Plasma Induzido por Laser (LIBS) e Espectroscopia na Região do Infravermelho Próximo (NIRS). Na primeira parte deste estudo, 64 amostras de mistura contendo açúcar, areia e pontos pretos (pp) foram preparadas, e modelos quimiométricos foram desenvolvidos e otimizados para determinação dos parâmetros Açúcar (%), Areia (%) e Pontos Pretos (pp/100g). Para avaliação do desempenho foram obtidos os valores de RMSEP dos modelos de previsão para Areia em 6,48, 1,29 e 5,96%, para Açúcar em 6,53, 1,23 e 5,99% e para Pontos Pretos em 7,74, 5,81 e 10,94 pp/100g de amostra, sendo cada dos valores obtidos com os dados do LIBS, NIR e Fusão de Dados, respectivamente. Na segunda parte desse estudo dados espectroanalíticos de 41.110 amostras de açúcares tipo Cristal, VHP, Refinado Amorfo e Refinado Granulado foram coletados via NIRS em diferentes regiões do Brasil e modelos de previsão mPLS e ANN foram desenvolvidas para os parâmetros Cinzas (% m/m), Cor ICUMSA (UI), Polarização (°Z) e Umidade (%). Modelos individuais mPLS e ANN foram comparados com modelos globais (contendo todos os tipos de açúcar), com a literatura e também com teste de proficiência aplicado em laboratórios do setor. Para os modelos ANN da calibração Global foi possível obter valores de RMSEP de 0,017 para cinzas, 43 para cor, 0,016 para umidade e 0,098 para Pol. Portanto, a técnica LIBS se mostrou promissora na avaliação de contaminantes no açúcar, podendo ser avaliada para desenvolvimento de aplicações de rotina com maior variabilidade espectral. A técnica NIRS para avaliação do açúcar demonstrou avançado estado de desenvolvimento e robustez pela variabilidade global e desempenho quando comparados com literatura e técnicas de laboratório. Por fim, a implementação de modelos globais ANN conferem maior estabilidade, robustez e precisão apresentando diversas vantagens sobre os modelos locais e modelos mPLS.

ABSTRACT

USE OF SPECTROANALYTICAL TECHNIQUES FOR THE EVALUATION OF SUGARCANE SUGAR.

Brazil is the largest sugarcane producer in the world, and its consolidated industry also places it in the position of the largest producer and exporter of sugar. With a highly variable process and also with large transported volumes, the control of tampering and analysis of its physical chemical parameters becomes a crucial step at various points in its supply chain, establishing quality standards and its final demand. Given the importance of control, in the laboratory scope, new techniques and applications are sought that enable quick, non-destructive analysis, not using harmful reagents and also easy to operate, and that this allows for quick actions that lead to improvements in the efficiency of the processes decrease in losses. Thus, the objective of this work is the evaluation and development of new applications for sugar analysis using the Laser Induced Plasma Optical Emission Spectrometry (LIBS) and Near Infrared Region Spectroscopy (NIRS) techniques. In the first part of this study 64 samples of mixture were prepared, and using chemometric tools, PLS models were developed and optimized to determine the parameters Sugar (%), Sand (%) and Black Spots (pp/100g). For performance evaluation, the RMSEP values were obtained for the prediction models for Sand at 6.48, 1.29 and 5.96 percent, for Sugar at 6.53, 1.23 and 5.99 percent and for Black Spots at 7.74, 5.81 and 10.94 bp/100g of sample, each of which is obtained with the LIBS, NIR and Data Fusion data respectively. In the second part of this study, spectroanalytical data from 41,110 samples of Crystal, VHP, Amorphous Refined and Granulated Refined sugars were collected via NIRS in different regions of Brazil and mPLS and ANN prediction models were developed for Ash (% m/m), Color ICUMSA (UI), Polarization (°Z) and Humidity (%). Individual mPLS and ANN models were compared with a global model (containing all types of sugar), with the literature and also with a proficiency test applied in industry laboratories. For the ANN models of the Global calibration, it was possible to obtain RMSEP values of 0.017 for ash, 43 for color, 0.016 for moisture and 0.098 for Pol. Therefore, the LIBS technique showed promise in the evaluation of contaminants in sugar, and can be evaluated for development of routine applications obtaining greater spectral variability. The NIRS technique for sugar evaluation demonstrated an advanced state of development and robustness due to global variability and performance when compared to literature and laboratory techniques. Finally, the implementation of global ANN models provide greater stability, robustness and precision, having several advantages over local models and mPLS models.

SUMÁRIO

1	INTRODUÇÃO.....	2
2	OBJETIVOS.....	8
2.1	Justificativas	8
3	REVISÃO BIBLIOGRÁFICA	10
3.1	Classificação do Açúcar	10
3.2	LIBS (Laser-Induced Breakdown Spectroscopy)	12
3.3	NIRS (Near Infrared Spectroscopy).....	13
3.4	Quimiometria e Calibração Multivariada.....	14
3.4.1	RMSEC (<i>Root Mean Square Error of Calibration</i>) e RMSEP (<i>Root Mean Square Error of Prediction</i>).....	16
3.4.2	Cross-Validação e RMSECV (<i>Root Mean Square of Cross Validation</i>)	17
3.4.3	Pré-tratamentos	19
3.4.4	<i>Partial Least Squares</i> (PLS).....	19
3.4.5	Modified Partial Least Squares (mPLS)	21
3.4.6	Artificial Neural Networks (ANN)	21
4	ESTUDO DE IDENTIFICAÇÃO E QUANTIFICAÇÃO DE ADULTERANTES NO AÇÚCAR EMPREGANDO LIBS E NIRS.....	26
4.1	Experimental.....	26
4.1.1	Obtenção das amostras	26
4.1.2	Preparo e obtenção dos dados espectroanalíticos LIBS.....	27
4.1.3	Preparo e obtenção de dados espectroanalíticos NIRS.....	29
4.1.4	Tratamento dos dados LIBS.....	29
4.2	Resultados e Discussões	30
4.2.1	Cálculo dos modelos iniciais	32
4.2.2	Seleção de Variáveis e Otimização dos Modelos	35
4.2.3	Avaliação de diferentes matrizes	41
4.2.4	Avaliação dos modelos utilizando Wilcoxon Signed-Rank Test para amostras pareadas.....	44
5	DESENVOLVIMENTO DE MODELOS DE PREVISÃO PARA QUANTIFICAÇÃO DE PARÂMETROS DO CONTROLE DE QUALIDADE DO AÇÚCAR EMPREGANDO NIRS.....	47
5.1	Experimental.....	47
5.1.1	Obtenção das Amostras.....	47
5.1.2	Métodos de referência.....	48
5.1.3	Avaliação das Metodologias de Referência	49
5.1.4	Preparo e obtenção dos dados espectroanalíticos	52

5.1.5	Fatores, Cross-Validação e Pré-tratamentos	53
5.2	Resultados e Discussões	55
5.2.1	Cálculo dos modelos de calibração.....	55
5.2.1.1	Açúcar VHP	56
5.2.1.2	Açúcar Cristal.....	58
5.2.1.3	Açúcar Refinado Amorfo	60
5.2.1.4	Açúcar Refinado Granulado.....	62
5.2.1.5	Açúcar Global	64
5.2.2	Desempenho comparativo dos modelos	67
6	CONSIDERAÇÕES FINAIS.....	76
7	REFERÊNCIAS	79

Capítulo 1
INTRODUÇÃO

1 INTRODUÇÃO

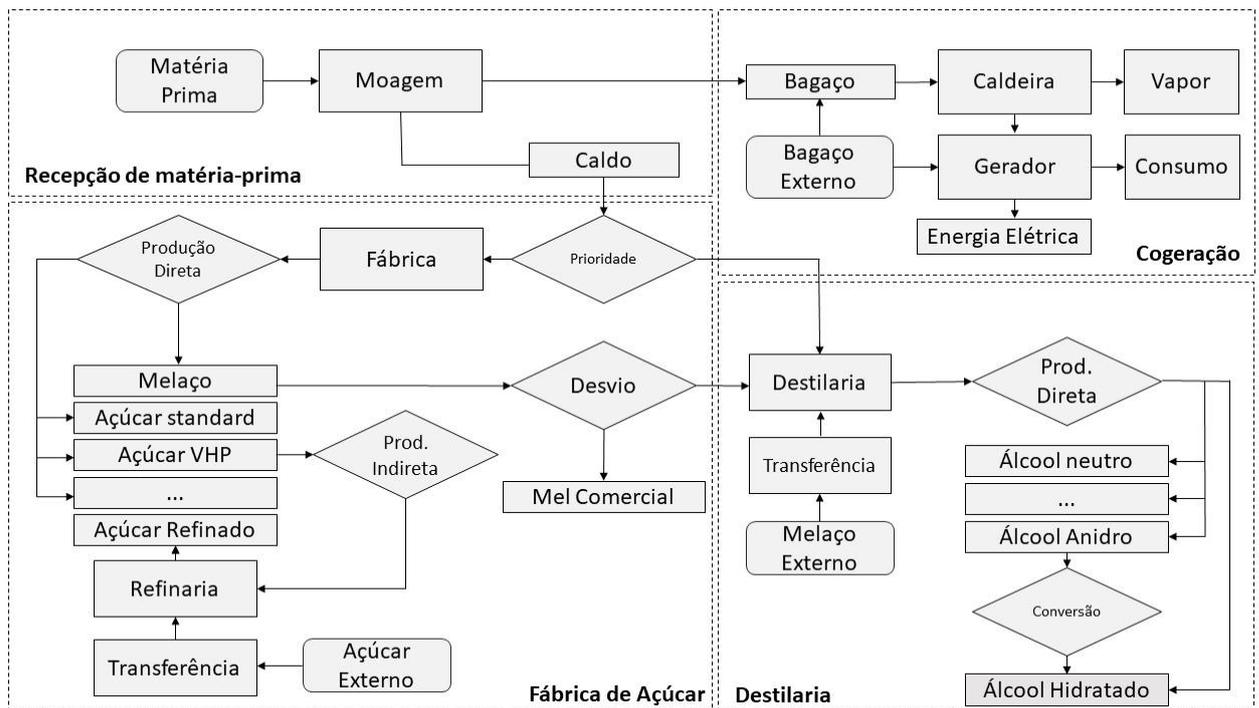
Fundada em 1993 a CETEC Equipamentos para Laboratório iniciou suas atividades com a prestação de serviços de manutenção e calibração de instrumentos analíticos. Em 1995 a CETEC participou do primeiro teste da tecnologia NIRS (*Near Infrared Spectroscopy*) realizado em uma Usina de cana-de-açúcar no Brasil. O referido teste foi executado na Usina São Martinho, maior Usina de Açúcar e Álcool do Brasil e do mundo, e a CETEC participou dos testes representando a fabricante Bran+Luebbe com o instrumento InfraAlyzer 450. Atualmente a CETEC é representante exclusiva para o setor bioenergético e demais seguimentos, das fabricantes Seal Analytical, ATAGO, Shimadzu e FOSS nas linhas de instrumentos SFA (*Segmented Flow Analyzers*), Analisadores Discretos, Polarímetros, Refratômetros, Espectrofotômetros, Cromatografia e Infravermelho Próximo. Ao longo da sua história se consolidou como prestadora de serviço nesse setor e hoje é considerada uma das grandes provedoras de novas tecnologias e desenvolvimento relacionado à instrumentação analítica para o setor sucroalcooleiro.

O Brasil é o maior produtor mundial de cana-de-açúcar, com uma participação de 39%, com cerca de 654 milhões de toneladas processadas na última safra (NEVES, 2017; ÚNICA, 2018; CONAB, 2021).

Na safra 20/21 foram 367 unidades produtoras em atividade no país. O Brasil é o maior produtor e exportador de açúcar do mundo, com 41,2 milhões de toneladas produzidas e 32,2 milhões exportadas no ciclo 2020/2021. Esses valores são equivalentes a 20% produção e 45% da exportação mundial, respectivamente. Além disso, é o segundo maior produtor global de etanol (ranking liderado pelos Estados Unidos). Na safra 2020/2021, o volume produzido atingiu 29,7 bilhões de litros (ÚNICA, 2018; CONAB, 2021)

O processo de produção sucroalcooleiro possui diferentes cadeias geradores de subprodutos e produtos, originando assim mais de 30 diferentes matrizes e precisam ser analisados mais de 100 parâmetros (orgânicos, inorgânicos e físico-químicos). Mais detalhes podem ser visualizados no esquema ilustrado na Figura 1.

Figura 1. Fluxograma de produtos e subprodutos da cadeia da cana-de-açúcar.



Fonte: Autor

Como observado na Figura 1, a cadeia de produção da cana-de-açúcar permite a avaliação de diferentes matrizes. Tais avaliações podem ser realizadas com o emprego de distintas técnicas espectroanalíticas devido a sua abrangência em parâmetros orgânicos e inorgânicos do processo. Desta forma, como matriz de trabalho desta dissertação de Mestrado Profissional foi escolhida a análise do açúcar, motivado pelas seguintes razões: (1) possibilidade de obtenção dos dados espectrais em conjunto com resultados de referência para proposição das calibrações com boa qualidade analítica (2) perspectivas de oportunidades e geração de valor, baseando-se na produção, rotina analítica, quantidade de análises por dia e tempo empregado na determinação dos diferentes parâmetros desse produto, diferentes pontos de amostragem (produção, armazenamento, transporte, transbordo) e (3) aplicabilidade das diferentes técnicas espectroanalíticas.

Os parâmetros físico-químicos do açúcar são muito importantes comercialmente e para etapa de produção, estabelecendo os padrões de qualidade e a demanda fim do mesmo. Dentro os diferentes parâmetros de importância, quatro deles são cruciais para classificação do produto, sendo: Cor (UI ICUMSA), Pol (°Z), Umidade (%) e Cinzas Condutimétricas (%) (MAPA, 2019).

A unidade fabril é responsável por gerar o laudo de seu produto, porém o mesmo pode passar por diferentes pontos de transbordo até chegar ao destino final.

Além disso, cada ponto de recebimento tem que garantir que o produto que adentra e sai de seu domínio tenha as especificações recomendadas pela legislação. Nesses momentos são cruciais as verificações, pois durante o transporte de um ponto intermodal ao outro o produto pode passar por adulterações. Não sendo incomuns atividades criminosas nesse sentido, conforme descrita em matéria intitulada “Açúcar vira areia com conivência de policiais” publicada pelo jornal O Estado de São Paulo em 2010. Neste caso, a areia, pela sua similaridade visual (Figura 2) com o açúcar bruto, é utilizada em atividades ilícitas. (TAVARES, 2010)

Figura 2. Açúcar bruto (esquerda) e areia (direita).



Fonte: Google Images

No contexto da indústria 4.0, nome que é dado para a atual tendência de automação e troca de dados em tecnologia de fabricação. A automação já é um tema recorrente e motivo da chamada “terceira revolução industrial”, que se dá na inclusão de sistemas ciber-físicos, internet das coisas, computação em nuvem e inteligência artificial. No âmbito laboratorial, em conjunto com os fatores citados anteriormente, busca-se técnicas capazes de realizar análises em um curto intervalo de tempo para controle de processos. Assim, propicia tomadas rápidas de ação e que conduzam a melhorias na eficiência e diminuição de perdas no processo, fornecendo condições para atender os requisitos de operacionalidade, aumento da produtividade analítica, diminuição dos insumos químicos e redução de riscos aos colaboradores (O’DONNELL, 2014).

Considerando as técnicas com capacidade de atender os requisitos e suas capacidades de obtenção de diferentes informações espectrais, foram escolhidas a

Espectroscopia na Região do Infravermelho Próximo (NIRS) e Espectrometria de Emissão Óptica com Plasma Induzido por Laser (LIBS).

A NIRS (*Near-Infrared Spectroscopy*) é uma técnica bastante difundida em uma grande variedade de indústrias, permitindo a avaliação qualitativa e quantitativa de diferentes produtos. A rotina analítica com esse instrumento traz diversas vantagens em comparação com os métodos convencionais que empregam via úmida, principalmente em seus custos analíticos mais baixos. Custos estes estimados em grande parte na relação com mão de obra, consumíveis e reagentes. Outras vantagens incluem riscos reduzidos quando materiais perigosos e procedimentos tecnicamente difíceis estão envolvidos, dessa forma sendo necessário menor habilidade técnica exigida dos analistas. Benefícios adicionais significativos também podem ser derivados pela aplicação de métodos NIRS para fins de controle de processo, sistemas de pagamento e a previsão de múltiplos constituintes a partir de medidas espectrais únicas (PASQUINI, 2003; BURNS, 2007).

Nas aplicações da cadeia de cana-de-açúcar, diversos trabalhos têm sido realizados a partir da década de 90. Os primeiros ainda em 1989 descreveram a capacidade de previsão de parâmetros de qualidade diretamente na cana desfibrada (BERDING, 1989). Com a evolução da técnica, foram desenvolvidos sistemas NIRS para medições *online*, tanto para Cana, como para Bagaço. Em laboratório, desde 1995 são apresentados desenvolvimentos no Brasil e em países como Austrália, Japão e África do Sul, com diversas aplicações para produtos da cadeia de produção do açúcar (O'SHEA, 2011).

A LIBS (*Laser-induced Breakdown Spectroscopy*) está encontrando muito rapidamente seu nicho entre as técnicas modernas de espectrometria devido ao seu potencial para lidar com amostras na forma sólida, líquida e gasosa (incluindo aerossóis), que podem ou não serem condutoras de maneira direta e praticamente não destrutiva. A LIBS tem sido aplicada para fins qualitativos e quantitativos em uma infinidade de matrizes de interesse em diversas áreas. Algumas dessas aplicações são exclusivas da LIBS e fazem uso dos atributos intrínsecos da técnica, como a capacidade de microanálise, possibilidade de análises remotas e natureza quase não destrutiva (SANTOS JÚNIOR, 2006; PASQUINI, 2007; CREMERS, 2013; COSTA, 2019).

A composição elementar é um indicador muito importante para compreender a qualidade, o valor nutricional e a autenticidade dos alimentos. A LIBS, como demais

técnicas de emissão, são técnicas de análise elementares, e informações sobre a composição elementar também podem ser usadas para a determinação de fraudes alimentares. As técnicas comuns utilizadas na análise de alimentos requerem muito tempo e, em algumas delas, é necessário o uso de certos produtos químicos perigosos que podem causar poluição ambiental. O LIBS, por outro lado, é um método in situ que pode analisar sólidos, líquidos e gases e fornece opções de detecção versáteis com processos mínimos de preparação de amostra. Dadas essas vantagens, pode-se concluir que a LIBS tem potencial para se tornar uma ferramenta analítica versátil para análise de alimentos (SEZER, 2017).

Nas aplicações da cadeia de cana-de-açúcar poucos trabalhos foram publicados nos últimos anos. ROMERA, J.P.R. et al. (2015) propuseram um método multivariado para determinação do teor de fibra em cana-de-açúcar. ANDRADE, D. F. et al. (2018) apresentaram estudos sobre a detecção de elementos químicos relacionados a impurezas lixiviadas de cana crua. SILVA, A. (2017) descreveu o potencial da utilização da técnica LIBS para caracterização e determinação de minerais em açúcares brasileiros.

Capítulo 2
OBJETIVOS E JUSTIFICATIVAS

2 OBJETIVOS

O principal objetivo desta dissertação é o desenvolvimento de novas aplicações para análise de Açúcar proveniente da cana-de-açúcar, utilizando técnicas espectroanalíticas. Para isso foi organizado um banco de dados espectrais de amostras de açúcar de diferentes tipos e origens utilizando as técnicas baseadas em laser e infravermelho, e a partir da união dessas informações, juntamente com ferramentas quimiométricas, criar modelos de calibração multivariada para classificação, detecção de adulterantes, quantificação de impurezas como pontos pretos e previsão de parâmetros físico-químicos.

2.1 Justificativas

A CETEC Equipamentos para Laboratório tem como visão “Ser uma empresa, de vendas de equipamentos, serviços de manutenção e calibração, sustentável atuando nos mercados sucroalcooleiro, alimentos, sucos, bebidas, agrícola e ambiental até 2022”. E para ser sustentável é imprescindível entender as demandas do mercado e oferecer soluções que permitam o desenvolvimento dos nossos produtos e expansão das possibilidades de negócio.

A maior demanda nos últimos anos se dá no interesse de automatizar processos, aumentar produtividade e obter respostas de maneira mais rápida e confiável. Dessa forma, se faz necessária a introdução de novos métodos, instrumentos analíticos e desenvolvimento de aplicações que permitam atingir esses pontos.

Assim, essa dissertação de mestrado se justifica na intenção de desenvolver novas aplicações que permitam atender as demandas analíticas desses mercados e possibilitar melhorias que as técnicas clássicas em suas limitações não permitem.

Capítulo 3
REVISÃO BIBLIOGRÁFICA

3 REVISÃO BIBLIOGRÁFICA

3.1 Classificação do Açúcar

Segundo a instrução normativa N° 42, de 13 de novembro de 2017 do Ministério de Estado da Agricultura, Pecuária e Abastecimento, é “Açúcar” o produto obtido a partir da cana-de-açúcar pertencente às cultivares provenientes da espécie *Saccharum Officinarium L.* através dos processos adequados. É constituído por cristais, com exceção do açúcar líquido. Para efeito do regulamento técnico, no artigo 2°, considera-se:

1. Cor ICUMSA: a cor de uma solução de açúcar em concentrações e pH definidos, cuja absorvância é medida em 420 nm (quatrocentos e vinte nanômetros), expressa em unidades ICUMSA (UI);
2. Polarização: a quantidade de sacarose aparente medida por polarimetria, em condições padronizadas, utilizando o peso normal de 26,000 g (vinte e seis gramas) em 100 ml (cem mililitros) de água a 20°C (vinte graus Celsius), e expressa em °Z (grau Zucker);
3. Cinzas condutimétricas: o teor de cinzas em uma amostra determinada pela medida da condutividade elétrica de uma solução aquosa da amostra de concentração conhecida, expressa em porcentagem (%);
4. Umidade: o teor de água livre encontrada no produto, medida por aquecimento a 105°C (cento e cinco graus Celsius), sendo expressa em gramas por 100g (cem gramas) do produto.
5. Pontos pretos: as partículas visíveis de coloração contrastantes com a dos cristais de açúcar podendo ser provenientes de açúcar caramelizado, açúcar carbonizado, fuligem, fagulhas da queima de cana, fibras da cana ou resíduos de incrustação dos equipamentos, expressas em número de pontos pretos por 100g (cem gramas);

Os artigos 5 e 6 dos Capítulo II estabelecem que os requisitos de qualidade do açúcar são definidos em função dos parâmetros estabelecidos nos Anexos I e II da Instrução Normativa sendo classificado em Grupos, Classes e Tipos, conforme o disposto na Tabela 1.

Tabela 1. Parâmetros de Qualidade do Açúcar do Grupo II (Vigente a partir de 1º de julho de 2019)

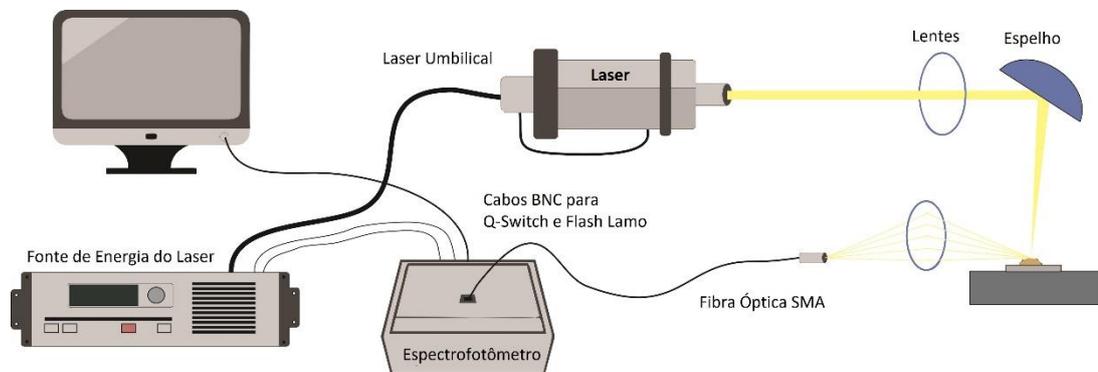
Classes	Tipos	Parâmetros					
		Polarização (°Z) mínimo	Umidade (%) máximo	Cor ICUMSA (UI) máximo	Cinzas Condutimétricas (%) máximo	Pontos Pretos (nº/100) máximo	Partículas Magnetizáveis (mg/kg) máximo
Branco	Cristal	99,50	0,10	400	0,10	20	15
	Ref. Amorfo	99,00	0,30	100	0,20	5	5
	Confeiteiro	99,00	0,30	150	0,20	5	5
	Ref. Granulado	99,80	0,05	60	0,20	5	5
Bruto	Demerara	96,00	1,25	5000	0,50	-	-
	VHP	99,00	0,25	2500	0,25	-	-
	VVHP	99,49	0,15	1000	0,15	-	-

Fonte: Instrução normativa N° 42, MAPA, 2017.

3.2 LIBS (Laser-Induced Breakdown Spectroscopy)

A LIBS tem como princípio a detecção de emissões de espécies excitadas, moléculas neutras ou ionizadas. A incidência de sucessivos pulsos de laser de curta duração (10 ns), focalizados em direção à amostra promovem a formação de micro plasma. Parte da amostra sofre ablação e interage com o plasma onde podem ocorrer, subsequentemente, processos de vaporização, atomização, ionização e excitação dos elementos presentes no plasma formado após a ablação. Cada elemento ou íon excitado possui linhas de emissão bem definidas, em comprimentos de onda específicos. Com o resfriamento do plasma as espécies excitadas começam a retornar ao seu estado fundamental, e assim emitem energia na forma de radiação em diversos comprimentos de onda. A radiação emitida passa por uma lente que a focaliza em fibras ópticas conectadas em um ou mais espectrômetros. A radiação é conduzida até o detector, que por sua vez, converte a radiação em sinal elétrico, que é registrado no computador na forma de espectro. O esquema ilustrado na Figura 3 mostra a instrumentação típica de um sistema LIBS (SANTOS JÚNIOR, 2006; PASQUINI, 2007; CREMERS, 2013; COSTA, 2019)

Figura 3. Esquema típico da instrumentação LIBS



Fonte: Adaptado de COSTA, 2019.

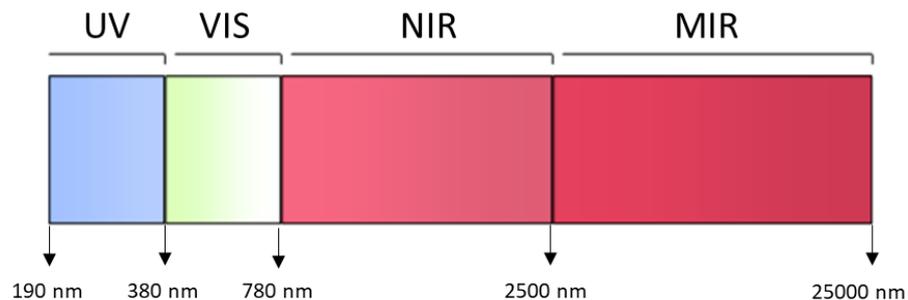
A configuração padrão de um espectrômetro LIBS deve ser constituído por um laser, um conjunto óptico, um detector e um computador para controle e processamento de dados. Neste sistema, um jogo de lentes convergentes focaliza um pulso de laser na amostra, promove a dissociação das moléculas do material em seus constituintes atômicos, devido à ruptura das ligações moleculares, e também em íons e elétrons devido à ruptura das ligações eletrônicas no ponto focal da lente. Em geral, a técnica LIBS emprega um feixe de laser pulsado com alta irradiância, da ordem de GW/cm^2 , que quando focalizado sobre a superfície da amostra, possibilita a formação

de um plasma de alta temperatura (e.g. 10000 a 20000 K). O comprimento de onda do laser também afeta significativamente a formação do plasma, tanto com respeito à criação dos elétrons iniciais como à interação plasma-laser. Geralmente, o espectro obtido na leitura das emissões é registrado nas regiões UV-Vis, mas vale ressaltar que emissões de raios-X também são possíveis, dependendo das características do laser incidente na amostra (SANTOS JÚNIOR, 2006; COSTA, 2019).

3.3 NIRS (Near Infrared Spectroscopy)

A NIRS é um tipo de espectroscopia que emprega energia de fótons ($h\nu$) no faixa de energia de $2,65 \times 10^{-19}$ a $7,96 \times 10^{-20}$ J, que corresponde à faixa de comprimento de onda de 780 a 2.500 nm (números de onda: 12.820 a 4.000 cm^{-1} , Figura 4). Essa faixa de energia é a necessária para promover moléculas até seus estados vibracionais mais excitados e inferiores aos valores típicos necessários para a excitação de elétrons em moléculas. Seu objetivo geral é investigar uma amostra para obter informações qualitativas e/ou quantitativas provenientes da interação de ondas eletromagnéticas no infravermelho próximo com seus constituintes (PASQUINI, 2003; BURNS, 2007).

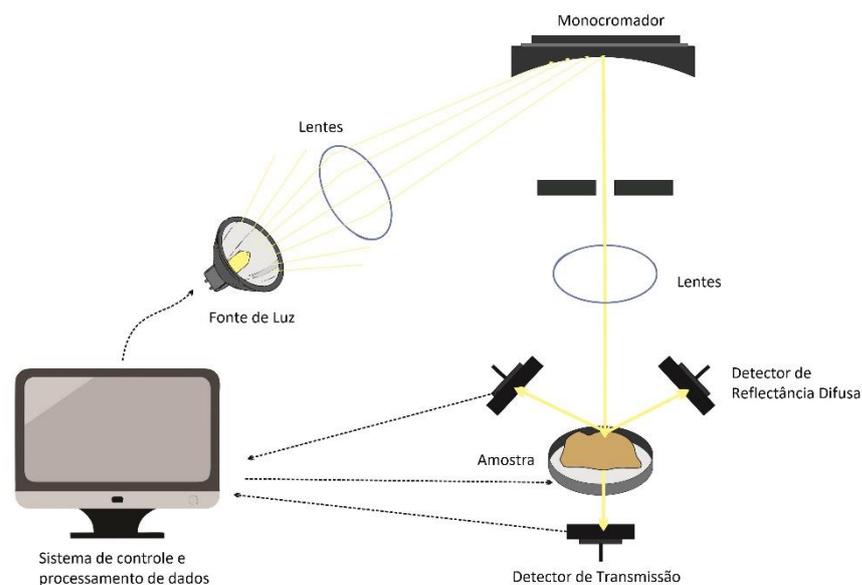
Figura 4. Espectro eletromagnético com faixas de comprimento de onda do UV até MIR.



Os métodos analíticos resultantes da utilização da região espectroscópica NIRS refletem suas características mais significativas, tais como: alta frequência de amostragem (um minuto ou menos por amostra), não destrutivo, não invasivo, com alta penetração do feixe de radiação de sondagem, adequado para uso em medidas em linha, aplicação quase universal, onde qualquer molécula contendo ligações CH, NH, SH ou OH pode ser detectada com exigências mínimas de preparação de amostras (PASQUINI, 2003)

Além disso, um instrumento NIRS pode se basear em refletância difusa ou transmitância, de forma que os detectores empregados com mais frequência para a região espectral NIRS são baseados em materiais fotocondutores de silício, PbS e InGaAs. Em particular, estes últimos possuem altas detectividade e velocidade de resposta. Juntamente com fontes de radiação de alta potência, tais como uma bobina de tungstênio ou uma lâmpada de halogênio é empregada pela maioria dos fabricantes. Os detectores podem transmitir uma alta relação sinal-ruído para medidas em NIRS. Este fato compensa parcialmente as menores intensidades das bandas de absorção NIRS. Instrumentos mais antigos utilizavam filtros para cobrir o comprimento de onda com poucos pontos de varredura, instrumentos modernos utilizam sistemas de monocromador que possuem resolução de até 0,5 nm, ou ainda sistemas baseados em interferometria (FT-NIR). O esquema ilustrado na Figura 5 mostra a configuração típica de um instrumento NIRS (PASQUINI, 2003; BURNS, 2007).

Figura 5. Esquema instrumentação NIRS



Fonte: Adaptado de PASQUINI, 2003

3.4 Quimiometria e Calibração Multivariada

A quimiometria denota a aplicação de métodos estatísticos e matemáticos para extrair informações de dados químicos e físicos (BURNS, 2007). Duas das metodologias quimiométricas mais importantes e amplamente utilizadas são a calibração e a classificação multivariadas.

A calibração multivariada visa encontrar relações quantitativas entre dois conjuntos de dados. Os dados são obtidos via técnicas espectroanalíticas e os valores

de referência são determinados por técnicas convencionais. O modelo de calibração obtido pode ser aplicado para prever com precisão os parâmetros desejados em amostras desconhecidas usando o método rápido, substituindo, portanto, o convencional. Diversos algoritmos, envolvendo técnicas não-lineares como *support vector machine* e redes neurais artificiais, ou técnicas lineares como regressão linear múltipla (MLR, *Multiple linear regression*) e de mínimos quadrados parciais (PLS, *Partial least squares*) são geralmente aplicadas para extrair as informações analíticas dos espectros (BURNS, 2007)

Além disso, análises de agrupamento e discriminantes podem ser aplicadas como classificações multivariadas para determinações qualitativas. Sendo que as análises de agrupamento podem identificar dados sem nenhuma estrutura de classes predefinida e as análises discriminantes classificam amostras desconhecidas em grupos.

Calibrações podem ser desenvolvidas com diversas finalidades, dentre elas duas se destacam, sendo a primeira delas o desenvolvimento de calibrações para estudos de viabilidade e a segunda as calibrações desenvolvidas para aplicação em rotina nas indústrias. Para os estudos de viabilidade é necessário apenas um número relativamente pequeno de amostras que são reunidas, analisadas e submetidas para dados de referência. As calibrações são desenvolvidas e os resultados publicados em artigos científicos ou apresentados em conferência. Nesses casos, normalmente não são feitas tentativas para determinar se tais calibrações realmente funcionam na rotina, e não há tal necessidade real, uma vez que o objetivo é demonstrar e provar a viabilidade. Já nas calibrações desenvolvidas para indústria podem ainda se subdividir em duas finalidades, o primeiro se destinando a aplicação em toda amostra que entra no laboratório, abrangendo toda a faixa de composição esperada e demais variáveis, tendo como exemplo o controle de processo do açúcar nas medições de cor, pol, umidade e cinzas. A forma de realizar a validação de tais calibrações é a partir de um conjunto de amostras utilizado para cálculo dos modelos, sendo testados imediatamente após a aplicação na análise diária, é submetido às estatísticas ubíquas, RMSEP, RSQ. A segunda subdivisão é a calibração desenvolvida para controle de qualidade, por exemplo, em moinhos de farinha ou fabricação de rações, onde a consistência do produto é essencial. Para este tipo de aplicação, a variação mínima é esperada nas amostras posteriormente testadas pela calibração (WILLIANS, 2017).

3.4.1 RMSEC (*Root Mean Square Error of Calibration*) e RMSEP (*Root Mean Square Error of Prediction*).

RMSEC é o desvio padrão devido as diferenças entre os valores reais (valores analíticos do laboratório primário ou de referência) e os valores previstos pelo modelo para amostras dentro do conjunto de calibração. Este parâmetro é uma indicação do erro residual total devido ao modelo calculado. A RMSEC diminuirá com o aumento do número de termos, ou LV (*latent variables*), permitirá que mais variações nos dados sejam explicadas ou “ajustadas”. A estatística RMSEC é uma estimativa útil da “melhor” precisão teórica obtida para um conjunto especificado de LV usado para desenvolver um modelo de calibração (BURNS, 2007) RMSEC é definida matematicamente pela equação 1.

$$RMSEC = \left(\frac{\sum_{i=1}^N (y_{i,ref} - y_{i,pred})^2}{N - K - 1} \right)^{1/2} \quad (1)$$

Onde, $y_{i,ref}$ são os valores do método de referência, $y_{i,pred}$ são os valores previstos, N é o número de amostras da matriz de dados e K é o número de variáveis latentes.

RMSEP é a medida mais simples e eficiente da incerteza nas previsões de um modelo. Este valor é uma medida da incerteza média que pode ser esperada ao prever novas amostras. RMSEP é expresso nas mesmas unidades que o parâmetro previsto. Esta medida é válida desde que as novas amostras sejam semelhantes às usadas para a calibração, caso contrário, o erro de previsão pode ser muito maior (DAVIES, 2006). RMSEP é definida matematicamente pela equação 2.

$$RMSEP = \left(\frac{\sum_{i=1}^N (y_{i,ref} - y_{i,pred})^2}{N} \right)^{1/2} \quad (2)$$

Onde, $y_{i,ref}$ são os valores do método de referência, $y_{i,pred}$ são os valores previstos, N é o número de amostras da matriz de dados.

3.4.2 Cross-Validação e RMSECV (*Root Mean Square of Cross Validation*)

O uso de validação cruzada é um método para determinar o “melhor” número de variáveis latentes para usar na construção do modelo de calibração. Este método é baseado em um algoritmo iterativo (repetitivo) que seleciona amostras de uma população de conjunto de amostras para desenvolver o modelo de calibração e, em seguida, prevê as amostras não selecionadas restantes. (BURNS, 2009). O RMSECV é determinado a partir dessa previsão e seu valor é utilizado como indicador de desempenho para escolha do melhor número de variáveis latentes. RMSECV é definida matematicamente pela equação 3.

$$RMSECV = \left(\frac{\sum_{i=1}^N (y_{i,ref} - y_{i,pred})^2}{N} \right)^{1/2} \quad (3)$$

Onde, $y_{i,ref}$ são os valores do método de referência, $y_{i,pred}$ são os valores previstos, N é o número de amostras da matriz de dados.

Dessa forma, é uma ferramenta muito útil que atende a duas funções críticas em quimiometria. Ele permite uma avaliação da complexidade ideal de um modelo, por exemplo, o número de componentes principais (PC) em um modelo PCA ou PCR, ou o número de variáveis latentes (LV) em um modelo PLS. E também permite estimar o desempenho de um modelo quando aplicado a dados desconhecidos.

Para um determinado conjunto de dados, a validação cruzada envolve uma série de experimentos, aqui chamados de experimentos de sub-validação, cada um dos quais envolve a remoção de um subconjunto de objetos de um conjunto de dados (o conjunto de teste), construção de um modelo usando os objetos restantes no conjunto de dados (o conjunto de construção do modelo) e a aplicação subsequente do modelo resultante aos objetos removidos. Dessa forma, cada experimento de sub-validação envolve testar um modelo com objetos que não foram usados para sua construção (EIGENVECTOR, 2020). Um procedimento de validação cruzada típico geralmente envolve mais de um experimento de sub-validação, cada um dos quais envolve a seleção de diferentes subconjuntos de amostras para construção e teste de modelo. Como existem vários métodos de modelagem diferentes em quimiometria, também existem vários métodos de validação cruzada diferentes, e estes variam com relação aos diferentes subconjuntos de amostra selecionados para esses experimentos de sub-validação. Para as descrições a seguir, na Tabela 2, N é o

número total de objetos no conjunto de dados e s é o número de divisões de dados especificadas para o procedimento de validação cruzada, que deve ser menor que $N/2$ (EIGENVECTOR, 2020).

Tabela 2. Descrições para quatro tipos de métodos de cross-validação com representação gráfica.

Tipo	Descrição	Representação Gráfica
Venetian Blinds	Cada conjunto de teste é determinado selecionando cada s objeto no conjunto de dados, começando nos objetos numerados de 1 a s .	
Contiguous Blocks	Cada conjunto de teste é determinado selecionando blocos contíguos de N/s objetos no conjunto de dados, começando no objeto número 1.	
Random Subsets	' s ' diferentes conjuntos de teste são determinados por meio da seleção aleatória de N/s objetos no conjunto de dados, de modo que nenhum objeto esteja em mais de um conjunto de teste. Este procedimento é repetido r vezes, onde r é o número de iterações. É a média dos resultados das iterações que é usada no relatório dos resultados da validação cruzada	
Leave-One-Out	Cada objeto único no conjunto de dados é usado como um conjunto de teste.	

Fonte: Adaptado de EIGENVECTOR, 2020.

3.4.3 Pré-tratamentos

O pré-tratamento de dados espectrais tornou-se parte integrante da modelagem quimiométrica. O objetivo do pré-tratamento é remover fenômenos físicos nos espectros, a fim de melhorar a regressão multivariada subsequente, modelo de classificação ou análise exploratória. Diversas técnicas de pré-processamento são conhecidas e aplicadas de acordo com as características espectrais obtidas pelo instrumento analítico. As técnicas mais amplamente utilizadas podem ser divididas em duas categorias: métodos de correção de dispersão e derivados espectrais (RINNAN, 2009).

O primeiro grupo de métodos de pré-tratamento corretivo de dispersão inclui: Correção de Dispersão Multiplicativa (MSC), MSC Inverso, MSC Estendido (EMSC), MSC Inverso Estendido, de-tendência, Variável Normal Padrão (SNV) e normalização. O grupo de derivação espectral pode ser representado por 2 tipos: derivados de Norris-Williams (NW) e filtros de derivadas polinomiais de Savitzky-Golay (SG). Ambos os métodos usam uma suavização dos espectros antes de calcular a derivada, a fim de diminuir o efeito prejudicial na razão sinal-ruído que as derivadas de diferenças finitas convencionais teriam. O objetivo da etapa de pré-tratamento pode ser um dos três: 1) para melhorar uma análise exploratória subsequente, 2) para melhorar um modelo de calibração bi linear subsequente (forçar os dados a obedecer à lei de Lambert-Beer); ou, 3) para melhorar um modelo de classificação subsequente (RINNAN, 2009).

3.4.4 *Partial Least Squares (PLS)*

A PLS é uma técnica de calibração multivariada linear consolidada e altamente difundida, sendo a técnica mais amplamente aplicada para a proposição de regressão. A literatura especializada apresenta milhares de artigos científicos sobre a PLS e suas aplicações. A popularidade é atribuída principalmente à sua capacidade de lidar com questões de singularidade, desenvolver modelos fáceis de interpretar e visualizar soluções que facilitam o entendimento da aplicação. Ela assume uma relação linear aproximada entre o parâmetro de qualidade e a intensidade de suas bandas de absorção e, portanto, é viável para dados espectrais que seguem a lei de Beer. Além disso, a PLS é adequada quando os dados são altamente dimensionais e multicolineares, o que costuma ser o caso com dados espectrais. O método é capaz

de lidar com não linearidades fracas no espaço \mathbf{X} , mas não possui essa propriedade para a relação entre \mathbf{X} e y (TANGE, 2015; NI, 2014)

A PLS é baseada em uma decomposição de variável latente de dois blocos de variáveis, as matrizes \mathbf{X} (dados espectrais) e \mathbf{Y} (referências). Uma relação entre esses dois blocos é descrita pela análise PLS por meio de uma relação interna de seus escores, maximizando a covariância entre essas variáveis internas (NI, 2014).

No método de PLS as matrizes \mathbf{X} e \mathbf{Y} são decompostas simultaneamente em uma soma de h variáveis latentes, como nas equações 4 e 5.

$$\mathbf{X} = \mathbf{T} \mathbf{P}' + \mathbf{E} = \sum t_h \mathbf{p}'_h + \mathbf{E} \quad (4)$$

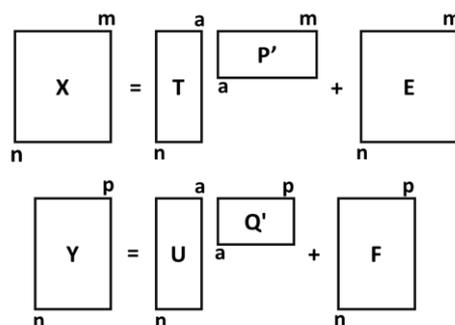
$$\mathbf{Y} = \mathbf{U} \mathbf{Q}' + \mathbf{F} = \sum u_h \mathbf{q}'_h + \mathbf{F} \quad (5)$$

Onde \mathbf{T} e \mathbf{P} são as matrizes de scores e loadings para \mathbf{X} , respectivamente; \mathbf{U} e \mathbf{Q} são as respectivas matrizes de scores e loadings para a matriz \mathbf{Y} . As matrizes \mathbf{E} e \mathbf{F} são os resíduos, graficamente podem ser descritos como na Figura 6. Uma correlação entre as matrizes \mathbf{X} e \mathbf{Y} é simplesmente uma relação linear obtida pelo coeficiente de regressão linear, tal como na equação 6.

$$\mathbf{U}_h = \mathbf{b}_h \mathbf{t}_h \quad (6)$$

Assumindo h de variáveis latentes, sendo que os valores de \mathbf{b}_h são agrupados na matriz diagonal \mathbf{B} , que contém os coeficientes de regressão entre a matriz de scores \mathbf{U} de \mathbf{Y} e a matriz de scores \mathbf{T} de \mathbf{X} . A relação linear melhor encontrada entre os scores dos dois blocos é calculada através de ajustes das variáveis latentes dos blocos \mathbf{X} e \mathbf{Y} (GELADI, 1986).

Figura 6. Representação gráfica das matrizes e vetores usados na PLS.



3.4.5 Modified Partial Least Squares (mPLS)

O algoritmo mPLS é encontrado nativo nos programas de desenvolvimento baseados no padrão ISI (Infrasoft Internacional), comumente utilizados nos instrumentos e programas da fabricante dinamarquesa FOSS. O algoritmo foi desenvolvido por Mark Westerhaus, que descreve uma alternativa de usar a correlação quadrática como pesos para construir fatores de MLR em combinação com a técnica PLS, porém no lugar de usar a covariância $x'y$ como peso de comprimento de onda, usando, o declive da regressão $x'y / x'x$. Depois de calcular os pesos dos comprimentos de onda, projetou-os no espaço espectral original. A PCA é projetada para explicar a maior variação espectral possível com cada fator. Cada fator PLS é um compromisso, tentando explicar a variação espectral enquanto tenta correlacionar altamente com a variável y , cada fator mPLS também é um compromisso, mas coloca menos ênfase na explicação da variação espectral. Se a maior parte da variação espectral for devido ao constituinte de calibração, a PLS deve fazer um bom trabalho. A mPLS é útil quando é necessário ignorar as principais fontes de variação espectral e se concentrar em regiões com variação espectral menor, mas correlações fortes (WESTERHAUS, 2014).

Em resumo, diferentemente do algoritmo PLS padrão que maximiza a covariância entre X e Y , a mPLS maximiza a correlação entre X e Y , permitindo maior flexibilidade quando é necessário lidar com menor variabilidade espectral (SHENK, 1991; WESTERHAUS, 2014).

3.4.6 Artificial Neural Networks (ANN)

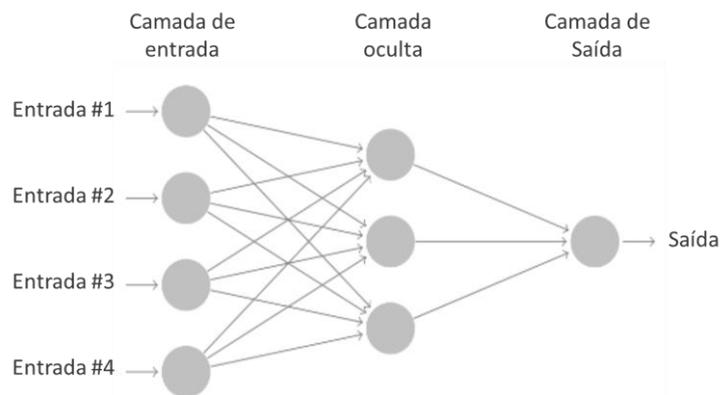
As redes neurais artificiais (ANN) são modelos computacionais inspirados no sistema nervoso humano que consistem em vários elementos de processamento que recebem entradas e entregam saídas com base em suas funções de ativação pré definidas.

Normalmente, uma ANN alimentada para frente (feed-forward), consistindo em três camadas (camada de entrada, camada oculta e camada de saída) como na Figura 7, é usada na aplicação espectroscópica prática. Para análise de dados espectrais, a ANN é expressa matematicamente como mostra a equação 7.

$$y_i = f\left[\sum_{t=1}^T \omega_t g\left(\sum_{j=1}^J W_{tj} X_{tj} + \varphi_t\right) + \varepsilon\right] + e_i \quad (7)$$

Onde T é o número de neurônios da camada oculta, W_{tj} denota os pesos conectando os neurônios da camada de entrada à camada oculta, ω_t representa o peso que conecta os neurônios da camada oculta e de saída, φ_t e ε são os vieses nas camadas oculta e de saída, respectivamente, enquanto e_i é o ruído. As duas funções de transferência, $f(\cdot)$ e $g(\cdot)$, denotam funções lineares e sigmóides, respectivamente. Sob certas suposições, esta rede neural é capaz de aproximar qualquer função com precisão arbitrária. Para calcular as informações de gradiente de forma eficiente para o problema de otimização não linear, um algoritmo de retropropagação é usado para otimizar os parâmetros em uma ANN ajustando os pesos para minimizar o erro de saída (NI, 2014).

Figura 7. ANN simplificado onde cada círculo representa um neurônio; neste exemplo, três neurônios ocultos são usados e apenas quatro entradas (representando variáveis espectrais) com uma saída.

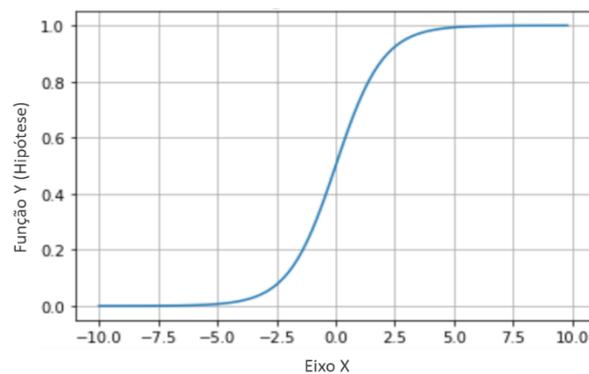


Em outras palavras, os pontos de dados de entrada (por exemplo, medições espectrais ou pré-processadas) são combinadas por pesos exatamente como em um modelo de regressão linear, conforme a equação 8.

$$\begin{aligned} x_1 * w_{11} + x_1 * w_{12} + \dots + x_{100} * w_{1100} &= \text{eno1} \\ x_1 * w_{21} + x_1 * w_{22} + \dots + x_{100} * w_{2100} &= \text{eno2} \\ x_1 * w_{31} + x_1 * w_{32} + \dots + x_{100} * w_{3100} &= \text{eno3} \end{aligned} \quad (8)$$

Considerando um espectro onde temos uma variável a cada 2nm de 850 a 1048nm correspondendo à 100 variáveis de entrada, então w_1 significa entrada para o neurônio oculto número 1 e w_2 corresponde ao segundo neurônio oculto, enquanto 100 significa peso de número 100. A não linearidade é normalmente tratada na camada oculta entre os dados de entrada (espectros) e a camada de saída (parâmetro). Suponha que temos três neurônios ocultos em nossa configuração; então calcularemos w_1 , w_2 e w_3 . Cada um deles será transformado por meio de uma função não linear que é a chave para a modelagem não linear de um modelo de ANN; pela seleção de uma função não linear adequada, o grau de não linearidade necessário será regulado por esta função de transferência. Na Figura 8, é mostrado um exemplo de uma função de transferência não linear sigmoide; uma infinidade de funções de transferência não linear pode ser usada com diferentes propósitos para fornecer soluções ideais em aplicações específicas (NOGAARD, 2013)

Figura 8. Função de transferência não-linear – função tangente hiperbólica sigmoideal.



A saída dos neurônios ocultos (sno_1 a sno_3) é então linearmente ponderada para fornecer um conjunto de "camada de entrada para a saída".

$$sno_1 * v_1 + sno_2 * v_2 + sno_3 * v_3 = \text{"camada de entrada para saída"} \quad (9)$$

Esta soma ponderada pode passar por uma função linear ou uma função não linear para aumentar ainda mais a flexibilidade na modelagem não linear. Para modelos lineares, é uma tarefa bastante simples estimar os coeficientes de regressão, por ex. através de uma estimativa de mínimos quadrados, minimizando os resíduos entre os valores de referência do laboratório e os valores previstos. Em um modelo linear, estimamos 100 coeficientes de regressão mais um deslocamento, resultando no total de 101 coeficientes. Na modelagem ANN, a parte de estimativa é mais

complexa; no caso apresentado, precisamos de 100 x 3 neurônios (w's) + 3 x 1 (v's) (excluindo vieses), além da complexidade na estimativa dos pesos em várias camadas e funções não lineares. Esse é o preço a ser pago por ter um modelo capaz de lidar com a não linearidade, e a estimativa dos coeficientes pode ser realizada por métodos de otimização a partir de análises numéricas; por exemplo: por gradiente descendente ou métodos mais avançados, modernos e sob medida (NORGAARD, 2013).

Ao treinar uma rede, o objetivo é fazer com que a mesma responda a um determinado espectro de entrada de maneira inteligente. Existem dois estados: a) um estado de treinamento ou aprendizado onde a ANN aprende como responder e b) um estado de predição onde a rede é aplicada a um espectro nunca visto antes. Neste estado de previsão, nenhuma correção para os pesos é feita. O treinamento e a aprendizagem ocorrem comparando a saída do parâmetro da rede com o valor do parâmetro de referência e, em seguida, alterando os pesos para se aproximar do valor de referência. Portanto, a rede aprende com o exemplo, assim como o cérebro poderia fazer. O aprendizado ocorre simplesmente mudando os pesos de forma sistemática (NORGAARD, 2013).

Capítulo 4
**ESTUDO DE IDENTIFICAÇÃO E
QUANTIFICAÇÃO DE ADULTERANTES NO
AÇÚCAR EMPREGANDO LIBS E NIRS**

4 ESTUDO DE IDENTIFICAÇÃO E QUANTIFICAÇÃO DE ADULTERANTES NO AÇÚCAR EMPREGANDO LIBS E NIRS

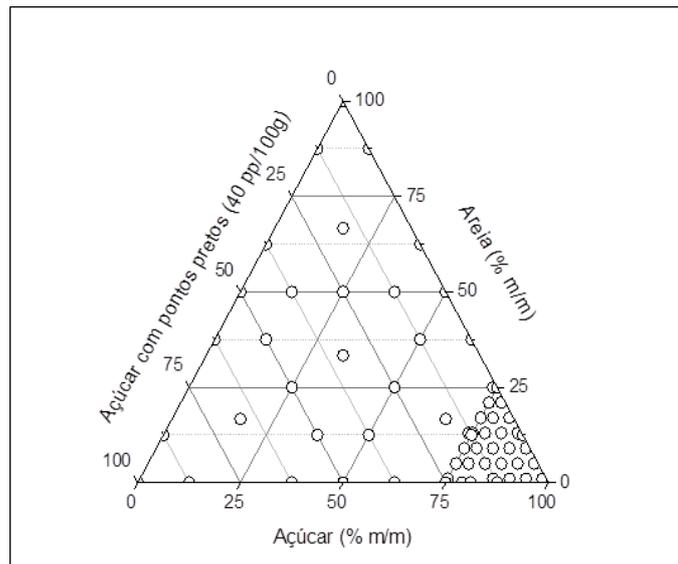
4.1 Experimental

4.1.1 Obtenção das amostras

Amostras adulteradas foram preparadas a partir da utilização de açúcar cristal, areia natural e pontos pretos, este último obtido a partir de partículas metálicas coletadas diretamente dos imãs de retenção no processo de fabricação do açúcar. Sessenta e quatro (64) amostras da mistura sólida foram preparadas por peneiração do material (35 mesh \varnothing - 0,5 mm). As porções com tamanho de partícula abaixo de 35 e acima de 45 mesh \varnothing foram descartadas por apresentarem características de poeira (35 mesh) e para otimizar a homogeneidade da mistura. As amostras foram preservadas em embalagens plásticas seladas e as misturas foram preparadas como recebidas para preservar as propriedades organolépticas (teor de umidade de 0,04%). As condições ambientais da preparação foram 26,6°C e RH 48,15%. Os materiais foram quarteados e homogeneizados, as proporções de massa definidas com um peso final de 40,000g e medidas em uma balança analítica Gehaka AG200 (São Paulo, BR).

O teor de Areia das amostras variou de 0 a 5% em peso para 21 amostras, 9-17% em peso para 20 amostras e entre 21 e 100% em peso para as demais amostras. Para PP o teor variou de 0 a 5 pp/100g para 37 amostras, 6 a 15 pp/100g para 16 amostras e 15 a 40 pp/100g para 11 amostras. As faixas para teores de areia foram definidas para contemplar uma situação de alíquotas com ausência de adulteração até alíquotas com apenas a presença de areia, já para pontos pretos foi extrapolado em 2 vezes a quantidade máxima permitida pela legislação como concentração máxima e demais concentrações distribuídas em toda a faixa de acordo com diagrama apresentado na Figura 9.

Figura 9. Diagrama de mistura ternária nas concentrações planejadas.

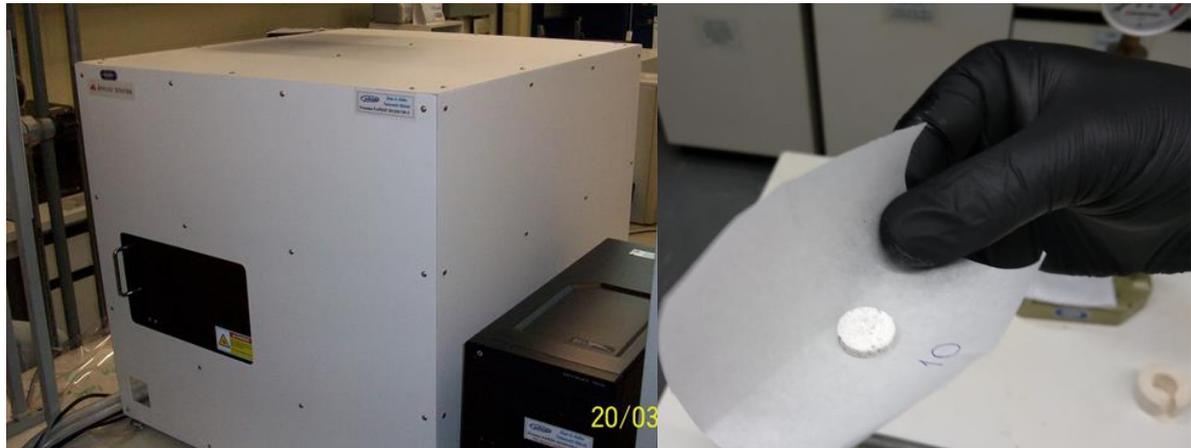


4.1.2 Preparo e obtenção dos dados espectroanalíticos LIBS

As amostras foram prensadas para formar pastilhas, utilizando prensa hidráulica, através da utilização de 0,5g de amostras previamente homogeneizadas. O procedimento de prensagem foi realizado utilizando a Prensa Hidráulica modelo SSP-10A da fabricante Shimadzu, empregando pressão de 60 KN durante 2 minutos, obtendo pastilhas de diâmetro médio de 15 mm e espessura média de 3 mm. Amostras com teores acima de 87,5% de areia foram descartadas do ensaio LIBS devido à incapacidade de formar pastilhas com consistência adequada.

Para obtenção dos espectros de emissão foi utilizado o Espectrômetro de Emissão Óptica com Plasma Induzido por Laser (LIBS) modelo J200 da fabricante Applied Spectra (Figura 10), com tipo de Laser Nd:YAG 1064nm, energia máxima do pulso de 100mJ e faixa espectral 186 a 1042nm.

Figura 10. (Esquerda) LIBS modelo J200 Applied Spectra (Direita) Pastilhas preparadas para leitura no instrumento LIBS.



As condições experimentais foram otimizadas para obtenção dos sinais das linhas de Fe, Si e C. Sendo escolhido o elemento Fe devido sua relação com os pontos pretos, Si com a Areia, e C com o Açúcar. Por meio de um planejamento fatorial foram realizados 11 experimentos utilizando uma amostra de concentrações intermediárias dos constituintes nas condições apresentadas na Tabela 3.

Tabela 3. Doehlert design usado para determinar as melhores condições operacionais do LIBS para os experimentos

Exp.	Laser pulse energy (mJ)	Spot size (μm)	Fluência		Delay time	
			Real (mJ/cm^2)	Valor codificado	Real (μs)	Valor codificado
1	80	75	1811	-0,9959	0,7	-1
2	69	50	3514	1,0041	0,7	-1
3	80	75	1811	-0,9959	1,7	1
4	69	50	3514	1,0041	1,7	1
5	53	50	2699	0	1,2	0
6	53	50	2699	0	1,2	0
7	53	50	2699	0	1,2	0
8	64	75	1448	-1,4222	1,2	0
9	75	50	3820	1,3635	1,2	0
10	53	50	2699	0	0,49	-1,414
11	53	50	2699	0	1,9	1,414

4.1.3 Preparo e obtenção de dados espectroanalíticos NIRS

Para obtenção dos espectros de infravermelho próximo foi utilizado o Espectrofotômetro de infravermelho próximo modelo DS2500 da fabricante dinamarquesa FOSS, com faixa espectral de 400 a 2500 nm.

Foram utilizadas 40,000g das amostras previamente homogeneizadas sem preparo e analisadas em duplicata. Para leitura das amostras foi utilizado o acessório *small cup*, em módulo de reflectância difusa, conforme Figura 11. A condição padrão de aquisição de espectros no instrumento foi definido realizando a leitura em 7 posições diferentes com 4 repetições por posição, obtendo assim o espectro médio gerado a partir de 28 subspectros.

Figura 11. Amostra de açúcar na célula de leitura Small Cup.



4.1.4 Tratamento dos dados LIBS

Os dados espectrais obtidos pela técnica LIBS foram normalizados com o emprego de doze modos de normalização. Essa etapa de normalização tem como finalidade compensar variações de sinal e diferenças na matriz da amostra. Desta forma, o objetivo final foi melhorar o poder de previsibilidade dos modelos de regressão. As estratégias de normalização são descritas por Castro & Pereira-Filho¹⁰, e a Tabela 4 mostra um resumo.

Tabela 4. Descrição dos 12 modos de normalizações aplicados.

Normalização	Descrição
1	É calculada a média de todos os espectros.
2	Cada espectro é normalizado pela sua norma individual. Em seguida é calculada a média.
3	Cada espectro é normalizado pela sua área. Os espectros normalizados posteriores são calculados em média;
4	Cada espectro é normalizado pelo seu sinal máximo. Os espectros normalizados posteriores são calculados em média;
5	Cada espectro é normalizado pelo sinal de C I 193,091 nm. Em seguida é calculada a média.
6	Cada espectro é normalizado pelo sinal de C I 247,856 nm. Em seguida é calculada a média.
7	É calculada a soma de todos os espectros.
8	Cada espectro é normalizado pela sua norma individual. Em seguida é calculada a soma.
9	Cada espectro é normalizado pela sua área. Os espectros normalizados posteriores são calculados a soma;
10	Cada espectro é normalizado pelo seu sinal máximo. Os espectros normalizados posteriores são calculados a soma;
11	Cada espectro é normalizado pelo sinal de C I 193,091 nm. Em seguida é calculada a soma.
12	Cada espectro é normalizado pelo sinal de C I 247,856 nm. Em seguida é calculada a soma.

Fonte: Adaptado de CASTRO & PEREIRA FILHO, 2019.

4.2 Resultados e Discussões

A partir dos resultados de desejabilidade, os experimentos 4 e 8 apresentaram as melhores condições experimentais. Assim, foi estabelecida uma condição intermediária entre os dois experimentos para o prosseguimento da parte

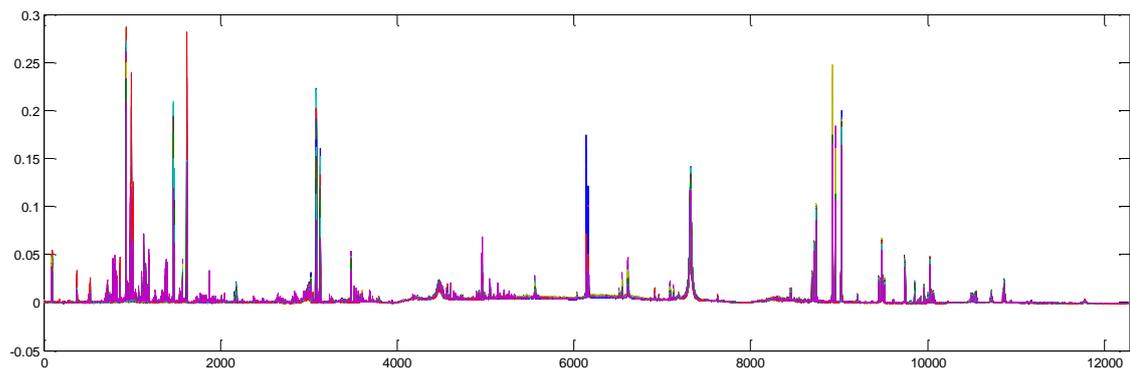
experimental, como mostra a Tabela 5.

Tabela 5. Condições experimentais definidas para elaboração dos experimentos LIBS.

Parâmetro	Condição Exp.
Laser Repetition Rate	5 Hz
Speed	1 mm/s
Delay time	1,2 s
Laser Pulse Energy	69 mJ
Spot Size	75
Fluence	2481

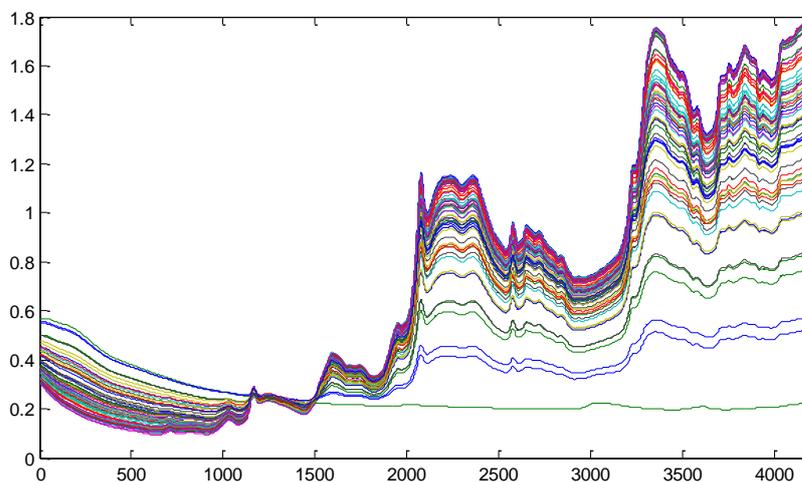
Na Figura 12 segue os espectros das 61 amostras obtidos no instrumento LIBS J200 a partir das condições experimentais descritas na Tabela 5.

Figura 12. Espectros das amostras de açúcar obtido pelo LIBS.



Na Figura 13 segue os espectros das 64 amostras obtidos no instrumento NIRS DS2500 nas condições descritas.

Figura 13. Espectros das amostras de açúcar obtido pelo NIRS.



4.2.1 Cálculo dos modelos iniciais

A partir das 12 normalizações, utilizando o programa MATLAB™ foi empregado a rotina PLS_ed para o cálculo dos modelos PLS e avaliação do desempenho em cada normalização a partir das figuras de mérito SLOPE, INTERCEPT, LV, RMSEC e RSQ para os diferentes parâmetros utilizados na fabricação das amostras, sendo eles Areia (%), Açúcar (%) e Pontos Pretos (PP/100g). As Tabelas de 6 a 8 mostram os resultados obtidos.

Tabela 6. Avaliação dos modelos PLS para as 12 normalizações no parâmetro Areia% (N=61).

Norma	SLOPE	INTCPT	LV	RMSEC	RSQ
1	0,8613	2,466	3	6,94	0,8613
2	0,8374	2,890	2	7,45	0,8374
3	0,8286	3,047	2	7,65	0,8286
4	0,8279	3,060	2	7,67	0,8279
5	0,8280	3,058	2	7,67	0,8280
6	0,8538	2,600	3	7,13	0,8538
7	0,8280	3,057	2	7,67	0,8280
8	0,8383	2,875	2	7,43	0,8383
9	0,8135	3,316	2	7,98	0,8134
10	0,8164	3,263	2	7,92	0,8164
11	0,7719	4,056	3	8,91	0,7719
12	0,7701	4,087	3	8,94	0,7701

Tabela 7. Avaliação dos modelos PLS para as 12 normalizações no parâmetro Açúcar% (N=61).

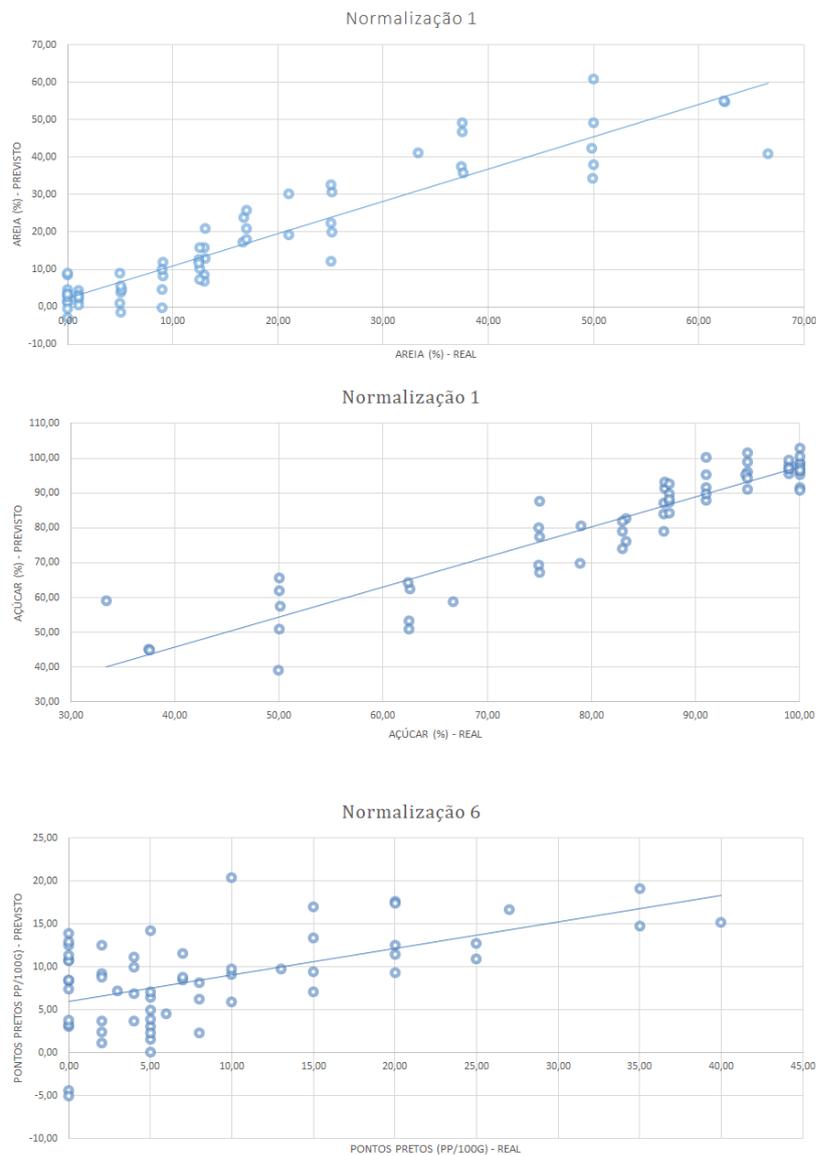
Normalização	SLOPE	INTCPT	LV	RMSEC	RSQ
1	0,8611	11,418	3	6,94	0,8613
2	0,8372	13,380	2	7,45	0,8374
3	0,8284	14,104	2	7,65	0,8286
4	0,8277	14,163	2	7,67	0,8279
5	0,8278	14,155	2	7,67	0,8280
6	0,8536	12,036	3	7,13	0,8538
7	0,8279	14,151	2	7,67	0,8280
8	0,8381	13,308	2	7,43	0,8383
9	0,8133	15,351	2	7,98	0,8134
10	0,8163	15,105	2	7,92	0,8164
11	0,7717	18,770	3	8,91	0,7719
12	0,7700	18,912	3	8,94	0,7701

Tabela 8. Avaliação dos modelos PLS para as 12 normalizações no parâmetro PP% (N=61).

Normalização	SLOPE	INTCPT	LV	RMSEC	RSQ
1	0,2999	6,039	3	8,32	0,3001
2	0,2291	6,651	2	8,66	0,2293
3	0,2100	6,816	2	8,77	0,2101
4	0,2421	6,539	2	8,59	0,2423
5	0,1993	6,908	2	8,82	0,1995
6	0,3075	5,974	2	8,21	0,3077
7	0,2169	6,756	2	8,73	0,2171
8	0,2413	6,546	2	8,59	0,2414
9	0,2427	6,533	2	8,58	0,2429
10	0,2547	6,430	2	8,51	0,2549
11	0,2281	6,659	3	8,74	0,2283
12	0,2401	6,556	3	8,67	0,2403

A partir do menor valor de RMSEC determinamos qual normalização melhor se ajustou ao conjunto de dados para obtenção do modelo de calibração, além disso, avaliamos também as melhores variáveis, além de definir a faixa de trabalho otimizado e por fim a possível proposição de um modelo univariado. Este modelo univariado seria proposto ao empregar apenas uma linha de emissão. Foram escolhidas a normalização 1 para os parâmetros Areia (%) e Açúcar (%) e a normalização 6 para o parâmetro PP (pp/100g). A Figura 14 mostra os gráficos de valores reais versus os previstos para os modelos escolhidos.

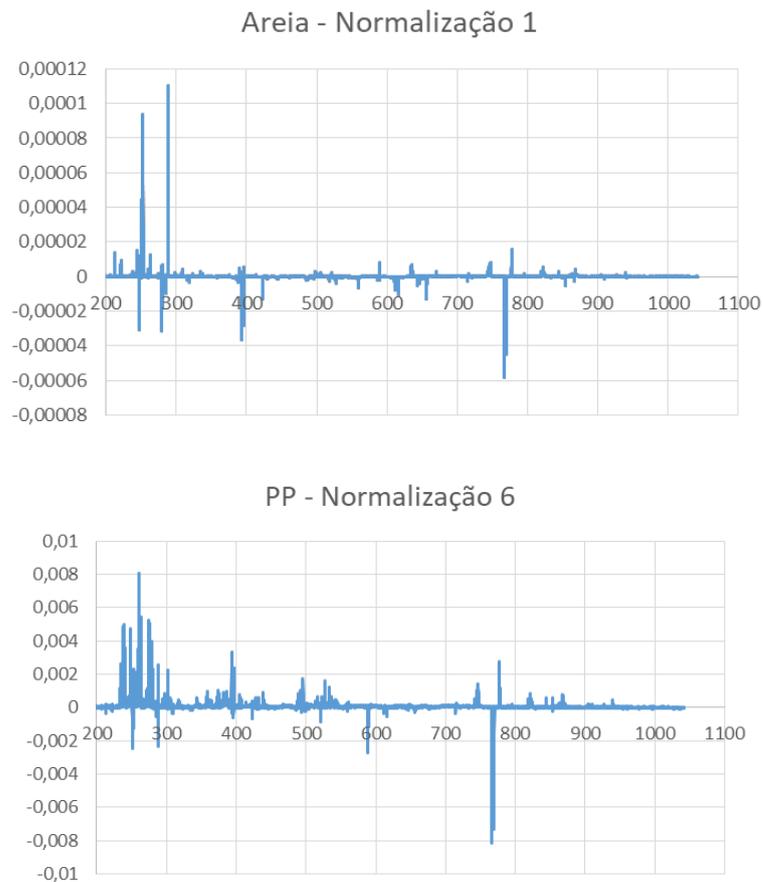
Figura 14. Valores reais versus os previstos para as 3 normalizações escolhidas nos parâmetros Areia (%), Açúcar (%) e Pontos Pretos (PP/100g).



4.2.2 Seleção de Variáveis e Otimização dos Modelos

A seleção de variáveis foi efetuada após escolher o modelo de calibração mais apropriado (menor valor de RMSEC) e quantidades adequadas de variáveis latentes. Além disso, foram observados os valores dos vetores de regressão que apresentaram, em módulo, as maiores influências nos modelos, como mostra a Figura 15.

Figura 15. Intensidade dos vetores de regressão obtidos através dos modelos calculados nas normalizações 1 (LV=3) e 6 (LV=2).



Diferentes seleções de variáveis foram realizadas para buscar a melhor correlação com a informação química da amostra. Por fim, foram avaliados também a redução das faixas de concentração e a elaboração de modelos univariados com a finalidade de reduzir o desvio e obter modelos robustos. A Tabela 9 traz em resumo a identificação da seleção e suas descrições.

Tabela 9. Descrição dos critérios de seleção de variáveis para otimização dos modelos.

Parâmetro	Seleção	Descrição	N de variáveis
Areia	SV1	Linhas de emissão Si	8
	SV2	Linhas de emissão Si e O	11
	SV3	30 Maiores Sinais	30
	SV4	60 Maiores Sinais	60
	SV5	605 Maiores Sinais	605
	Faixa	Conc. Máxima 37,5%	-
	Univariado	Maior Sinal (288,21 nm)	1
Açúcar	SV1	Linhas de emissão C e O	5
	SV2	30 Maiores Sinais	30
	SV3	60 Maiores Sinais	60
	SV4	605 Maiores Sinais	605
	Faixa	Conc. Mínima 75%	-
	Univariado	Maior Sinal (288,21 nm)	1
	PP	SV1	Linhas de emissão Fe
SV2		61 Maiores Sinais	61
SV3		167 Maiores Sinais	167
SV4		1260 Maiores Sinais	1260
Univariado		Maior Sinal (259,96 nm)	1

A partir das diferentes seleções, utilizando o programa MATLAB™ foi empregado a rotina PLS_ed para calcular os modelos PLS e avaliar o desempenho em cada normalização e otimização. Seguindo nas Tabelas 10, 11 e 12 os resultados obtidos.

Tabela 10. Avaliação dos modelos PLS para seleção de variáveis no parâmetro Açúcar%.

Modelo	Seleção	N	SLOPE	INTCPT	LV	RMSEC	RSQ
PLS	Norm1	61	0,861	11,418	3	6,94	0,8613
PLS	SV1	61	0,512	40,137	3	13,03	0,5119
PLS	SV2	61	0,852	12,204	2	7,12	0,8517
PLS	SV3	61	0,842	13,012	2	7,35	0,8419
PLS	SV4	61	0,864	11,188	3	6,87	0,8641
LS	Uni	48	0,774	18,53	-	8,63	0,7750
PLS	Faixa	53	0,765	21,292	4	3,69	0,7645

Tabela 11. Avaliação dos modelos PLS para seleção de variáveis no parâmetro Areia%.

Seleção	N	SLOPE	INTCPT	LV	RMSEC	RSQ
Norm1	61	0,861	2,466	3	6,94	0,8613
SV1	61	0,837	2,889	1	7,39	0,8375
SV2	61	0,851	2,654	2	7,14	0,8507
SV3	61	0,852	2,636	2	7,12	0,8517
SV4	61	0,842	2,811	2	7,35	0,8419
SV5	61	0,864	2,416	3	6,87	0,8641
LS	61	0,774	4,000	-	8,62	0,7750
Faixa	53	0,853	1,787	3	4,12	0,8527

Tabela 12. Avaliação dos modelos PLS para seleção de variáveis no parâmetro PP (pp/100g).

Seleção	N	SLOPE	INTCPT	LV	RMSEC	RSQ
Norm6	61	0,307	5,974	2	8,20	0,3077
SV1	61	0,157	7,271	2	9,05	0,1574
SV2	61	0,290	6,126	2	8,31	0,2900
SV3	61	0,256	6,422	2	8,50	0,2558
SV4	61	0,301	6,032	3	8,31	0,3009
LS	61	0,211	10,094	-	9,52	0,0345

As mesmas 64 amostras nas condições descritas anteriormente foram analisadas via NIRS, com o programa MATLAB™ foi empregada a rotina PLS_ed para obtenção dos modelos PLS e utilizando o programa WinISI™ foi empregada rotina nativa do programa para confecção dos modelos mPLS. Os espectros passaram por pré-tratamento com objetivo de melhorar a previsibilidade dos modelos de regressão, utilizando a rotina nir_sgolay para obtenção do espectro em primeira derivada no programa MATLAB™ e utilizando as configurações nativas do programa WinISI™ para obtenção dos espectros em 1 derivada, com correções de SNV e *Detrend* e aplicação de *Smoothing* e *Gap*. Os parâmetros descritos anteriormente foram utilizados para a avaliação dos modelos para Areia (%), Açúcar (%) e Pontos Pretos (PP/100g) nas tabelas 13 a 15 respectivamente.

Tabela 13. Avaliação dos modelos PLS e mPLS para o parâmetro Areia%.

Modelo	Pré-tratamento	N	SLOPE	INTCPT	LV	RMSEC	RSQ
PLS	Derivada = 1	64	0,9901	0,1708	3	2,656	0,989
	Derivada = 1						
mPLS	SNV e Detrend Smoting = 10 Gap = 10	64	0,9754	0,5216	4	2,148	0,993

Tabela 14. Avaliação dos modelos PLS e mPLS para o parâmetro Açúcar%.

Modelo	Pré-tratamento	N	SLOPE	INTCPT	LV	RMSEC	RSQ
PLS	Derivada = 1	64	0,9919	0,6399	4	2,283	0,992
	Derivada = 1						
mPLS	SNV e Detrend Smoting = 10 Gap = 10	64	0,9752	1,9498	4	2,16	0,9928

Tabela 15. Avaliação dos modelos PLS e mPLS para o parâmetro PP(pp/100g).

Modelo	Pré-tratamento	N	SLOPE	INTCPT	LV	RMSEC	RSQ
PLS	Derivada = 1	64	0,2203	6,4359	4	8,861	0,222
mPLS	Derivada = 1 SNV e Detrend Smoting = 10 Gap = 10	64	0,3969	4,2771	4	7,776	0,406

Foi realizada a fusão dos dados espectroanalíticos a partir da inserção das variáveis obtidas por infravermelho na matriz dos dados de emissão atômica com variáveis selecionadas em etapa anterior. Para o pré-processamento dos dados foi utilizada a rotina `data_pre` no programa MATLAB™ com intuito de auto-escalar as intensidades de sinal de cada técnica, e também através do programa MATLAB™ foi empregado a rotina `PLS_ed` para obter os modelos. Tabela 16 apresenta os modelos e compara os individuais.

Tabela 16. Comparação dos modelos PLS obtidos através da fusão de dados com individuais.

Parâmetro	Dados	Modelo	N	SLOPE	INTCPT	LV	RMSEC	RSQ
Areia (%)	Norm 1 - SV5	PLS	61	0,8641	2,4162	3	6,874	0,864
	NIRS PLS	PLS	64	0,9901	0,1708	3	2,563	0,989
	Fusão	PLS	61	0,9847	0,2721	2	2,287	0,985
Açúcar (%)	Norm 1 - SV4	PLS	61	0,8639	11,188	3	6,874	0,864
	NIRS PLS	PLS	64	0,9919	0,6399	4	2,258	0,992
	Fusão	PLS	61	0,9856	1,1739	3	2,214	0,986
PP (pp/100g)	Norm 6 - SV4	PLS	61	0,30066	6,03250	3	8,319	0,301
	NIRS PLS	PLS	64	0,22389	6,40268	4	8,861	0,222
	Fusão	PLS	61	0,46032	4,65314	4	7,372	0,461

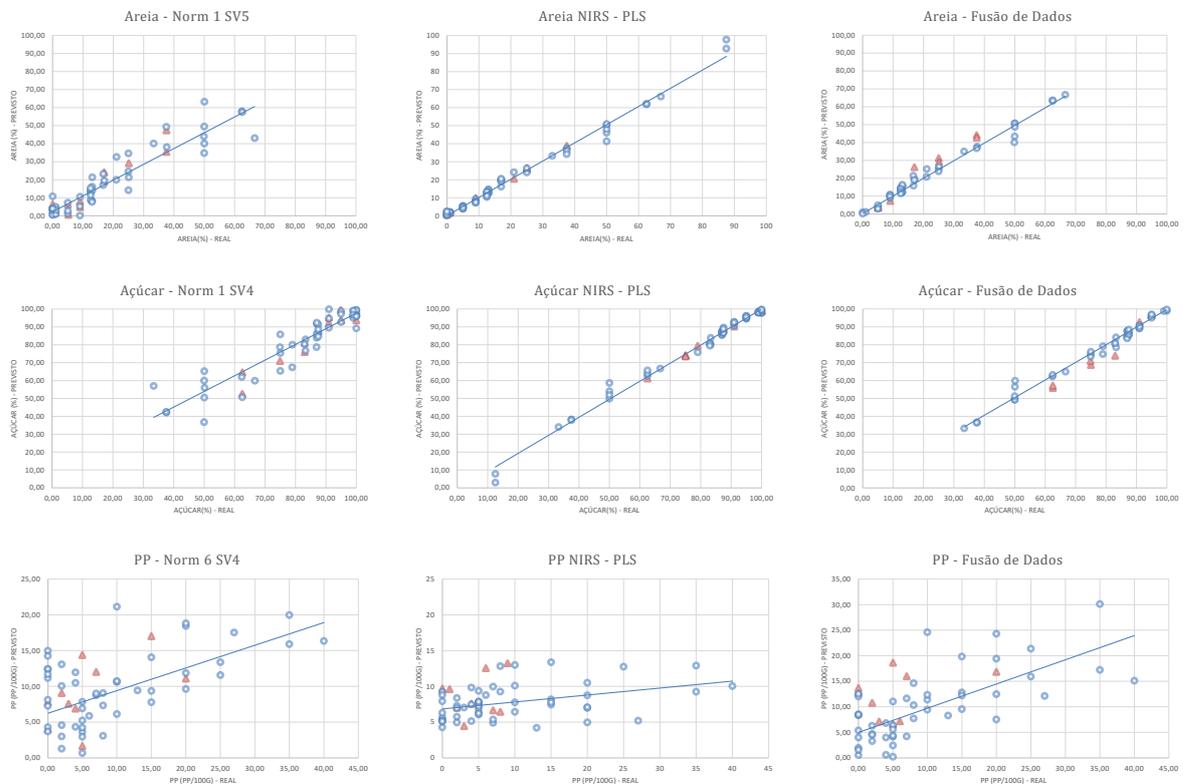
Depois que o pré-processamento foi realizado e seguido pela otimização do modelo, o próximo passo é validação do modelo de calibração. O objetivo aqui é escolher a mais exata e precisa possível e estimar o desempenho em futuras amostras. Se uma quantidade suficiente de amostras de calibração estiver disponível, o melhor método para selecionar e validar um modelo é dividir o conjunto de calibração em dois subconjuntos. Um conjunto é empregado para construir todos os modelos a serem considerados. O segundo conjunto é empregado para escolher o melhor modelo em termos de exatidão e precisão. A melhor estimativa do desempenho futuro de um modelo de calibração é o RMSEP, que se diferencia do RMSEC pela aplicação do modelo no conjunto não pertencente ao grupo de calibração. (BURNS, 2007)

O algoritmo Kennard-Stone a partir da rotina `kestone.m` executada no programa MATLAB™ foi utilizado para seleção das amostras adequadas aos conjuntos de calibração e validação. Tabela 17 mostra os modelos e compara com os individuais e na Figura 16 são apresentados os gráficos de regressão para cada modelo e sua validação.

Tabela 17. Comparação dos modelos utilizando grupos de calibração e validação.

Parâmetro	Modelo	Calibração			Validação			
		N	RMSEC	RSQ	N	BIAS	RMSEP	RSQVal
Areia (%)	Norm1_SV5	51	6,284	0,8762	10	-1,85	6,485	0,9091
	Sand_NIR	54	2,392	0,9886	10	-0,88	1,295	0,9983
	Sand_Norm1_F	51	2,147	0,9851	10	-1,85	5,954	0,9778
Açúcar (%)	Norm1_SV4	51	6,284	0,8762	10	1,86	6,533	0,9091
	Sugar_NIR	54	2,255	0,9887	10	0,90	1,227	0,9983
	Sugar_Norm1_F	51	2,147	0,9851	10	1,86	5,985	0,9778
PP (pp/100g)	Norm6_SV4	51	7,884	0,3166	10	-3,04	7,741	0,3136
	PP_NIR	54	8,661	0,1625	10	-3,89	5,814	0,0415
	PP_Nor6_F	51	7,017	0,4739	10	-3,98	10,941	0,1656

Figura 16. Regressão da calibração (círculos azuis) e validação (triângulos vermelhos) para os modelos otimizados na técnica LIBS, NIRS e Fusão de dados nos parâmetros Areia (%), Açúcar (%) e Pontos Pretos (PP/100g).



4.2.3 Avaliação de diferentes matrizes

Após a elaboração dos modelos, novas amostras com diferentes matrizes foram testadas para verificar a robustez e capacidade desses modelos de prever concentrações de areia em diferentes tipos de açúcar que não foram treinados. Para a técnica LIBS foram elaboradas 5 novas amostras com composição e resultados das leituras e referência apresentados na Tabela 18.

Tabela 18. Resultados previstos para novas misturas

Amostra	Composição	Ref. (Areia%)	Pred. (Areia%)
1	VHP	0,00	0,00*
2	VHP + Areia	10,00	0,00*
3	VHP+ Cristal	0,00	0,00*
4	Cristal	0,00	1,55
5	Cristal + Areia	10,00	5,92

*Valores negativos expressos como zero.

Para avaliar o efeito de matriz na técnica NIRS foram analisadas 10 amostras de diferentes tipos de açúcar com ausência de contaminação e previstos os resultados de Areia (%) a partir do modelo calculado Sand_NIR, os resultados previstos foram expressos na Tabela 19.

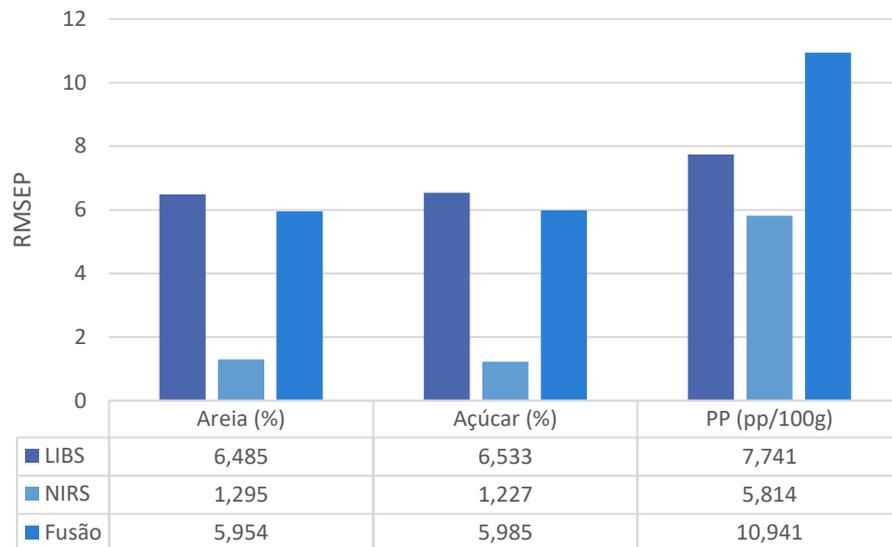
Tabela 19. Resultados de Areia (%) previstos nos modelos NIR para diferentes tipos de açúcar.

Amostra	Cristal	VHP	Refinado	Esperado
1	0,00	13,19	0,00	0,00
2	3,38	32,93	0,00	0,00
3	8,76	24,01	0,00	0,00
4	0,00	41,10	0,00	0,00
5	0,00	34,26	0,00	0,00
6	0,00	33,42	0,00	0,00
7	0,00	26,08	0,00	0,00
8	0,30	40,94	2,75	0,00
9	0,00	35,52	0,13	0,00
10	0,00	35,56	0,00	0,00

*Valores negativos expressos como zero.

Os dados espectroanalíticos das 64 amostras foram obtidos via LIBS e NIRS. Depois da avaliação dos 12 tipos de normalização para os espectros do LIBS foram escolhidas as que obtiveram menor RMSEC para os parâmetros Areia (%), Açúcar (%) e Pontos Pretos (pp/100g). Foi avaliado a melhor normalização para os dados LIBS obtidos e após os modelos passaram otimização através de seleção de variáveis. O mesmo ocorreu para o NIRS com etapa de pré-tratamento e posteriormente confecção dos modelos para os mesmos parâmetros. Por fim, foi realizada a fusão de dados apenas com etapa de auto escalonamento das diferentes variáveis espectrais obtidas (absorbância e emissão). Após a otimização dos modelos, grupos de calibração e validação foram selecionados e os valores de RMSEC da validação são comparados na Figura 17.

Figura 17. Gráficos de barras com valores de RMSEP comparativos para os parâmetros Areia (%), Açúcar (%) e PP (pp/100g) nos modelos obtidos via LIBS, NIRS e Fusão de dados.

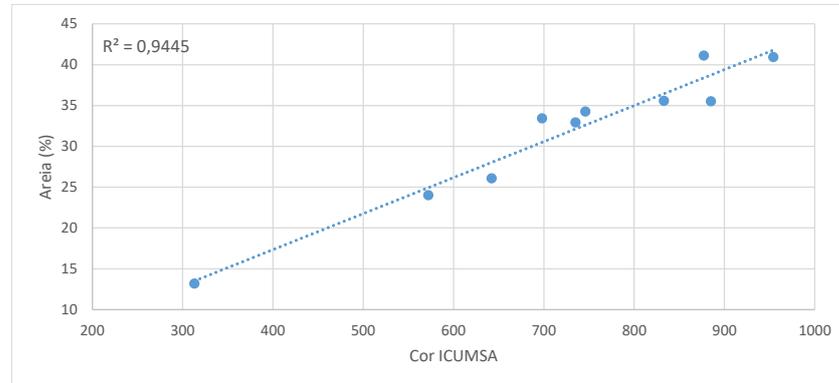


Apesar da técnica LIBS estar se tornando uma opção viável para análises de alimentos, até a redação deste trabalho não foram encontrados na literatura nenhuma publicação relacionada à análise de açúcar, seja para determinação de parâmetros de controle ou ainda para determinação de contaminantes.

Mesmo tendo sido observado desvios maiores do que o NIRS, a LIBS deverá apresentar maior seletividade à areia uma vez que possui linhas de emissão específicas para o Si, devendo ser uma opção tecnicamente mais robusta. O NIRS por sua vez se aproveitou da correlação entre outros parâmetros, como a redução do teor de sacarose ao acréscimo de areia e também a alteração da cor na amostra para obter melhores desvios e correlação, porém não há regiões de absorção das ligações Si-OH ou Si-C na região do infravermelho próximo, sendo assim menos robusto às variações da matriz. Como mostrado nas tabelas 18 e 19 nas duas técnicas foi observado influência da matriz das amostras com dificuldade de previsão ao variar dos tipos de açúcares, sugerindo que para obtenção de aplicações de rotina deverá contemplar diferentes tipos de açúcar.

Para evidenciar da utilização da correlação com outros parâmetros para determinação de areia (%) no instrumento NIRS, utilizamos as amostras de açúcar VHP analisadas na Tabela 19 e correlacionamos os resultados de areia obtidos pelo modelo com os resultados de Cor das amostras, e obtendo correlação de 0,95. Na Figura 18 é apresentada a relação.

Figura 18. Gráfico de dispersão Cor ICUMS versus Areia a partir dos resultados das amostras de açúcar VHP da Tabela 19.



4.2.4 Avaliação dos modelos utilizando Wilcoxon Signed-Rank Test para amostras pareadas

Ao utilizarmos pré-tratamentos, algoritmos ou ainda técnicas diferentes para construção de modelos, e evitar avaliações subjetivas na análise dos resultados, é recomendado realizar comparação estatística entre eles (VOET, 1994). Dessa forma, para comparar o desempenho entre os modelos foi utilizado o Wilcoxon Signed-Rank Test para amostras pareadas (WILCOXON, 1945). VOET, 1994 propôs utilização de testes estatísticos para comparação do RMSEP se baseando na avaliação dos resíduos de previsão dos dois modelos usando como hipóteses:

- i) Hipótese Nula (**H₀**): $RMSEP_{Método1} = RMSEP_{Método2}$
- ii) Hipótese Alternativa (**H₁**): $RMSEP_{Método1} \neq RMSEP_{Método2}$

H₀ é aceita se o p-valor encontrado for superior ao nível de significância (α) adotado (0,05), indicando que não há diferença na exatidão dos métodos. O teste foi aplicado nas diferentes situações descritas abaixo para garantir se realmente havia diferença significativa nos valores de RMSEP (ou RMSEC) obtidos, sendo exibidos os modelos comparados e p-valores obtidos. Com o nível de significância adotado ($\alpha = 0,05$), são avaliadas as evidências para rejeitar ou aceitar H₀ determinando se os dois métodos são equivalentes ou não.

No item 4.2.2 foi realizado a seleção de variáveis e otimização dos modelos, nas tabelas 10 a 12 foram realizadas diferentes seleções e os modelos com menores RMSEC de cada parâmetro foram escolhidos para as etapas subsequentes de

otimização, foi aplicado o teste comparando o modelo inicial com cada seleção de variável, e apesar de termos escolhido o menor valor de RMSEC, na tabela 20 são apresentados os resultados do teste que sugerem que os valores de RMSEC para os algoritmos são equivalentes.

Tabela 20. p-valores obtidos através do teste de Wilconxon para modelos com seleção de variáveis.

Parâmetro	Modelo	p-valor	Ho	Resultado
Areia (%)	SV1	0,6691	Aceita	Equivalentes
	SV2	0,9285	Aceita	Equivalentes
	SV3	0,4752	Aceita	Equivalentes
	SV4	0,4047	Aceita	Equivalentes
	SV5	0,835	Aceita	Equivalentes
Açúcar (%)	SV1	0,9114	Aceita	Equivalentes
	SV2	0,4704	Aceita	Equivalentes
	SV3	0,4088	Aceita	Equivalentes
	SV4	0,8322	Aceita	Equivalentes
Pontos Pretos (pp/100g)	SV1	0,7821	Aceita	Equivalentes
	SV2	0,8801	Aceita	Equivalentes
	SV3	0,6167	Aceita	Equivalentes
	SV4	0,4752	Aceita	Equivalentes

Da mesma forma nas tabelas 13 a 15 foram apresentados os modelos para os diferentes parâmetros obtidos pelo NIRS utilizando algoritmos PLS e mPLS com valores de RMSEC ligeiramente menores para os modelos mPLS, na tabela 21 são apresentados os resultados do teste que sugerem que apesar de serem menores os valores de RMSEC para os dois algoritmos são equivalentes.

Tabela 21. p-valores obtidos através do teste de Wilconxon para diferentes modelos a partir de algoritmos diferentes.

Modelos	p-valor	Ho	Resultado
Tabela 13	0,8910	Aceita	Equivalentes
Tabela 14	0,2876	Aceita	Equivalentes
Tabela 15	0,0731	Aceita	Equivalentes

Capítulo 5

**DESENVOLVIMENTO DE MODELOS DE
PREVISÃO PARA QUANTIFICAÇÃO DE
PARÂMETROS DE CONTROLE DE
QUALIDADE DO AÇÚCAR EMPREGANDO
NIRS**

5 DESENVOLVIMENTO DE MODELOS DE PREVISÃO PARA QUANTIFICAÇÃO DE PARÂMETROS DO CONTROLE DE QUALIDADE DO AÇÚCAR EMPREGANDO NIRS

5.1 Experimental

5.1.1 Obtenção das Amostras

Nesta parte da dissertação foram usados espectros obtidos de diferentes Usinas espalhadas em 4 estados do Brasil: São Paulo, Minas Gerais, Mato Grosso e Tocantins (Figura 19). Em todos os casos foram obtidos espectros com o modelo NIRS DS2500 alocados nos laboratórios de 14 Usinas. Para avaliação e desenvolvimento dos modelos foram utilizados 4 tipos de açúcar: tipo Refinado Amorfo e Refinado Granulado produzidos em uma Usina com refinaria anexa, açúcar do tipo Cristal produzidos em seis destas Usinas e o tipo VHP produzido por dez outras.

Figura 19. Distribuição regional dos instrumentos para coleta de açúcar na região centro-sudeste do Brasil.



As amostras de açúcar, em material sólido cristalino foram obtidas após o processo industrial de centrifugação e secagem, sendo caracterizado como produto final do processo. Conforme rotina analítica do laboratório industrial, as amostras foram coletadas do processo e enviadas diretamente para o laboratório para análise em momento simultâneo à aquisição dos espectros, não havendo assim etapa de preparo ou armazenamento das amostras. As amostras foram submetidas a um processo de homogeneização (manual ou por quarteador de amostras) sendo realizada a aquisição espectral e em seguida as análises de referência nos

parâmetros Cinzas (%), Umidade (%), Pol (°Z) e Cor ICUMSA (UI) conforme métodos de referência. As amostras foram analisadas em temperatura ambiente com resultados reportados como recebido (base) totalizando 41.110 amostras, distribuídas conforme Tabela 20, sendo utilizado 80% das amostras para cálculo dos modelos e 20% para validação.

Tabela 22. Quantidade de amostras por tipo de açúcar.

Tipo de Açúcar	Quantidade de Amostras
VHP	16.185
Cristal	17.507
Refinado Amorfo	3.573
Refinado Granulado	3.845

5.1.2 Métodos de referência

Em todos os laboratórios participantes foram utilizados os mesmos métodos de referência para análise das amostras obtidas. Sendo a ICUMSA o órgão responsável pela uniformização dos métodos para análise de açúcar, cada parâmetro que segue foi analisado conforme descrição e método ICUMSA indicado.

A Cor ICUMSA é um índice de absorvância de uma solução açucarada, em concentrações e pH definidos, cuja absorvância é medida em 420 nm (quatrocentos e vinte nanômetros) de acordo com o 'Determination of Sugar Solution Color at pH 7.0 by the MOPS Buffer Method according to ICUMSA Method GS9/1/2/3-8'. Resultados são expressos em UI, unidade ICUMSA.

A Polarização de uma solução açucarada é determinada pelo desvio provocado pelos constituintes desta solução (sacarose entre outros) no plano da luz polarizada de um instrumento polarímetro, sendo quantidade de sacarose aparente medida totalizando o peso normal de 26,000 g (vinte e seis gramas) em 100 ml (cem mililitros) de água a 20°C (vinte graus Celsius), de acordo com "Polarisation of Raw Sugar by Polarimetry according ICUMSA Method GS 1/2/3/9-1 (2011)". Resultados são expressos em °Z, (grau Zucker) escala internacional de açúcar.

Determina-se cinzas o teor de sais solúveis ionizados presentes em uma solução açucarada através de medidas em unidade de condutividade elétrica. O método baseia-se na grandeza da condutividade elétrica para determinar o teor de cinzas no açúcar, de acordo com 'Conductivity Ash in Refined Sugar Products and

Plantation White Sugar according ICUMSA Method GS 2/3/9-17 (2011). Resultados são expressos em porcentagem (%).

Teor de umidade, mais corretamente perda por secagem, determinada através de gravimetria por secagem em condições específicas. Sendo utilizado o determinador de umidade por secagem com infravermelho ajustado a partir do método 'The Determination Of Sugar Moisture by Loss On Drying – Official ICUMSA Method GS 2/1/3/9-15 (2007)'. Resultados são expressos em gramas por 100g do produto (g/100g).

5.1.3 Avaliação das Metodologias de Referência

Diversas fontes de erro são introduzidas nos dados de calibração produzidos pelo método analítico primário. Para cada técnica usada como método primário, existem erros inerentes à medição e estes erros devem ser considerados ao desenvolver o modelo de calibração. O erro aleatório nos dados químicos deve ser reduzido ao nível mínimo possível para realizar a calibração ideal (BURNS, 2007). Quando utilizamos diversas origens de dados para confecção de uma calibração robusta, devemos considerar a reprodução destes métodos primários em diferentes origens seja para minimizar os erros da calibração, mas também para avaliar os erros de previsão obtidos pelos modelos.

Uma forma confiável e objetiva para avaliar diversos laboratórios é através dos dados obtidos por programas de ensaios de proficiência, que através de comparações interlaboratoriais são usados para determinar o desempenho para ensaios ou medições específicas (THOMPSON, 1993; ISO 13528, 2015). Desde o ano 2000 o Centro de Tecnologia Canavieira realiza programas de ensaios de proficiência para análise de açúcar nas usinas do Brasil, e com fim das atividades do setor de laboratórios do Centro em 2016, o ensaio passou a ser aplicado pela Consul-Lab – Consultoria para Laboratórios. A mesma gentilmente forneceu os dados do “Ensaio de Proficiência de Açúcar, 1ª Rodada” de 2019 para ser calculado o valor de SEL (Standard Error of Laboratory) e comparado, em conjunto com os limites de reprodutibilidade estabelecidos pelos organizadores do ensaio, com os modelos de calibração obtidos nessa dissertação.

Por ser um serviço prestado pela empresa, a Consul-Lab garante sigilo absoluto quanto ao nome e resultados dos laboratórios, sendo adotado um sistema

de codificação numérico para identificar o laboratório participante e foi disponibilizado apenas os dados de relatório para esta análise.

O valor de SEL pode ser determinado utilizando uma amostra analisada em replicata por um ou por mais laboratórios (BURNS, 2007). A média do valor analítico para as réplicas de uma única amostra é determinada pela equação 11.

$$ym = \frac{\sum_{i=1}^N y_i}{N} \quad (11)$$

Onde o valor y_i é o resultado determinado por cada laboratório ou réplica e N o número de laboratórios participantes ou réplicas analisadas. Sendo valor de SEL obtido através da equação 12.

$$SEL = \left\{ \frac{\sum_{i=1}^N (y_i - ym)^2}{N} \right\}^{1/2} \quad (12)$$

Onde o valor y_i é o resultado determinado por cada laboratório ou réplica, ym a média do valor analítico determinado para as réplicas e N o número de laboratórios participantes ou réplicas analisadas.

A reprodutibilidade foi definida como a diferença entre dois resultados de ensaios, individuais e independentes, obtidos por operadores diferentes, com a mesma aparelhagem, trabalhando em laboratórios distintos e em amostras idênticas, a longo prazo, com a execução correta e normal deste método, pode exceder o limite somente em um caso em 20. Os limites de reprodutibilidade, R, utilizados nos cálculos foram obtidos do histórico de ensaios de proficiência realizados desde 2000 no Centro de Tecnologia Canavieira e na Consul-LAB (CONSUL-LAB, 2019).

Para realização do ensaio foram enviadas três amostras de açúcar cristal branco (B19, B20 e B21), três amostras de Açúcar VVHP (VV19, VV20 e VV21), e três amostras de açúcar VHP (V16, V17 e V18) para a realização das análises seguindo as metodologias conforme item 5.1.2.

Nas Tabelas 21 a 23 são obtidos os resumos estatísticos para os resultados obtidos nos ensaios para Cor ICUMSA, Cinzas e Polarização, conforme o tipo de açúcar e quantidade de participantes em conjunto com o valor de SEL calculado através dos resultados individuais de cada participante e o limite de reprodutibilidade estabelecido pelo fornecedor do ensaio.

Tabela 23. Avaliação para o parâmetro Cor ICUMSA (UI).

Amostra	Participantes	Mediana	Média	SEL	R
B19	49	105	104	9	10
B20	49	153	152	8	15
B21	49	163	162	11	15
VV19	17	331	332	18	38
VV20	17	517	515	17	38
VV21	16	414	411	10	38
V16	50	872	869	38	50
V17	50	974	971	27	50
V18	50	765	773	36	50

Tabela 24. Avaliação para o parâmetro Polarização ($^{\circ}Z$).

Amostra	Participantes	Mediana	Média	SEL	R
B19	49	99,83	99,82	0,10	0,070
B20	49	99,79	99,78	0,09	0,070
B21	49	99,78	99,76	0,13	0,070
VV19	17	99,65	99,64	0,12	0,10
VV20	17	99,55	99,57	0,09	0,10
VV21	16	99,66	99,65	0,05	0,10
V16	50	99,41	99,43	0,10	0,14
V17	50	99,37	99,36	0,09	0,17
V18	50	99,44	99,44	0,10	0,17

Tabela 25. Avaliação para o parâmetro Cinzas (% m/m)

Amostra	Participantes	Mediana	Média	SEL	R
B19	49	0,060	0,058	0,009	0,010
B20	49	0,032	0,031	0,004	0,010
B21	49	0,036	0,035	0,007	0,010
VV19	17	0,103	0,104	0,005	0,010
VV20	17	0,141	0,140	0,006	0,010
VV21	16	0,090	0,090	0,004	0,010
V16	50	0,100	0,100	0,006	0,017
V17	50	0,061	0,062	0,004	0,015
V18	50	0,094	0,094	0,006	0,017

O parâmetro Umidade não é avaliado neste ensaio. Variações causadas devido as diferentes condições de temperatura, umidade ambiente e armazenagem podem levar à uma alteração nas amostras do ensaio, invalidando assim a avaliação.

5.1.4 Preparo e obtenção dos dados espectroanalíticos

Para obtenção dos espectros de infravermelho próximo foi utilizado o Espectrofotômetro de infravermelho próximo modelo DS2500 da fabricante dinamarquesa FOSS com faixa espectral de 400 a 2500 nm.

Foram utilizadas aproximadamente 100g das amostras previamente homogeneizadas sem preparo. Para leitura das amostras foi utilizado o acessório *Large cup*, em módulo de reflectância difusa, conforme Figura 20. A condição padrão de aquisição de espectros no instrumento foi definido realizando a leitura em 8 posições diferentes com 4 repetições por posição, obtendo assim o espectro médio gerado a partir de 32 subspectros.

Figura 20. (Esquerda) NIRS modelo DS2500 da fabricante FOSS. (Direita) Amostra de açúcar na célula de leitura Large Cup.



5.1.5 Fatores, Cross-Validação e Pré-tratamentos

O programa FOSSCalibrator™ foi utilizado para cálculo dos modelos de calibração. Em sua configuração nativa o programa nos permite a seleção do tipo de algoritmo de cross-validação para otimização das variáveis latentes, ou fatores. Também permite aplicar diferentes tipos de pré-tratamentos, derivativos e selecionar as regiões ou variáveis de interesse.

Para os modelos mPLS foram definidas as configurações conforme Tabela 24. Também selecionada a faixa espectral completa do instrumento com variáveis obtidas na resolução de 0,5nm, sendo excluídos regiões iniciais e finais de cada detector (Vis-NIR) a fim de evitar tais regiões com excesso de ruído.

Tabela 26. Parâmetros de calibração para modelos mPLS.

Fatores e Cross Validação		
Modelo	mPLS	
Tipo de Cross Validação	Venetian Blinds	
Número de Blocos de Amostras	4	
Faixa de Fatores	1 – 16 – 1	
Pré-tratamentos		
Tipo	SNV e Detrend	
Derivativos	Derivative	1
	Gap	16
	Smooth	16
	Smooth 2	1
Seleção de Variáveis		
Faixa Espectral	407-1091,5 (0,5)	
	1107,5 – 2491,5 (0,5)	

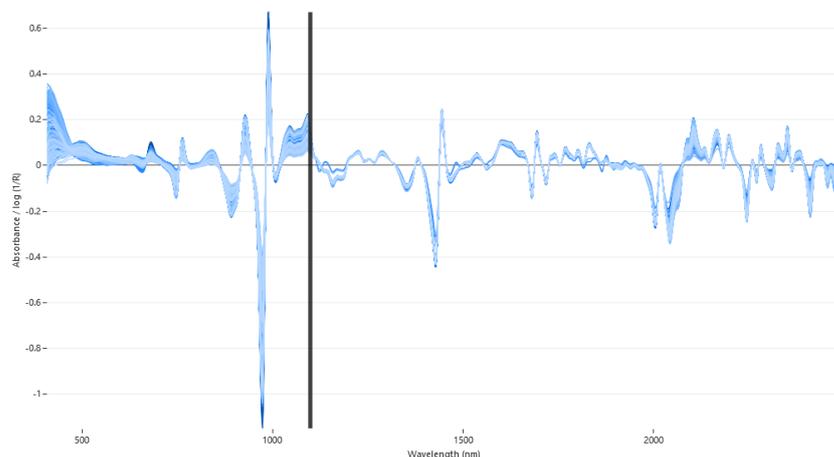
Para cálculo dos modelos ANN também foi utilizado programa FOSSCalibrator™ que além das configurações de cross-validação, pré-tratamentos e seleção, nos permitiu a seleção dos parâmetros de complexidade para otimização das iterações no algoritmo de redes neurais artificiais, conforme Tabela 25.

Tabela 27. Parâmetros de calibração para modelos ANN.

Fatores e Cross Validação		
Modelo	ANN	
Tipo de Cross Validação	Venetian Blinds	
Número de Blocos de Amostras	4	
Fatores	Input	20
	Hidden	3
Complexidade		
Método de Seleção	Automático	
Input Neurons	20 – 22 -1	
Hidden Neurons	3 – 5 – 1	
Seeds	3	
Pré-tratamentos		
Tipo	SNV e Detrend	
Derivativos	Derivative	1
	Gap	16
	Smooth	16
	Smooth 2	1
Seleção de Variáveis		
Faixa Espectral	407-1091,5 (0,5)	
	1107,5 – 2491,5 (0,5)	

Na Figura 21 temos um exemplo de espectros de açúcar obtidos pelo NIRS DS2500 sendo aplicados os pré-tratamentos com correção de espalhamento e derivativos.

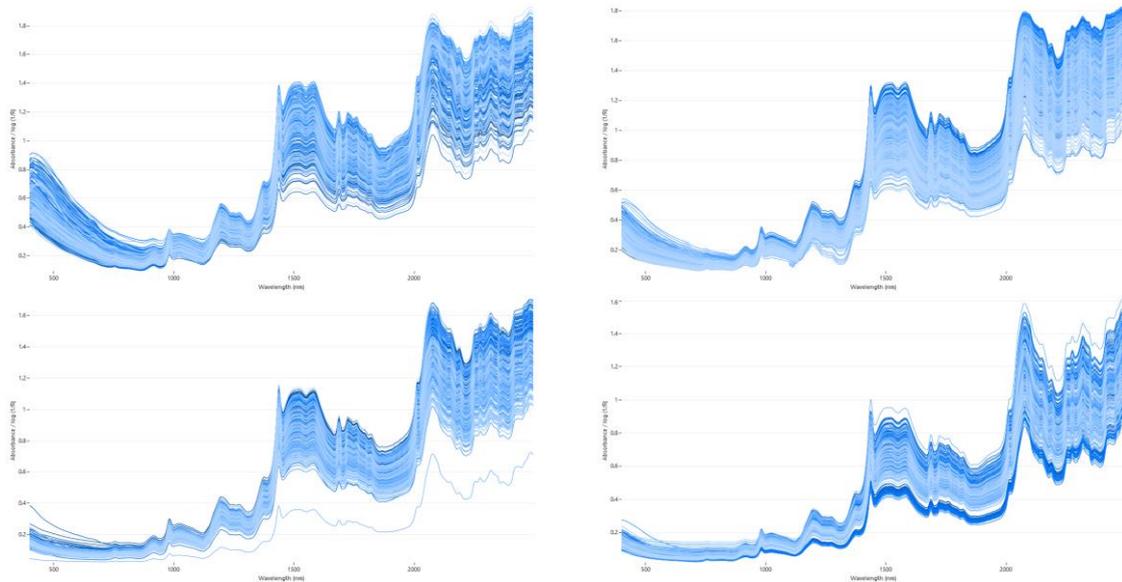
Figura 21. Espectros de açúcar utilizando pré-tratamento de SNV (*only scale*) e *detrend* com derivativos em 1, 16, 16, 1.



5.2 Resultados e Discussões

Na Figura 22 são apresentados os espectros das amostras de açúcar VHP, Cristal, Refinado Granulado e Refinado Amorfo obtidos no instrumento NIRS DS2500 nas condições estabelecidas.

Figura 22. Espectros de Açúcar obtidos pelo NIRS DS2500



5.2.1 Cálculo dos modelos de calibração

A partir dos dados espectrais obtidos e os resultados das análises físico-químicas realizadas pelos laboratórios foi possível o cálculo dos modelos de previsão para os parâmetros Cinzas (% m/m), Cor ICUMSA (UI), Umidade (%) e Polarização (°Z). Foram calculados modelos para os diferentes tipos de açúcar, sendo eles VHP, Cristal, Refinado Amorfo e Refinado Granulado, e último foi calculado modelos Global com a junção dos dados espectrais de todos os tipos de açúcar.

Para desenvolvimento dos modelos de previsão foram utilizados os programas FOSSCalibrator™ e WINISI™, sendo utilizados os pré-tratamentos e configurações de acordo com o item 5.1.5.

O conjunto de amostras foi dividido em dois grupos, sendo utilizado 80% das amostras para confecção dos modelos e 20% para validação. Os grupos foram divididos utilizando o algoritmo de seleção nativo do programa mantendo a distribuição temporal e de faixa dos parâmetros. Conjuntos ou parâmetros com pequenas quantidades amostras foram calculados apenas modelos utilizando mPLS, foi evitado utilizar ANN nesses grupos para evitar sobreajuste e um comparativo errôneo.

Os resumos estatísticos e desempenho dos modelos são descritos para cada tipo de açúcar nos itens que se seguem.

5.2.1.1 Açúcar VHP

Na Tabela 26 são apresentadas as descrições estatísticas dos grupos de validação e calibração obtidos para açúcar VHP, reportando o número de amostras utilizadas (N), valores máximos e mínimos nos grupos, média e desvio padrão por parâmetro.

Tabela 28. Estatísticas dos grupos de calibração e validação de Açúcar VHP

Parâmetro	Grupo	N	Min	Max	Média	D. Padrão
Cinzas	Calibração	10113	0,01	0,03	0,10	0,03
	Validação	2494	0,01	0,25	0,10	0,03
Cor	Calibração	5065	237	2787	797	269
	Validação	1222	292	1961	799	264
Umidade	Calibração	6357	0,02	0,32	0,09	0,03
	Validação	1540	0,02	0,26	0,08	0,03
Pol	Calibração	4456	98,24	99,92	99,32	0,17
	Validação	1092	98,56	99,86	99,32	0,17

De acordo com o item 5.1.6 foram desenvolvidos dois modelos para cada parâmetro do açúcar VHP, sendo um modelo mPLS e outro ANN e avaliadas seus desempenhos através das figuras de RMSEC, RMSEP, Slope, Intercet, BIAS e RSQ com valores descritos na Tabela 27.

Tabela 29. Desempenho de calibração e validação para Açúcar VHP

Parâm.	Modelo	RMSEC	RMSECV	RMSEP	Slope/Inter	Bias	RSQ	LV
Cinzas	mPLS	0,018	0,019	0,018	0,99/0,00	0,00	0,65	16
	ANN	0,017	0,018	0,018	0,99/0,00	0,00	0,68	20
Cor	mPLS	92	93	93	1,01/-0,82	5	0,88	16
	ANN	74	78	77	1,01/-6,49	7	0,2	22
Umidade	mPLS	0,018	0,018	0,019	0,98/0,00	0,00	0,71	16
	ANN	0,017	0,018	0,018	0,99/0,00	0,00	0,74	21
Pol	mPLS	0,123	0,126	0,124	1,01/-0,75	0,00	0,49	16
	ANN	0,117	0,125	0,126	0,94/5,93	0,00	0,48	21

Nas Figuras 23 e 24 são apresentados os gráficos de dispersão dos resultados analisados versus previstos para os grupos de calibração e validação dos diferentes parâmetros nos modelos mPLS e ANN, respectivamente.

Figura 23. Regressão para os modelos mPLS, calibração (círculo azul) e validação (triângulo laranja) para parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol no açúcar VHP.

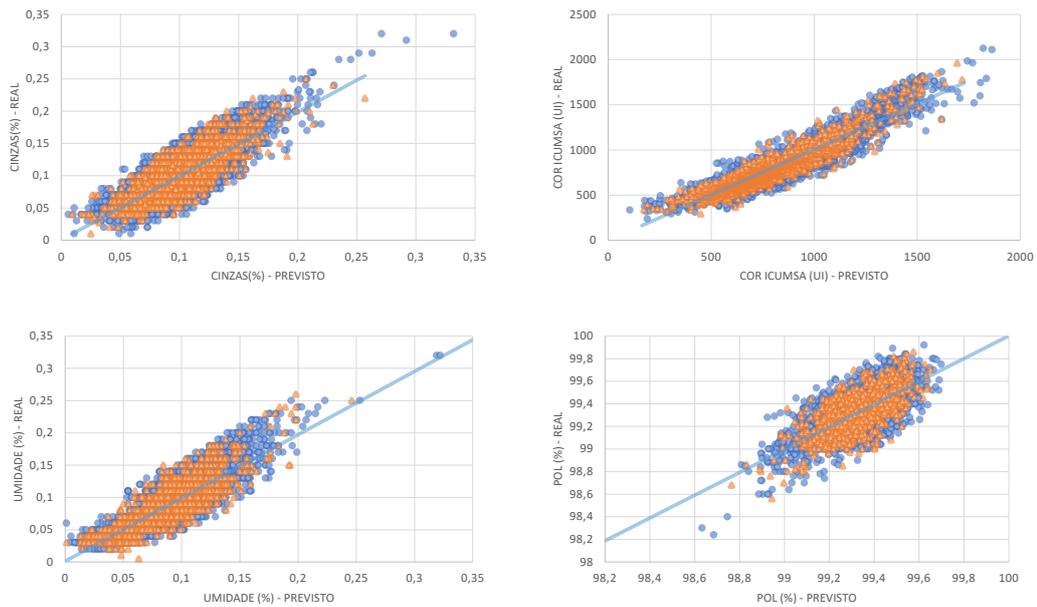
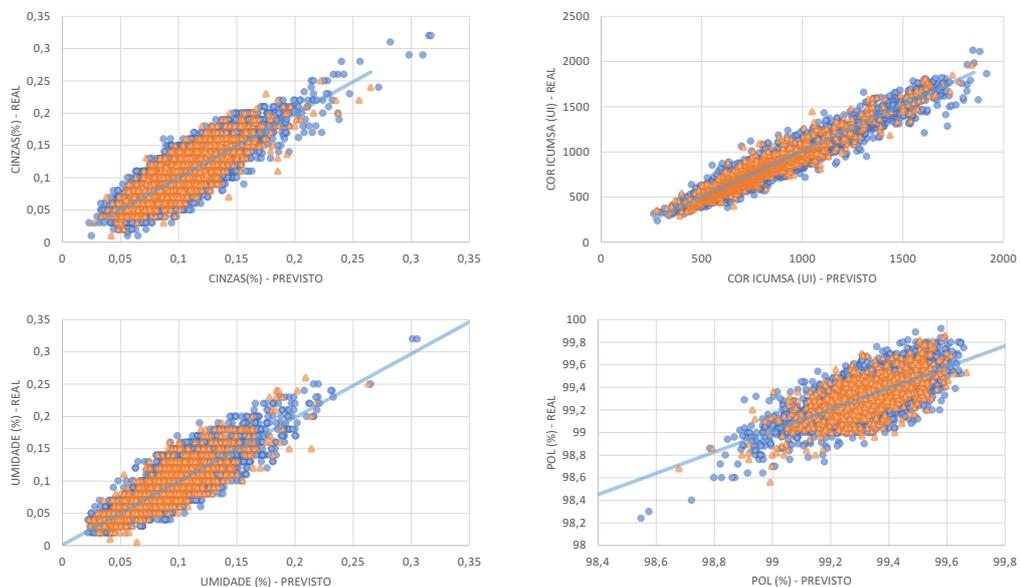


Figura 24. Regressão para os modelos ANN calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol no açúcar VHP.



5.2.1.2 Açúcar Cristal

Na Tabela 28 são apresentadas as descrições estatísticas dos grupos de validação e calibração para açúcar Cristal, reportando o número de amostras utilizadas (N), valores máximos e mínimos nos grupos, média e desvio padrão por parâmetro.

Tabela 30. Estatísticas dos grupos de calibração e validação de Açúcar Cristal

Parâmetro	Grupo	N	Min	Max	Média	D. Padrão
Cinzas	Calibração	3191	0,01	0,12	0,035	0,015
	Validação	770	0,01	0,12	0,035	0,016
Cor	Calibração	11987	41	735	166	86
	Validação	3019	58	582	164	86
Umidade	Calibração	3222	0,01	0,08	0,04	0,011
	Validação	781	0,01	0,08	0,04	0,012
Pol	Calibração	3044	99,45	99,98	99,77	0,08
	Validação	744	99,46	99,96	99,78	0,08

Os modelos mPLS e ANN são mostrados na Tabela 29.

Tabela 31. Desempenho de calibração e validação para Açúcar Cristal

Parâm.	Modelo	RMSEC	RMSECV	RMSEP	Slope/Inter	Bias	RSQ	LV
Cinzas	mPLS	0,007	0,008	0,008	1,02/0,00	0,00	0,78	16
	ANN	0,007	0,008	0,008	0,99/0,00	0,00	0,78	16
Cor	mPLS	15	15	15	1,00/-0,31	0,140	0,97	16
	ANN	13	13	13	1,00/0,08	0,033	0,98	22
Umidade	mPLS	0,007	0,008	0,008	0,99/0,00	0,001	0,58	16
	ANN	0,007	0,007	0,008	0,96/0,00	0,00	0,57	20
Pol	mPLS	0,033	0,035	0,037	0,98/2,36	0,00	0,79	16
	ANN	0,029	0,033	0,035	0,99/1,21	0,00	0,82	21

Nas Figuras 25 e 26 são apresentados os gráficos de dispersão dos resultados analisados versus previstos para os grupos de calibração e validação dos diferentes parâmetros nos modelos mPLS e ANN, respectivamente

Figura 25. Regressão para os modelos mPLS, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol no açúcar Cristal.

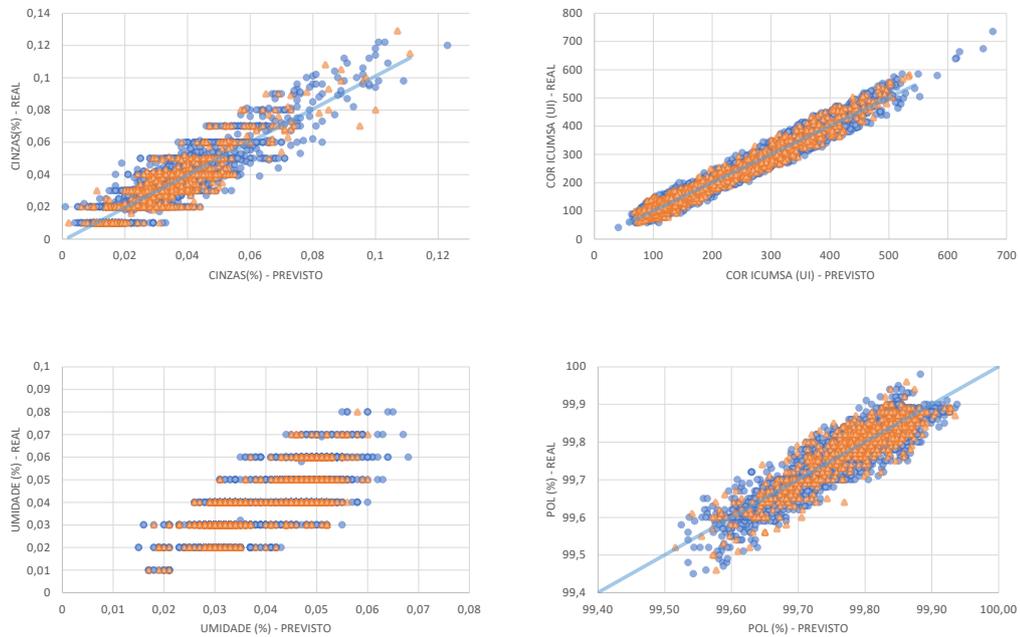
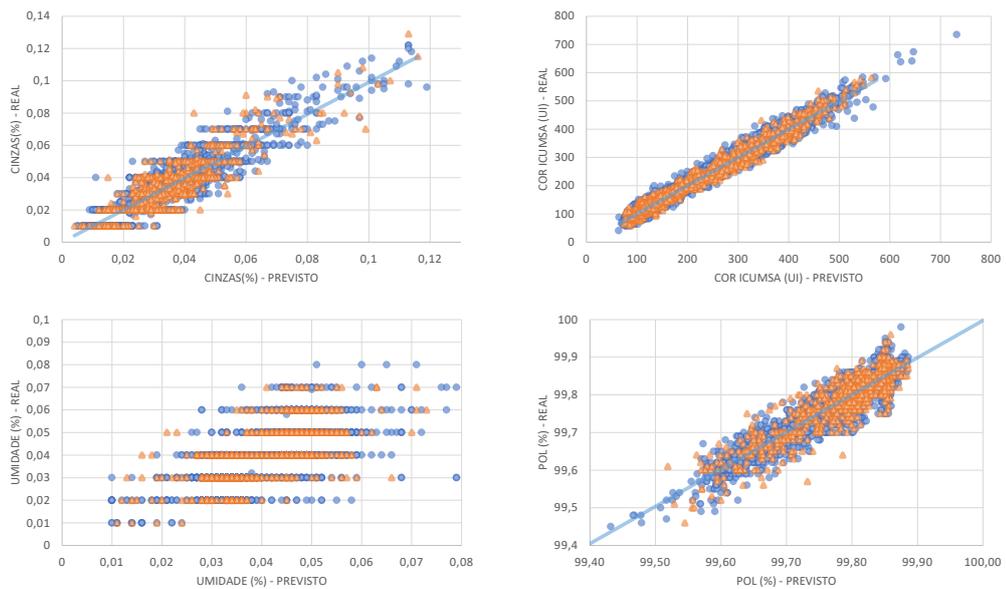


Figura 26. Regressão para os modelos ANN, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol no açúcar Cristal



5.2.1.3 Açúcar Refinado Amorfo

Na Tabela 30 são apresentadas as descrições estatísticas dos grupos de validação e calibração para açúcar Refinado Amorfo, reportando o número de amostras utilizadas (N), valores máximos e mínimos nos grupos, média e desvio padrão por parâmetro.

Tabela 32. Estatísticas dos grupos de calibração e validação de Açúcar Ref. Amorfo

Parâmetro	Grupo	N	Min	Max	Média	D. Padrão
Cinzas	Calibração	282	0,03	0,22	0,10	0,026
	Validação	72	0,04	0,19	0,10	0,025
Cor	Calibração	2151	24	173	72	19
	Validação	549	24	164	72	20
Umidade	Calibração	268	0,04	0,24	0,15	0,03
	Validação	69	0,08	0,21	0,14	0,03
Pol	Calibração	248	99,08	99,60	99,30	0,084
	Validação	61	99,08	99,48	99,29	0,088

Os modelos mPLS e ANN são mostrados na Tabela 31.

Tabela 33. Desempenho de calibração e validação para Açúcar Refinado Amorfo

Parâm.	Modelo	RMSEC	RMSECV	RMSEP	Slope/Inter	Bias	RSQ	LV
Cinzas	mPLS	0,009	0,0112	0,014	0,93/0,01	0,00	0,70	15
	ANN	-	-	-	-	-	-	-
Cor	mPLS	5	6	5	1,02/-1,48	-0,06	0,93	16
	ANN	5	6	5	1,01/-1,48	-0,12	0,93	22
Umidade	mPLS	0,019	0,020	0,021	1,15/-0,03	-0,00	0,57	5
	ANN	-	-	-	-	-	-	-
Pol	mPLS	0,067	0,073	0,086	0,54/45,67	-0,00	0,10	5
	ANN	-	-	-	-	-	-	-

Nas Figuras 27 e 28 são apresentados os gráficos de dispersão dos resultados analisados versus previstos para os grupos de calibração e validação dos diferentes parâmetros nos modelos mPLS e ANN, respectivamente.

Figura 27. Regressão para os modelos mPLS, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol no açúcar Refinado Amorfo.

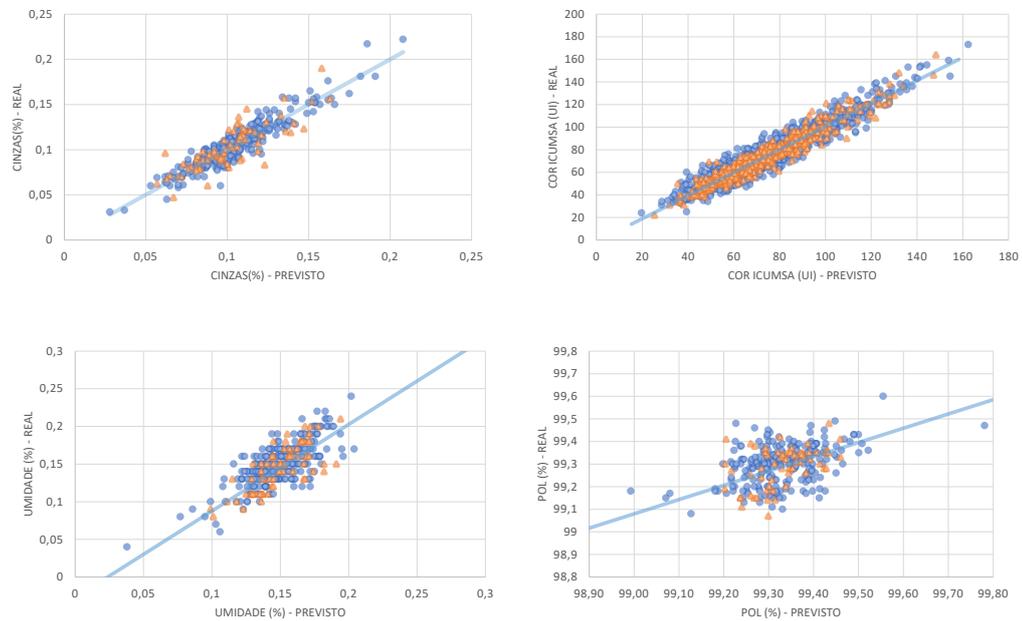
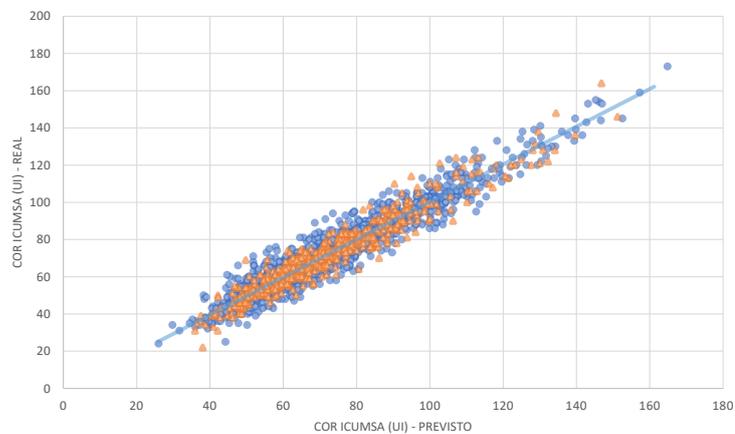


Figura 28. Regressão para os modelos ANN, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cor no açúcar Refinado Amorfo.



5.2.1.4 Açúcar Refinado Granulado

Na Tabela 32 são apresentadas as descrições estatísticas dos grupos de validação e calibração para açúcar Refinado Granulado, reportando o número de amostras utilizadas (N), valores máximos e mínimos nos grupos, média e desvio padrão por parâmetro.

Tabela 34. Estatísticas dos grupos de calibração e validação de Açúcar Refinado Granulado

Parâmetro	Grupo	N	Min	Max	Média	D. Padrão
Cinzas	Calibração	217	0,00	0,02	0,007	0,003
	Validação	51	0,00	0,02	0,008	0,003
Cor	Calibração	2330	7	124	32	14
	Validação	602	7	113	31	13
Umidade	Calibração	1060	0,00	0,06	0,026	0,009
	Validação	263	0,00	0,05	0,027	0,01
Pol	Calibração	185	99,89	99,97	99,92	0,014
	Validação	43	99,89	99,96	99,92	0,012

Os modelos mPLS e ANN são mostrados na Tabela 33.

Tabela 35. Desempenho de calibração e validação para Açúcar Refinado Granulado

Parâm.	Modelo	RMSEC	RMSECV	RMSEP	Slope/Inter	Bias	RSQ	LV
Cinzas	mPLS	0,001	0,002	0,002	1,01/0,00	0,00	0,72	15
	ANN	-	-	-	-	-	-	-
Cor	mPLS	4	4	4	0,98/0,40	-0,35	0,91	16
	ANN	3	4	4	0,97/0,50	-0,45	0,91	21
Umidade	mPLS	0,008	0,009	0,009	1,16/0,00	0,00	0,25	9
	ANN	0,008	0,009	0,009	0,76/0,01	0,00	0,16	20
Pol	mPLS	0,015	0,015	0,016	0,78/22,0	0,00	0,019	2
	ANN	-	-	-	-	-	-	-

Nas Figuras 29 e 30 são apresentados os gráficos de dispersão dos resultados analisados versus previstos para os grupos de calibração e validação dos diferentes parâmetros nos modelos mPLS e ANN, respectivamente

Figura 29. Regressão para os modelos mPLS, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol no açúcar Refinado Granulado.

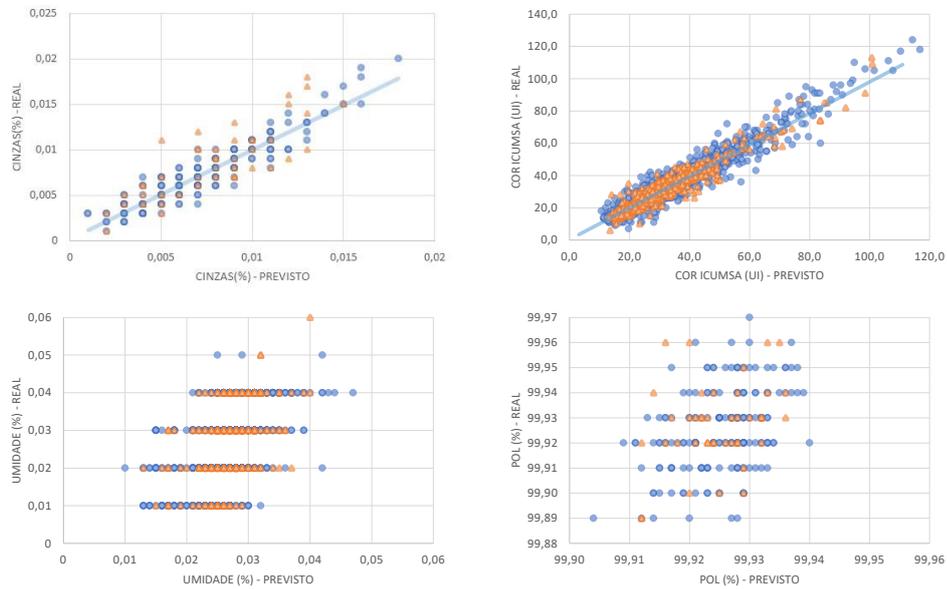
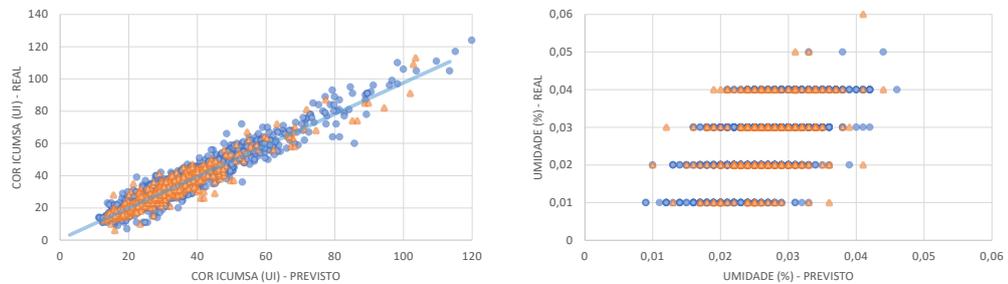


Figura 30. Regressão para os modelos ANN, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cor e (B) Umidade no açúcar Refinado Granulado.



5.2.1.5 Açúcar Global

Na Tabela 34 são apresentadas as descrições estatísticas dos grupos de validação e calibração para o conjunto de todas os tipos de açúcar, chamados aqui conjunto Global, reportando o número de amostras utilizadas (N), valores máximos e mínimos nos grupos, média e desvio padrão por parâmetro.

Tabela 36. Estatísticas dos grupos de calibração e validação para dados Globais

Tipo	Parâmetro	Grupo	N	Min	Max	Média	D. Padrão
Global (Todos)	Cinzas	Calibração	13803	0,00	0,32	0,09	0,04
		Validação	3387	0,00	0,25	0,08	0,04
	Cor	Calibração	21533	7	2787	288	318
		Validação	5392	7	1961	283	316
	Umidade	Calibração	10907	0,00	0,32	0,07	0,04
		Validação	2653	0,00	0,26	0,07	0,04
	Pol	Calibração	7933	98,24	99,98	99,51	0,27
		Validação	1940	98,56	99,96	99,51	0,27

Além de variabilidade espectral, com produtos de características que podem ser bastante distintas, o agrupamento dos diferentes tipos de açúcar introduz faixas que podem ser extensas ou ainda apresentar relações de linearidade ou não-linearidade entre si.

Os modelos mPLS e ANN são mostrados na Tabela 35.

Tabela 37. Desempenho de calibração e validação para dados globais de Açúcar

Parâm.	Modelo	RMSEC	RMSECV	RMSEP	Slope/Inter	Bias	RSQ	LV
Cinzas	mPLS	0,017	0,018	0,017	1,00/0,00	0,00	0,82	16
	ANN	0,016	0,017	0,017	1,00/00	0,00	0,83	22
Cor	mPLS	71	72	71	1,00/0,87	2	0,95	16
	ANN	43	43	43	1,01/-2,19	1	0,98	21
Umidade	mPLS	0,017	0,017	0,017	0,99/0,00	0,00	0,81	16
	ANN	0,015	0,016	0,016	0,98/0,00	0,00	0,83	22
Pol	mPLS	0,101	0,102	0,101	1,01/-1,22	0,00	0,86	16
	ANN	0,096	0,101	0,098	1,01/-0,82	0,00	0,87	20

Nas Figuras 31 e 32 são apresentados os gráficos de dispersão dos resultados analisados versus previstos para os grupos de calibração e validação dos diferentes parâmetros nos modelos mPLS e ANN respectivamente

Figura 31. Regressão para os modelos mPLS, , calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol para dados globais de açúcar.

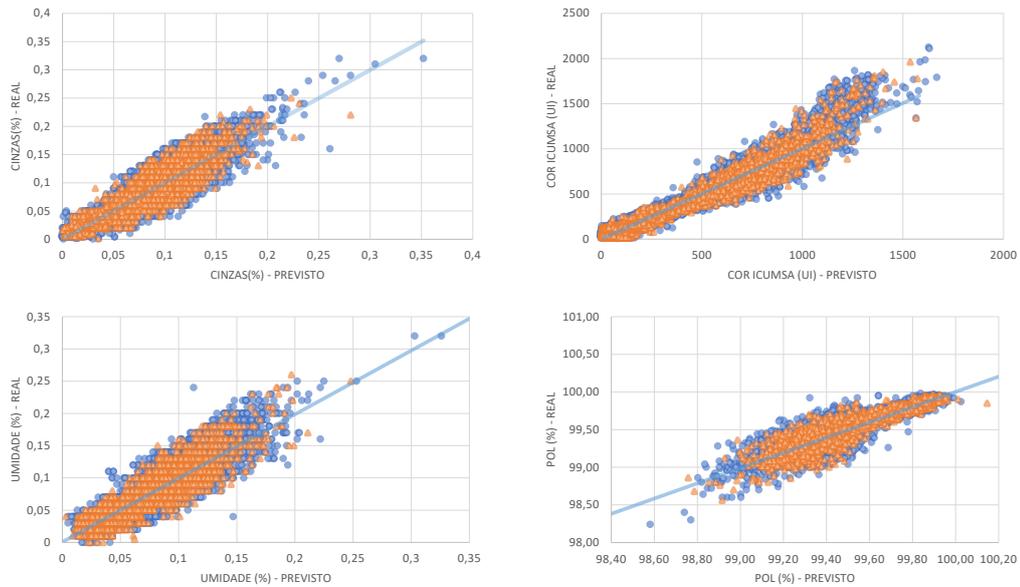
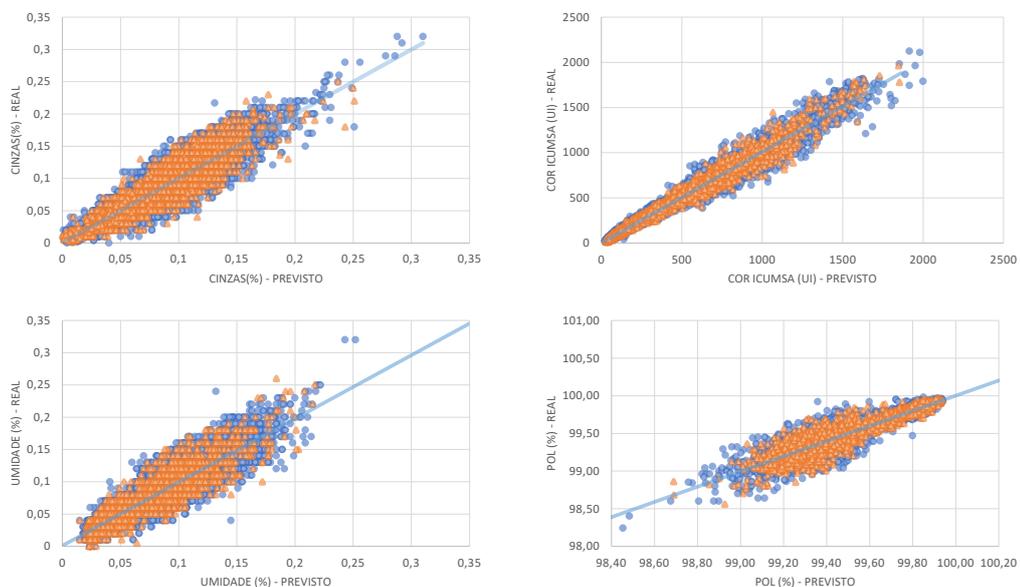


Figura 32. Regressão para os modelos ANN, calibração (círculo azul) e validação (triângulo laranja) para os parâmetros (a) Cinzas, (b) Cor, (c) Umidade e (d) Pol para dados globais de açúcar.



Após avaliar o desempenho global com grupo de validação contendo todos os tipos de açúcar, utilizamos as mesmas amostras constituintes nos grupos de validação individual (que em conjunto formam o grupo de validação global) para avaliação do desempenho do modelo global individualmente em cada tipo de açúcar.

Na Tabela 36 são apresentados os desempenhos por tipo e parâmetro através das figuras de mérito RMSEP, Slope, Intercept, BIAS e RSQ.

Tabela 38. Desempenho dos modelos globais nos tipos de Açúcar individuais

Tipo	Parâmetro	Modelo	RMSEP	Slope/Inter	Bias	RSQ
VHP	Cinzas	mPLS	0,019	1,00/0,00	0,00	0,64
		ANN	0,018	0,99/0,00	0,00	0,66
	Cor	mPLS	125	1,19/-149	7	0,80
		ANN	82	1,01/-5,49	6	0,90
	Umidade	mPLS	0,020	1,00/0,00	0,00	0,70
		ANN	0,019	0,99/0,0	0,00	0,71
	Pol	mPLS	0,126	0,97/3,47	0,00	0,47
		ANN	0,123	0,98/1,51	0,00	0,49
Cristal	Cinzas	mPLS	0,012	0,79/0,00	0,00	0,57
		ANN	0,010	0,90/0,00	0,00	0,62
	Cor	mPLS	42	0,74/44,15	2	0,87
		ANN	19	1,00/-0,11	1	0,95
	Umidade	mPLS	0,011	0,52/0,02	0,00	0,23
		ANN	0,010	0,75/0,01	0,00	0,37
	Pol	mPLS	0,047	0,78/22,02	0,00	0,72
		ANN	0,044	0,90/9,77	0,00	0,72
Refinado Amorfo	Cinzas	mPLS	0,020	0,94/0,01	0,00	0,40
		ANN	0,019	0,80/0,02	0,00	0,49
	Cor	mPLS	50	0,12/63,07	-5	0,09
		ANN	9	0,84/0,12	1	0,84
	Umidade	mPLS	0,026	0,76/0,03	0,00	0,30
		ANN	0,024	0,84/0,02	0,00	0,39
	Pol	mPLS	0,081	0,63/36,5	-0,04	0,22
		ANN	0,087	0,54/44,9	-0,03	0,08
Refinado Granulado	Cinzas	mPLS	0,012	-0,03/0,00	0,00	0,01
		ANN	0,005	0,08/0,00	0,00	0,01
	Cor	mPLS	49	-0,10/34,8	-4	0,13
		ANN	7,5	1,1/-9,95	-6	0,67
	Umidade	mPLS	0,012	0,30/-0,01	0,00	0,07
		ANN	0,010	0,68/0,00	0,00	0,11
	Pol	mPLS	0,048	0,08/91,49	0,03	0,06
		ANN	0,099	0,11/89,06	0,03	0,10

5.2.2 Desempenho comparativo dos modelos

Após o cálculo dos modelos, o desempenho de cada conjunto nos diferentes algoritmos mPLS e ANN são comparados nas Figuras 33 a 36 sendo apresentados os gráficos comparativos dos valores de RMSEP para cada tipo de açúcar nos parâmetros Cor ICUMSA, Cinzas, Umidade e Pol respectivamente.

Figura 33. Gráfico de barras comparativo dos valores de RMSEP dos modelos para Cor ICUMSA.

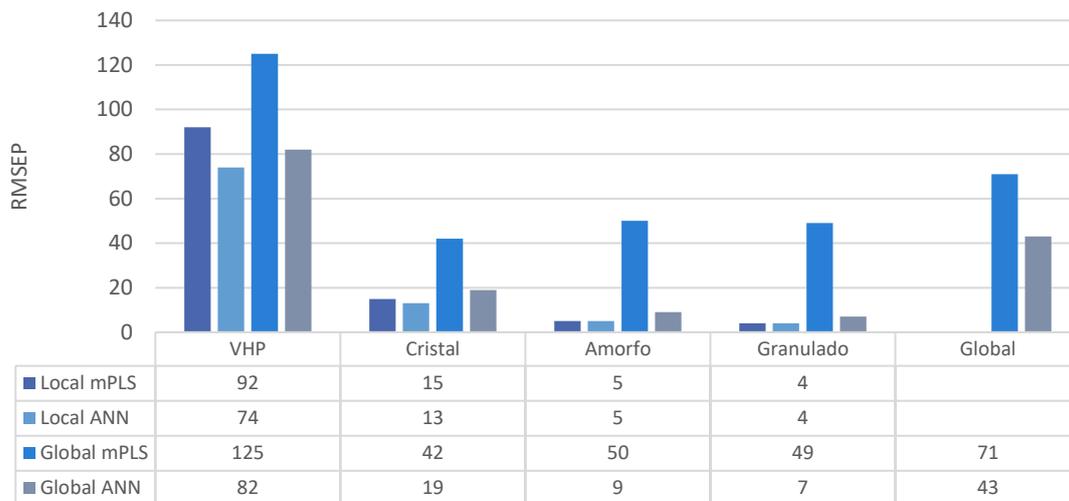


Figura 34. Gráfico de barras comparativo dos valores de RMSEP dos modelos para Cinzas.

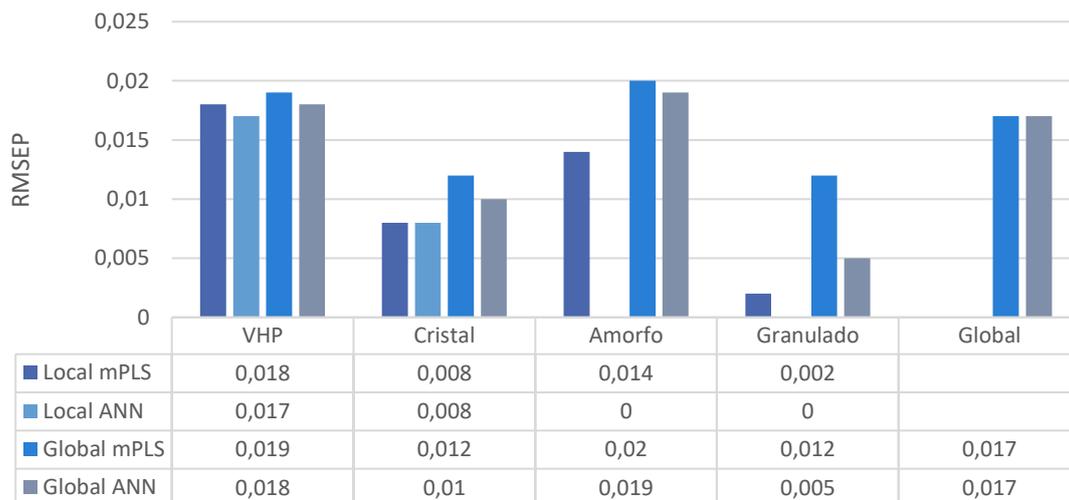


Figura 35. Gráfico de barras comparativo dos valores de RMSEP dos modelos para Umidade.

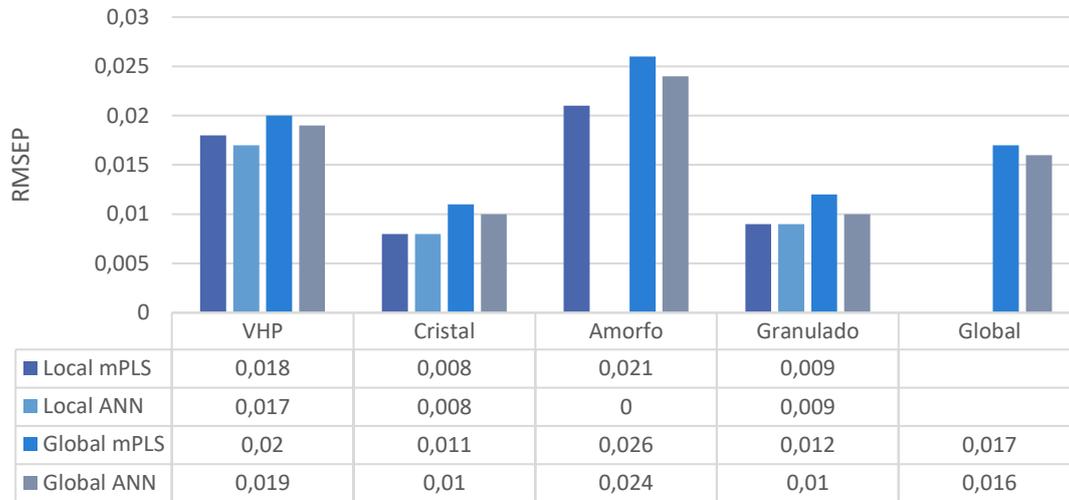
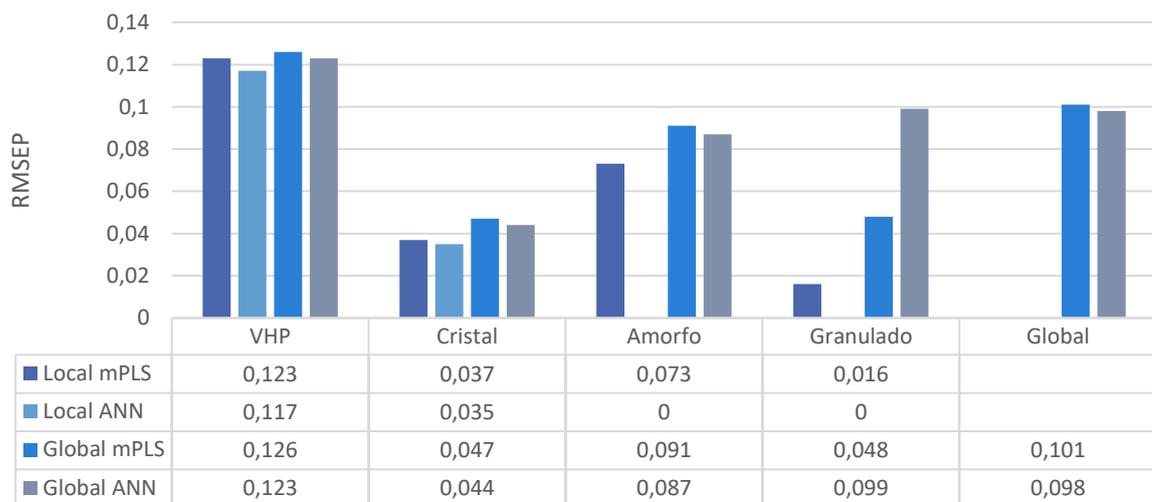


Figura 36. Gráfico de barras comparativo dos valores de RMSEP dos modelos para Pol.



Para os parâmetros com bancos de dados robustos em faixas limitadas ou inteiramente lineares os modelos ANN e mPLS apresentaram localmente, ou seja, apenas com amostras do mesmo tipo, desempenhos similares ou pouco superior para os modelos ANN. Porém quando observamos parâmetros com faixa extensas ou não lineares, como exemplo de Cor para açúcar VHP a melhora do desempenho se torna significativa. Por sua vez a utilização de um único modelo global quando validado globalmente e em cada conjunto de dados o desempenho dos modelos ANN é superior aos modelos mPLS, pela capacidade dos modelos ANN lidar com as não linearidades fornecidas ao conjunto pelos diferentes tipos de açúcar.

No item 5.1.3 foram apresentados os dados obtidos através de teste de proficiência para diferentes tipos de açúcar e sendo calculados os valores de SEL e estabelecidos os limites de reprodutibilidade para cada parâmetro. Os gráficos comparativos com os modelos obtidos são apresentados nas Figuras 37 a 39.

Figura 37. Gráfico de barras comparativo dos valores de RMSEP entre os modelos, SEL e R obtidos pelo ensaio de proficiência para o parâmetro Cor ICUMSA.

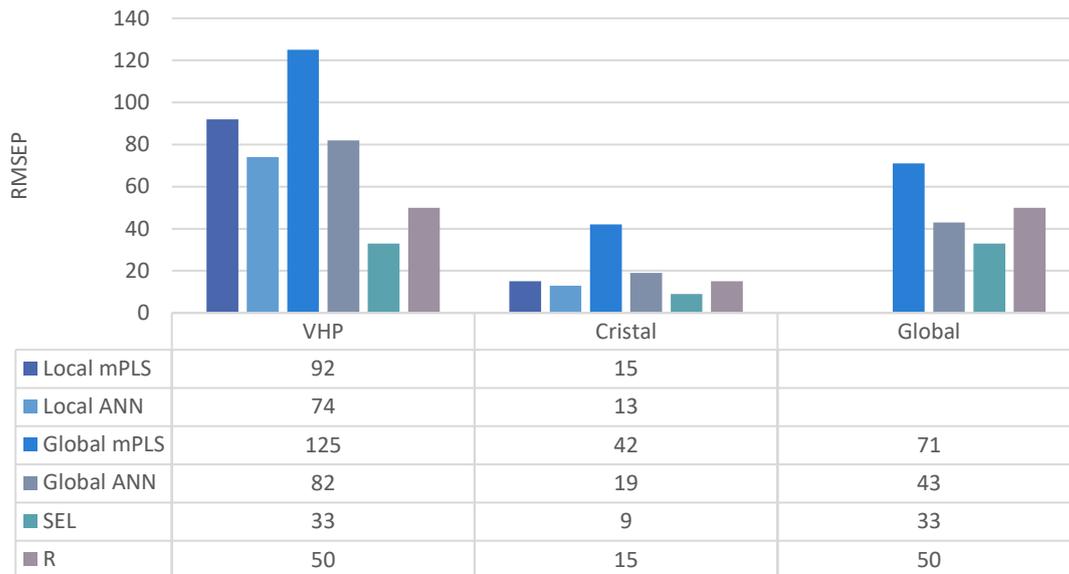


Figura 38. Gráfico de barras comparativo dos valores de RMSEP entre os modelos, SEL e R obtidos pelo ensaio de proficiência para o parâmetro Cinzas.

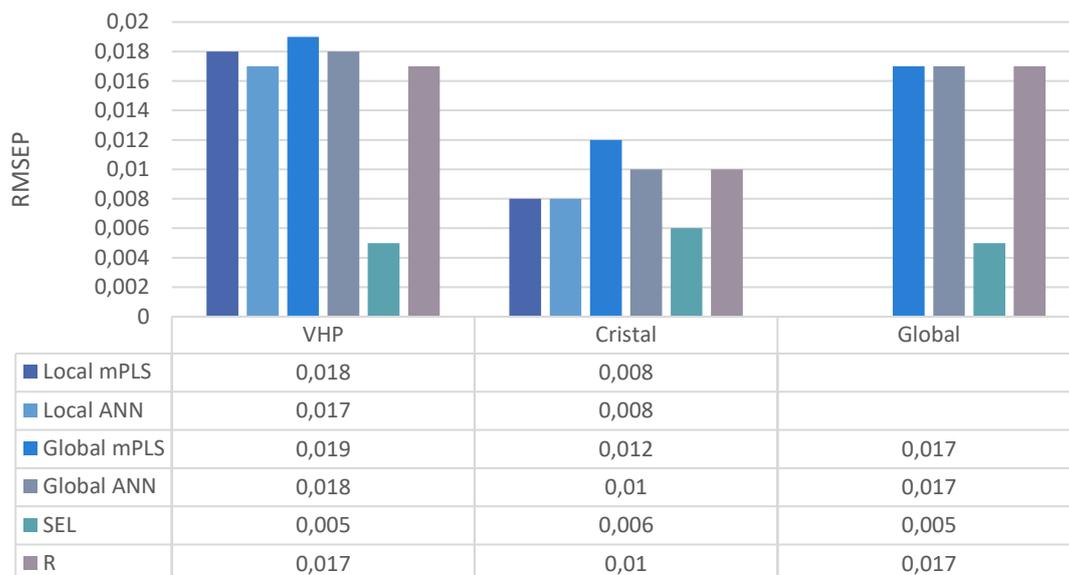
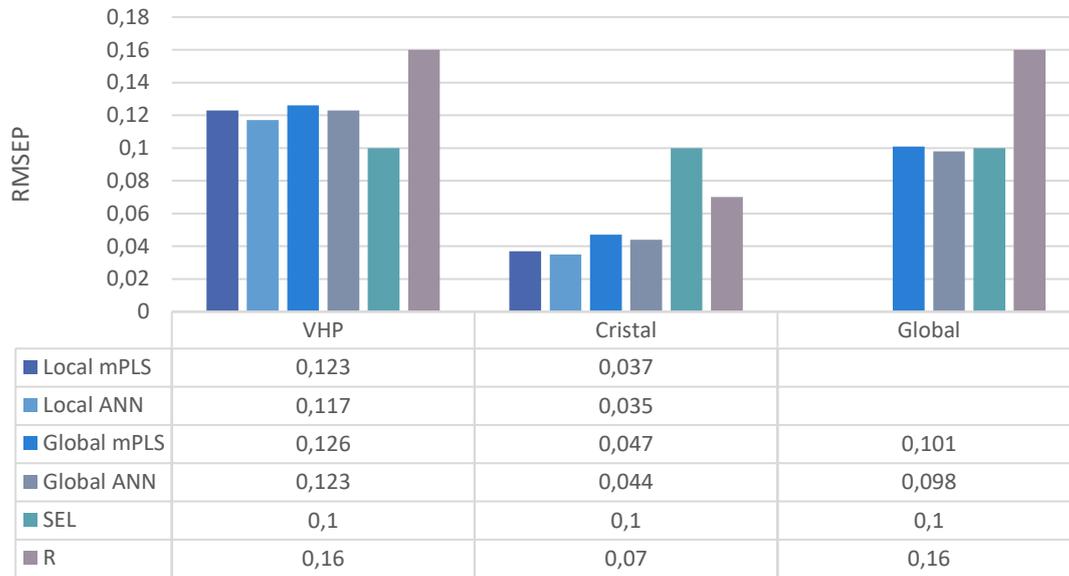


Figura 39. Gráfico de barras comparativo dos valores de RMSEP entre os modelos, SEL e R obtidos pelo ensaio de proficiência para o parâmetro Pol.



Apesar da consolidação da técnica NIRS para análise de alimentos e sua utilização em etapas dos processos sucroenergéticos desde a década de 90, se tratando da determinação dos parâmetros físico-químicos para classificação de diferentes tipos açúcares não são encontradas grandes quantidades de publicações na literatura. NADASEN, 2013 realizou desenvolvimentos de prova de conceito, porém em seu trabalho não é determinado nenhum valor de desempenho (RMSEC, RMSECV ou RMSEP) para avaliação, sendo apenas demonstrados os valores de correlação, sendo assim inválidos para essa comparação. O'SHEA, 2011 realizou a comparação entre dois instrumentos, sendo eles o FOSS InfraXact e Brucker MPA, de forma robusta foram coletadas elevado número de espectros em faixa extensa para açúcar bruto, em todos os seus parâmetros. Os valores de RMSEP obtidos para os quatro parâmetros são comparados aos valores de RMSECV obtidos na literatura para os dois instrumentos e apresentados nas Figuras 40 a 43. Até a consolidação deste trabalho não foram encontradas publicações que demonstrassem o desempenho de modelos com dados obtidos via NIRS para açúcar Cristal Refinado Amorfo ou Refinado Granulado.

Figura 40. Gráfico de barras comparativo dos valores de RMSEP entre os modelos e literatura para Cor ICUMSA.

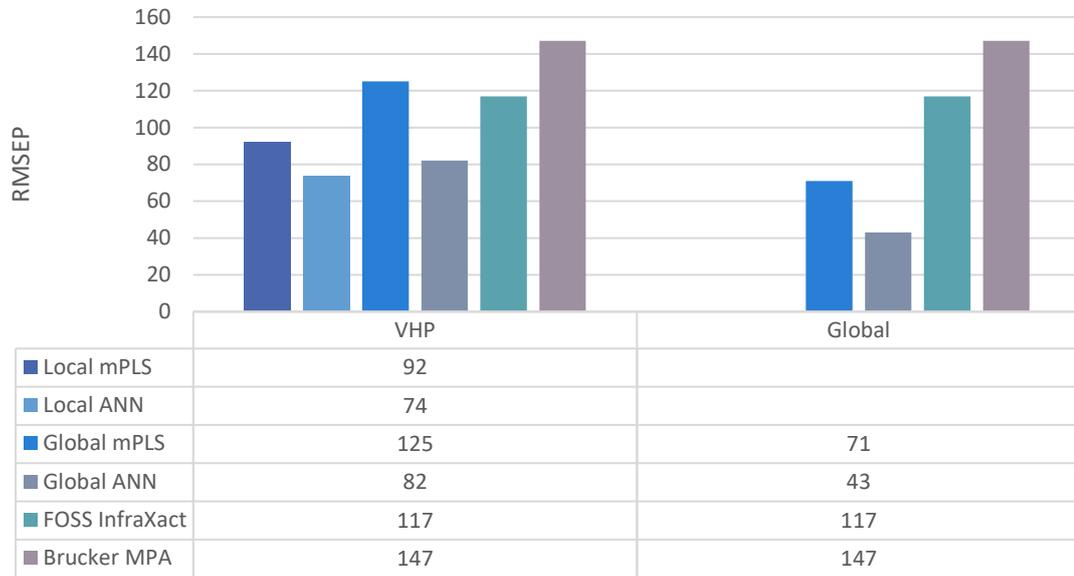


Figura 41. Gráfico de barras comparativo dos valores de RMSEP entre os modelos e literatura para Cinzas.

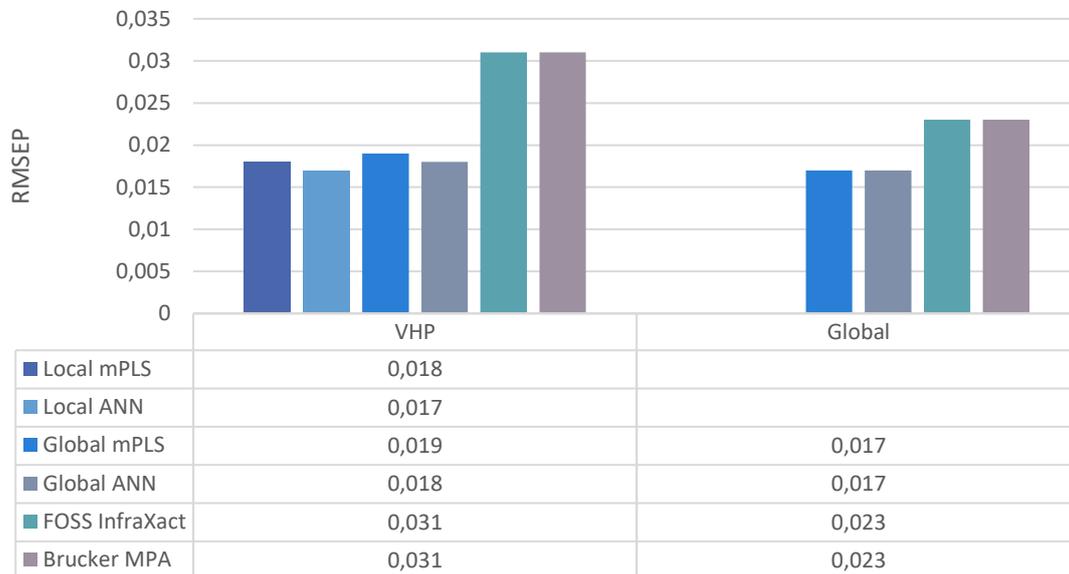


Figura 42. Gráfico de barras comparativo dos valores de RMSEP entre os modelos e literatura para Umididade.

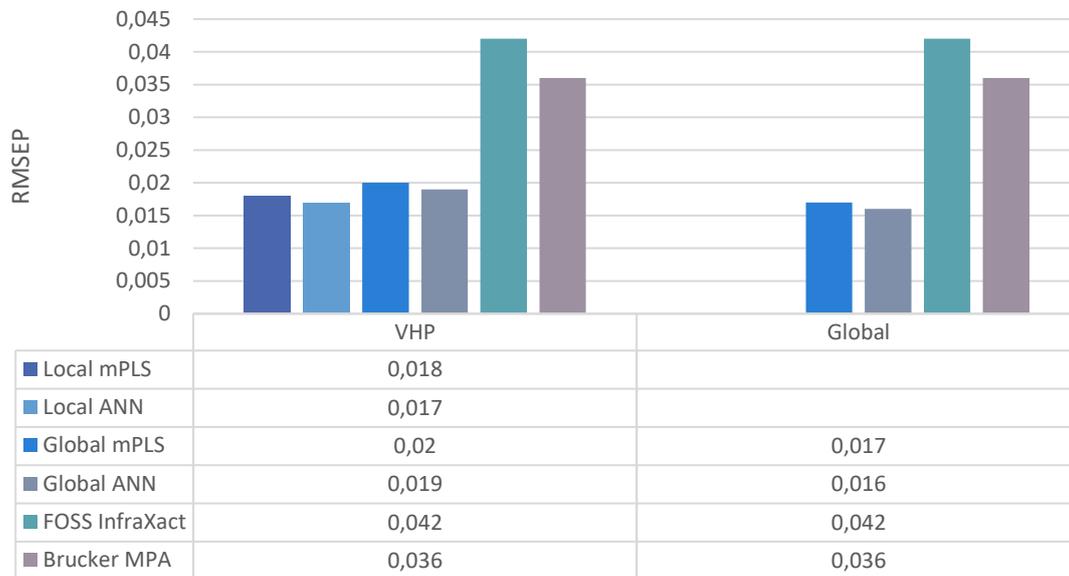
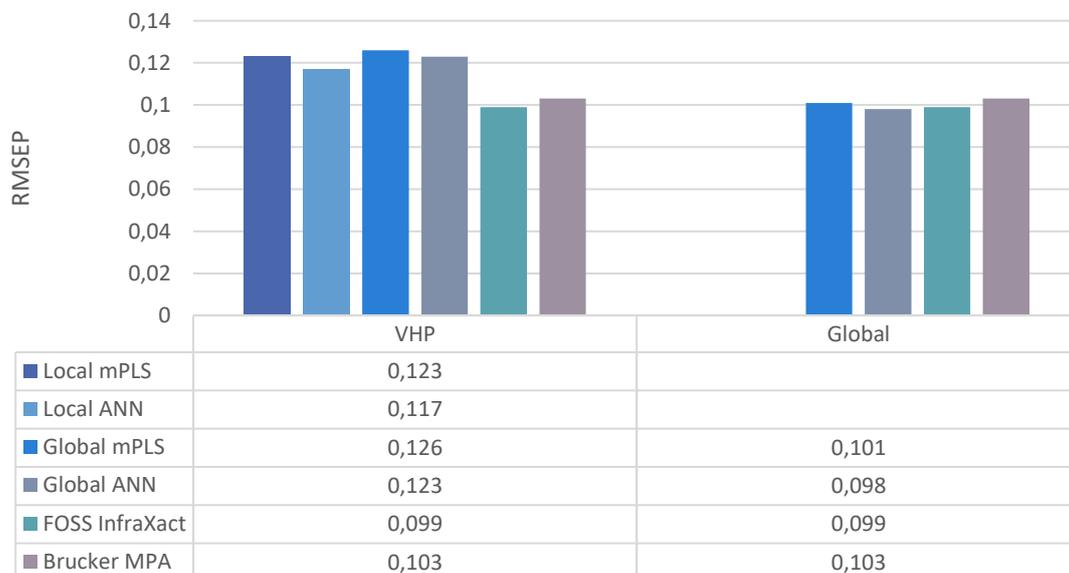


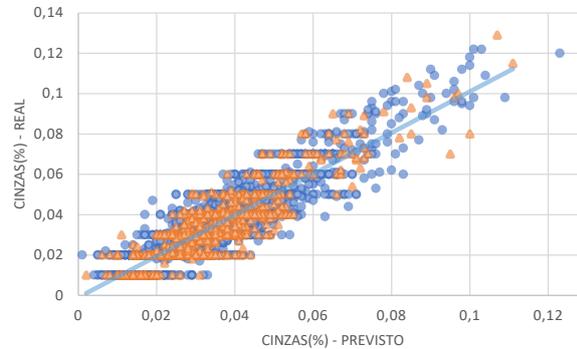
Figura 43. Gráfico de barras comparativo dos valores de RMSEP entre os modelos e literatura para Pol.



Nos gráficos de dispersão obtidos pelos resultados de laboratório versus previstos, principalmente para os parâmetros Cinzas (% m/m) e Umididade (%) foi possível observar que o NIRS detecta uma variação não observada na técnica de referência, demonstrado na Figura 44. Ao avaliar os dados foi possível determinar que esse é um efeito causado pelo arredondamento digital realizado pelos aparelhos de processamento dos dados dessas medições, uma vez que uma parte dos dados que possuía 3 casas decimais obtiveram menor dispersão, enquanto nos dados que

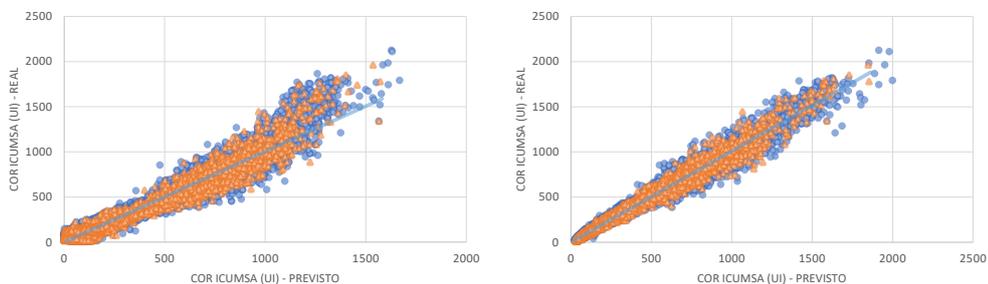
possuíam 2 casas decimais houve um efeito de dispersão com repetição em unidade de 2 decimais para a medição.

Figura 44. Gráfico de dispersão laboratório versus previsto para o parâmetro cinzas.



Quando temos grupos robustos com faixas menores ou inteiramente lineares os modelos ANN obtiveram desempenho pouco melhor ou igual aos modelos mPLS. Porém, principalmente no parâmetro Cor para açúcar VHP ou dados globais, foi observado um desempenho muito superior dos modelos ANN devido às características de não linearidade do parâmetro entre as faixas dos diferentes produtos ou ainda na faixa mais extensa do mesmo produto, como o VHP. Na Figura 45 é demonstrado o gráfico de dispersão para os modelos mPLS e ANN dos modelos de Cor ICUMSA para açúcar global, com destaque para a perda de linearidade principalmente nas extremidades do modelo mPLS, e conseqüentemente nas avaliações de RMSEP obtendo valores significativamente superiores.

Figura 45. Gráficos de dispersão laboratório versus previsto para Cor ICUMSA (UI) utilizando os dados globais nos modelos mPLS (esquerda) e ANN (direita).

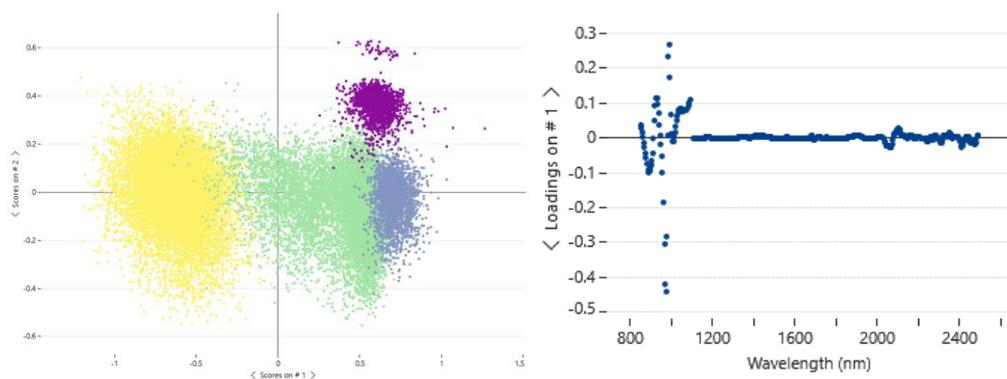


Ao avaliar o desempenho dos modelos mPLS e ANN, por tipo de açúcar ou globalmente, obtivemos desvios e robustez que permitem a aplicação em rotina por tipo ou utilizando apenas um modelo global. Porém determinar qual a melhor rotina de operação e aplicação destes modelos vai além do valor de RMSEP. Uma unidade fabril que possui uma produção constante de um tipo de açúcar por longos períodos,

com variações dentro das faixas cobertas pelos modelos, pode se beneficiar de calibrações por tipo com localmente desempenho superior à uma calibração global, porém processos com diferentes linhas de produção ou ainda locais de recebimento de cargas com grande variação do produto podem ter dúvidas em qual melhor modelo utilizar nessa rotina, dessa forma se beneficia em utilizar o modelo global.

Um modelo ANN global com toda as variações do produto tem uma grande capacidade em seu aprendizado de máquina em entender à qual padrão melhor se encaixa a amostra medida e calcular os pesos e regressões que retornem o resultado mais próximo ao real, independentemente do tipo ou classificação prévia. Mas ainda existe a possibilidade da utilização de determinações hierárquicas, de forma que o sistema primeiro aplica um modelo de discriminação utilizando uma análise de componentes principais e diferentes métodos discriminantes como distâncias de Mahalanobis, máximos de residuais e afins, para então determinar a qual grupo de amostras aquela a ser medida melhor se encaixa e aplicar em seguida o modelo de previsão adequado. Ao fazermos uma análise de componentes principais (PCA) nas amostras dos diferentes tipos de açúcares utilizadas para cálculo dos modelos, foi possível ver uma clara classificação e diferenciação no espaço das PC's entre elas, demonstrado na Figura 46.

Figura 46. (a) PC1 e PC2 para Açúcar Cristal (Verde), Açúcar VHP (Amarelo), Açúcar Granulado (Azul) e Açúcar Amorfo (Roxo) (b) Loading #1 da PCA



A clara diferenciação dos tipos de açúcar na PCA é um indicativo da possibilidade de diferenciação a partir dos espectros obtidos via NIRS e dessa forma a aplicabilidade na confecção de modelos de classificação e também de discriminantes para utilização na classificação em rotina com calibrações globais.

Capítulo 6
CONSIDERAÇÕES FINAIS

6 CONSIDERAÇÕES FINAIS

As técnicas LIBS, NIRS e a fusão de dados espectroanalíticos se mostraram promissoras na determinação de contaminantes em amostras de açúcar. Para a areia (%) foi possível obter altas correlações e desvios que possibilitam o monitoramento de fraudes. Para pontos pretos (pp/100p) houve menores correlações, consequência da dificuldade em obter homogeneidade nas superfícies de leituras, uma vez que o LIBS e o NIRS não penetram na amostra, dessa forma devendo ser efetuados mais estudos de preparo e apresentação das amostras para essa finalidade. Apesar de ter obtido desvios maiores do que o NIRS, o LIBS deverá apresentar maior seletividade à areia uma vez que possui linhas de emissão específicas para o Si, devendo ser uma opção tecnicamente mais robusta. O NIRS por sua vez se aproveitou da correlação entre outros parâmetros, como a redução do teor de sacarose ao acréscimo de areia e também a alteração da cor na amostra para obter melhores desvios e correlação, porém não há regiões de absorção das ligações Si-OH no infravermelho próximo, sendo assim menos robusto às variações da matriz. Nas duas técnicas foi observado influência da matriz das amostras com dificuldade de previsão ao variar dos tipos de açúcares, sugerindo que para obtenção de aplicações de rotina deverá contemplar diferentes tipos de açúcar.

Para o LIBS, além do desenvolvimento de bancos de dados robustos para aplicações em instrumentos de bancada, uma vez que as duas técnicas permitem utilização de instrumentos de campo/processo, existe um grande potencial em se avaliar a possibilidade de determinação/identificação de contaminantes diretamente no processo ou esteiras de transbordo, atuando assim como um alarme para tomada de ações em tempo real. Outro potencial descrito na literatura é a caracterização e determinação de minerais no açúcar via LIBS.

A técnica NIRS se mostrou eficiente em determinar Cinzas (% m/m), Cor ICUMSA (UI), Umidade (%) e Polarização ($^{\circ}$ Z) em amostras de Açúcar VHP, Açúcar Cristal, Açúcar Refinado Amorfo e Açúcar Refinado Granulado. Quando comparado com a literatura os desenvolvimentos atuais obtiveram melhor desempenho, obtendo desvios equiparados às técnicas de referência avaliadas em diferentes laboratórios através de teste de proficiência. Foram desenvolvidos e avaliados modelos mPLS e ANN para cada tipo de açúcar. Foi desenvolvido um modelo global multi-produto capaz de determinar os resultados utilizando apenas um modelo para cada parâmetro.

A partir de banco de dados robustos e com grande quantidade de amostras os modelos mPLS e ANN obtiveram desempenho similar, ou com ANN levemente superior, em faixas lineares ou restritas, já em parâmetros de faixa extensa que não apresentam linearidade os modelos ANN se mostraram mais eficientes. Para a utilização de modelos multi-produtos a ANN obteve menores desvios e ajustes do que os modelos mPLS. As aplicações NIR pela sua robustez e desempenho podem ser consideradas aplicações prontas para utilização em rotina, com capacidade de determinar globalmente os diferentes parâmetros com rapidez e acurácia nos diferentes tipos e em diferentes regiões.

Para NIRS, além da utilização para aplicações em instrumentos de bancada, existe um grande potencial em se avaliar a possibilidade de determinação diretamente no processo ou esteiras de transbordo, atuando assim para tomada de ações em tempo real. Existe o potencial de utilização da técnica NIRS também para determinação de demais parâmetros com açúcares redutores, amido, dextrana, turbidez, compostos fenólicos e sulfito.

Com o sucesso no desenvolvimento de tais aplicações e implementação na rotina de laboratórios em Usinas, e também entrepostos de recebimento, é possível alterar a rotina de trabalho em relação à produtividade analítica e também em curto período transformar o investimento nesse tipo de tecnologia em retornos financeiros através da diminuição das perdas e ganhos no processo. As técnicas LIBS e NIRS estão em total sintonia com a indústria 4.0 possibilitando resultados rápidos, conectividade e suas intervenções nos processos garantem sua rápida viabilidade, pois evitar uma adulteração ou gerar ganhos de eficiência na indústria da ordem de 0,1 a 0,5% podem se traduzir em centenas de milhares de reais.

Capítulo 7
REFERÊNCIAS BIBLIOGRÁFICAS

7 REFERÊNCIAS

1. ANDRADE, D. F.; GUEDES, W. N.; PEREIRA, F. M. V. Detection of chemical elements related to impurities leached from raw sugarcane: Use of laser-induced breakdown spectroscopy (LIBS) and chemometrics. **Microchemical Journal**, n. 137, p. 443-448, 2018.
 2. BERDING N.; Brotherton G.A.; LEBROCQ D.G.; SKINNER J.C. Analysis of sugar cane in clonal evaluation trials by near infra-red reflectance (NIR) spectroscopy. **Proceedings of the Australian Society of Sugar Cane Technologists** 11, 8–15, 1989.
 3. BURNS, D.A.; CIURCZAK, E.W. **Handbook of near-infrared analysis**. 3rd ed. CRC Press, 2007 (Practical spectroscopy; 35).
 4. CASTRO, J. P.; Pereira-Filho, E. R.; Twelve different types of data normalization for the proposition of classification, univariate and multivariate regression models for the direct analyses of alloys by laser-induced breakdown spectroscopy (LIBS). **J. Anal. At. Spectrom.**, 2016, 31, 2005–2014.
 5. CONAB, Acomp. safra bras. cana, v. 7- Safra 2020/21, n. 4 - Quarto levantamento, Brasília, p. 1-57 maio de 2021.
 6. CONSUL-LAB. Ensaio de Proficiência em Análises de Açúcar, 1ª Rodada, Piracicaba, 2019.
 7. COSTA, V. C. et al. Laser Induced-Breakdown Spectroscopy (LIBS): Histórico, fundamentos, aplicações e potencialidades. **Quim. Nova**, Vol. 42, No. 5, 527-545, 2019.
 8. CREMERS, D.A.; RADZEMSKI, L.J. **Handbook of Laser-Induced Breakdown Spectroscopy**. 2nd ed. WILEY, 2013.
 9. DAVIES, A.M.C and FEARN, T. 2006. Back to basics: calibration statistics. **Spectroscopy Europe**. 18: 31-32.
 10. DE OLIVEIRA, R. R.; DAS NEVES, L. S.; LIMA, K. M. G. Experimental Design, Near-Infrared Spectroscopy, and Multivariate Calibration: Na Advanced Project in a Chemometrics Course. **J. Chemical Education**. 89, 1566-1571, 2012.
 11. GELADI, P; KOWALSKI B.R. Partial Least-Squares Regression: A tutorial. **Analytica Chimica Acta**, 185, 1-17, 1986.
 12. International Organization for Standardization (ISO); ISO 13528: Statistical methods for use in proficiency testing by interlaboratory comparisons. Geneva, 2015.
 13. MAPA, **Projeções do Agronegócio: Brasil 2017/18 a 2027/28 projeções de longo prazo** / Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Política Agrícola. MAPA/ACE, Brasília, 2018.
-

-
14. NADASEN T AND NAICKER K, Practical application of near-infrared spectroscopy to predict quality of solid sugar. **Proc S Afr Sug Technol Ass**, Issue 86, pag. 114 – 120, 2013.
 15. NEVES, M.F. **O setor sucroenergético em 2030: dimensões, investimentos e uma agenda estratégica**. Confederação Nacional da Indústria, Brasília, 2017.
 16. NI, W.; NORGAARD, L.; MORUP, M. Non-linear calibration models for near infrared spectroscopy, **Analytica Chimica Acta**, Volume 813, Pages 1-14, 2014.
 17. NORGAARD, L; LAGERHOLM, M; WESTERHAUS, M. Artificial Neural Networks and Near Infrared Spectroscopy – A case Study on protein content in whole wheat grain, **Infocus**, Issue 1, April 2013.
 18. O'DONNELL, C. P. et al. (eds.), **Process Analytical Technology for the Food Industry**, Food Engineering Series, Springer Science+Business Media, New York 2014.
 19. O'SHEA MG et al. Developing laboratory near infra-red (NIR) instruments for the analysis of sugar factory products. **Proc. Aust. Sugar Cane Technol.**, v, 33, 2011.
 20. PASQUINI, Celio et al. Laser Induced Breakdown Spectroscopy. **J. Braz. Chem. Soc.**, São Paulo, v. 18, n. 3, p. 463-512, 2007.
 21. PASQUINI, Celio et al. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications, **J. Braz. Chem. Soc.**, São Paulo, v. 14, n. 2, p. 198-219, 2003.
 22. RINNAN, Å.; VAN DEN BERG, F.; ENGELSON, S. B. Review of the most common pre-processing techniques for near-infrared spectra, **TrAC Trends in Analytical Chemistry**, Volume 28, Issue 10, 1201-1222, 2009.
 23. ROMERA, J.P.R.; BARSANELLI, P. L.; PEREIRA, F. M. V. Expedient prediction of fiber content in sugar cane: An analytical possibility with LIBS and chemometrics. **Fuel**, n. 166, p. 473-476, 2016.
 24. SANTOS JUNIOR, D. et al. Espectrometria de emissão óptica com plasma induzido por laser (LIBS) – fundamentos, aplicações e perspectivas. **Revista Analytica**, n. 24, p. 72-81, ago./set. 2006.
 25. SEZER, B.; BILGE, G.; BOYACI, I. H. Capabilities and Limitations of LIBS in food analysis. **Trends in Analytical Chemistry**, n. 97, p. 345-353, 2017.
 26. SILVA, A. F. Caracterização e determinação de minerais em amostras de açúcares brasileiros. Dissertação (Mestrado) – ESALQ-USP, 133p, Piracicaba, 2017.
 27. SHENK J.; WESTERHAUS, M. Populations Structuring of Near Infrared Spectra and Modified Partial Least Squares Regression. **Crop Science**, Vol. 31, Issue 6, p 1548-1555, 1991.
-

-
28. TANGE, R. et al. Application of support vector regression for simultaneous modelling of near infrared spectra from multiple process steps. **Journal of Near Infrared Spectroscopy**, 23, 75-84, 2015.
 29. TAVARES, B., GODOY, M. **Açúcar vira areia com convivência de policiais**, O Estado de São Paulo, 25 de junho de 2010. Disponível em <<https://sao-paulo.estadao.com.br/noticias/geral,acucar-vira-areia-com-convencia-de-policiais-imp-,571814>>.
 30. THOMPSON, M; WOOD, R. The international harmonized protocol for the proficiency testing of (chemical) analytical laboratories, **AOAC**, September 1993.
 31. Using Cross-Validation, **Eigenvector**, 2020. Disponível em: <https://www.wiki.eigenvector.com/index.php?title=Using_Cross-Validation> Acesso em maio, 2021.
 32. WESTERHAUS, M. Eastern Analytical Symposium Award for Outstanding Achievements in near Infrared Spectroscopy: My Contributions to near Infrared Spectroscopy. **NIR News**, Vol 25, Issue 8, 2014.
 33. WILLIAMS, P.; DARDENNE, P.; FLINN, P. Tutorial: Items to be included in a report on a near infrared spectroscopy Project. **Journal of Near Infrared Spectroscopy**. 25: (2), 85-90, 2017.
 34. WILLIAMS, P.; DARDENNE, P.; FLINN, P. Tutorial: Items to be included in a report on a near infrared spectroscopy project, **Journal of Near Infrared Spectroscopy**, 25(2), 85-90, 2017.
-