

UNIVERSIDADE FEDERAL DE JUIZ DE FORA

INSTITUTO DE CIÊNCIAS EXATAS

DEPARTAMENTO DE QUÍMICA

Lucas Mattos Duarte

Modelagem multivariada para determinação de propriedades físico-químicas de petróleo e para quantificação de edulcorantes artificiais em adoçantes de mesa

Juiz de Fora

2015

Lucas Mattos Duarte

Modelagem multivariada para determinação de propriedades físico-químicas de petróleo e para quantificação de edulcorantes artificiais em adoçantes de mesa

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal de Juiz de Fora como requisito parcial para a obtenção do título de Mestre em Química

Orientador: Marcone Augusto Leal de Oliveira

Coorientadora: Celly Mieko Shinohara Izumi

Juiz de Fora
2015

“A vida é muito curta para ser pequena.”

(Benjamin Disraeli)

AGRADECIMENTOS

Seria impossível chegar até aqui sozinho, até porque seria vazio e sem graça. Logo, agradeço: A DEUS, por ter me despertado para a vida que vale a pena ser vivida, me mostrar que o AMOR é a base de tudo e ter me dado paz e ânimo nos momentos difíceis.

À minha querida família. Em especial à minha amada mãe, que sempre persistiu para que eu trilhasse o caminho do conhecimento, mesmo quando minha única expectativa era o futebol e a diversão. Ao meu melhor amigo e pai, pelas conversas, conselhos e exemplo de nobreza de caráter (meu ídolo). À Bibi, por ser minha melhor amiga e através de jeitinho singular, despertar em mim o sentimento mais puro de amor e cuidado (você é a pérola que DEUS me incumbiu de cuidar). Ao meu irmão Danilo, que com seu jeito peculiar, divertido e destrambelhado sempre foi meu maior parceiro (mesmo sendo o maior fã do meu futebol, nunca te perdoei nas “canetas” – Amigos, amigos, futebol à parte).

Ao meu querido amigo e orientador Marcone. Sou grato por sempre ter acreditado em mim e por ser importantíssimo no meu desenvolvimento científico e profissional. Você é um grande exemplo de humildade e otimismo. Obrigado pelas conversas, conselhos e cervejas.

À professora Celly, por ter aceitado me coorientar e ter contribuído na parte dos Edulcorantes.

Ao grupo de Quimiometria do Petróleo pela confiança, oportunidades e ensinamentos. Especialmente ao Paulo (e ae caboclo), que me ensinou muito sobre calibração multivariada e sempre atendeu prontamente às minhas requisições.

Ao meu querido grupo de pesquisa GQAQ. Aprendi muito com vocês e sou muito grato a todos.

Ao professor Rafael Arromba, pelos conselhos, motivação e carisma, esse que nos contagia.

Ao partido, constituído pelo presidente Sandro (“olha o corte em dotorzinho, hoje tem, ahhh hoje tem”), “príncipe” Pedro e Marcelão. Sandrão obrigado pela forte parceria, você é um grande amigo. Quanta saudade eu sentirei do moçambicano mais brasileiro de todos, lembrarei para sempre das brincadeiras e piadas meu “camarrada”. Grande Marcelo, obrigado por me ensinar a ter serenidade. Muito obrigado pelas inúmeras cervejas e momentos de intensa alegria!

Especialmente à camarada Brenda, que com carinho e paciência me ensinou a pensar e trabalhar como Químico Analítico. Tenho segurança em dizer, que sem suas contribuições o trabalho não

seria realizado. Agradeço também ao Thiago pela amizade, discussões científicas e por me ensinar a utilizar o NIR.

À Maressa e ao professor Gustavo (UFSC), pelas análises no LC-MS. Esses resultados foram importantíssimos para o fechamento do trabalho.

Ao professor Alexandre Leitão, por ter me dado a primeira oportunidade na área científica, em um momento de muita desilusão durante a graduação. Como o senhor bem dizia, eu fui mais um que o senhor “salvou” para a ciência. Especialmente ao Ary, que foi o maior motivador para minha inserção na área científica. Sua motivação e dedicação pela ciência me inspiraram. Obrigado pelas discussões científicas e parcerias (te devo um chopp).

Aos primos e tios, pela amizade, cuidado e pelos bons momentos. DEUS me abençoou com pessoas maravilhosas como vocês. Amo a todos!

Ao meu querido grupo de estudo bíblico (Marcelo, Ed, Alex, Barão, Jonathan, Aline, Dani, Lu e Maria). Obrigado pelos momentos de comunhão, aprendizado e muita diversão. Vocês são espetaculares!

Aos queridos amigos:

Barão, que com sua peculiar mentalidade é capaz de me alegrar e irritar ao mesmo tempo. Obrigado pela fidelidade e parceria.

Kenzi (frango forever), por tantos cafés, conversas e parceria. Obrigado por ter aberto as portas da i9 (celf) para que eu virasse noites estudando e escrevendo relatórios.

Dudu (ueiii papito), pelas orações, amizade e conselhos.

Arthur, pela amizade mas principalmente pelos materiais emprestados. Você é o cara mais caprichoso na construção de resumos que eu conheço.

Leandra, pela amizade e principalmente por rir de minhas piadas sem graça.

Pestana, Willian, Vavá, Fábio, Ulisses e Betinho por grande amizade e parceria.

Da PIB, local onde fui muito abençoado.

RESUMO

O desenvolvimento de metodologias de análise que sejam rápidas, robustas e de baixo custo é um dos grandes focos da química analítica moderna. A aplicação de técnicas espectroscópicas associadas à modelagens multivariadas, com o intuito de realizar inferências quantitativas, vem apresentando grande crescimento ao longo dos últimos anos. O presente trabalho visa o desenvolvimento de metodologias analíticas, por meio da aplicação de diferentes técnicas espectroscópicas, aliadas à regressão por Mínimos Quadrados Parciais (PLS), para duas aplicações diferentes. A primeira consiste na determinação de algumas propriedades físico-químicas de petróleo e a segunda na determinação da concentração de quatro edulcorantes artificiais em amostras de adoçantes de mesa. Em relação ao petróleo, este é formado basicamente por diferentes famílias de hidrocarbonetos, apresentando uma enorme gama de compostos. A composição complexa da matriz faz com que sua avaliação envolva um grande número de análises físico-químicas, o que torna sua caracterização de alto custo e demorada, podendo chegar a um ano. Nessa primeira parte, foram construídos modelos de PLS a partir de dados de Ressonância Magnética Nuclear para o núcleo de Hidrogênio ($^1\text{H-RMN}$), para determinar as propriedades gravidade API, resíduo de carbono (RC) e temperatura inicial de aparecimento de cristais (TIAC), além de simular a curva de temperaturas de ebulição para as frações destiladas (SimDis). As curvas obtidas por essa metodologia foram chamadas de “curva de Pontos de Ebulição Estimadas por Quimiometria (PEEQ)”. Na segunda parte do trabalho, foram construídos modelos multivariados pela aplicação de PLS aos dados de espectroscopia Raman e de Infravermelho Próximo (NIR), para quantificação de Aspartame (ASP), Ciclamato (CIC), Sacarina (SAC) e Acesulfame-K (ACSK), em amostra de adoçante de mesa em pó. Para tanto, o método de referência utilizado para fornecer a informação de concentração referentes aos analitos, foi a Cromatografia Líquida com detecção por Espectrometria de Massas (LC-MS). Os modelos de NIR-PLS foram mais eficazes na previsão de ASP, CIC e SAC, enquanto o modelo de Raman-PLS respondeu melhor para o ACSK. Todos os modelos, pra ambas as partes, foram avaliados com base nos seguintes parâmetros de modelagem: número de Variáveis Latentes (LVs), Coeficiente de Correlação para a Validação Cruzada e para Previsão (R^2_{vc} e R^2_p , respectivamente), Raiz quadrada do Erro Quadrático Médio para Validação Cruzada e para Previsão (RMSECV e RMSEP, respectivamente). Por fim, os resíduos de cada modelo construído foram submetidos a testes estatísticos para avaliação de erros sistemáticos e erros de tendência.

Palavras-chave: Calibração Multivariada. PLS. Espectroscopia. Petróleo. Edulcorantes

ABSTRACT

The development of analytical methods that are fast, robust and low cost is one of the major focuses of modern analytical chemistry in order to perform quantitative inferences. The application of spectroscopic techniques associated with multivariate modeling has shown significant growth over the last years. This work aims at the development of analytical methodologies, by application of different spectroscopic techniques combined with regression by Partial Least Squares (PLS), for two different applications. The former involves the determination of some physicochemical properties of crude oil and the latter the determination of concentration of four artificial sweeteners in solid tabletop sweeteners samples. Petroleum is basically formed by different hydrocarbons families with a huge range of compounds. Due to the complex composition of the matrix a large number of high costs and time-consuming physicochemical analyses are required. In this first part, PLS models were built from ^1H Nuclear Magnetic Resonance data (^1H -RMN), in order to determine the following properties: API gravity, Carbon Residue (CR), Wax Appearance Temperature (WAT) and Simulated Distillation (SimDis). The curves obtained by this methodology were called *Boiling Points Estimated by Chemometrics (BPEC)*. In the second part of this work, multivariate models were built by the application of PLS algorithm to Raman spectroscopy data and Near Infrared (NIR) spectroscopy to quantify Aspartame (ASP), Cyclamate (CIC), Saccharin (SAC) and Acesulfame-K (ACSK) in tabletop sweetener powder forms. Therefore, the reference method used to provide information regarding the concentration of analytes was the Liquid Chromatography coupled with Mass Spectrometry (LC-MS). The NIR-PLS models were more effective for the prediction of ASP, CIC and SAC, while Raman-PLS model presented better results to ACSK. All models, for both studies, were evaluated based on the following modelling parameters: number of Latent Variables (LVs), correlation coefficient for Cross Validation and Prediction (R^2_{vc} and R^2_{p} , respectively), Root Mean Square Error for Cross Validation and Prediction (RMSECV and RMSEP, respectively). Finally, the residuals of each model built were submitted to statistical tests for evaluation of systematic errors and tendency errors.

Keywords: Multivariate Calibration. PLS. Spectroscopy. Petroleum. Sweeteners

LISTA DE FIGURAS

Figura 1: Princípio do PLS	21
Figura 2: Esquema dos espalhamentos Raman e Rayleigh	30
Figura 3: Representação esquemática dos modelos harmônico (A) e anarmônico (B) para a energia potencial de uma molécula diatômica, onde <i>de</i> refere-se à distância de equilíbrio	31
Figura 4: Níveis de energia dos núcleos de hidrogênio quando aplicado um campo magnético de magnitude B_0	33
Figura 5: Esquema do mecanismo de separação por CZE em contra-fluxo.....	36
Figura 6: Espectro de ^1H -RMN para um óleo cru	40
Figura 7: Conjunto espectral de ^1H -RMN para 92 amostras de petróleo	41
Figura 8: Conjunto espectral de ^1H -RMN, alinhado, para 92 amostras de petróleo	42
Figura 9: Efeito do alinhamento nos espectros de ^1H -RMN para amostras de petróleo, pelo uso da ferramenta <i>icoshift</i> : (a) bruto e (b) após o alinhamento.....	42
Figura 10: PCA para o conjunto espectral sem o alinhamento (a) e após o alinhamento (b)...	43
Figura 11: (a) conjunto espectral de ^1H -RMN após a retirada do outlier (b) PCA após a retirada do outlier.....	44
Figura 12: Loadings PCA realizada sem o alinhamento para o conjunto espectral	45
Figura 13: Otimização do número de variáveis latentes para construção de modelo de calibração, para a propriedade API.....	47
Figura 14: Correlação entre os valores de referência e os previstos pelo modelo PLS, para a propriedade API.....	48
Figura 15: Gráfico de <i>leverage</i> , que demarcam a influência das amostras na construção do modelo, para a propriedade API.....	49
Figura 16: Gráfico dos resíduos que permite a identificação de amostras que não tiveram boa previsão.....	49
Figura 17: Histograma de permutação utilizando uma função linear (a) e uma quadrática (b) para descrever os resíduos do modelo construído, para a propriedade API.....	50
Figura 18: Otimização do número de variáveis latentes para construção de modelo de calibração, para a propriedade RC.....	52
Figura 19: Correlação entre os valores de referência e os previstos pelo modelo PLS, para a propriedade RC.....	53
Figura 20: Gráfico de leverages, que demarcam a influência das amostras na construção do modelo, para a propriedade RC	54
Figura 21: Gráfico dos resíduos para a propriedade RC	54
Figura 22: Histograma de permutação utilizando uma função linear (a) e uma quadrática (b) para descrever os resíduos do modelo construído, para a propriedade RC	55
Figura 23: Coeficiente de regressão para o modelo de ^1H -RMN-PLS com 9LVs (a) e o espectro médio para o conjunto amostral (b).....	57

Figura 24: Otimização do número de variáveis latentes para construção de modelo de calibração, para a propriedade TIAC.....	58
Figura 25: Correlação entre os valores de referência e os previstos pelo modelo PLS, para a propriedade TIAC.....	58
Figura 26: Gráfico de leverages, que demarcam a influência das amostras na construção do modelo, para a propriedade TIAC	59
Figura 27: Gráfico dos resíduos para a propriedade TIAC	60
Figura 28: Otimização do número de variáveis latentes para construção de modelo de calibração, para o corte referente a 5% em massa.....	63
Figura 29: Correlação entre os valores de referência e os previstos pelo modelo PLS, para o corte referente a 5% em massa	64
Figura 30: Gráfico dos resíduos para o corte referente a 5% em massa da curva SimDis para amostras de petróleo	64
Figura 31: Histograma de permutação utilizando uma função linear (a) e uma quadrática (b) para descrever os resíduos do modelo construído, para o corte referente a 5% em massa	65
Figura 32: Monitoramento de parâmetros de modelagem ao longo dos cortes para: (a) correlação para validação cruzada e predição; (b) número de LVs utilizadas; (c) RMSEP% ..	66
Figura 33: Monitoramento dos resíduos deixados pelos modelos construídos ao longo dos cortes, por meio de testes para erros sistemáticos (teste de bias).....	67
Figura 34: Monitoramento dos resíduos deixados pelos modelos construídos ao longo dos cortes, por meio de testes para erros de tendência (teste de permutação), aplicando funções lineares e quadráticas.....	67
Figura 35: Comparação entre a SimDis, obtida por cromatografia gasosa de acordo com a ASTM e a PEEQ, estimada pela metodologia ¹ H-RMN-PLS, para uma amostra com API 17	68
Figura 36: Comparação entre a SimDis, obtida por cromatografia gasosa de acordo com a ASTM e a PEEQ, estimada pela metodologia ¹ H-RMN-PLS, para uma amostra com API 29,5	68
Figura 37: Fórmula estrutural plana e valores de pka para os quatro edulcorantes analisados nesse trabalho	72
Figura 38: Eletroferograma para mistura de padrões dos analitos em estudo.....	77
Figura 39: Eletroferograma das amostras em que foram aplicadas a metodologia desenvolvida por Fernandes. Uma amostra sólida (a), nomeada AM1 e duas líquidas, (b) e (c)	78
Figura 40: Esquema para a estratégia de ação adotada nessa etapa do trabalho	81
Figura 41: Espectro FT-Raman para os padrões puros de ACSR, ASP, CIC e SAC	82
Figura 42: Espectros Raman agrupados para as 30 amostras de trabalho (a) e espectro Raman para a lactose (b).....	83
Figura 43: Aproximação feita na região espectral compreendida entre 2840 e 3000 cm ⁻¹ , para melhor visualização do possível <i>outlier</i>	84
Figura 44: PCA para o conjunto de espectros Raman, referente às 30 amostras de trabalho ..	84

Figura 45: PCA para o novo conjunto de espectros Raman, referente as 29 amostras de trabalho (a) e o gráfico de <i>loading</i> mostrando que a primeira PC descreveu 99,61% da variabilidade dos dados originais (b)	86
Figura 46: Espectros NIR agrupados para as 30 amostras de trabalho	87
Figura 47: PCA para o conjunto de espectros NIR, referente as 30 amostras de trabalho	87
Figura 48: Espectro médio para as 30 amostras (superior) e gráfico de loadings para a PC1 (inferior), que descreveu 99,98% das variáveis originais.....	88
Figura 49: Otimização do número de variáveis latentes para construção de modelo de calibração, com dados de Raman (a) e NIR (b), para o analito Aspartame.....	93
Figura 50: Correlação entre os valores de referência e os previstos pelos modelos PLS, com dados de Raman (a) e NIR, para o analito ASP.....	93
Figura 51: Gráfico dos resíduos dos modelos PLS, com dados de Raman (a) e NIR (b), para o analito ASP.....	94
Figura 52: Histograma de permutação utilizando uma função linear, (a) e (c) e, uma quadrática, (b) e (d), para descrever os resíduos dos modelos construídos, para o analito ASP. Os gráficos superiores referem-se ao modelo de raman e os inferiores ao modelo de NIR.....	95
Figura 53: Coeficiente de regressão para os modelos de Raman-PLS (superior) com 3LVs e NIR-PLS (inferior) com 4LVs.....	95
Figura 54: Otimização do número de variáveis latentes para construção de modelo de calibração, com dados de Raman (a) e NIR (b), para o analito Ciclamato	97
Figura 55: Correlação entre os valores de referência e os previstos pelos modelos PLS, com dados de Raman (a) e NIR, para o analito CIC	98
Figura 56: Gráfico dos resíduos dos modelos PLS, com dados de Raman (a) e NIR (b), para o analito CIC.....	98
Figura 57: Histograma de permutação utilizando uma função linear, (a) e (c) e, uma quadrática, (b) e (d), para descrever os resíduos dos modelos construídos, para o analito CIC. Os gráficos superiores referem-se ao modelo de Raman e os inferiores ao modelo de NIR.....	100
Figura 58: Coeficiente de regressão para os modelos de Raman-PLS (superior) com 5LVs e NIR-PLS (inferior) com 4LVs.....	100
Figura 59: Otimização do número de variáveis latentes para construção de modelo de calibração, com dados de Raman (a) e NIR (b), para o analito Sacarina	102
Figura 60: Correlação entre os valores de referência e os previstos pelos modelos PLS, com dados de Raman (a) e NIR (b), para o analito SAC	102
Figura 61: Gráfico dos resíduos dos modelos PLS, com dados de Raman (a) e NIR (b), para SAC	103
Figura 62: Histograma de permutação utilizando uma função linear, (a) e (c) e, uma quadrática, (b) e (d), para descrever os resíduos dos modelos construídos, para o analito SAC. Os gráficos superiores referem-se ao modelo de Raman e os inferiores ao modelo de NIR.....	104
Figura 63: Coeficiente de regressão para os modelos de Raman-PLS (superior) com 4LVs e NIR-PLS (inferior) com 4LVs.....	105
Figura 64: Otimização do número de variáveis latentes para construção de modelo de calibração, com dados de Raman (a) e NIR (b), para ACSK	106

Figura 65: Correlação entre os valores de referência e os previstos pelos modelos PLS, com dados de Raman (a) e NIR (b), para o analito ACSK.....	107
Figura 66: Gráfico dos resíduos dos modelos PLS, com dados de Raman (a) e NIR (b), para o analito ACSK.....	108
Figura 67: Histograma de permutação utilizando uma função linear, (a) e (c) e, uma quadrática, (b) e (d), para descrever os resíduos dos modelos construídos, para o analito ACSK. Os gráficos superiores referem-se ao modelo de Raman e os inferiores ao modelo de NIR.....	108
Figura 68: Coeficiente de regressão para os modelos de Raman-PLS (superior) com 2LVs e NIR-PLS (inferior) com 3LVs.....	109
Figura 69: Correlação entre os valores de referência e os previstos pelos modelos PLS, com dados de NIR, para o analito ACSK. Evidenciando que três amostras de validação aparecem antes do primeiro ponto de calibração	120

LISTA DE TABELAS

Tabela 1: Principais deslocamentos químicos referentes a compostos alifáticos, naftênicos e aromáticos presentes em um óleo cru (Masili et al. 2012)	41
Tabela 2: Variância explicada para os modelos de PCA aplicados a diferentes momentos	45
Tabela 3: Classificação de petróleo de acordo com a escala API	46
Tabela 4: Tabela resumo contendo os principais parâmetros de modelagem para a gravidade API.....	51
Tabela 5: Tabela contendo os principais parâmetros de modelagem para o RC.....	55
Tabela 6: Avaliação das amostras com alto valor de leverage, para modelagem da propriedade TIAC.....	59
Tabela 7: Tabela resumo contendo os principais parâmetros de modelagem para a TIAC	61
Tabela 8: Variação da temperatura referente ao ponto de ebulição (PE) com relação ao percentual em massa do corte, para as amostras PET1 e PET2	62
Tabela 9: Erros relativos para os cortes do PI até 90% em massa de destilado, para as amostras de API 17 e 29,5	69
Tabela 10: Cálculo das massas dos edulcorantes na amostra, AM1, via curvas de adição de padrão (a) e (b).	79
Tabela 11: Tabela resumo da avaliação dos pressupostos para as curvas externas referentes aos quatro analitos	90
Tabela 12: Quantificação e comparação da concentração de dois analitos, ASP e CIC, nas 5 primeiras amostras de trabalho, por meio de curvas externas e internas com dois pontos	91
Tabela 13: Comparação estatística entre as metodologias, CE e LC-MS, para quantificação dos analitos nas amostras	91
Tabela 14: Principais parâmetros de modelagem para o analito ASP.....	96
Tabela 15: Valores medidos e preditos e intervalo de confiança para o analito ASP.....	96
Tabela 16: Principais parâmetro de modelagem para o analito CIC	101
Tabela 17: Valores medidos e preditos e intervalo de confiança para o analito CIC.....	101
Tabela 18: Principais parâmetros de modelagem para o analito SAC	105
Tabela 19: Valores medidos e preditos e intervalo de confiança para o analito SAC.....	105
Tabela 20: Principais parâmetros de modelagem para o analito ACSK	109
Tabela 21: Valores medidos e preditos e intervalo de confiança para o analito ACSK.....	110
Tabela 22: Valores medidos e preditos e intervalo de confiança para as propriedades modeladas (API, TIAC e RC).....	117
Tabela 23: Quantificação e comparação da concentração dos analitos nas 30 amostras de trabalho, por meio de curvas externas e internas com dois pontos.....	118
Tabela 24: Comparação entre CE e LC-MS para a concentração dos analitos nas 30 amostras de trabalho	119

ABREVIACÕES E SIGLAS

ABE - Ácido Benzoico

ACSK - Acesulfame-K

ANVISA - Agência Nacional de Vigilância Sanitária

API - *American Petroleum Institute*

ASP - Aspartame

ASTM - *American Society for Testing and Materials*

CE - Eletroforese Capilar (do inglês, Capillary Electrophoresis)

CG - Cromatografia Gasosa

CIC - Ciclamato

CZE - Eletroforese Capilar de Zona (do inglês, Capillary Zone Electrophoresis)

DSC – Calorimetria Exploratória Diferencial (do inglês, Differential Scanning Calorimetry)

EOF - Fluxo Electroosmótico (do inglês, Electroosmotic Flow)

FAO - *Food and Agriculture Organization*

FID - Decaimento Livre de Indução (do inglês, Free Induction Decay)

FT - Transformada de Fourier (do inglês, Fourier Transform)

GA - Algoritmos Genéticos (do inglês, Genetic Algorithms)

¹H-RMN - Ressonância Magnética Nuclear para o núcleo de Hidrogênio

¹³C-RMN - Ressonância Magnética Nuclear para o núcleo de Carbono

icoshift - interval-correlation-shifting

IDA - Ingestão Diária Aceitável

LC-MS - Cromatografia Líquida acoplada a Espectrometria de Massas (do inglês, Liquid Chromatography coupled with Mass Spectrometry)

LD - Limite de Detecção

LQ - Limite de Quantificação

LVs - Variáveis Latentes (do inglês, Latent Variables)

MIR - Infravermelho Médio (do inglês - Mid Infrared)

MLR - Regressão Linear Múltipla (do inglês, Multiple Linear Regression)

MP - Mistura de Padrões

MSC - Correção do Espalhamento Multiplicativo (do inglês, Multiple Scatter Correction)

NIR - Infravermelho Próximo (Near Infrared)

PC - Componente Principal (do inglês, Principal Component)

PCA - Análise por Componentes Principais (do inglês, Principal Component Analysis)

PCR - Regressão por Componentes Principais (do inglês, Principal Component Regression)

PEEQ - Ponto de Ebulição Estimado por Quimiometria

PEV - Curva de Pontos de Ebulição Verdadeiros

PI - Padrão Interno

PLS - Mínimos Quadrados Parciais (do inglês, Partial Least Squares)

PS - Pico de Sistema

R^2_p - Coeficiente de determinação para previsão

R^2_{vc} - Coeficiente de determinação para validação cruzada

RC -Resíduo de Carbono

RMSECV - Raiz quadrada para o erro médio de validação cruzada (do inglês, Root mean Squares of Cross-Validation)

rf - Radiofrequência

RMSEP - Raiz quadrada para o erro médio de previsão (do inglês, Root mean Squares of Prediction)

SAC - Sacarina

SimDis - Destilação Simulada

SNV - Transformação Normal Padrão de Variação (do inglês, Standard Normal Variate)

SVD - Decomposição em valores singulares (do inglês, Single Value Decomposition)

SVR - Regressão por vetores de Suporte (do inglês, Support Vectors Regression)

TBS - Tetraborato de Sódio

TIAC -Temperatura Inicial de Aparecimento de Cristais

TMS - Tetrametilsilano

μ_{ef} - Mobilidade Efetiva

μ_{ep} - Mobilidade Eletroforética

μ_i - Mobilidade Iônica

μ_{osm} - Mobilidade Electrosmótica

SUMÁRIO

1.	QUIMIOMETRIA E TÉCNICAS INSTRUMENTAIS	17
1.1	INTRODUÇÃO	17
1.2	OBJETIVO GERAL	17
1.3	ESTRUTURA	18
1.4	METODOLOGIAS QUIMIOMÉTRICAS	19
1.4.1	ANÁLISE POR COMPONENTES PRINCIPAIS (sigla do inglês, PCA)	19
1.4.2	REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS (sigla do inglês, PLS ou PLSR)	20
1.5	CONSTRUÇÃO DE UM MODELO DE CALIBRAÇÃO MULTIVARIADA	22
1.5.1	ANÁLISE DO MODELO MULTIVARIADO	26
1.5.1.1	PRINCIPAIS FIGURAS DE MÉRITO	26
1.5.1.2	ANÁLISE RESIDUAL	27
1.6	TÉCNICAS INSTRUMENTAIS	28
1.6.1	ESPECTROSCOPIA RAMAN	28
1.6.2	ESPECTROSCOPIA VIBRACIONAL NA REGIÃO DO INFRAVERMELHO PRÓXIMO (sigla do inglês, NIR)	31
1.6.3	ESPECTROSCOPIA DE RESSONÂNCIA MAGNÉTICA NUCLEAR PARA O NÚCLEO DE HIDROGÊNIO (sigla do inglês, ¹ H-RMN)	32
1.6.4	ELETROFORESE CAPILAR DE ZONA (sigla do inglês, CZE)	34
2.	PETRÓLEO	37
2.1	INTRODUÇÃO	37
2.2	OBJETIVOS ESPECÍFICOS	38
2.3	PARTE EXPERIMENTAL	38
2.3.1	METODOLOGIA E ESTRATÉGIA DE AÇÃO	38
2.3.2	INSTRUMENTAÇÃO E SOFTWARES	39
2.4	RESULTADOS	40
2.4.1	ANÁLISE DO CONJUNTO AMOSTRAL	40
2.4.2	DETERMINAÇÃO DA GRAVIDADE API	46
2.4.3	RESÍDUO DE CARBONO (%m/m)	51
2.4.4	TEMPERATURA INICIAL DE APARECIMENTO DE CRISTAIS (°C)	57
2.4.5	DESTILAÇÃO SIMULADA	61
2.5	CONCLUSÕES	69
2.6	PERSPECTIVAS FUTURAS	70
3.	EDULCORANTES	71
3.1	INTRODUÇÃO	71

3.2	OBJETIVOS ESPECÍFICOS	73
3.3	PARTE EXPERIMENTAL.....	73
3.3.1	ELETROFORESE CAPILAR.....	73
3.3.2	CROMATOGRAFIA LÍQUIDA COM DETECÇÃO POR ESPECTROMETRIA DE MASSAS (sigla do inglês, LC-MS).....	75
3.3.3	ESPECTROSCOPIA RAMAN NIR	75
3.4	RESULTADOS	77
3.4.1	ELETROFORESE CAPILAR – REPRODUÇÃO E APLICAÇÃO DA METODOLOGIA.....	77
3.4.2	ESTRATÉGIA DE AÇÃO	80
3.4.3	ESPECTROSCOPIAS	81
3.4.4	ADQUIRINDO AS CONCENTRAÇÕES PELO MÉTODO DE REFERÊNCIA: COMPARAÇÃO ENTRE CE-DAD E LC-MS	89
3.4.5	MODELAGEM MULTIVARIADA PARA DETERMINAÇÃO DOS EDULCORANTES ARTIFICIAIS	92
3.4.5.1	ASPARTAME.....	92
3.4.5.2	CICLAMATO	96
3.4.5.3	SACARINA.....	101
3.4.5.4	ACESULFAME-K	106
3.5	CONCLUSÕES	110
3.6	PERSPECTIVAS FUTURAS	111
	REFERÊNCIAS BIBLIOGRÁFICAS	112
	APÊNDICES	117
	A. INTERVALOS DE CONFIANÇA PARA OS MODELOS CONSTRUÍDOS	117
	B. QUANTIFICAÇÃO DOS ANALITOS NAS AMOSTRAS POR CURVA EXTERNA E INTERNA	118
	C. COMPARAÇÃO ENTRE CE E LC-MS.....	119
	D. EVIDÊNCIA DE MÁ SEPARAÇÃO DO CONJUNTO AMOSTRAL FEITO PELO KENNARD-STONE	120

1. QUIMIOMETRIA E TÉCNICAS INSTRUMENTAIS

1.1 INTRODUÇÃO

De modo geral, em uma análise quantitativa em amostras que possuem matrizes complexas, não lidamos somente com a etapa da medida instrumental e posteriormente o tratamento de dados mas, na maioria das vezes, são necessários alguns passos de pré-tratamento destas. Um dos pontos chave de uma análise quantitativa é o preparo da amostra e posterior eliminação de interferentes, sendo este talvez, o aspecto mais crítico do processo de análise química (Skoog et al. 2005).

Extrações, purificações e reações químicas são alguns dos artifícios utilizados por químicos para tentar isolar o(s) componente(s) de interesse de outros presentes na matriz, os quais poderão interferir no procedimento de determinação analítica. Após essas primeiras etapas de pré-processamento da amostra, ocorre o emprego de técnicas de separação acopladas a poderosos sistemas que tem por finalidade separar e detectar o(s) componente(s) de interesse (Lanças 2008).

No entanto, alternativas como as mencionadas anteriormente podem ser de alto custo, laboriosas e muito dispendiosas no que diz respeito ao tempo total de análise, além de muitas delas não compactuarem com as premissas da química verde. Nesse contexto, existe um apelo analítico para que esforços sejam direcionados à minimização de etapas ao longo de processo de determinação de analitos de interesse nas mais diversas amostras, sem que ocorra perda da confiabilidade dos resultados.

A aplicação de técnicas espectroscópicas associadas à modelagens multivariadas, com o intuito de realizar inferências quantitativas, vem apresentando grande crescimento e consolidação ao longo dos últimos anos e, conseqüentemente, despertando o interesse de grandes companhias de diversos segmentos, como por exemplo a petrolífera e a alimentícia. A principal contribuição e promessa desse tipo de metodologia é que a partir de medidas espectroscópicas diretas, isto é, sem ou com o mínimo de preparo de amostra, se torne possível extrair informações quantitativas a respeito do componente ou propriedade de interesse, o que viabiliza uma análise química rápida, de baixo custo e sustentável.

1.2 OBJETIVO GERAL

O cerne deste trabalho se fundamenta na construção, validação e aplicação de modelos multivariados, por PLS, a partir de dados oriundos de ¹H-RMN para determinação de algumas

propriedades de petróleo e Raman e NIR para a quantificação dos edulcorantes artificiais aspartame, ciclamato, sacarina e acesulfame-K, em amostras de adoçante de mesa em pó.

1.3 ESTRUTURA

Essa dissertação está dividida em três partes. A primeira consiste da abordagem teórica das principais ferramentas quimiométricas utilizadas no decorrer do trabalho, bem como uma breve discussão teórica, realizada por meio de uma revisão da literatura, sobre as principais técnicas instrumentais utilizadas.

A segunda parte, refere-se ao desenvolvimento de metodologias multivariadas, pela aplicação de PLS a dados oriundos de $^1\text{H-RMN}$ ($^1\text{H-RMN-PLS}$) para determinação da gravidade API, Resíduos de Carbono (RC) e Temperatura Inicial de Aparecimento de Cristais (TIAC). Nessa parte também foram construídos modelos para cada percentual correspondente à fração de destilado na curva SimDis, com o objetivo de prever essas temperaturas e por fim plotar uma curva confrontando a mesma. Essa curva adquirida através da modelagem foi chamada de PEEQ, isto é, Pontos de Ebulição Estimados por Quimiometria.

A última parte, referente aos edulcorantes, começa com a reprodução do método desenvolvido por Fernandes e colaboradores, seguida de sua aplicação a amostras comerciais. É, então, mostrado que a metodologia desenvolvida por Eletroforese Capilar (CE) não foi satisfatória para a quantificação dos edulcorantes de interesse em presença de lactose na matriz. Logo, como condição de contorno, utilizou-se respostas oriundas da Cromatografia Líquida com detecção por Espectrometria de Massas (LC-MS) como referência para a construção dos modelos. Os analitos foram estimados por modelagem multivariada quando aplicado PLS aos dados oriundos das espectroscopias Raman e de NIR.

É importante destacar que para a análise dos edulcorantes, a ideia inicial era utilizar as respostas de concentração para os analitos, a partir da CE e portanto, os objetivos e toda a sistemática foram descritas com essa técnica para metodologia de referência na construção dos modelos multivariados. Os objetivos foram mantidos da mesma forma, no entanto, durante a discussão dos resultados será enfatizada a nova estratégia de análise em função dos contratempos. Por esse motivo, a teoria sobre a técnica de LC-MS não foi abordada na primeira parte, sendo somente apresentadas as condições de análise na parte experimental.

1.4 METODOLOGIAS QUIMIOMÉTRICAS

1.4.1 ANÁLISE POR COMPONENTES PRINCIPAIS (sigla do inglês, PCA)

A PCA, do inglês *Principal Component Analysis*, foi descrita primeiramente por Karl Pearson em 1901 em seu artigo intitulado “On lines and planes of closest fit to system of points in space” (Pearson 1901). Nesse artigo o autor discutiu a representação gráfica dos dados que melhor representavam os dados originais. Apesar de sua descrição no início do século, seu uso só foi difundido após a publicação do clássico artigo de Hotteling nos anos 30, época em que ficou marcada como a revolução no uso dos métodos multivariados (Hotteling 1933; Anh & Magi 2009).

Provavelmente a PCA é a técnica de estatística multivariada mais amplamente utilizada em quimiometria, e por causa da importância das medidas multivariadas na área química, é vista por muitos como uma das mais poderosas técnicas de análise de dados (Breton 2007). O principal objetivo da PCA é reduzir a dimensionalidade de um conjunto de dados, preservando ao mesmo tempo o máximo da informação. Isto é feito calculando-se combinações lineares das variáveis originais (Neto et al. 2006).

A PCA consiste na decomposição da matriz de dados originais, X , que relaciona um grande conjunto de variáveis (valores de absorvância em cada comprimento de onda, por exemplo) aos objetos (número de amostras, por exemplo). Essa redução se dá pela introdução de novas variáveis ao sistema, sendo que estas são escritas como uma combinação linear das variáveis originais. As novas variáveis são chamadas de Componentes Principais, ou simplesmente PC (abreviação do inglês). O objetivo dessas PC's é explicar a variabilidade dos dados, sendo que a primeira PC (PC1) é responsável pela maior porcentagem de explicação, a segunda (PC2), que deve ser ortogonal à PC1, explica o que não foi explicado pela primeira e assim sucessivamente (Massart et al. 1991).

As novas matrizes geradas, oriundas da decomposição da matriz original, X , são chamadas de *scores* e *loadings*, representadas matricialmente pelas letras T e P , respectivamente. Essa decomposição pode ser realizada por vários algoritmos diferentes, dos quais os mais utilizados são: NIPALS, do inglês *Non-Linear Iterative Partial Least Squares*, e o SVD, do inglês *Single Value Decomposition*, sendo que ambos levam ao mesmo resultado, dentro da precisão computacional. Além das matrizes de *scores* e *loadings* que são geradas pela decomposição, uma matriz de erros, E , também é gerada, sendo que essa engloba as variâncias não explicadas. O que foi dito anteriormente pode ser descrito matematicamente pela equação (1).

$$X = TP' + E \quad (1)$$

Cada matriz de *scores* consiste de uma série de vetores coluna e a matriz de *loadings* a vetores linhas; estes podem ser denotados por t_a e p_a , respectivamente, e cada componente principal por a (1, 2, 3, ..., A). Após a aplicação da PCA, as variáveis originais são reduzidas ao número de componentes principais relevantes para descrever a variabilidade dos dados. Então, os dados originais são ditos como modelados matematicamente pelas PC's. Logo, usando A PC's, é possível estabelecer um modelo para cada elemento de X , da seguinte forma (Brereton 2003):

$$x_{ij} = \sum_{i=1}^A t_{ia} p'_{aj} + e_{ij} = \hat{x}_{ij} + e_{ij} \quad (2)$$

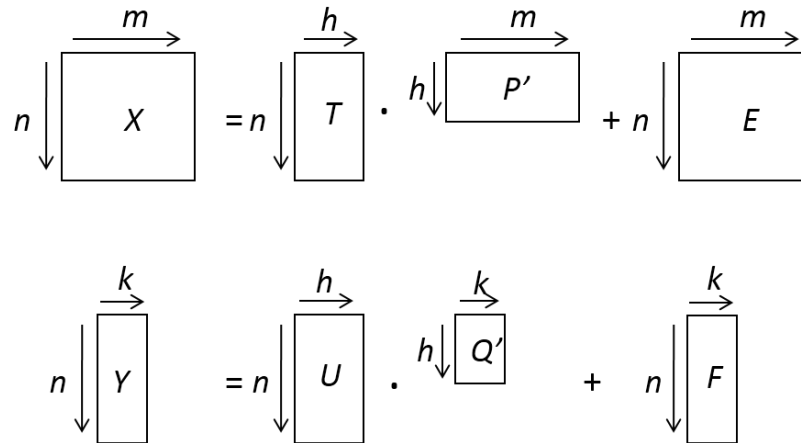
1.4.2 REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS (sigla do inglês, PLS ou PLSR)

O trabalho pioneiro com PLS foi feito no final dos anos sessenta por Herman Wold na área de econometria. Contudo, o uso dessa abordagem para aplicações químicas, iniciada pelos grupos de Svante Wold e Harald Martens, só veio no final dos anos setenta após a primeira publicação de Kowalski e seus colaboradores (Geladi & Kowalski 1986).

O PLS tem sua forma mais simples utilizada em diversas áreas da química e tecnologia, isto é, na forma preditiva de dois blocos. Este é um método para relacionar duas matrizes de dados, X e Y , através de um modelo linear e que vai além das regressões tradicionais, as quais também modelam as estruturas de X e Y ; onde X contém as variáveis (respostas instrumentais) e o Y o valor da concentração para o analito ou propriedade de interesse. O PLS deriva sua grande diferença e utilidade a partir de sua habilidade de analisar dados, a priori desfavoráveis, isto é, ruidosos, apresentando multicolinearidades ou até mesmo com variáveis incompletas em ambos, X e Y (Wold et al. 2001).

Diferentemente da PCR (Regressão por Componentes Principais), que considera os valores de Y livres de erros, o PLS não decompõe e modela exclusivamente as variáveis x , referente aos espectros, mas também leva em consideração a informação contida na matriz de resposta Y , para o cálculo das variáveis latentes (também podendo ser chamadas de fatores ou componentes principais). Sendo assim, as matrizes X e Y são decompostas simultaneamente em uma soma de h variáveis latentes, o que pode ser melhor entendido pela análise da Figura 1 (Brereton 2003; Filgueiras 2011):

Figura 1: Princípio do PLS



FONTE: (Filgueiras, 2011)

A matriz X , contendo n amostras e m variáveis, é decomposta em uma matriz de *scores* ($n \times h$), pois sabe-se que estes se reportam aos objetos, e uma matriz de *loadings* ($h \times m$), sendo este m é referente às variáveis originais (também chamadas de manifestas (Massart et al. 1991)). Da mesma forma a matriz de respostas, Y , com m respostas para k fatores de interesse, também é decomposta. É possível ver, por meio das equações 3 e 4, o que foi dito anteriormente:

$$X_{(n \times m)} = T_{(n \times h)} \cdot P'_{(h \times m)} + E_{(n \times m)} = \sum_h t_i \cdot p'_i + e_i \quad (4)$$

$$Y_{(n \times k)} = U_{(n \times h)} \cdot Q'_{(h \times k)} + F_{(n \times k)} = \sum_h u_i \cdot q'_i + f_i \quad (5)$$

onde T e U referem-se aos *scores* das matrizes X e Y , respectivamente, e P e Q aos *loadings* também de X e Y , enquanto E e F referem-se aos resíduos. O produto de T por P se aproxima dos dados espectrais, enquanto que o produto de U por Q às concentrações verdadeiras.

A correlação entre X e Y é obtida por meio da correlação entre suas matrizes de *scores*, através do coeficiente de regressão linear, como descrito na equação a seguir:

$$u_h = b_h \cdot t_h \quad (6) \quad \text{ou} \quad U_{n \times h} = B_{h \times h} \cdot T_{n \times h} \quad (6')$$

onde b_h , para h variáveis latentes, refere-se a cada termo da matriz de coeficientes de regressão, B (que é uma matriz diagonal), entre as matrizes de *scores* de Y e X , a saber U e T , respectivamente. Dessa forma, pode-se calcular a matriz B pela pseudo-inversa da matriz T :

$$\mathbf{B}_{h_x h} = \left((\mathbf{T}'_{(h_x n)} \cdot \mathbf{T}_{(n x h)}) \right)^{-1} \cdot \mathbf{T}'_{(h_x n)} \cdot \mathbf{U}_{(n x h)} \quad (7)$$

Dessa forma, a matriz Y poderá ser estimada através das informações contidas em B , por meio da equação 8:

$$\mathbf{Y}_{(n x k)} = \mathbf{T}_{(n x h)} \cdot \mathbf{B}_{(h x h)} \cdot \mathbf{Q}'_{(h x k)} + \mathbf{F}_{(n x k)} \quad (8)$$

Uma vez realizados todos os procedimentos de análise das figuras de mérito, conforme será descrito na seção 1.5.1.1, esse modelo contendo sua respectiva matriz B , poderá ser utilizado para a predição de propriedades de novas amostras. Sejam adquiridos espectros de g amostras, onde $T^*_{(g x h)}$ sejam seus *scores*, o vetor contendo os valores das propriedades, $Y^*_{(g x k)}$, será encontrado através da equação:

$$\mathbf{Y}^*_{(g x k)} = \mathbf{T}^*_{(g x h)} \cdot \mathbf{B}_{(h x h)} \cdot \mathbf{Q}'_{(h x k)} \quad (9)$$

Esse tipo de modelagem descrito até o momento (Brereton 2003; Parisotto 2007; Filgueiras 2011), foi discutido de modo genérico para a regressão por mínimos quadrados parciais, pois existem diferenças entre as abordagens para algumas variações do PLS, principalmente no que diz respeito as respostas oferecidas por eles. Por exemplo, para o PLS1, Q' , que foi chamada de matriz durante todo procedimento, trata-se de um vetor e, conseqüentemente, Y também, pois apenas uma resposta estará associada a cada modelagem. O PLS2 é uma extensão do PLS1 e existe uma pequena diferença conceitual, no que diz respeito a resposta da modelagem. Como no PLS2 tem-se, de fato, uma matriz para Q' , pode-se obter respostas simultâneas em uma mesma análise, em conseqüência de Y também ser uma matriz. Isso pode trazer grande vantagem, visto que, por exemplo, todas as concentrações dos analitos requeridos podem ser estimados através de apenas um único modelo de regressão (Brereton 2000).

1.5 CONSTRUÇÃO DE UM MODELO DE CALIBRAÇÃO MULTIVARIADA

A calibração, talvez o maior pilar da quimiometria, trata-se da determinação de uma relação matemática que envolva o uso de um tipo de medida para predizer o valor de um parâmetro de base ou propriedade. Por mais de um século, químicos analíticos vêm realizando

calibrações com o objetivo de determinar a concentração de analitos em diferentes amostras (Brereton 2007).

A calibração busca relacionar as propriedades medidas (variáveis independentes) com as propriedades a serem estimadas (variáveis dependentes). Em sua forma mais simples de uso para um químico analítico, a variável dependente diz respeito a concentração de um único composto e a variável independente a resposta instrumental, seja ela oriunda de métodos espectrométricos ou cromatográficos. Na calibração univariada, usando a espectrofotometria como exemplo de geração de dados, busca-se a determinação da concentração pelo uso de um único comprimento de onda; em geral as concentrações de compostos de interesse numa amostra são determinadas através da resolução de um sistema de equações simultâneas obtidos pela lei de Beer em tantos comprimentos de onda quantos forem os analitos. Caso de misturas binárias simples, muitas vezes bons resultados podem ser obtidos por esse método. Entretanto, quando se passa para amostras reais, onde existem interferentes desconhecidos, a aplicação desse método pode resultar em sérios erros de estimativa (Ferreira et al. 1999; Brereton 2003).

Fazendo uso da calibração multivariada, mais de um comprimento de onda é utilizado permitindo a correção de interferências espectrais e de outros efeitos de matriz, tais como interações químicas, que podem levar a pequenas variações espectrais não-lineares com respeito à concentração (Kalivas 2009). A maior e mais simples vantagem de se usar toda informação medida, é a redução do ruído obtida pelo uso de mais medidas (redundantes) do mesmo fenômeno. Essa melhora pode ser conseguida pela substituição de todas as variáveis medidas por uma nova classe, conhecida como *componentes principais* ou, dependendo do modelo, também conhecida como *variáveis latentes*; essa minimização da influência dos ruídos, produzindo maior robustez, é uma das vantagens mais claras da análise multivariada. Outra interessante vantagem do uso de dados multivariados e seus modelos, é que os sinais não seletivos (sobrepostos por interferentes) podem tornar-se seletivos pelo uso dos artifícios matemáticos utilizados na modelagem, uma vez que é possível incorporar os interferentes no modelo de calibração (Bro 2003).

Para a realização de uma modelagem multivariada primeiramente precisamos de um conjunto significativo de amostras (n), no qual tenhamos conhecimento da concentração do analito em questão ou do valor da propriedade físico-química a ser estimada. A ASTM, formalmente conhecida como *American Society for Testing and Materials*, segundo a norma E-1655-05, reaprova em 2012 um documento de práticas padrões para a realização criteriosa de uma análise quantitativa multivariada. Originalmente essa normativa foi confeccionada para

aplicação da modelagem multivariada a dados provenientes de espectroscopia na região do infravermelho, seja próximo ou médio, mas os passos utilizados para a construção de um modelo desse tipo é praticamente o mesmo para a maioria das espectroscopias (ASTM 2012). Sendo assim, as etapas para a construção de um modelo multivariado são:

- Seleção do conjunto de calibração - Este também é conhecido como conjunto de treinamento. Para que essas amostras sejam incluídas no conjunto, a premissa de existência do valor da propriedade de interesse ou concentração do analito, medidos pelo método de referência, não pode ser violada. Esse conjunto deve ser representativo em todas as fontes de variação;
- Aquisição dos sinais de resposta – Várias fontes podem fornecer as informações referentes à variável independente. Dados espectroscópicos são os mais comuns para esse fim;
- Pré-processamento dos dados – Um passo anterior à construção do modelo propriamente dito é, na maioria das vezes, a necessidade de uma transformação dos dados originais, pois esses podem oferecer contribuições indesejadas que implicarão na diminuição da capacidade preditiva do modelo. Essas contribuições não desejadas podem surgir de várias fontes, como por exemplo, interferências físicas e químicas, bem como de mudanças na condição experimental. Basicamente, essas transformações espectrais consistem de tratamento e pré-processamento dos dados espectrais, o que em última instância, gerará modelos mais parcimoniosos (de Noord 1994). Talvez o método mais simples e mais utilizado seja o da primeira e segunda derivadas, que consiste na remoção de termos constantes e lineares, respectivamente, que podem contribuir para o deslocamento da linha base (Savitzky & Golay 1964). Outros pré-processamentos também vem sendo muito utilizados, como a Correção do Espalhamento da Luz (sigla do inglês, MSC), que usa regressão linear das variáveis espectrais versus o espectro médio e corrige simultaneamente os efeitos aditivos e de espalhamento da luz (Isaksson & Naes 1988), a Normalização, que permite usar os dados adquiridos em uma mesma faixa de amplitude de sinal e os métodos aplicados para se comparar variáveis com diferentes dimensões, a saber, Centrar na Média e Autoescalonamento (Parisotto 2007);

O pré-processamento utilizado nesse trabalho foi a transformação padrão normal de variação (do inglês, Standard Normal Variate - SNV), que tem grande aplicabilidade para

correção de linha de base (Barnes et al. 1989; Ribeiro 2009) e também atua corrigindo o efeito de espalhamento multiplicativo, de maneira análoga ao MSC. A correção de cada espectro é realizada pela normalização com relação ao seu próprio desvio padrão, não necessitando de um espectro ideal. As equações abaixo mostram isso. O espectro médio é dado por:

$$\bar{X}_i = \sum_{j=1}^p \frac{X_i}{p} \quad (10)$$

O espectro corrigido é:

$$X_{i(SNV)} = \frac{X_i - \bar{X}_i}{\sqrt{\frac{\sum_{j=1}^p (X_i - \bar{X}_i)^2}{p - 1}}} \quad (11)$$

onde X se refere a uma matriz contendo n espectros cada um com p comprimentos de onda e \bar{X}_i é a média do vetor que representa o espectro i da matriz original, X .

- Construção do modelo matemático – Nesse momento é preciso definir, com base no tipo de dados, o modelo que melhor estabeleça uma relação entre o sinal instrumental e a propriedade requerida. Existem numerosos métodos para esse fim, sendo que os principais e que dão base para os mais sofisticados são: Regressão Linear Múltipla (sigla do inglês, MLR), Regressão por Componentes Principais (sigla do inglês, PCR) e o mais amplamente utilizado, PLS, que já foi descrito no item 1.4.2. Normalmente esse modelo é construído sobre 2/3 do universo amostral, sendo que o restante é utilizado para validá-lo.
- Validação do modelo – Uma vez que o universo amostral foi dividido em amostras de calibração e validação, essa etapa pode ser dividida em duas partes:
 - A primeira trata-se de uma validação interna, aplicada sobre o conjunto de calibração. O método de validação cruzada, por meio de um procedimento conhecido como *leave one out*, consiste de um processo iterativo no qual uma parcela do conjunto de calibração, que foi previamente segmentado, é utilizado para validá-lo. O parâmetro de finalização do cálculo é a minimização do erro residual, que é dado pela variável RMSECV (sigla do inglês, *Root Mean Squared Error of Cross-Validation*);

- A segunda refere-se ao processo que é, de fato, responsável pela resposta preditiva do modelo. Para isso um novo conjunto de amostras é utilizado, isto é, que não participaram do conjunto de calibração; estes também devem possuir espectros e valores de concentração ou propriedade físico-química conhecida (referem-se aos 1/3 restantes do universo amostral).
- Avaliação das figuras de mérito – A validação de um procedimento analítico deve ser certificado através de parâmetros conhecidos como figuras de mérito. Para dados de primeira ordem, como os que serão utilizados nesse trabalho, a exatidão, precisão, robustez e tendência são similares para a análise univariada (ordem zero). Linearidade, sensibilidade, razão sinal ruído, seletividade e intervalo de confiança necessitam de uma abordagem diferenciada e específica para dados multivariados (Valderrama et al. 2009).
- Aplicação do modelo - Uma vez analisadas essas figuras de mérito, o modelo pode ser empregado para estimar as propriedades requeridas em amostras desconhecidas, a partir de seus espectros.

1.5.1 ANÁLISE DO MODELO MULTIVARIADO

1.5.1.1 PRINCIPAIS FIGURAS DE MÉRITO

Normalmente, a primeira figura de mérito a ser analisada é a exatidão, por meio do RMSEP (do inglês, *Root Mean Squares Error of Prediction*), que é uma média ao longo do número de amostras utilizadas para testar o modelo; sua forma matemática é dada pela equação 12. Entretanto, este é um parâmetro global que incorpora tanto erros sistemáticos quanto aleatórios. Sendo assim, outras estimativas de figuras de mérito para o modelo também devem ser realizadas, como por exemplo a explicitação dos resultados por meio de intervalo de confiança (Olivieri et al. 2006; Valderrama et al. 2009).

O RMSECV, para a validação cruzada e o RMSEP, para a previsão são dados por:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_{i_p} - \hat{y}_{i_p})^2}{n_v}} \quad (12) \quad e \quad RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_{i_{vc}} - \hat{y}_{i_{vc}})^2}{n_{vc}}} \quad (13)$$

Onde y_i e \hat{y}_i correspondem aos valores de referência e previstos pelo modelo, respectivamente e n_v é o número de amostras de validação; os índices vc e p referem-se a

validação cruzada e previsão, respectivamente. Outro importante parâmetro que é sempre expresso para inferir sobre a qualidade do modelo é o coeficiente de determinação, que permite a avaliação de ajuste linear entre os valores medidos pelo método de referência e os previstos pela modelagem. O cálculo desse parâmetro se dá pela equação:

$$R^2 = 1 - \frac{\sum_i (y_r - y_p)^2}{\sum_i (y_r - \bar{y})^2} \quad (14)$$

onde, y_r e y_p se referem aos valores de referência e previsto, enquanto o \bar{y} é a média do vetor contendo os valores previstos pela modelagem. O ajuste dos dados será considerado melhor, quando mais próximo de 1 for o valor calculado. Vale ressaltar que, pelo fato da validação consistir em uma etapa interna (validação cruzada) e outra externa (com um conjunto não participante da calibração), o modelo final contará com dois valores de R^2 (ASTM 2012; Filgueiras 2014).

1.5.1.2 ANÁLISE RESIDUAL

Uma outra avaliação importante para garantir a qualidade e validade do modelo construído é quanto aos resíduos deixados pelo mesmo. Os resíduos na calibração multivariada também são utilizados como uma estimativa do erro experimental, sendo esse dado por:

$$e_i = y_i - \hat{y}_i \quad (15)$$

Os erros sistemáticos afetam a estimativa sempre no mesmo sentido, gerando resultados acima ou abaixo do esperado. A ASTM E-1655 traz um teste para identificação desse tipo de erro, denominado teste de *bias*. Como se trata de um teste de *student*, o valor calculado é dado por:

$$bias = \frac{\sum_{i=1}^n (y_r - y_p)}{n} \quad (16)$$

Como trata-se de um teste bilateral as duas hipóteses são anunciadas:

$$H_0: bias = 0$$

$$H_a: bias \neq 0$$

Logo o valor de t calculado para comparação com o valor t tabelado é dado por:

$$t_{calc} = \frac{|bias|\sqrt{n}}{SVD} \quad (17)$$

onde SVD é o desvio padrão dos erros. Logo, se o valor de t_{calc} for menor que o valor de t_{tab} , com $n-1$ graus de liberdade para o nível de confiança de 95% (utilizado nesse trabalho), não há

evidências de erros sistemáticos, pois o valor está na região de aceitação de H_0 (Filgueiras 2014).

Em um modelo corretamente ajustado espera-se que os resíduos sejam uniformes em magnitude e que se distribuam normalmente em torno de zero. Um ajuste pobre do modelo reflete tendências em seus resíduos, onde a informação relevante não tenha sido incorporada ao modelo. A presença dessas tendências nos resíduos de regressão podem ser identificadas, muitas vezes por meio de uma análise visual do gráfico de resíduos contra os valores de referência. Contudo, nem sempre uma análise visual é suficiente para assegurar ou não a presença de tendência no modelo, sendo necessário testes estatísticos para essa avaliação.

O teste de permutação não-paramétrico permite a avaliação de tendência no modelo ajustado. Uma grande vantagem desse teste é que ele é livre de distribuição, o que significa que os pressupostos de normalidade e homocedasticidade não são requeridos. Esse teste é baseado na aleatoriedade dos valores presentes no vetor Y (valores de referência) enquanto a ordem dos dados de X (variáveis instrumentais) é mantida constante. A hipótese nula (H_0) garante que os resíduos (e_i) são independentes de y_i , enquanto a hipótese alternativa (H_a) diz que os resíduos estão relacionados com y_i de acordo com a equação:

$$e_i = g(y_i) + \epsilon_i \quad (18)$$

onde $g(y_i)$ é uma função polinomial que é postulante a modelar a relação entre os resíduos e os valores de referência. Caso os resíduos sejam modeláveis, por uma função do primeiro (b_1 é o coeficiente do polinômio) ou segundo grau (b_2 é o coeficiente do polinômio) por exemplo, dentro do nível de significância trabalhado (nesse caso 0,05), dizemos que o modelo é tendencioso e este é um indício de um ajuste duvidoso, tornando o modelo questionável e sendo importante uma avaliação mais criteriosa do mesmo. e não trata-se de um bom ajuste (Filgueiras et al. 2014).

1.6 TÉCNICAS INSTRUMENTAIS

1.6.1 ESPECTROSCOPIA RAMAN

De maneira genérica, quando se refere à Espectroscopia, imediatamente deve-se correlacionar com a interação entre radiação eletromagnética e a matéria (Oliveira 2001), onde o principal objetivo é a determinação dos níveis de energia de átomos ou moléculas. No caso de moléculas, a região espectral onde as transições são observadas depende dos tipos de níveis envolvidos, sejam eles Eletrônicos, Vibracionais ou Rotacionais. As formas de se obter o espectro referente às frequências vibracionais são basicamente duas: por absorção de radiação

eletromagnética na região do infravermelho ou pelo espalhamento inelástico de radiação após a irradiação de radiação monocromática contra a amostra, o que caracteriza o efeito Raman (Sala 2008a).

O efeito Raman foi previsto teoricamente por Smekal (1923) e descoberto experimentalmente por Raman (Raman & Krishnan 1928). Quando a radiação eletromagnética atinge a molécula, seja na região do visível ou infravermelho, ocorre o espalhamento da mesma, mas de duas formas distintas. A primeira é dita como espalhamento elástico ou Rayleigh, onde os fótons de radiação incidente apenas mudam de direção e a segunda é conhecida como inelástico ou Raman, onde além de mudar de direção o fóton é espalhado com energia diferente da inicial (Sala 2008b).

No Raman, diferentemente do infravermelho, não é necessário que haja uma polaridade intrínseca da molécula, mas a atividade está ligada a variação do Momento Dipolo Induzido (\vec{P}) na molécula pelo Campo Elétrico (\vec{E}) da radiação incidente. A presença do campo causa na molécula um deslocamento da nuvem eletrônica em relação aos núcleos; a propriedade das cargas se rearranjarem sob a ação de um campo constitui a Polarizabilidade (α). A relação entre essas grandezas é dada por:

$$\vec{P} = \alpha \cdot \vec{E} \quad (19)$$

Para haver atividade no efeito Raman, o momento dipolo de transição deve ser diferente de zero, como no infravermelho, contudo o operador utilizado na integral é do momento de dipolo induzido, que pode ser escrito em termos da polarizabilidade devido à proporcionalidade, como pode ser visto abaixo:

$$(\alpha_{ij})_{mn} = \int \psi_m \alpha_{ij} \psi_n d\tau \quad (i, j \text{ são } x, y \text{ ou } z) \quad (20)$$

onde ψ_m e ψ_n são as funções de onda dos estados m e n , que possuem diferentes energias e, α_{ij} são os componentes do tensor polarizabilidade que podem ser desenvolvidos em série de potência adquirindo a seguinte forma:

$$(\alpha_{ij})_{mn} = (\alpha_{ij})_0 \int \psi_m \psi_n d\tau + \left(\frac{d\alpha_{ij}}{dq} \right)_0 \int \psi_m q \psi_n d\tau \quad (21)$$

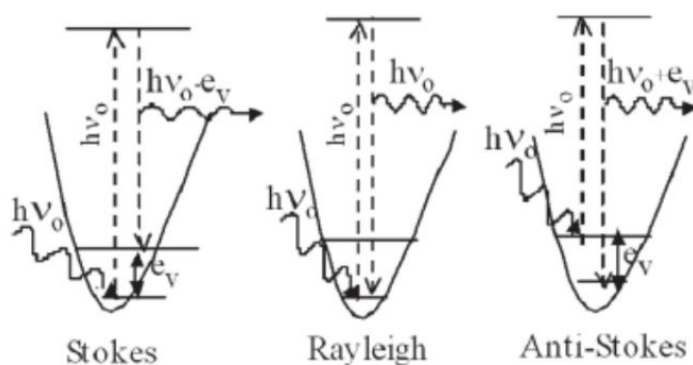
No espalhamento Stokes ou anti-Stokes, os estados vibracionais m e n são diferentes, logo a primeira integral do segundo membro é zero, devido a ortogonalidade entre ψ_m e ψ_n .

Para $m = n$ o primeiro termo corresponde ao espalhamento Rayleigh. Para o segundo termo ser diferente de zero é necessário que duas condições sejam satisfeitas:

- $\left(\frac{d\alpha_{ij}}{dq}\right)_0 \neq 0$ – isto é, ao menos um dos componentes do tensor de polarizabilidade deve variar com pequena vibração em torno da posição de equilíbrio;
- $\int \psi_m q \psi_n d\tau \neq 0$ – para que essa integral seja diferente de zero o produto $\psi_m q \psi_n$ deve ser função par. A regra de seleção para o oscilador harmônico é $\nu = \pm 1$, onde o sinal “+” vale para Stokes e o “-” para anti-Stokes.

O mecanismo de espalhamento Raman pode ser representado pelo esquema a seguir:

Figura 2: Esquema dos espalhamentos Raman e Rayleigh



Fonte: (Sala 2008b)

No espalhamento Stokes o fóton, com energia $h\nu_0$, encontra a molécula no estado vibracional fundamental e após a colisão a leva para um estado intermediário (ou virtual), no fim do processo decaindo para um estado vibracional excitado (e_ν), originando a radiação espalhada com energia $h\nu_0 - e_\nu$. No espalhamento Rayleigh (elástico), após a interação do fóton com a molécula, esta volta ao estado fundamental de energia vibracional e , portanto, sem modificação na energia no fóton incidente. Já no anti-Stokes, o fóton já encontra a molécula no estado excitado e após a interação o mesmo retorna ao estado fundamental com energia $h\nu_0 + e_\nu$.

1.6.2 ESPECTROSCOPIA VIBRACIONAL NA REGIÃO DO INFRAVERMELHO PRÓXIMO (sigla do inglês, NIR)

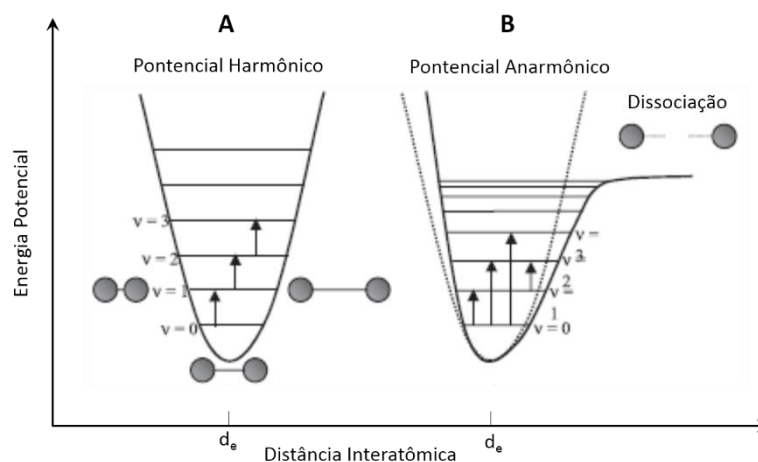
O modelo mais simples, que permite fácil entendimento do fenômeno, é o clássico onde são consideradas massas (correspondentes aos núcleos) pontuais ligadas por molas desprezíveis (correspondentes às ligações químicas)(Sala 2008a). A lei de Hooke estabelece que a energia para esse tipo de sistema é dada por:

$$E = \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} \quad (22)$$

onde μ é a massa reduzida, k é a constante de força e h é a constante de Planck.

A vibração molecular pode ser descrita por um modelo simplificado admitindo um oscilador harmônico para o qual a energia potencial (V) é uma função do deslocamento dos núcleos em questão (x), o que pode ser visto abaixo:

Figura 3: Representação esquemática dos modelos harmônico (A) e anarmônico (B) para a energia potencial de uma molécula diatômica, onde d_e refere-se à distância de equilíbrio



Fonte: ADAPTADA DE (Pasquini, 2003)

A Figura 3(A) mostra o comportamento da energia potencial como função do deslocamento dos núcleos, no entanto essa abordagem falha, pois sistemas moleculares não podem assumir perfis de energia contínua prevista pelo modelo de massa-mola clássico. Quanticamente, através da aplicação da equação de Schrodinger para tal sistema, é possível chegar-se a conclusão de que os níveis de energia são discretos e dados pela equação:

$$E_v = \left(v + \frac{1}{2} \right) h\nu \quad (23)$$

onde

$$\nu = \frac{1}{2} \sqrt{\frac{k}{\mu}} \quad (24)$$

Uma restrição é imposta ao modelo do oscilador harmônico quântico, o qual permite apenas transições entre níveis de energia adjacentes, $\Delta\nu = \pm 1$. Para que ocorra absorção e posterior transição entre os níveis de energia vibracionais, o fóton deve ter energia equivalente a $\Delta E = E_{\nu_2} - E_{\nu_1} = \Delta\nu h\nu$. Esse modelo pode ajudar no entendimento da espectroscopia vibracional, entretanto para NIR não é possível utilizá-lo, pois as transições requeridas para o fenômeno estudado são superiores a ± 1 , isto é, utilizando esse modelo as principais bandas de sobretons e combinação não podem ser entendidas.

A Figura 3 mostra um modelo mais realístico para uma molécula diatômica, onde são considerados os comportamentos não-ideais do oscilador levando em conta a repulsão entre as nuvens eletrônicas quando ocorre a aproximação dos núcleos atômicos e os comportamentos variáveis da força de ligação quando os átomos se afastam um do outro. De fato, em uma molécula real, o grande afastamento dos núcleos pode levar a ruptura da ligação.

Aplicando mecânica quântica a equação de Morse tem-se os novos níveis de energia, agora levando em consideração a anarmonicidade, como é mostrado na equação abaixo:

$$E_\nu = \left(\nu + \frac{1}{2}\right) h\nu - \chi_m h\nu \left(\nu + \frac{1}{2}\right)^2 \quad (25)$$

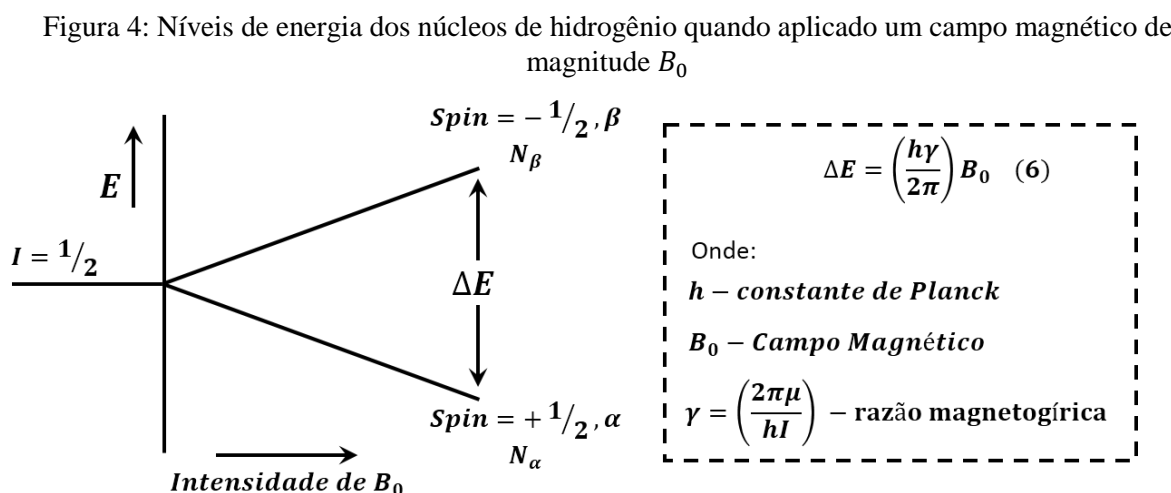
onde χ_m é a constante de anarmonicidade. O modelo do oscilador anarmônico também impõe algumas restrições sobre os possíveis estados de energia das moléculas, entretanto, ele prevê a ocorrência de transições com $\Delta\nu = 2$ ou superior e a existência de bandas de combinação entre os modos vibracionais. Esses dois tipos de bandas são os tipos de absorção mais comuns que ocorrem na região de NIR (Pasquini 2003; Burns & Ciurcsak 2008).

1.6.3 ESPECTROSCOPIA DE RESSONÂNCIA MAGNÉTICA NUCLEAR PARA O NÚCLEO DE HIDROGÊNIO (sigla do inglês, $^1\text{H-RMN}$)

A espectroscopia de Ressonância Magnética Nuclear (do inglês, RMN), é basicamente uma forma espectroscópica de absorção de radiação eletromagnética na região de radiofrequências (rf), sob certas condições, quando submetido à ação de um campo magnético. Essa absorção é função de determinados núcleos da molécula e o espectro gerado é um registro gráfico das frequências dos picos de absorção contra suas intensidades (Silverstein et al. 2007).

Dentre as propriedades fundamentais que os núcleos possuem, o momento magnético (μ) e o spin nuclear (I) são importantíssimos para descrição do fenômeno de RMN. Pode-se obter facilmente espectros dos núcleos com número de spin igual a $1/2$ e, portanto, distribuição de carga esférica e uniforme; núcleos com spin igual a I ou superiores têm distribuição de carga não-esférica. Essa assimetria é descrita por um momento elétrico quadrupolar que afeta o tempo de relaxação.

No caso do Hidrogênio, que possui spin igual a $1/2$, existem dois níveis de energia com um pequeno excesso de população no estado de energia mais baixo, pois segundo a distribuição de Boltzmann $N_\alpha > N_\beta$. Os estados de energia α e β são separados por uma diferença de energia ΔE , conforme mostrado na figura abaixo:



Fonte: ADAPTADA DE (Silverstein et al. 2007)

A razão magnetogírica é uma constante nuclear fundamental. Após a aplicação do campo magnético o núcleo começa a mudar de direção sobre seu próprio eixo de rotação com frequência angular ω ; trata-se de um movimento de precessão. Um conjunto de núcleos equivalentes em movimento aleatório de precessão em torno do eixo-z (direção de B_0) tem magnetização resultante na direção z, sem resultante no plano xy.

Quando a radiofrequência (ν_1) aplicada equivale à frequência de precessão dos núcleos equivalentes (chamada frequência de Larmor, ν_L em MHz), ocorre o efeito de ressonância e a relação básica da RMN pode ser descrita por:

$$\nu_L = \nu_1 = \left(\frac{\gamma}{2\pi}\right) B_0 \quad (26)$$

No espectrômetro pulsado, a amostra é submetida ao campo magnético e irradiada com pulsos de rf de alta potência. Esse pulso excita simultaneamente todos os núcleos da amostra, que após o relaxamento produz o decaimento livre de indução (do inglês, FID), que é a soma da radiação de todos os núcleos durante o tempo da emissão. A informação contida no FID, que é em função do tempo, é convertida para o domínio da frequência pela transformada de Fourier e, então, o espectro no domínio da frequência é obtido (Silverstein et al. 2007)b.

1.6.4 ELETROFORESE CAPILAR DE ZONA (sigla do inglês, CZE)

Eletroforese Capilar (sigla do inglês, CE) é uma técnica de separação baseada na migração diferenciada de compostos neutros, íons solvatados e espécies ionizáveis, em um campo elétrico. Dentre vários modos de separação possíveis em CE, no presente trabalho foi utilizado a CZE. A utilização de tubos com dimensões capilares (15-100 μm de d.i.), na maioria das vezes constituídos de sílica fundida, soluciona talvez a maior limitação da técnica que consiste na dificuldade de dissipação de calor pela passagem de corrente através do meio condutor, conhecido como Efeito Joule.

A parede interna do capilar é constituída de sílica fundida, sendo que essa é caracterizada pela presença de vários grupos silanóis (Si-OH), os quais, em média, apresentam caráter ácido. Em contato com o meio aquoso, esses grupos silanóis podem ser desprotonados tornando a superfície do capilar negativamente carregada (Si-O⁻); outra hipótese para o aparecimento de carga na superfície é devido a adsorção de íons solvatados. O fato é, que após a formação de uma dupla camada elétrica e posterior aplicação de um campo elétrico tangenciando à superfície, ocorre a formação de um movimento unilateral de íons em direção ao eletrodo de carga oposta. Durante essa migração, os íons transportam moléculas de água, induzindo um fluxo de solução como um todo, conhecido como fluxo eletrosmótico (do inglês, EOF).

A existência do EOF tem importantes implicações na CE, como principalmente a automação da técnica, permitindo a introdução da amostra e detecção diretamente no capilar. Quando a velocidade eletrosmótica é de grande magnitude, o fluxo é responsável pela condução dos solutos, sem distinção de carga, em direção ao detector. Assim, a análise simultânea de amostras contendo tanto solutos catiônicos, neutros ou aniônicos é possível, embora não haja discriminação temporal entre solutos neutros. Além disso, as características de alta eficiência da técnica estão, em parte, vinculadas ao perfil radial da velocidade eletrosmótica.

Para soluções de eletrólitos compostos de um ácido ou base fracas, existem pelo menos duas espécies em equilíbrio, a molécula não ionizada (com mobilidade zero) e a base ou ácido conjugado, cada qual com um valor de mobilidade. Assim, da mesma forma que íons simples são caracterizados por um valor de mobilidade iônica (μ_i), o conceito de mobilidade efetiva é utilizado para descrever a migração de eletrólitos fracos. A migração de qualquer substância i como um único soluto, presente na solução em diferentes formas j , em um campo elétrico pode ser entendida com base no conceito de mobilidade efetiva (μ_{ef}):

$$(\mu_{ef})_i = \sum (\mu_j \alpha_j) \quad (27)$$

onde α_j é a função de distribuição e μ_j que se refere a mobilidade iônica dada por:

$$\mu_j = \frac{|z_j|e}{6\pi\eta r_j} \quad (28)$$

sendo z_j a carga do íon, e é a carga elementar, η é a viscosidade do meio e r_j é o raio iônico efetivo. As funções de distribuições dependem da magnitude das constantes de dissociação do soluto (pKa) e determinam a predominância relativa das várias espécies num intervalo de pH.

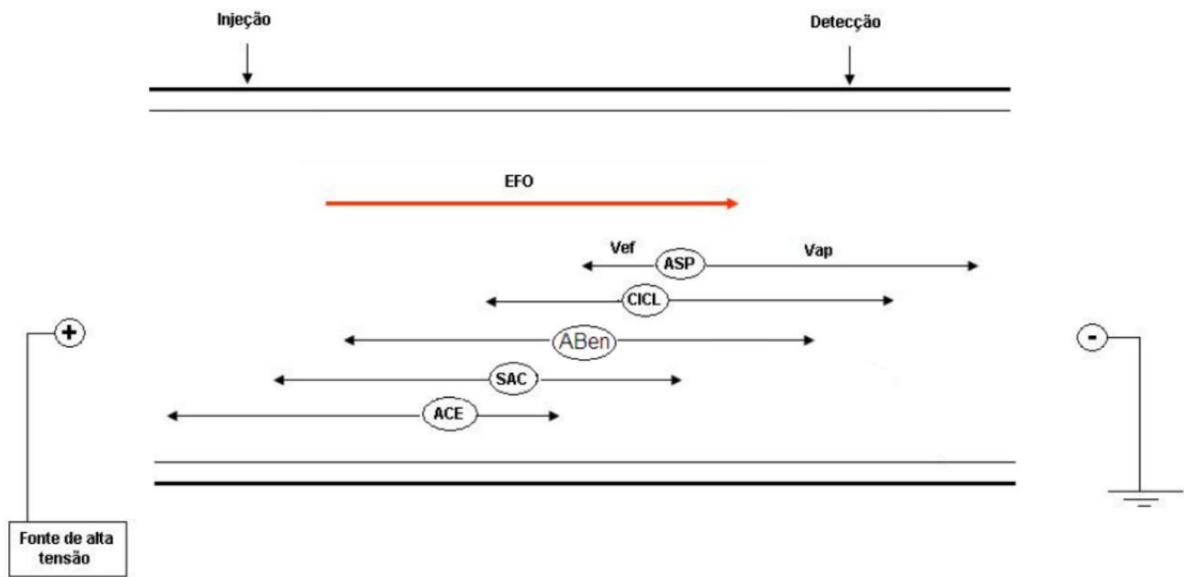
Na presença do EOF, o tempo de migração do soluto i , do ponto de injeção até a posição do detector (L_{det}), é determinado por sua velocidade aparente de migração (v_i), que é a combinação vetorial da velocidade eletroosmótica (v_{osm}) e eletroforética (v_{ep}), de acordo com:

$$t_i = \frac{L_{det}}{v_i} = \frac{L_{det}}{v_{osm} + v_{ep}} = \frac{L_{det}L_{tot}}{(\mu_{osm} + \mu_{ep})V} \quad (29)$$

onde μ_{osm} e μ_{ep} são as mobilidades eletroosmótica e eletroforética, respectivamente, V é o potencial aplicado e L_{tot} é o comprimento total do capilar (Tavares 1996; Tavares 1997).

A Figura 5 a seguir ilustra o modo de separação, por CZE, que será utilizado nesse trabalho. Como trata-se de uma separação em contra-fluxo, pois no pH trabalhado (9,15) todos os analitos se comportam como ânions, as velocidades eletroosmótica e eletroforética tem a mesma direção, porém sentidos contrários. Isso faz com que a espécie de menor mobilidade efetiva alcance o detector primeiro, visto que o EOF exercerá sobre esta maior influência no arraste; em outras palavras a velocidade aparente ($v_{apa} = v_{osm} - v_{ep}$) será maior para o ASP do que para os outros analitos e a ordem de eluição para os analitos obedecendo ao mecanismo CZE é: ASP, CYC, SAC e ACSK.

Figura 5: Esquema do mecanismo de separação por CZE em contra-fluxo



Fonte: (Oliveira 2010).

2. PETRÓLEO

2.1 INTRODUÇÃO

Petróleo, que literalmente significa “óleo de rocha”, é o termo utilizado para descrever uma infinidade de fluidos ricos em hidrocarbonetos que foram acumulados durante muitos anos em reservatórios subterrâneos. O petróleo, que também é chamado de óleo cru, varia drasticamente em cor, odor e viscosidade, o que reflete a diversidade de sua origem (Speight, 2002; Farah 2012). Quimicamente, os petróleos são formados, em maior parte de sua constituição, por diferentes famílias de hidrocarbonetos – parafínicos, naftênicos ou aromáticos – o que lhes confere diferentes características. A enorme gama de compostos presentes no petróleo implica em uma complexa caracterização, que é feita pela checagem de mais de vinte parâmetros, sendo estes dados por diversas propriedades físico-químicas, como por exemplo: densidade, viscosidade, índice de refração, entre outros; isso culmina em um longo tempo de avaliação, podendo chegar a um ano.

Os métodos atuais padrão para caracterizar óleos brutos são baseados principalmente em uma série de elaborados e dispendiosos testes desenvolvidos pela American Standards for Testing and Materials (ASTM). Como uma alternativa muito mais rápida, técnicas envolvendo espectroscopia molecular vêm sendo propostas, com o principal argumento de que esses novos métodos teoricamente refletem, qualitativamente e quantitativamente, a composição completa da amostra. A maioria dos estudos de caracterização de óleos crus envolvem a técnica de espectroscopia vibracional na região do infravermelho médio (MIR) e próximo (NIR). Além das técnicas espectroscópicas vibracionais mais usuais para esse fim, um grande avanço nas pesquisas, envolvendo estudos com RMN vem acontecendo (de Peinder et al. 2009). Entretanto, a utilização dessas técnicas instrumentais modernas gera um grande número de dados, o que respalda o uso de ferramentas quimiométricas no auxílio de previsão de propriedades físico-químicas do petróleo.

Talvez a grande dificuldade de se modelar usando dados oriundos de RMN, é que esses, nem sempre, estão prontamente adequados para a análise utilizando os métodos quimiométricos chamados “bilineares”. Espectros de RMN são geralmente uma sequência de sinais de resposta, aproximadamente picos escritos por lorentzianas, diferindo na forma, intensidade e posição. Uma vez que sabemos que essas três entidades do espectro trazem informações importantes, mudanças na posição do pico de um mesmo analito entre amostras, não está de acordo com os pressupostos de bilinearidade e assim deterioram a modelagem quimiométrica. Apesar das normalizações, o desalinhamento dos espectros ainda ocorre e essa é a razão pela qual um

alinhamento, *a posteriori*, é requerido. O algoritmo *icoshift*, do inglês *interval-correlation-shifting*, alinha independentemente cada espectro de RMN ao marcado (que pode opcionalmente ser um sinal atual ou um sintético como uma média ou mediana) pela maximização da correlação entre os intervalos definidos pelo usuário (Savorani et al. 2010); esse algoritmo foi usado nesse trabalho.

2.2 OBJETIVOS ESPECÍFICOS

Como foi dito anteriormente, hoje em dia a caracterização do petróleo pode demorar de seis meses até um ano e isso torna o processo, de maneira geral, oneroso. Outro motivo crucial é que para realizar todos os ensaios físico-químicos referentes a caracterização do óleo bruto, é necessário um grande volume de amostra, sendo esse por volta de quarenta litros. Diante desse quadro, os principais objetivos para essa etapa são:

- Utilização de pequeno volume de amostra para a realização dos ensaios espectroscópicos;
- Utilização da Espectroscopia de Ressonância Magnética nuclear para o núcleo de Hidrogênio (^1H -RMN), que confere aquisições espectrais muito rápidas devido ao menor tempo de relaxação do núcleo – por volta de 3 minutos – o que implica no aumento da frequência analítica;
- Construção de modelos quimiométricos para previsão de propriedades físico-químicas de petróleo. Para esse trabalho as seguintes propriedades foram estimadas por meio da modelagem: Gravidade API, Resíduos de Carbono (RC), Temperatura Inicial de Aparecimento de Cristais (TIAC) e a Destilação Simulada (SimDis);
- Validação dos modelos, por meio da estimativa das figuras de mérito;
- Aplicação dos modelos construídos em novas amostras.

2.3 PARTE EXPERIMENTAL

2.3.1 METODOLOGIA E ESTRATÉGIA DE AÇÃO

Quando as amostras chegam ao laboratório de petróleo da UFES (Labpetro-UFES), de cada uma delas são recolhidos 5 frascos distintos, sendo que o primeiro se destina as análises básicas de rotina, isto é, checagem das propriedades físico-químicas para classificação prévia e programação dos passos posteriores da amostra. O segundo frasco serve como checagem do

primeiro, sendo realizadas as mesmas análises físico-químicas. Os outros três frascos se destinam aos ensaios espectroscópicos e de destilação simulada.

A estocagem desses frascos se diferencia pelas características de densidade de cada amostra; essa densidade é expressa em termos do grau API (API – sigla em homenagem ao *American Petroleum Institute*), que é uma escala arbitrária para densidade de petróleos e seus derivados. Os óleos mais leves (de grau API acima de 30) são estocadas sob refrigeração, em uma geladeira comum, para que não haja perda das frações leves do petróleo. As demais amostras, com API abaixo de 30, são armazenadas em condições ambiente.

Especialmente para o caso da aquisição de espectros de ^1H -RMN e de ^{13}C -RMN, primeiramente as amostras com grau API abaixo de 30, sejam moderadas ou pesadas, são aquecidas em estufa a 60°C durante aproximadamente 20 minutos para que haja uma maior homogeneização e melhor fluidez das mesmas. Especificamente para a aquisição dos espectros de ^1H -RMN, após o aquecimento, as amostras foram preparadas dissolvendo-se uma massa correspondente a 5% em diclorometano- d_2 (CD_2Cl_2), totalizando um volume de $600\mu\text{L}$. Uma homogeneização branda é realizada, para que não haja adesão de amostra na parede do frasco e consecutiva perda, uma vez que trabalha-se com pouca quantidade de amostra, cerca de 20 a 30mg. Como última etapa, antecedente a análise instrumental, essa solução preparada é transferida para o tubo de RMN, de 5mm no caso do ^1H -RMN.

Para a realização desse trabalho, um universo de 92 amostras estava disponível, mas nem todas puderam ser utilizadas para modelagem, visto que algumas tiveram problemas na aquisição dos espectros e/ou na obtenção dos valores da propriedade físico-química, via metodologia de referência. Diante disso, um número diferente de amostras foi utilizado para a predição das diferentes propriedades que serão abordadas a seguir.

2.3.2 INSTRUMENTAÇÃO E SOFTWARES

Após o preparo das amostras, as análises de ^1H -RMN foram realizadas no espectrômetro da marca Varian, modelo VRMNS 400, operando com um campo magnético de 9,4 T, utilizando uma sonda 5 mm BroadBand $^1\text{H}/^{19}\text{F}/\text{X}$, a 25°C ; Tetrametilsilano (TMS) foi utilizado como referencial interno. Em resumo, os parâmetros instrumentais foram: Frequência de 399,73 MHz, janela espectral de 6410,3 Hz, tempo de Aquisição de 2,556 s, tempo espera de 1 s, pulso de 90° (11,7 μseg) e número de transientes igual a 64. Essas análises foram realizadas pela doutoranda Emanuele Catarina, responsável pela parte experimental de RMN do projeto de quimiometria do petróleo.

Para o desenvolvimento das modelagens, envolvendo PCA e PLS, foi utilizado o software MATLAB, versão R2012, com as principais rotinas escritas e adaptadas para o tipo de trabalho sendo desenvolvida, em primeira instância, pelo coordenador do projeto de quimiometria Dr. Paulo Filgueiras. O software da Microsoft, Excel versão 2013, também foi muito utilizado para a manipulação dos dados brutos oriundos da espectroscopia e dos métodos de referência.

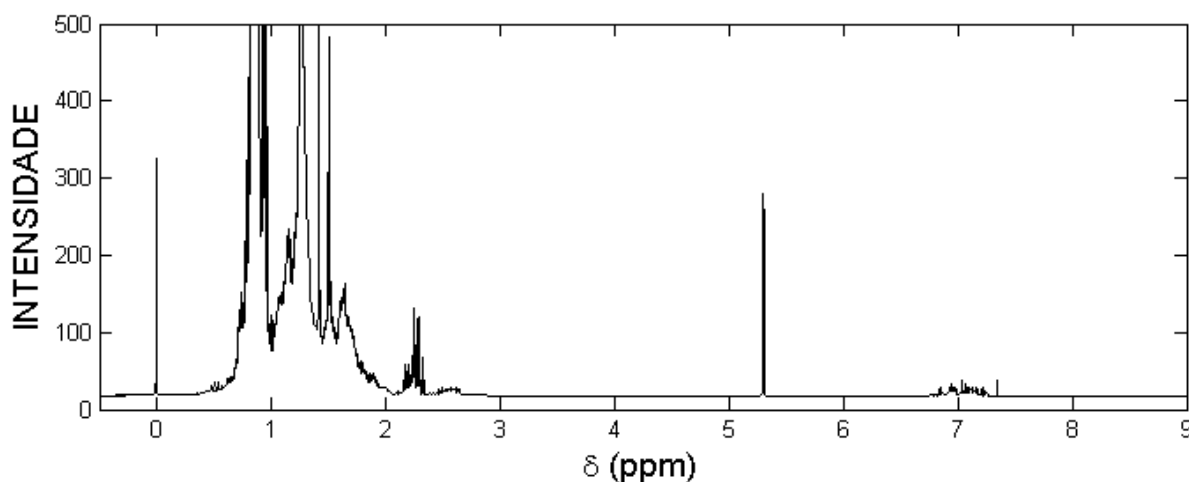
2.4 RESULTADOS

2.4.1 ANÁLISE DO CONJUNTO AMOSTRAL

Conforme discutido anteriormente, o petróleo é constituído por uma enorme gama de compostos e mesmo utilizando uma técnica espectroscópica poderosa para elucidação estrutural, como é o caso do RMN, não é possível obter informações a nível molecular especificamente para cada constituinte da matriz. Sendo assim, os espectros de ^1H -RMN obtidos podem ser discutidos de acordo com as principais regiões que trazem informação, levando em consideração os principais tipos de hidrocarbonetos existentes no petróleo.

Na Figura 6 é mostrado o perfil espectral para um óleo cru e na Tabela 1 as informações referentes aos deslocamentos químicos dos principais grupamentos responsáveis por tais sinais. Vale ressaltar que diclorometano, deslocamento químico em 5,32 ppm, foi o solvente escolhido pelo fato de seu sinal aparecer na região em que a amostra não apresenta sinal, portanto eliminando problemas referentes a sobreposição.

Figura 6: Espectro de ^1H -RMN para um óleo cru



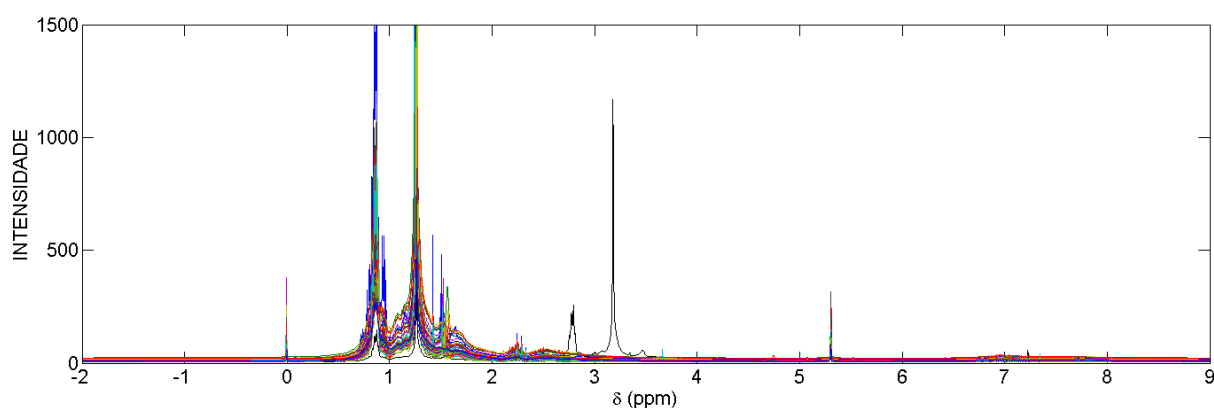
Fonte: ELABORADA PELO PRÓPRIO AUTOR

Tabela 1: Principais deslocamentos químicos referentes a compostos alifáticos, naftênicos e aromáticos presentes em um óleo cru (Masili et al. 2012)

Alifáticos	ppm	Naftênicos	ppm	Aromáticos	ppm
CH ₃	0,5 – 1,0	CH _{2(ax)} aromático	1,1 – 1,5	Benzeno	7,2
CH ₃ β-aromático	1,2 – 1,3	CH _{2(eq)} aromático	1,6 – 2,0	Mono-aromáticos	6,3 – 7,2
CH ₂	1,0 – 1,4	CH α-aromático	2,5 – 2,8	Di-aromáticos	7,2 – 8,3
CH ₂ β-aromático	1,5 – 1,6	Cicloexano	1,4 – 1,5	Poli-aromáticos	8,3 – 9,3
CH	1,4 – 1,6	Ciclopentano	1,4 – 1,5		
CH ₃ α-aromático	2,0 – 2,5				
CH ₂ α-aromático	2,5 – 2,8				
CH α-aromático	2,8 – 3,1				

Fonte: ELABORADA PELO PRÓPRIO AUTOR

A Figura 7 mostra o agrupamento de 92 espectros de ¹H-RMN, referentes aos diferentes óleos disponíveis. Analisando essa figura é possível perceber que esses espectros são qualitativamente similares, visto que os mesmos apresentam um perfil semelhante. Entretanto, pode-se notar que para uma determinada amostra, seu espectro apresenta claramente ao menos dois picos em posições diferentes dos principais apresentados por todas as outras amostras do conjunto, sendo o de maior intensidade na região de 3,2 ppm.

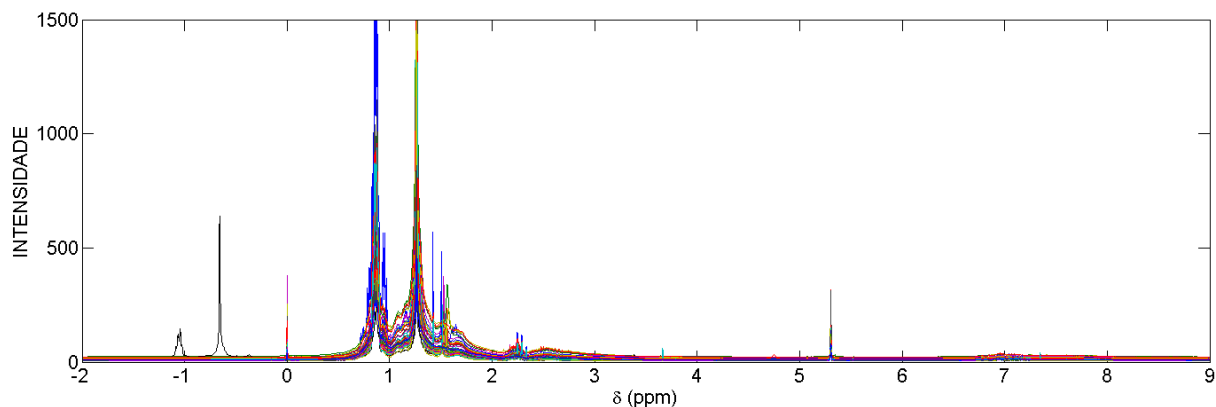
Figura 7: Conjunto espectral de ¹H-RMN para 92 amostras de petróleo

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Como descrito na introdução desse capítulo, o alinhamento é fundamental para se trabalhar com dados de RMN, quando o intuito é modelagem. Sendo assim, esse procedimento foi realizado utilizando o algoritmo *icoshift*, onde o espectro foi dividido em 50 intervalos e após isso ocorreu a reconstrução do sinal por meio da ferramenta FFT; todas essas etapas são internas ao algoritmo descrito e disponibilizado por Savorani e colaboradores (Savorani et al.

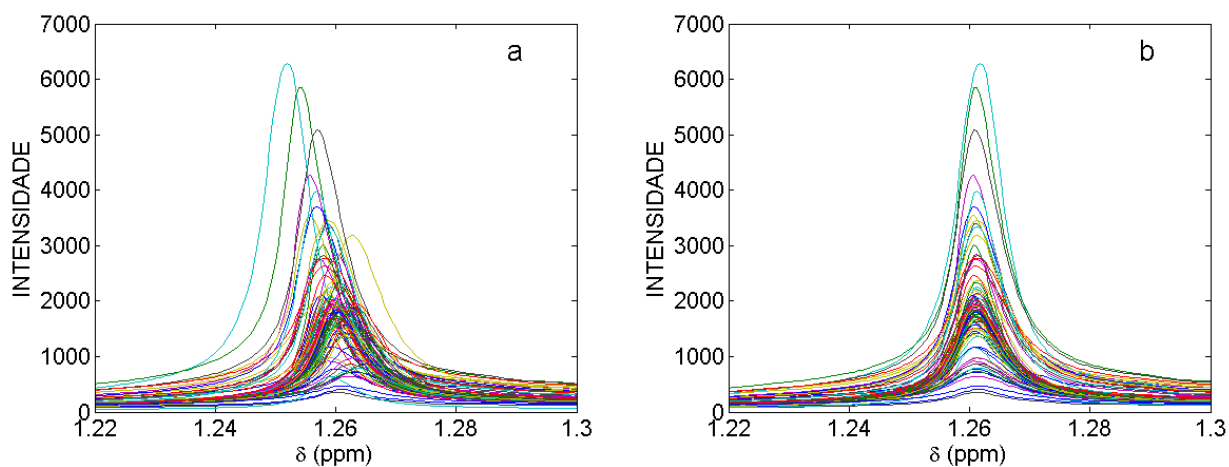
2010). O conjunto espectral após o alinhamento pode ser visto na Figura 8 e o efeito visual do alinhamento na Figura 9.

Figura 8: Conjunto espectral de ^1H -RMN, alinhado, para 92 amostras de petróleo



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 9: Efeito do alinhamento nos espectros de ^1H -RMN para amostras de petróleo, pelo uso da ferramenta *icoshift*: (a) bruto e (b) após o alinhamento



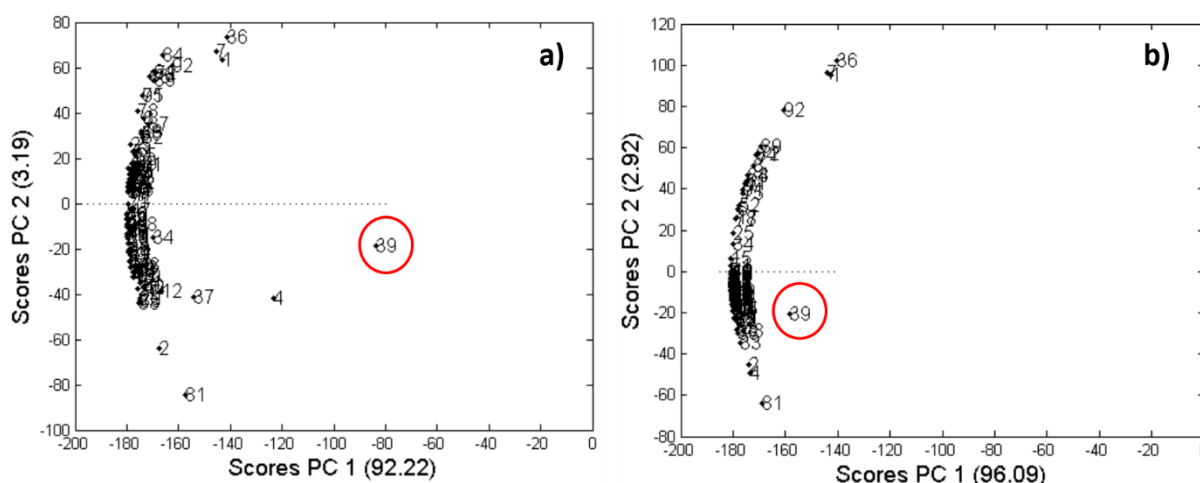
Fonte: ELABORADA PELO PRÓPRIO AUTOR

O algoritmo se mostra eficiente para a finalidade de alinhamento, no entanto é preciso enfatizar que a relevância química é sempre mais importante que a matemática e quando analisamos o que ocorreu com o sinal em 3,2 ppm (sinal duvidoso) após o alinhamento, pode-se notar que ele foi apenas transladado para outra região do espectro; local esse que não apresenta significância química se tratando do espectro de ^1H -RMN, pois a faixa para os deslocamentos químicos referente a esse núcleo está compreendida entre 0 e 14 ppm. Por conta disso, uma avaliação minuciosa com relação ao conjunto amostral foi realizada, cujo objetivo

era a verificação de possíveis *outliers* (amostra anômala) espectrais, sendo essa amostra a principal candidata. Para tanto, a análise por componentes principais foi realizada.

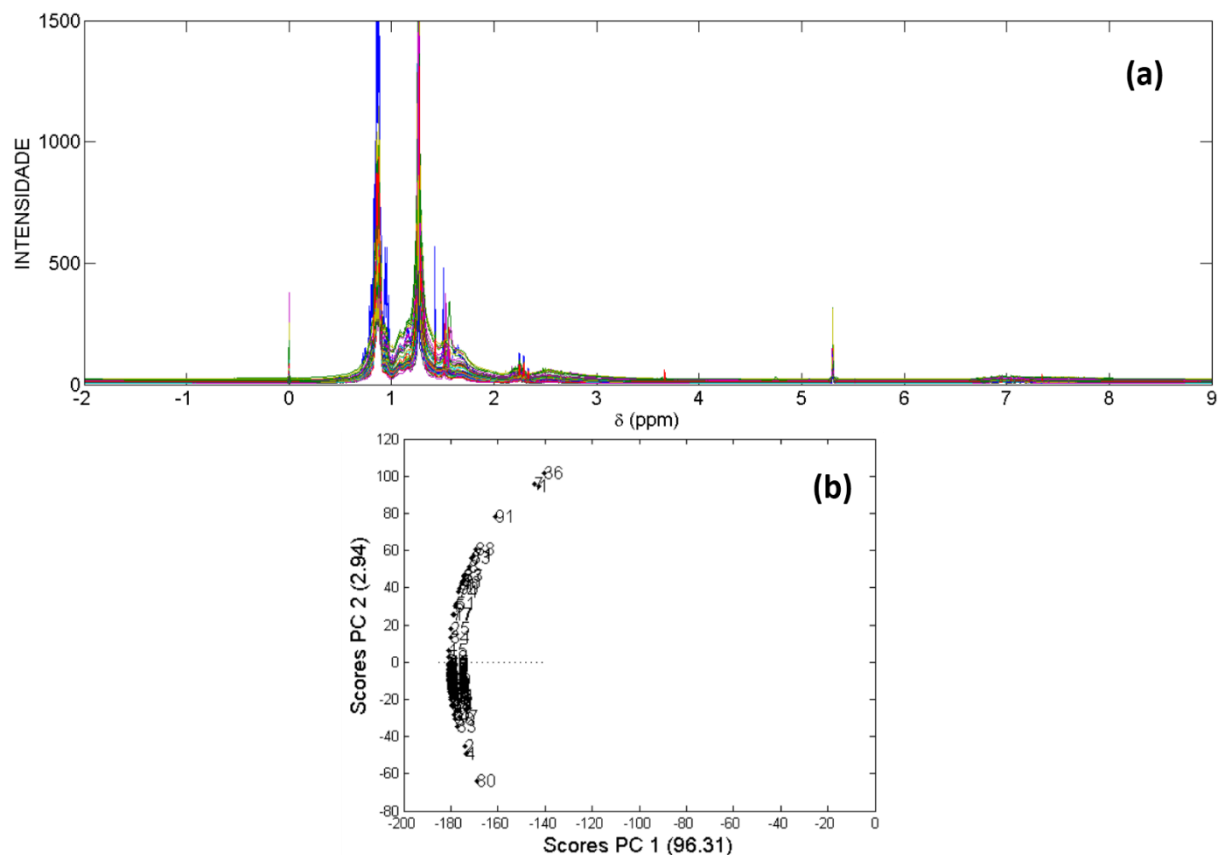
Primeiramente foi realizada uma PCA no conjunto espectral bruto, isto é, sem o alinhamento. O único tratamento realizado no conjunto espectral foi o pré-processamento pela ferramenta SNV, discutida na seção 1.5, pelo fato de todas as modelagens que foram feitas adiante utilizaram esse mesmo procedimento, logo se faz necessário aplicar nas PCAs para que a discussão seja padronizada. Nessa primeira PCA (Figura 10(a)) é possível perceber que algumas amostras se distanciam do principal grupo, o que já era de se esperar, visto que o conjunto de 92 amostras é composto por óleos de diferentes poços, variando de extra leves (API 54) a extra pesados (API 11,2), no entanto, a amostra de número 39, circulado em vermelho, chama atenção por ser a que mais se distancia das demais, além de estar sozinha.

Figura 10: PCA para o conjunto espectral sem o alinhamento (a) e após o alinhamento (b)



Fonte: ELABORADA PELO PRÓPRIO AUTOR

A Figura 10(b) mostra a PCA realizada após o alinhamento do conjunto espectral e apesar do maior agrupamento das amostras, confirmando concomitantemente com a Figura 9 que o alinhamento é importante, a amostra de número 39 continua fora, sozinha e bem distante do maior agrupamento. Analisando esses fatos, essa amostra tem a maior probabilidade de ser responsável pelo sinal problema apresentado nas Figura 7 e Figura 8, sendo então, caso se confirme essa hipótese, um *outlier*. Para a confirmação dessa hipótese, a amostra de número 39 foi retirada do conjunto amostral e uma PCA foi realizada, bem como plotado o novo conjunto de espectros que compõe, a partir de agora, o universo amostral a ser utilizado para as modelagens, o que pode ser visto nas Figura 11 (a) e (b).

Figura 11: (a) conjunto espectral de ^1H -RMN após a retirada do outlier (b) PCA após a retirada do outlier

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Pela figura acima pode-se confirmar que a amostra de número 39 era a anômala procurada, pois o sinal problema desapareceu na Figura 11(a) e amostra que aparecia sozinha na Figura 10(b) já não se encontra presente na Figura 11(b). A retirada dessa amostra é muito importante, pois sua permanência poderia acarretar problemas futuros na modelagem, afetando diretamente a capacidade preditiva do modelo. Uma possível explicação para tal sinal é a contaminação com algum solvente durante a manipulação das amostras, anteriormente à análise instrumental para a aquisição dos espectros.

Outro aspecto interessante da PCA que se deve comentar, é com relação ao aumento da variância explicada pelos modelos construídos. A Tabela 2 mostra que após o alinhamento e a retirada do outlier, houve um aumento significativo na explicação da primeira PC, passando de 92,22 para 96,31%. O percentual total foi praticamente o mesmo, no entanto, quanto maior a explicação das variáveis originais, por meio da PC1, menor será a probabilidade de sobreajustes do modelo e mais fiel este será à natureza dos dados.

O gráfico da Figura 12 mostra os locais do espectro de ^1H -RMN em que as novas variáveis deram mais peso para a construção do modelo. A PC1, responsável por maior explicação, apresenta um perfil mais fino e dá maior importância à região de maior informação química do espectro, sendo essa responsável pelos principais deslocamentos químicos de hidrocarbonetos. A PC2 se apresenta com um perfil um pouco mais largo, também dando relevância à mesma região que a PC1 e a PC3, por descrever o restante das variáveis originais já aparenta um perfil com mais ruídos.

Tabela 2: Variância explicada para os modelos de PCA aplicados a diferentes momentos

	Variância Explicada			
	PC1	PC2	PC3	Total
CAM-b	92.22	3.19	2.33	97.74
CAM-ali	96.09	2.92	0.53	99.54
CAM-so	96.31	2.94	0.54	99.79

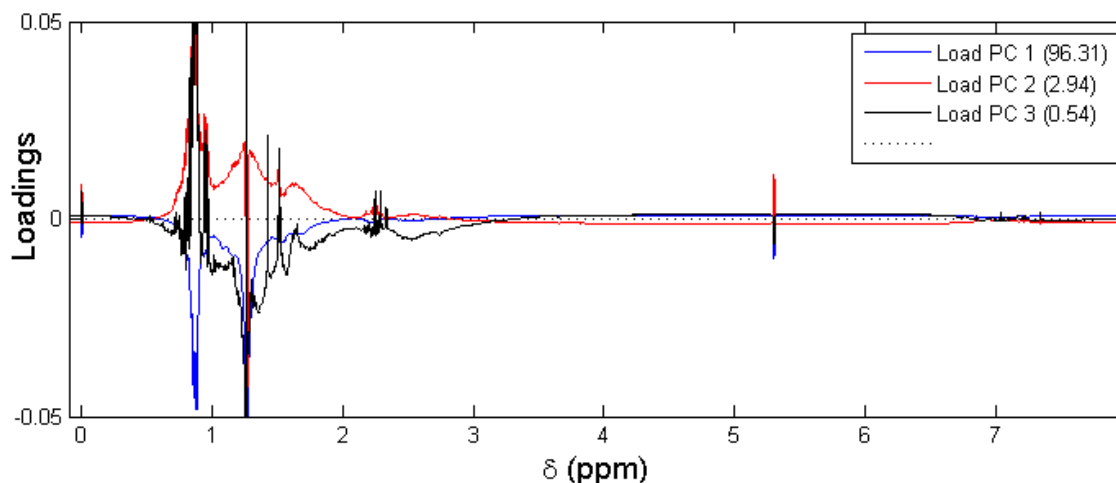
CAM-b - Conjunto amostral bruto

CAM-ali - Conjunto amostral alinhado

CAM-so - Conjunto amostral após a retirada do outlier do conjunto alinhado

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 12: Loadings PCA realizada sem o alinhamento para o conjunto espectral



Fonte: ELABORADA PELO PRÓPRIO AUTOR

2.4.2 DETERMINAÇÃO DA GRAVIDADE API

O primeiro parâmetro, ou propriedade físico-química de petróleo modelada via metodologia proposta por esse trabalho, foi a gravidade API. O método de referência para essa análise é descrito pela ASTM D4052 ou ISO 12185. O API é uma escala arbitrária para medir a densidade de petróleos e seus derivados e quanto mais denso for o óleo, menor será seu valor de API e vice-versa, o que pode ser explicado pela equação a seguir (Farah 2012):

$$API = \frac{141,5}{\rho(60^{\circ}F)} - 131,5$$

A primeira classificação dos petróleos normalmente é feita com relação a gravidade API e sua primeira importância se dá ao fato de poder diferenciar os óleos, de extra-leves a extra-pesados, por esse índice adimensional (Dalmaschio 2012). Outro ponto interessante é que conhecendo as características de fluidez durante o escoamento dentro de dutos e tubulações, pois a viscosidade que seria a propriedade diretamente relacionada a esse fato tem uma grande dependência com a temperatura, pode-se agilizar no processo de tomada de decisão para destinar esse óleo para futuro processamento e refino, visto que as refinarias são adequadas para processar determinados tipos de óleo. A tabela de classificação dos petróleos de acordo com a escala API é dada a seguir:

Tabela 3: Classificação de petróleo de acordo com a escala API

Petróleo	API	Densidade 20/40 °C
Extra-Leve	> 40,0	< 0.821
Leve	40.0 - 33.0	0.821 - 0.857
Médio	33.0 - 27.0	0.857 - 0.889
Pesado	27,0 - 19.0	0.889 - 0.937
Extra-Pesado	19.0 - 15.0	0.937 - 0.962
Asfáltico	< 15.0	>0.962

Fonte: (Dalmaschio 2012)

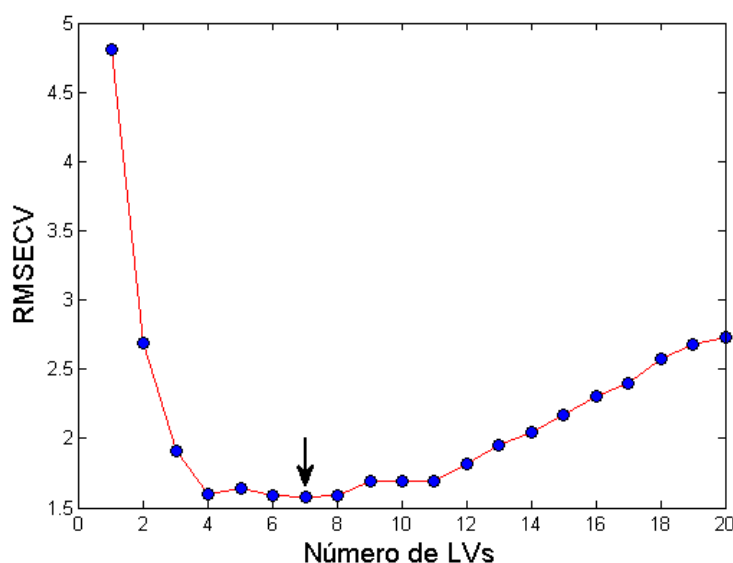
Uma vez de posse dos valores de referência para a propriedade API e dos espectros das amostras, é possível realizar a modelagem. O passo posterior à aquisição dos espectros, é a separação do universo amostral (85 nesse caso) em dois conjuntos, isto é, calibração e predição. Para essa primeira propriedade foi utilizado o algoritmo de separação Kennard-Stone. Nessa separação, 70% das amostras (60 das 85) foram utilizadas para calibração e 30% (25 das 85) ficam para predição. É importante ressaltar que essa etapa é crucial para a construção do

modelo, bem como para a etapa de validação, visto que não se recomenda a seleção de amostras para validar que não sejam contempladas pelo modelo construído.

Separado o conjunto, as amostras de validação são guardadas para a aplicação no segundo momento; o objetivo agora é construir o modelo de calibração. Os dados foram pré-processados pela Transformação Padrão Normal de Variação (sigla do inglês, SNV), que tem por objetivo a correção da linha de base espectral; que é realizada pela normalização de cada espectro para o seu próprio desvio padrão.

A etapa seguinte trata-se da validação interna, conforme descrito na seção 1.5. O objetivo agora é escolher o número ótimo de variáveis latentes em que ocorra a minimização do RMSECV e para essa análise foi utilizado o método da validação cruzada. A princípio, como não se sabia o número ótimo de variáveis latentes, efetuou-se o cálculo com 20. O gráfico para otimização das variáveis latentes é dado a seguir:

Figura 13: Otimização do número de variáveis latentes para construção de modelo de calibração, para a propriedade API



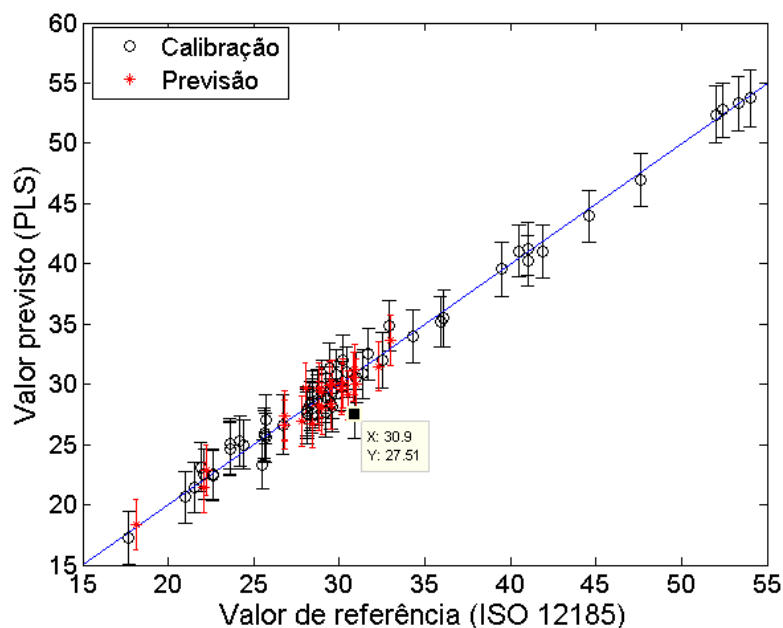
Fonte: ELABORADA PELO PRÓPRIO AUTOR

O gráfico acima indica que o número ótimo de variáveis latentes é 7, isto é, a quantidade de novas variáveis para minimizar o erro de validação cruzada; 4, 6 e 8 também fornecem valores baixos de RMSECV, no entanto, 7 é número ótimo que fornecerá um modelo com maior capacidade preditiva. O número de LVs foi então otimizado para 7 e deu-se continuidade ao processo, partindo para a etapa de validação externa, quando um conjunto não participante da

calibração é utilizado para, de fato, realizar a validação e avaliar a capacidade preditiva do modelo construído.

O gráfico de correlação entre os valores de referência (adquiridos pela metodologia oficial) e os valores previstos pelo modelo construído pode ser visto a seguir:

Figura 14: Correlação entre os valores de referência e os previstos pelo modelo PLS, para a propriedade API



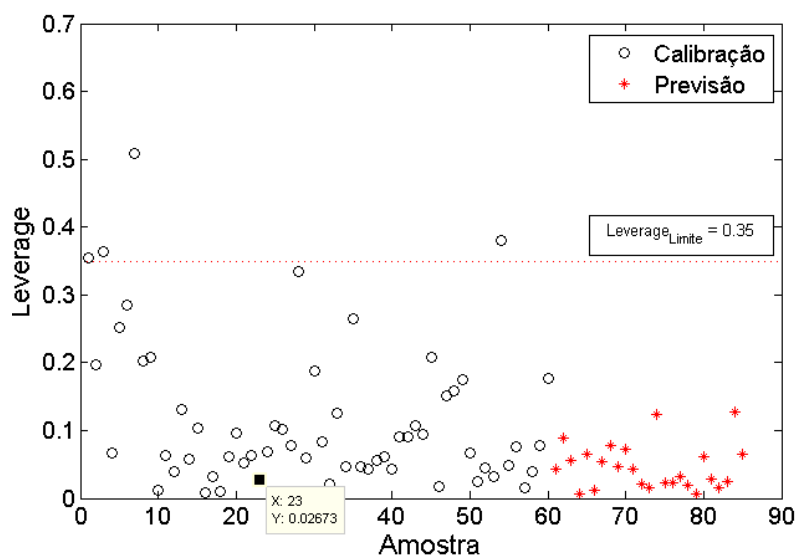
Fonte: ELABORADA PELO PRÓPRIO AUTOR

Pelo gráfico da Figura 14 é possível perceber uma boa correlação entre os dados medidos e previstos, visto que o coeficiente de correlação para a previsão foi de 0,939 e para a calibração foi de 0,966. Outro ponto a se destacar na Figura 14 é a amostra com valor de referência 30,9, pois ela não é contemplada pela curva nem em seu intervalo de confiança (dado pela barra o delimita), o que confere um valor de resíduo mais alto. Então uma avaliação mais crítica para averiguação dessa amostra foi realizada.

Com o objetivo de checar mais a fundo a influência dessa amostra, que faz parte do conjunto de calibração, o gráfico de *leverage* pelas amostras é plotado na Figura 15. A *leverage* traz informação de influência da amostra no modelo de regressão, portanto, se uma amostra apresenta alto valor de *leverage*, quer dizer que ela influencia muito na construção do modelo. Em contrapartida, a verificação da presença de amostras anômalas não pode ser realizada somente com base no valor de *leverage*, mas também o gráfico de resíduos deve ser investigado.

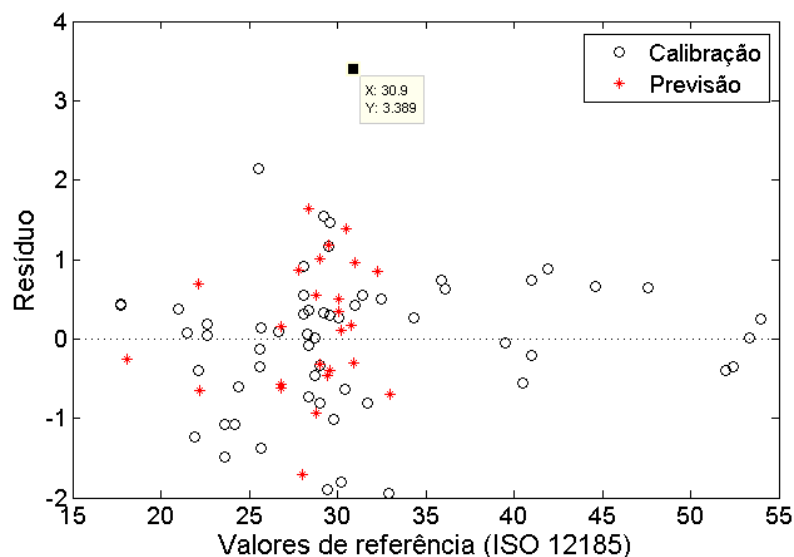
Logo, amostras com alto valor de *leverage* e alto valor de resíduo, devem ser descartadas, para que a qualidade do modelo seja assegurada. O gráfico de resíduos é apresentado na Figura 16.

Figura 15: Gráfico de *leverage*, que demarcam a influência das amostras na construção do modelo, para a propriedade API



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 16: Gráfico dos resíduos que permite a identificação de amostras que não tiveram boa previsão



Fonte: ELABORADA PELO PRÓPRIO AUTOR

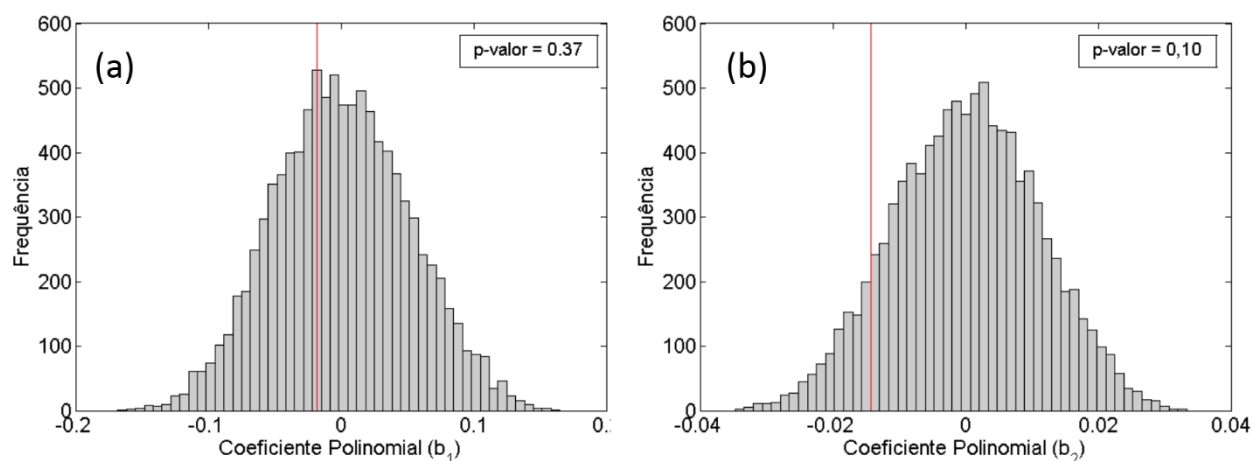
A linha pontilhada no gráfico referente a Figura 15, corresponde ao valor limite para a *leverage*, que é 0,36. A amostra destacada, de número 23 (eixo-x) está abaixo desse limite (*Leverage* = 0,027) e portanto influencia pouco na construção do modelo, consequentemente

não devendo ser excluída, mesmo apresentando o mais alto valor de resíduo do conjunto, o que pode ser visto na Figura 16; o intervalo de confiança para essa previsão foi de 28,86 a 32,94 ($30,9 \pm 2,04$) e o valor estimado pelo modelo foi de 27,5.

A exatidão do modelo foi verificada pelo cálculo do RMSEP, apresentando o valor igual a 0,816, que é perfeitamente aceitável pela indústria petrolífera brasileira que preconiza erros menores que 1 para essa propriedade físico-química. Esse limite de aceitação está atrelado ao fato das refinarias se adequarem para processar o óleo de acordo com sua procedência e especificações, das quais, a gravidade API é um dos principais parâmetros de caracterização do óleo para a tomadas de decisão e desfecho de todo processo, da logística ao refino.

Outro ponto interessante na abordagem desse trabalho é a análise pós-modelagem, que trata da avaliação dos resíduos, conforme discutido no item 1.5.1.2. Pela análise visual da Figura 16 é muito subjetiva a afirmação de presença ou ausência de erros sistemáticos e aleatórios, então os resíduos foram submetidos aos referidos testes estatísticos para esse tipo de análise. De acordo com o teste de bias, não há presença de erros sistemáticos, pois o valor obtido é menor que o tabelado, no nível de significância de 0,05 ($t_{cal} = 0,86 < t_{tab(\alpha=0,05,n-1)} = 2,06$). Para o teste de permutação, tentou-se descrever os resíduos por duas diferentes funções, uma linear e uma quadrática e as Figura 17 (a) e (b) mostram o resultado:

Figura 17: Histograma de permutação utilizando uma função linear (a) e uma quadrática (b) para descrever os resíduos do modelo construído, para a propriedade API



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Os gráficos (a) e (b) da Figura 17 mostram a distribuição dos coeficientes linear (b_1) e quadrático (b_2) permutados para avaliação do modelo. Como já era esperado, os coeficientes se distribuem normalmente com maior frequência em torno de zero. A linha vertical vermelha

indica o valor dos coeficientes da função usada para tentar descrever os resíduos. Para ambos os casos, utilizando uma função linear ou quadrática, pode-se afirmar que não existem erros de tendência no modelo para o nível de significância de 0,05, visto que os valores encontrados para o teste de permutação foram superiores a 0,05, a saber 0,37 e 0,10, para função linear e quadrática, respectivamente. A tabela a seguir mostra resumidamente os principais parâmetros para a modelagem da propriedade gravidade API e no Apêndice A os valores medidos e previstos são confrontados, além de apresentar os intervalos de confiança para cada amostra.

Tabela 4: Tabela resumo contendo os principais parâmetros de modelagem para a gravidade API

API	
R^2_{vc}	0,966
R^2_p	0,939
RMSECV	1,575
RMSEP	0,816
E_{sist}	$t_{cal} = 0,86 < t_{tab} = 2,06$
$E_{aleat}(b1)$	p-valor = 0,37 > 0,05
$E_{aleat}(b2)$	p-valor = 0,10 > 0,05

Fonte: ELABORADA PELO PRÓPRIO AUTOR

2.4.3 RESÍDUO DE CARBONO (%m/m)

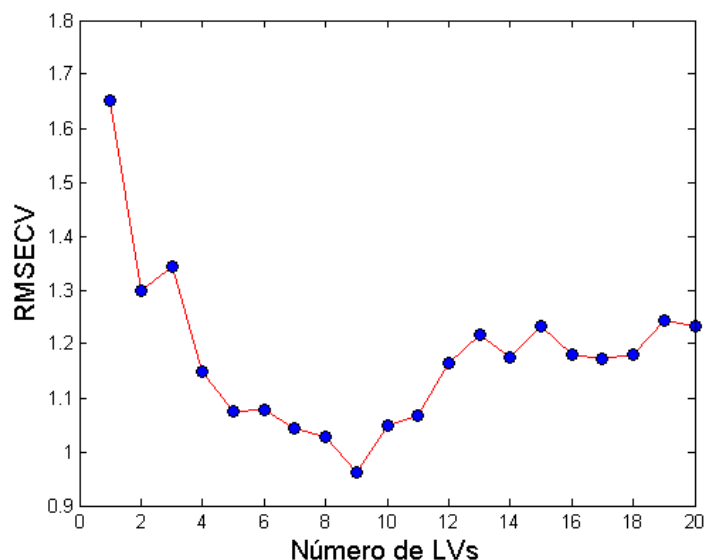
A segunda propriedade físico-química modelada foi o Resíduo de Carbono. Sua metodologia oficial é dada pela ASTM D4530. A determinação do resíduo de carbono envolve a pirólise do produto e a vaporização dos formados pela reação, sob atmosfera inerte. A quantidade de material carbonáceo residual é medida em termos do percentual em massa do resíduo em relação à amostra. Esse é um parâmetro de grande importância, pois correlaciona o resíduo de carbono obtido com a possibilidade de depósitos no injetor e câmara de combustão de motores (Nascimento et al. 2014).

As 77 amostras disponíveis para essa propriedade, apresentam a faixa de percentual em massa variando de 0,07 a 9,8 %. O conjunto espectral para esse conjunto é muito semelhante ao da Figura 11 (a), pois as amostras foram submetidas ao mesmo tipo de alinhamento e pré-processamento para correção da linha de base (SNV), logo não será mostrado novamente. Assim, de posse dos espectros e dos valores de percentual em massa para o RC, para cada amostra, é possível construir o modelo de calibração multivariada.

O universo amostral foi dividido nos conjuntos de calibração (70%) e validação (30%), pelo algoritmo já mencionado, Kennard-Stone. Utilizando o procedimento de validação-

cruzada para a construção do modelo de calibração, com base na minimização de seu erro quadrático médio, RMSECV, obteve-se o gráfico para otimização do número de variáveis latentes, conforme a Figura 18:

Figura 18: Otimização do número de variáveis latentes para construção de modelo de calibração, para a propriedade RC

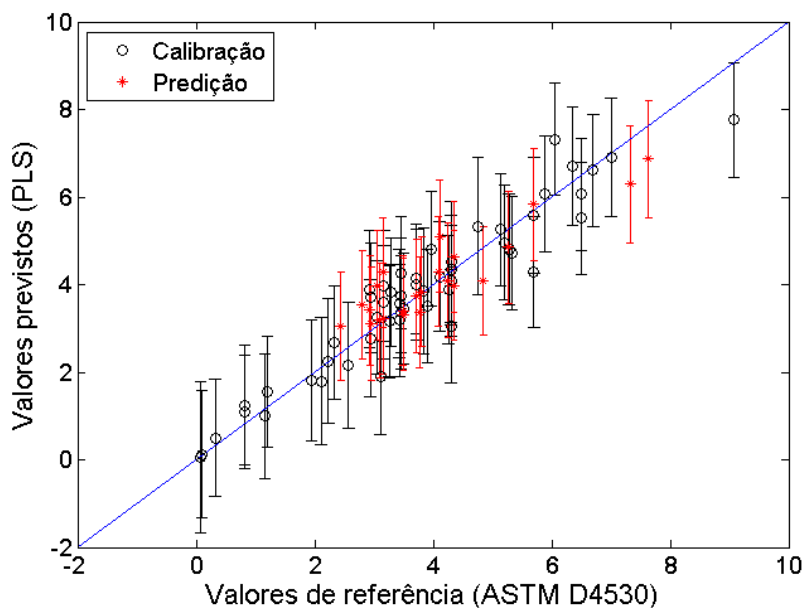


Fonte: ELABORADA PELO PRÓPRIO AUTOR

Através da Figura 18 pôde-se otimizar o número de LVs para a construção do modelo de calibração, o que foi feito redefinindo esse valor para 9. Esse novo número de LVs trata-se de um valor consideravelmente alto, pois é muito discutido na literatura que se for utilizado um número excessivo de LVs, existe a possibilidade de se modelar ruído, pois o modelo pode estar utilizando de locais do espectro em que não faça sentido químico. Esse problema será melhor avaliado ao final da modelagem quando ocorrer a análise do coeficiente de regressão com relação as principais regiões do espectro, onde será possível observar os principais locais em que o modelo atribuiu considerável peso e importância. No entanto, para esse caso, a escolha das 9 LVs é com base na minimização do RMSECV, como já havia sido feito para a modelagem da gravidade API. Na etapa da validação, as amostras que não participaram da calibração são utilizadas. O gráfico na Figura 19 mostra a correlação entre os valores medidos e previstos pela modelagem. Pelo aspecto do gráfico, para essa propriedade, é possível notar que ocorre uma diminuição da linearidade e das correlações quando comparamos com o gráfico de medidos versus previstos para a propriedade API (Figura 14). Para essa e outras propriedades, a modelagem utilizando ferramentas lineares, torna-se um pouco mais complexa, pois os dados

de petróleo se apresentam com características de desvios de linearidades, o que torna interessante a utilização de modelos não lineares (como é possível ver o aumento de suas aplicações na literatura). Entretanto, o PLS também fornece resultados consideráveis como é mostrado nos resultados obtidos nesse trabalho.

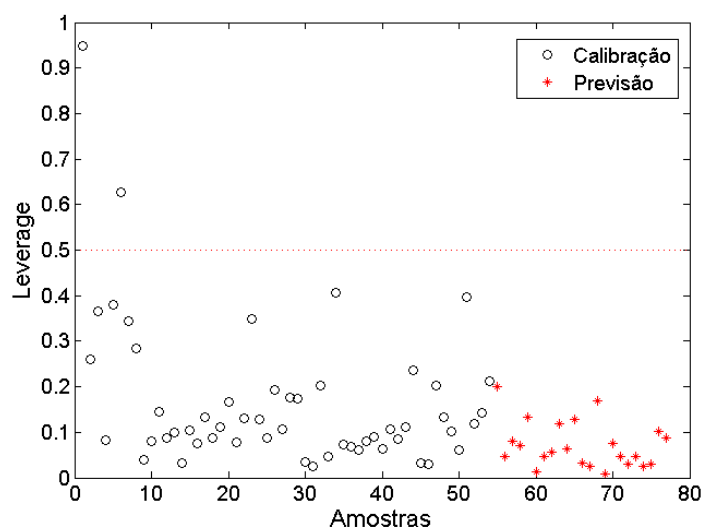
Figura 19: Correlação entre os valores de referência e os previstos pelo modelo PLS, para a propriedade RC



Fonte: ELABORADA PELO PRÓPRIO AUTOR

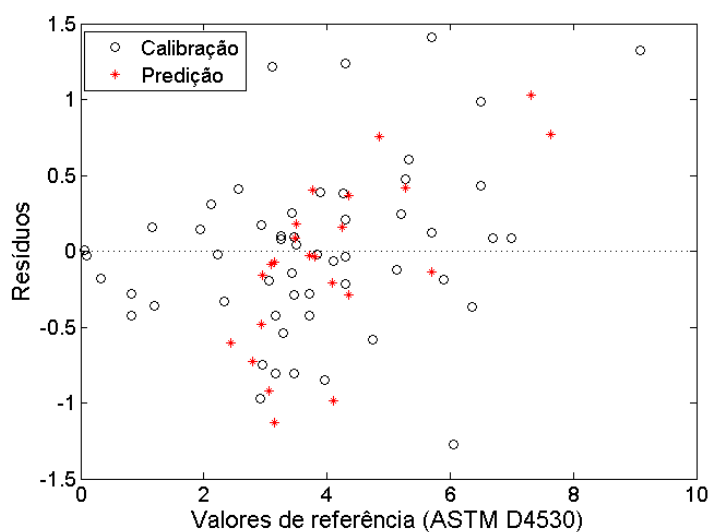
Todas as amostras utilizadas para essa modelagem encontram seus valores dentro do intervalo de confiança calculado, mesmo que no limite para algumas, como é o caso da amostra de valor 9,8 para o conjunto de calibração, apresentada na Figura 19. Sendo assim, como não existem amostras fora dos limites, não se faz necessária a busca por *outliers*, até porque somente duas amostras tem *leverages* acima do valor limite, sendo que essas apresentam baixos valores de resíduo; as amostras com *leverage* 0,9495 e 0,6268 apresentam resíduos de 0,0102 ($V_{med} = 0,0700$ e $V_{prev} = 0,0598$) e 0,5813 ($V_{med} = 4,7500$ e $V_{prev} = 5,3313$) respectivamente, sendo o *leverage* limite igual a 0,5. As figuras 20 e 21 mostram os gráficos de *leverage* e dos resíduos.

Figura 20: Gráfico de leverages, que demarcam a influência das amostras na construção do modelo, para a propriedade RC



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 21: Gráfico dos resíduos para a propriedade RC

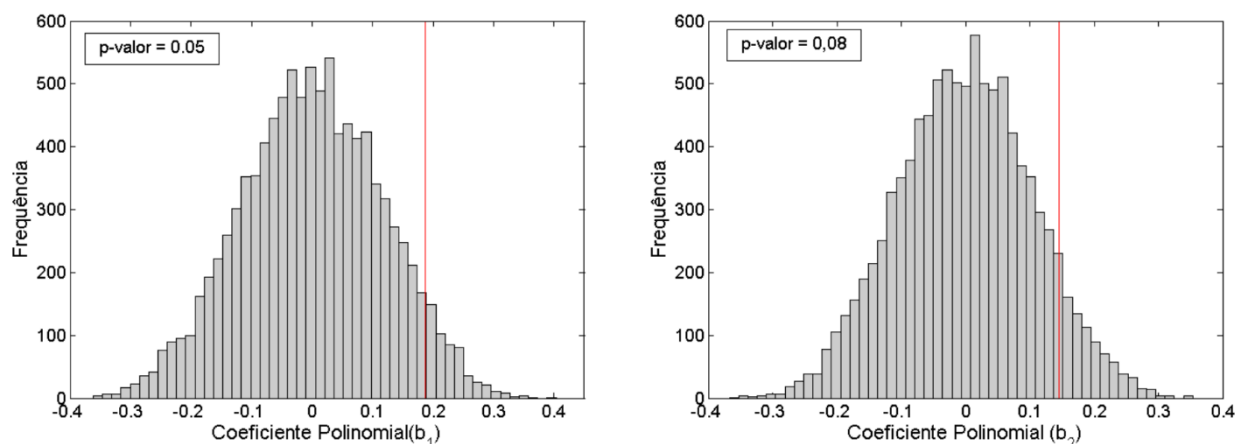


Fonte: ELABORADA PELO PRÓPRIO AUTOR

A exatidão do modelo foi avaliada por meio do RMSEP que apresentou um valor de 0,56. A correlação para os valores de previsão foi melhor que para os de calibração, sendo 0,736 e 0,843 os valores para esses parâmetros do modelo, respectivamente. A avaliação dos resíduos foi realizada pelos testes de bias e permutação. Empregando o teste t para avaliação dos erros sistemáticos, observou-se que não há evidência desse tipo de erro para o nível de confiança de 95% ($t_{cal} = 0,63 < t_{tab(\alpha=0,05,n-1)} = 2,07$). Para a avaliação de tendências, foi tentado descrever os resíduos por meio de funções lineares e quadráticas. Realizando uma análise visual da Figura

21, os resíduos da previsão aparentam adotar uma postura linear, mas pelo teste de permutação utilizando uma função do primeiro grau, foi possível perceber que de acordo com as 10.000 permutações estabelecidas, o número de vezes em que se conseguiu descrever os resíduos não foi tão significativa ao ponto de superar a margem de 5%, o que pode ser basicamente explicado pelo p-valor ter sido igual o superior a 0,05. Utilizando uma função quadrática também não foi possível modelar os resíduos, o que pode ser confirmado pelos gráficos abaixo:

Figura 22: Histograma de permutação utilizando uma função linear (a) e uma quadrática (b) para descrever os resíduos do modelo construído, para a propriedade RC



Fonte: ELABORADA PELO PRÓPRIO AUTOR

A Tabela 5 traz os principais parâmetros do modelo construído, evidenciando principalmente que os modelos não apresentam erros sistemáticos ou de tendência. O Apêndice A traz a comparação direta entre os valores medidos e previstos pelo modelo, bem como os intervalos de confiança calculados para cada amostra.

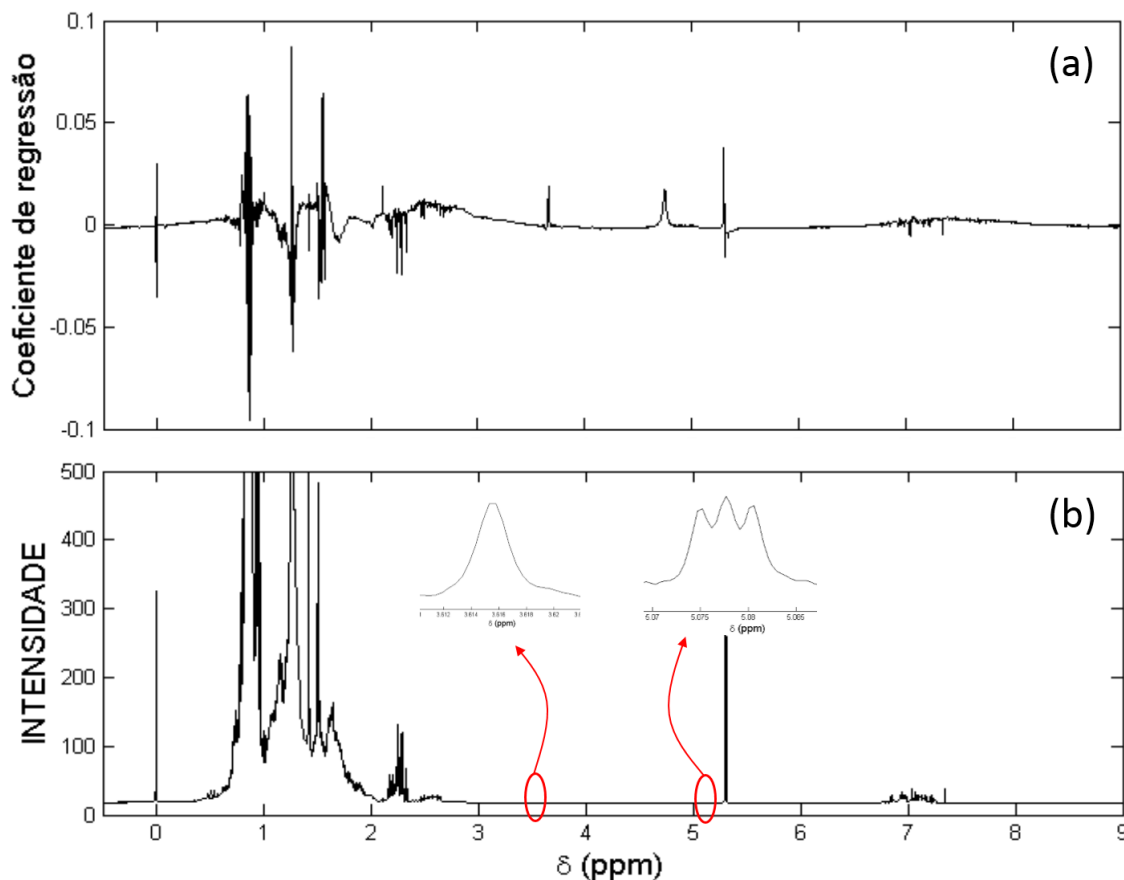
Tabela 5: Tabela contendo os principais parâmetros de modelagem para o RC

RC	
R^2_{vc}	0,736
R^2_p	0,843
RMSECV	0,962
RMSEP	0,559
E_{sist}	$t_{cal} = 0,63 < t_{tab} = 2,07$
$E_{aleat}(b1)$	p-valor = 0,05 > 0,05
$E_{aleat}(b2)$	p-valor = 0,08 > 0,05

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Como levantado a discussão anterior, definir corretamente o número de variáveis latentes para descrever as variáveis originais é de suma importância e tem grande implicação no resultado final da modelagem, isto é, na capacidade preditiva do modelo. Os trabalhos encontrados na literatura para aplicação de PLS em amostras de petróleo bruto, apresentam para suas modelagens, um número considerável de LVs, como é visto nos trabalhos de Abbas e Masili (Abbas et al. 2012; Masili et al. 2012). A Figura 23 mostra os gráficos do coeficiente de regressão (a) e do espectro médio de ^1H -RMN para o conjunto amostral. Analisando essa figura, é possível perceber que o modelo construído com 9 LVs atribui a maior importância à região dos principais deslocamentos químicos referentes aos hidrocarbonetos, isto é, onde temos as principais contribuições para núcleos de hidrogênio pertencentes aos grupamentos metílicos, metilênicos e metínicos. Entretanto, como o número de LVs é consideravelmente elevado, é dado peso também a regiões em que a informação não é muito interessante, como por exemplo nos sinais referentes ao TMS (0 ppm), às impurezas contidas no solvente (5,32 ppm) e a dois outros sinais que pareciam imperceptíveis, mas que com uma aproximação foi possível identificá-los. O primeiro sinal amplificado se apresentando como singlete, a que o modelo dedicou importância em sua construção, refere-se ao deslocamento químico em 3,616 ppm e o segundo, se apresentando em aspecto de um tripleto, em 5,078 ppm. Atribuí-los de maneira inequívoca é impossível com os recursos utilizados nesse trabalho, pois o petróleo apresenta uma enorme gama de compostos contendo os mais variados grupos funcionais, entretanto pode-se afirmar que estes não são os principais sinais a que o modelo deveria atribuir grande importância, sendo esse um problema referente ao considerável número de LVs utilizadas. Uma saída para problemas desse tipo é a aplicação de algoritmos de seleção de variáveis antes da realização da regressão multivariada. Algoritmos Genéticos e Redes Neurais vêm se apresentando como interessantes possibilidades para essa finalidade e muitos trabalhos são encontrados na literatura.

Figura 23: Coeficiente de regressão para o modelo de ^1H -RMN-PLS com 9LVs (a) e o espectro médio para o conjunto amostral (b)



Fonte: ELABORADA PELO PRÓPRIO AUTOR

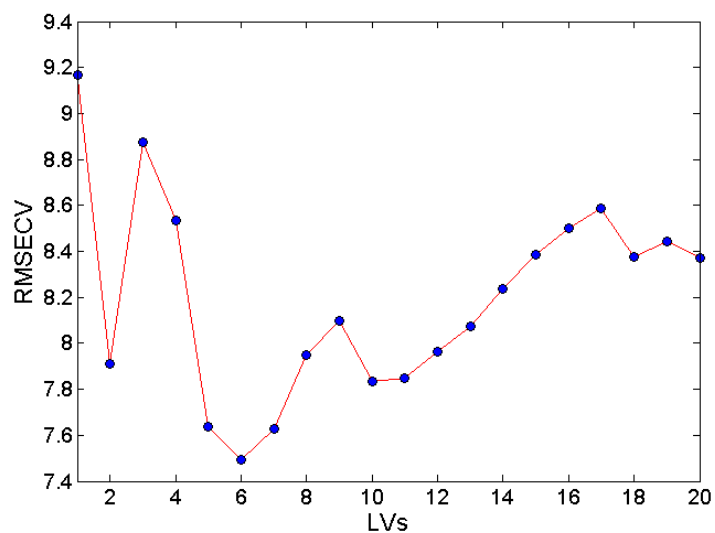
2.4.4 TEMPERATURA INICIAL DE APARECIMENTO DE CRISTAIS ($^{\circ}\text{C}$)

A TIAC, Temperatura de Aparecimento de Cristais, tem na Calorimetria Exploratória Diferencial (sigla do inglês, DSC) sua metodologia de referência. Essa também é uma propriedade importantíssima para a indústria petrolífera pois a partir dela é possível prever em qual temperatura as parafinas do petróleo começam a cristalizar, o que pode ocasionar obstrução nas paredes dos dutos em instalações de produção e em tubulações de transporte durante a produção de óleo e gás (Gomes 2009).

Para essa propriedade o universo amostral foi de 70 amostras, que da mesma forma como nas modelagens anteriores, tiveram seus espectros de ^1H -RMN agrupados para a realização do alinhamento pela ferramenta *icoshift*, seguido do pré-tratamento para correção de linha base (SNV) e então separadas em dois conjuntos pelo algoritmo Kennard-Stone, em 70% para calibração e os 30% restantes para validar o modelo. A validação interna consistiu na

otimização das LVs com base na minimização do RMSECV, o que pode ser visto na figura abaixo:

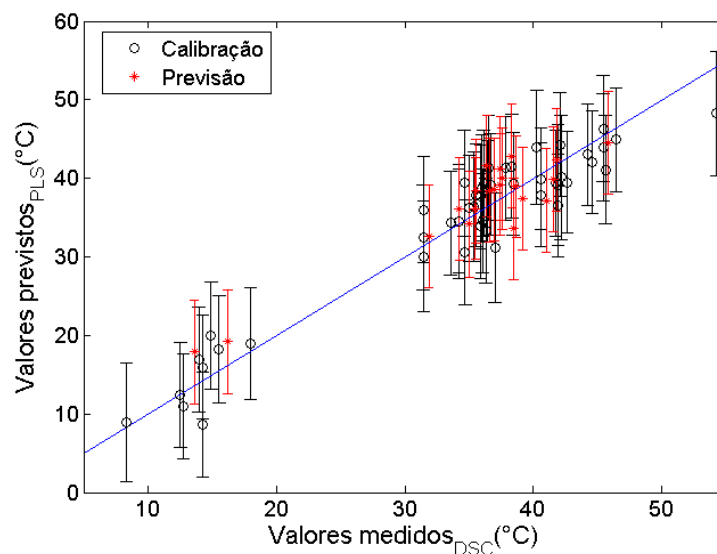
Figura 24: Otimização do número de variáveis latentes para construção de modelo de calibração, para a propriedade TIAC



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Com a otimização das LVs para 6, construiu-se o modelo e a correlação entre os valores de temperatura medidos pelo DSC e estimados pelo modelo de $^1\text{H-RMN-PLS}$ podem ser vistos na Figura 25:

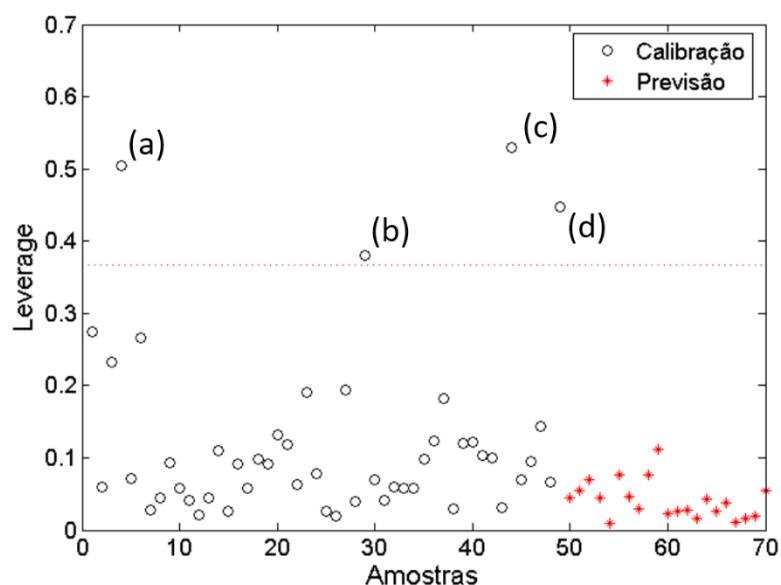
Figura 25: Correlação entre os valores de referência e os previstos pelo modelo PLS, para a propriedade TIAC



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Para essa modelagem também não há amostras como possíveis *outliers*, visto que todas são contempladas pelo modelo em seu intervalo de confiança e, além disso, as que apresentam valor de *leverage* acima do limite estão com baixo valor resíduo e assim não contribuirão para o deterioramento do modelo, o que pode ser visto na Figura 26 e Tabela 6.

Figura 26: Gráfico de leverages, que demarcam a influência das amostras na construção do modelo, para a propriedade TIAC



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Tabela 6: Avaliação das amostras com alto valor de leverage, para modelagem da propriedade TIAC

	Leverage	y_{med}	y_{prev}	 e_i 	IC	
A	0.5	42.15	40.13	2.02	32.24	48.01
B	0.38	8.32	8.95	0.63	1.40	16.50
C	0.53	54.28	48.27	6.01	40.32	56.22
D	0.45	41.94	39.18	2.76	31.45	46.92

leverage_{limite} = 0.37; |e_i| - módulo do resíduo

y_{med} - valor medido; y_{prev} - valor previsto

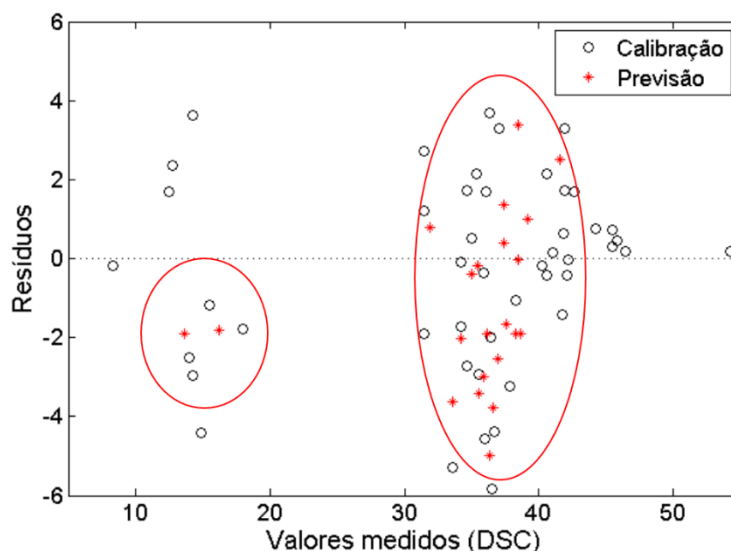
IC - intervalo de confiança, vai do primeiro valor ao segundo

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Entretanto, analisando a Figura 25 e o gráfico dos resíduos na Figura 27 pode-se notar claramente que existem dois conjuntos para o dados de previsão, isto é, duas amostras com TIAC abaixo de 20°C e um conjunto maior de amostras com TIAC entre 30 e 50°C. Matematicamente falando, pode-se olhar esses dois conjuntos como duas densidades de pontos

e pensar que para uni-los apenas uma reta é suficiente, logo, tentar descrever esses resíduos por funções lineares e/ou quadráticas não faz sentido, então o teste de permutação não foi realizado para essa propriedade. Para erros sistemáticos o teste de bias mostrou que não há evidências deste com 95% de confiança ($t_{cal} = 1,70 < t_{tab}(\alpha=0,05,n-1) = 2,08$).

Figura 27: Gráfico dos resíduos para a propriedade TIAC



Fonte: ELABORADA PELO PRÓPRIO AUTOR

No entanto, quando é feita a comparação direta entre os valores medidos e preditos, conforme mostra no Apêndice A, percebe-se grande proximidade entre eles, sendo isso o mais importante para a indústria petrolífera e de grande importância para a tomada de decisão.

Pelas duas últimas figuras é possível perceber a lacuna, tanto para amostras de calibração quanto para validação, de amostras com valores para essa propriedade entre 20 e 30°C. Uma saída para conseguir um modelo em que suas amostras e consequentemente seus resíduos se distribuam de maneira aleatória, é a obtenção de um novo conjunto de amostras que apresentem valores para essa propriedade ao longo de toda extensão da curva de calibração, sendo intercalada por amostras de validação. Porém, quando é feita a comparação direta entre os valores medidos e preditos, conforme mostra no Apêndice A, percebe-se significativa proximidade entre eles, sendo isso o mais importante para a indústria petrolífera e de grande importância para a tomada de decisão. A tabela 7 traz um resumo para os principais parâmetros do modelo desenvolvido para TIAC.

Tabela 7: Tabela resumo contendo os principais parâmetros de modelagem para a TIAC

TIAC	
R^2_{vc}	0,534
R^2_p	0,875
RMSECV	7,495
RMSEP	2,797
E_{sist}	$t_{cal} = 1,70 < t_{tab} = 2.08$

Fonte: ELABORADA PELO PRÓPRIO AUTOR

2.4.5 DESTILAÇÃO SIMULADA

Na indústria do petróleo, a destilação é o processo de separação mais importante utilizado. Informações experimentais relacionadas com a distribuição dos pontos de ebulição dos componentes de óleos crus e de espécies químicas oriundas do processo de refino do petróleo, são consideradas essenciais para o monitoramento, controle e garantia da qualidade do produto. A destilação simulada por cromatografia gasosa (SimDis) surgiu no início dos anos 60 para ser utilizada na avaliação da curva de volatilidade de petróleos e de suas frações destiladas (Ferreira & Neto 2005). A utilização de pouca quantidade de amostra e a rapidez na obtenção dos resultados são fatores que favoreceram a ampla difusão desta técnica como uma alternativa aos métodos convencionais de determinação da curva de destilação dos pontos de ebulição verdadeiros (PEV) - ASTM D2892 e D5236.

A SimDis é uma das mais importantes análises realizadas para a caracterização do óleo bruto, pois é a partir dela que se conhece os teores dos cortes que esse petróleo irá produzir, sendo essa informação de suma importância para a valoração do mesmo. Nesse caso, não se trata somente da construção de modelos para determinar uma boa correlação entre valores medidos e previstos e a obtenção de um baixo valor de RMSEP, mas o crucial, é a comparação entre as curvas geradas pela SimDis e através da modelagem, sendo essa última denominada nesse trabalho como PEEQ (curva de Pontos de Ebulição Estimados por Quimiometria).

Para esse estudo dispúnhamos de 74 amostras e para cada uma delas foi obtida, por meio da cromatografia gasosa (método de referência), as temperaturas de ebulição para os cortes de 5 em 5% em massa; com esses valores de temperatura nos correlacionados percentuais de corte das frações é possível plotar a curva SimDis. Na Tabela 8, são mostradas as temperaturas de ebulição, referentes aos percentuais em massa de cada corte, para duas amostras que terão suas curvas comparadas, isto é, SimDis (referência) versus PEEQ (prevista pela modelagem). A

amostra denominada PET1 refere-se a um petróleo intermediário (API 29,5), enquanto a PET2 a um petróleo pesado (API 17).

Tabela 8: Variação da temperatura referente ao ponto de ebulição (PE) com relação ao percentual em massa do corte, para as amostras PET1 e PET2

	PET1	PET2
% m/m	PE (°C)	PE (°C)
PI	-20.50	59.90
5.00	90.60	196.90
10.00	132.80	253.80
15.00	172.90	295.60
20.00	213.20	329.60
25.00	247.10	361.90
30.00	280.00	391.60
35.00	311.10	417.30
40.00	343.30	437.80
45.00	376.20	461.30
50.00	407.70	487.10
55.00	433.10	511.70
60.00	459.70	536.20
65.00	491.30	562.70
70.00	523.30	589.20
75.00	577.50	591.80
80.00	596.30	645.20
85.00	637.30	676.40
90.00	685.80	708.30
95.00	744.2	744.3
PF	750	750

PET1 - API 29.5

PET2 - API 17

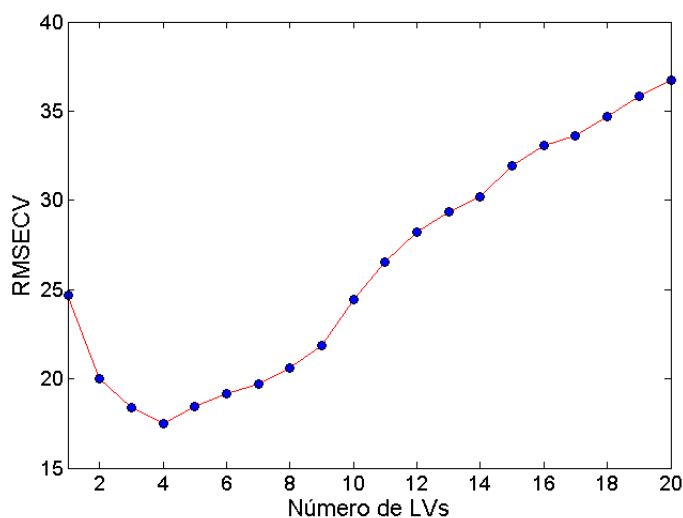
Fonte: ELABORADA PELO PRÓPRIO AUTOR

As modelagens, via ¹H-RMN-PLS, foram realizadas do ponto inicial (PI) até 90% em massa referente ao corte, isso devido aos cortes de 95% e para o ponto final (PF) não apresentarem variabilidade entre as amostras, o que torna a modelagem inviável; por esse fato a Tabela 8 encontra-se hachurada para 95% e para o PF, enfatizando que esses cortes não foram modelados. Foram então construídos 19 modelos, um para cada corte, do PI à 90% em massa.

Para explanação da estratégia de ação na construção desses modelos, será mostrado como foi realizado para o corte referente a 5% em massa; as demais informações relacionadas aos outros cortes encontram-se nas tabelas ao final dessa seção. A sistemática seguiu o mesmo protocolo utilizado para a modelagem das propriedades mostradas nos itens anteriores. Após o

agrupamento dos espectros e valores de referência para formar a matriz X e o vetor Y , respectivamente, o alinhamento e a correção de linha base foram feitas nos espectros. Em seguida as amostras foram separadas nos conjuntos de calibração e validação, com a proporção de 70 e 30%, respectivamente. A otimização das LVs foi feita com base na minimização do RMSECV, pelo método iterativo *leave one out* e, através da figura abaixo pôde-se definir o número de LVs para a construção do modelo referente a 5% em massa:

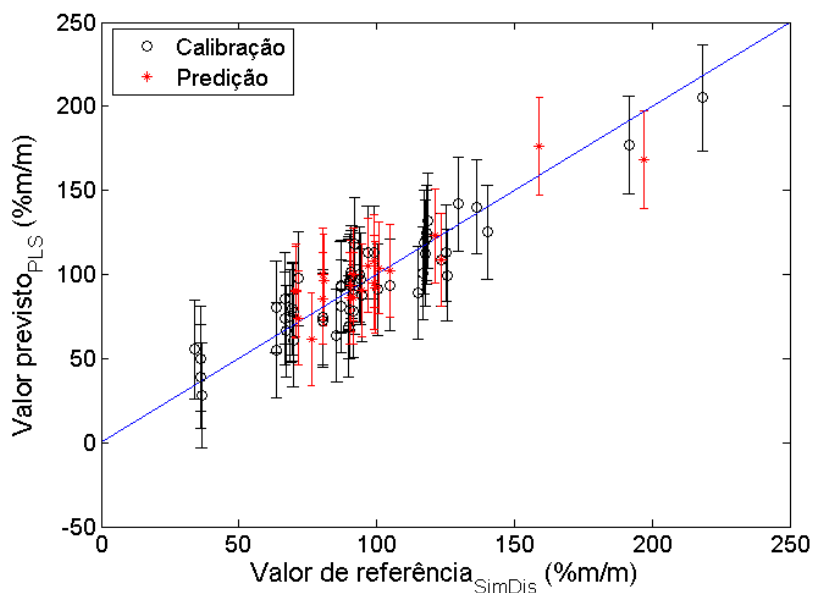
Figura 28: Otimização do número de variáveis latentes para construção de modelo de calibração, para o corte referente a 5% em massa



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Otimizando as LVs para 4, passou-se para a etapa de validação. Utilizando o conjunto que não participou da calibração foi realizada a validação externa e o gráfico de correlação entre os valores medidos e previstos é dado pela Figura 29. Os valores de correlação para a validação cruzada e para a predição foram de 0,73 e 0,82, respectivamente. Todas as temperaturas foram estimadas pelo modelo e apresentaram seus valores dentro dos intervalos de confiança calculados e sendo assim a busca por *outliers* torna-se de menor importância. O RMSECV e o RMSEP apresentaram valores 17,50 e 12,35, entretanto para que fosse monitorado a evolução dos erros, principalmente o de predição, também foi calculado o RMSEP %, que para este caso foi 12,14.

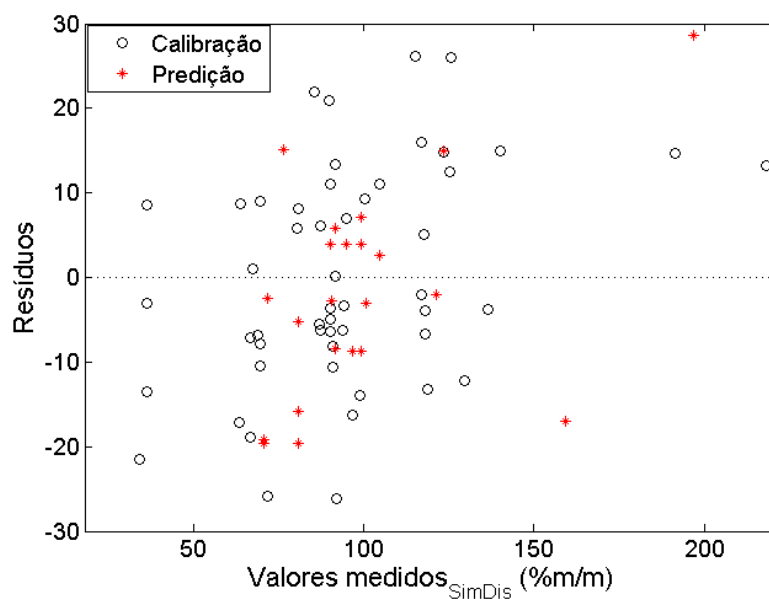
Figura 29: Correlação entre os valores de referência e os previstos pelo modelo PLS, para o corte referente a 5% em massa



Fonte: ELABORADA PELO PRÓPRIO AUTOR

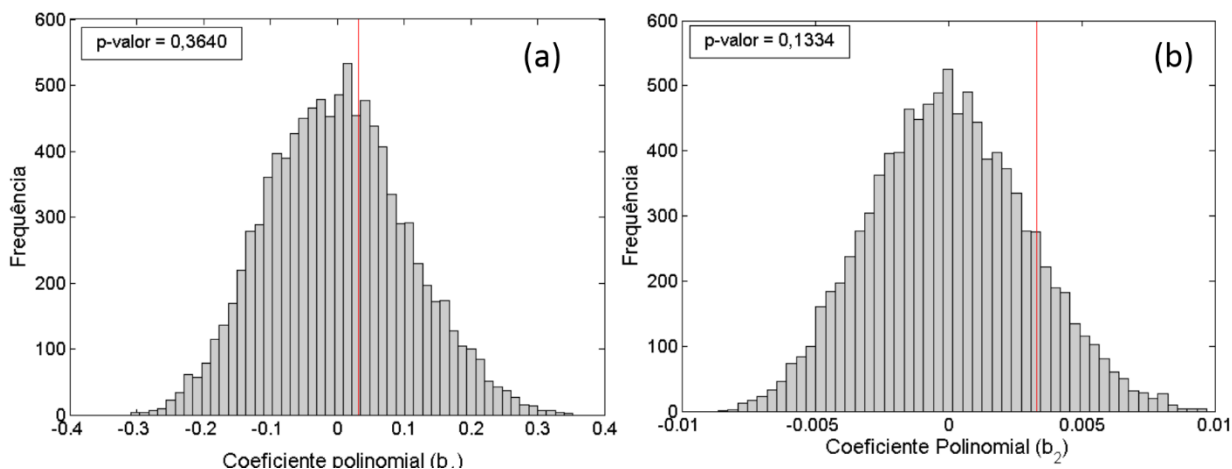
Os resíduos deixados pelo modelo foram avaliados e para o teste de bias, não há evidências de erros sistemáticos no nível de confiança de 95% ($t_{cal} = 0,80 < t_{tab(\alpha=0,05,n-1)} = 2,08$). O teste de permutação (Figura 31) para avaliação de tendências nos resíduos mostrou que os resíduos (Figura 30) se comportam aleatoriamente, como pode ser vistos nas figuras a seguir:

Figura 30: Gráfico dos resíduos para o corte referente a 5% em massa da curva SimDis para amostras de petróleo



Fonte: ELABORADA PELO PRÓPRIO AUTOR

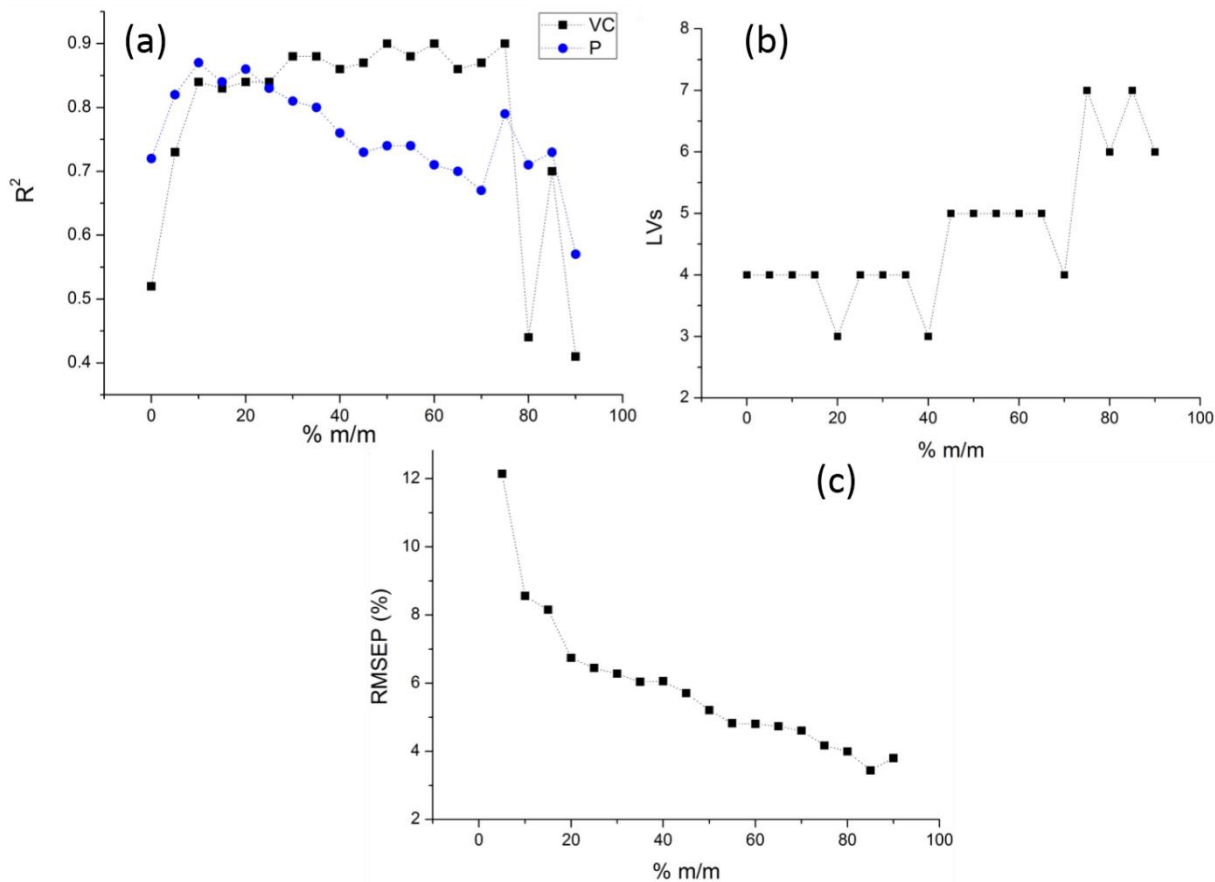
Figura 31: Histograma de permutação utilizando uma função linear (a) e uma quadrática (b) para descrever os resíduos do modelo construído, para o corte referente a 5% em massa



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Com o objetivo de tornar mais prática a análise dos outros principais parâmetros dos modelos construídos, foram plotados gráficos para o acompanhamento dos seguintes parâmetros: R^2_{vc} , R^2_p , LVs e RMSEP%. Na Figura 32(a) pode-se perceber que o R^2_{vc} apresenta valores baixos para as modelagens referentes aos dois primeiros cortes, daí em diante fica com um aspecto mais constante (acima de 0,8) até o corte de 75% em massa e partir desse ponto passa a apresentar baixíssima correlação, o que é acompanhado de um aumento significativo no número de LVs utilizados para construir os modelos (Figura 32(b)). A curva para o R^2_p se apresenta de maneira semelhante à R^2_{vc} , no entanto as subidas e descidas dos valores são menos abruptas. Para ambas as correlações ocorrem uma diminuição à medida que ocorre a modelagem dos últimos cortes da destilação, visto que os valores estão cada vez mais próximos, o que diminui a variabilidade, chegando à impossibilidade de modelagem, como é o caso do corte referente à 95% em massa e do PF. No entanto, mesmo aumentando o número de LVs e diminuindo o R^2_p ao longo dos cortes, o RMSEP% diminui consideravelmente, o que é interessante para comparação da SimDis e da PEEQ.

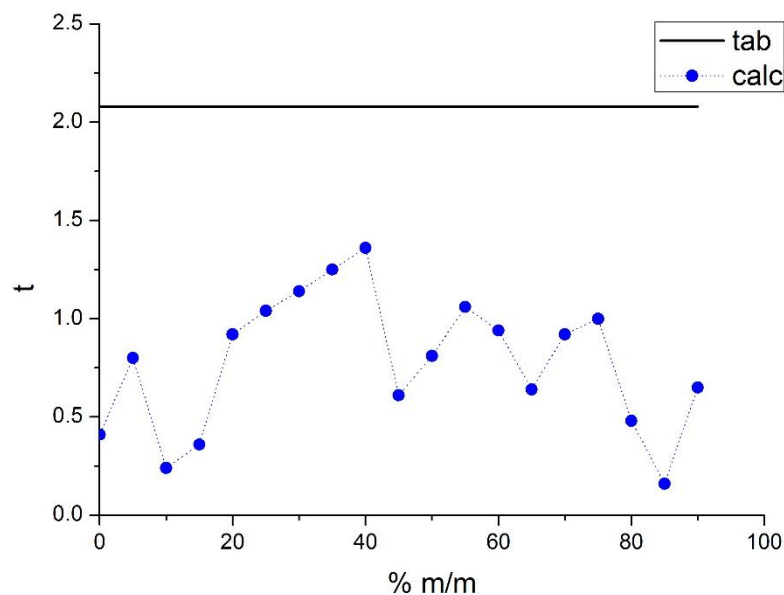
Figura 32: Monitoramento de parâmetros de modelagem ao longo dos cortes para: (a) correlação para validação cruzada e predição; (b) número de LVs utilizadas; (c) RMSEP%



Fonte: ELABORADA PELO PRÓPRIO AUTOR

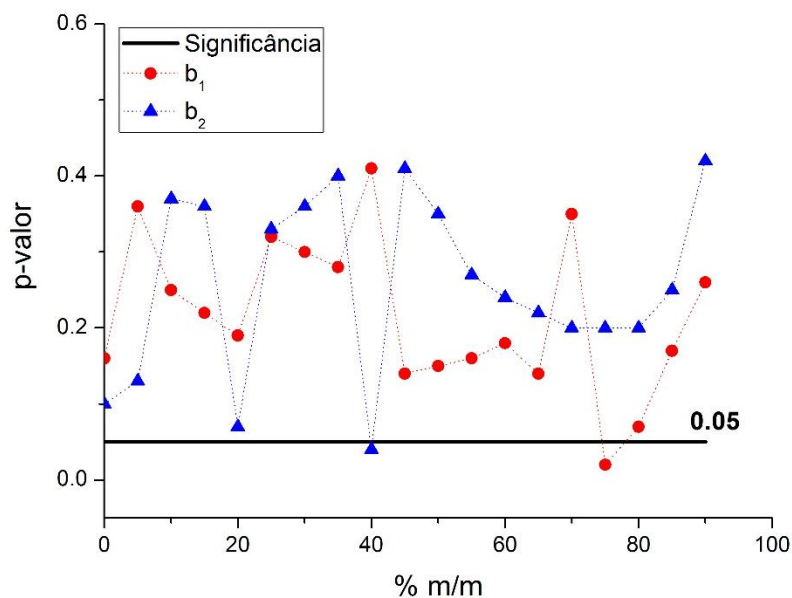
Outro monitoramento realizado ao longo das modelagens levou em consideração os principais testes estatísticos para avaliação dos resíduos. Se tratando do teste de bias, nenhuma modelagem apresentou erros sistemáticos, como pode ser visualizado na Figura 33, visto que todos os valores de t_{calc} , ao longo dos cortes, ficaram abaixo do limite dado pelo valor de *student* tabelado para 95% de confiança. Aplicando o teste de permutação aos resíduos de cada modelo, nota-se que para os cortes referentes a 40 e 75% em massa, é possível modelá-los pela utilização de funções. A Figura 34 mostra que os resíduos deixados pelo modelo construído, para o corte de 40%, apresentam perfil quadrático ao nível de 95% de confiança, uma vez que seu p-valor para a utilização da função de segundo grau é inferior a significância de 0,05, marcado pelo limite inferior apresentado no gráfico. O mesmo ocorre para o corte de 75% em massa, sendo esses resíduos modeláveis por uma função linear.

Figura 33: Monitoramento dos resíduos deixados pelos modelos construídos ao longo dos cortes, por meio de testes para erros sistemáticos (teste de bias)



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 34: Monitoramento dos resíduos deixados pelos modelos construídos ao longo dos cortes, por meio de testes para erros de tendência (teste de permutação), aplicando funções lineares e quadráticas

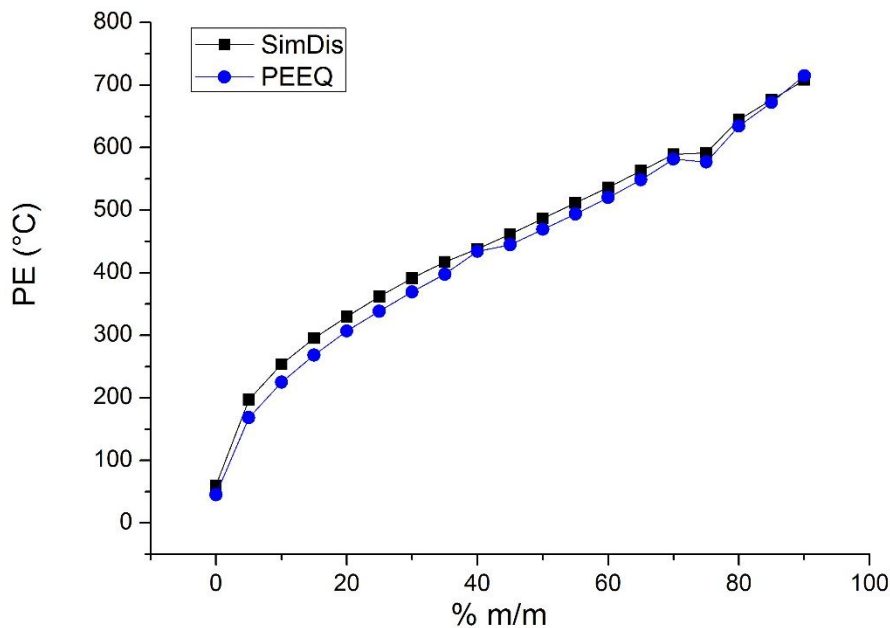


Fonte: ELABORADA PELO PRÓPRIO AUTOR

Mesmo dois pontos da curva de PE se apresentando como falhos nos testes para erros de tendência, os valores estimados pelos modelos de ¹H-RMN-PLS foram próximos aos de referência, pois para a maioria dos cortes os erros relativos estiveram abaixo de 5% (Tabela 9), o que possibilita uma comparação efetiva entre a PEEQ e a SimDis. Essas curvas foram plotadas

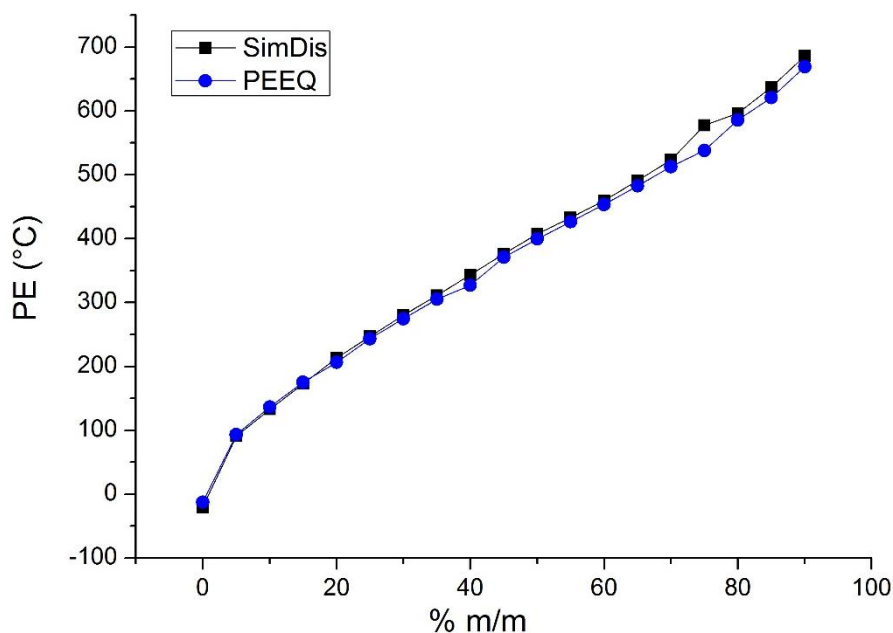
e se mostraram semelhantes entre si (analisar em concomitância com a Tabela 9), principalmente a partir do segundo ponto, o que torna a metodologia desenvolvida nesse trabalho como potencial a ser utilizada para estimar as temperaturas de ebulição dos cortes de petróleo, a partir de uma análise por ¹H-RMN.

Figura 35: Comparação entre a SimDis, obtida por cromatografia gasosa de acordo com a ASTM e a PEEQ, estimada pela metodologia ¹H-RMN-PLS, para uma amostra com API 17



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 36: Comparação entre a SimDis, obtida por cromatografia gasosa de acordo com a ASTM e a PEEQ, estimada pela metodologia ¹H-RMN-PLS, para uma amostra com API 29,5



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Tabela 9: Erros relativos para os cortes do PI até 90% em massa de destilado, para as amostras de API 17 e 29,5

% m/m	API 29,5 API 17,0	
	Er (%)	Er (%)
PI	39.62	24.59
5.00	3.08	14.53
10.00	2.92	11.31
15.00	1.38	9.20
20.00	3.21	6.88
25.00	1.53	6.45
30.00	1.89	5.66
35.00	1.89	4.70
40.00	4.72	0.83
45.00	1.44	3.56
50.00	1.97	3.64
55.00	1.55	3.47
60.00	1.32	2.97
65.00	1.76	2.52
70.00	2.03	1.27
75.00	6.83	2.51
80.00	1.78	1.68
85.00	2.58	0.62
90.00	2.45	0.92
95.00	-	-
PF	-	-

Fonte: ELABORADA PELO PRÓPRIO AUTOR

2.5 CONCLUSÕES

Os modelos construídos para gravidade API, TIAC e RC foram validados e apresentaram grande proximidade com os valores obtidos pelos respectivos métodos de referência. A realização de testes para avaliação dos resíduos, dá aos modelos maior confiabilidade e robustez, tornando-os aptos a serem aplicados a um novo conjunto de amostras com o objetivo de previsão das propriedades de interesse.

Mediante a metodologia desenvolvida, foi possível criar modelos para cada corte da SimDis e após a validação dos mesmos, plotar uma curva de Pontos de Ebulição Estimados por Quimiometria (PEEQ). A SimDis, realizada por CG, reduz o tempo e custo do processo de destilação convencional para construção da PEV. A PEEQ simulou com grande proximidade a curva gerada por SimDis, que agora pode ser estimada pela simples aquisição de um espectro

de ^1H -RMN, cujo tempo de análise ocorre por volta de 3 minutos devido ao baixo tempo de relaxação para o núcleo de hidrogênio.

Nessa parte do trabalho foram construídos modelos multivariados com o objetivo de previsão de importantes propriedades físico-químicas, que fazem parte do portfólio de análises para caracterização de um óleo bruto. A metodologia proposta, evidencia que a partir do espectro de ^1H -RMN do petróleo bruto, é possível estimar diferentes propriedades físico-químicas, quando aplicada modelagem quimiométrica. Espera-se que essas metodologias possam causar impacto na indústria petrolífera, visto que estas são rápidas e necessitam de pouquíssima quantidade de amostra, além de apresentarem grande viabilidade econômica e serem quimicamente menos poluentes, uma vez que não utilizam solventes orgânicos.

2.6 PERSPECTIVAS FUTURAS

- Pretende-se aplicar as metodologias desenvolvidas nesse trabalho a um novo conjunto de amostras;
- Teste e aplicação de algoritmos de seleção de variáveis, aos dados de ^1H -RMN, como um passo anterior a realização da modelagem propriamente dita, como por exemplo algoritmos genéticos (GA);
- Aplicação de modelos não-lineares, visto que os dados de petróleo se apresentam desvios da linearidade e na literatura muitos trabalhos vêm mostrando que esses modelos fornecem melhores resultados;
- Publicação dos resultados obtidos nesse trabalho em revistas da área

3. EDULCORANTES

3.1 INTRODUÇÃO

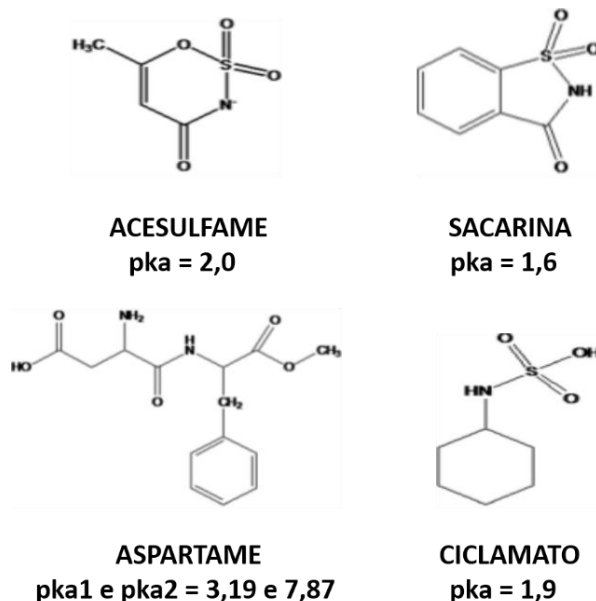
No início das décadas de 60 e 70, o consumo de alimentos industrializados no país apresentava apenas uma ligeira tendência de aumento, o que nas décadas seguintes foi intensificado de maneira abrupta (Barros 2008; Contreras 2011). Os alimentos industrializados surgiram proporcionando redução de custos e aumento do tempo de conservação. A indústria passou a utilizar diversos compostos capazes de conferir aos alimentos algumas características necessárias, sendo estes: os antioxidantes, aromatizantes, adoçantes entre outros. Esses compostos são conhecidos como aditivos e podem ser de origem natural ou sintética. A utilização e identificação de cada aditivo são regidas por normas e códigos específicos (Fatibello-filho et al. 1996).

Segundo a *Food and Agriculture Organization* (do inglês, FAO), por meio da Codex Alimentarius, aditivos são “substâncias não consumidas normalmente como alimentos propriamente ditos, podendo ter ou não valor nutricional, cujo propósito da adição a diferentes gêneros alimentícios é tecnológico (incluindo organoléptico), no que diz respeito ao fabrico, transformação, preparação, tratamento, embalagem, acondicionamento...” (CODEX 2013). No Brasil, a regulamentação, controle e fiscalização de aditivos alimentares e coadjuvantes de tecnologia de fabricação são feitos pela Agência Nacional de Vigilância Sanitária (Anvisa), conforme disposto na Lei nº 9782, de 26 de Janeiro de 1999. A principal discussão sobre o emprego de aditivos na produção de alimentos resulta da controvérsia entre a necessidade e segurança de seu uso (ANVISA 2014).

Dentre os vários tipos de aditivos existentes, os edulcorantes destacam-se por conferirem sabor doce aos produtos alimentícios. Compostos dessa classe tem um alto poder adoçante, sendo que esse pode ser até 450 vezes superior ao da sacarose e, em virtude disso, muitas vezes estes são chamados de “adoçantes de alta intensidade” (Cantarelli et al. 2008). Pelo fato de serem poderosos adoçantes e, além disso, serem classificados como não-nutritivos, os edulcorantes tem se apresentado como uma moderna alternativa não-calórica à sacarose nos mais diferentes gêneros alimentícios (Armenta et al. 2004b).

Existe uma ampla gama de edulcorantes conhecidos, sejam de origem natural ou sintética. Esse trabalho está focado nos adoçantes artificiais, devido a sua maior empregabilidade de modo conjunto, isto é, na forma de *blends*. Dentre os muitos existentes, o maior foco das pesquisas atuais está em quatro principais, sendo eles demonstrados no esquema da Figura 37.

Figura 37: Fórmula estrutural plana e valores de pka para os quatro edulcorantes analisados nesse trabalho



Fonte: ADAPTADA DE (Fernandes et al., 2013)

Hoje em dia, a tendência comum da indústria de alimentos é usar combinações de edulcorantes, também chamados de *blends*, pois é comprovado que o efeito sinérgico entre os adoçantes trazem muitos benefícios, principalmente no que diz respeito ao perfil sensorial. Além disso, essa abordagem dos *blends*, mantém a administração de cada adoçante bem abaixo da Ingestão Diária Aceitável (IDA) e dentro dos limites legais estabelecidos pela legislação (He et al. 2012; Llamas et al. 2008; Pierini et al. 2013; Fernandes et al. 2013).

Uma grande variedade de métodos, baseados em diferentes princípios, tem sido empregados para a determinação dos compostos acima mencionados em diferentes gêneros alimentícios. Zygler e colaboradores (Zygler et al. 2009), elucidam em seu artigo de revisão as principais metodologias aplicadas para uma análise simultânea desses analitos. Cromatografia Líquida de Alta Eficiência (do inglês, HPLC) é a técnica mais frequentemente utilizada para esse tipo de análise, devido à ampla disponibilidade de colunas e um expressivo portfólio de detectores. Além de outros tipos de cromatografias que vem sendo empregadas para a determinação quantitativa desses analitos, a CE também se mostra uma opção atrativa e alguns trabalhos são reportados na literatura (Zygler et al. 2009; Fernandes et al. 2013).

Uma alternativa aos métodos de separação é se trabalhar com a determinação simultânea de edulcorantes por meio da espectrometria ou espectroscopia assistida por modelagens quimiométricas. Alguns trabalhos têm sido encontrados para determinação de até dois edulcorantes (Armenta et al. 2004a; Turak & Ozgur 2013; Khurana et al. 2008; Cantarelli et al.

2009). As principais vantagens dessa metodologia são: análise de baixo custo e sustentável (visto que na maioria das vezes não é necessária etapa de pré-tratamento e também não utiliza solventes de alto grau de pureza, como por exemplo as técnicas de separação no emprego das fases móveis), rápida e não destrutiva.

3.2 OBJETIVOS ESPECÍFICOS

- Reprodução do trabalho realizado por Fernandes e colaboradores (Fernandes et al. 2013);
- Aplicação da metodologia reproduzida em amostras com diferentes matrizes, sendo que diferentes produtos *light* e *diet* apresentam em sua composição ao menos um dos edulcorantes em estudo;
- Aquisição dos espectros Raman dos padrões puros (analitos), com o objetivo de reconhecimento de bandas características nos espectros das amostras comerciais;
- Aquisição dos espectros Raman das amostras reais que tiveram os analitos quantificados via método de referência (Eletroforese Capilar);
- Construção dos modelos multivariados, a priori via PLS, para determinação ASP, CIC, SAC e ACSK;
- Validação dos modelos, avaliando as principais figuras de mérito;

3.3 PARTE EXPERIMENTAL

3.3.1 ELETROFORESE CAPILAR

Para realização das análises foi utilizado um equipamento comercial de Eletroforese Capilar da marca Agilent, equipado com fonte de alta tensão ($\pm 30\text{kV}$), detector de arranjo de diodos (sigla do inglês, DAD) e contendo o software interno, ChemStation 6.0, para aquisição e tratamento dos dados.

Utilizou-se capilar de sílica fundida com revestimento externo de poliimida com dimensões de 50 μm de diâmetro interno (d.i.) e 375 μm de diâmetro externo (d.e.). A janela de detecção foi realizada pela retirada do revestimento de poliimida, por um processo de aquecimento local, a 40 cm do início do capilar, para um comprimento total de 48,5 cm. O capilar foi condicionado, para o primeiro uso, com um *flush* de solução de NaOH 1 mol.L⁻¹

durante 30 min, água deionizada por 10 min e 15 min do eletrólito de corrida. Entre as corridas, o capilar foi recondicionado para limpeza através do *flush* de NaOH 0,02 mol.L⁻¹, água deionizada e eletrólito por 2 minutos cada.

Todos os reagentes utilizados foram adquiridos com grau analítico. Os padrões de Ácido Benzóico (ABE), Aspartame (ASP), Ciclamato (CIC), Acesulfame-K (ACSK) e Sacarina (SAC), foram adquiridos junto à empresa Sigma-Aldrich. O padrão interno (PI), Levofloxacina (LEVO), foi conseguido por meio de uma doação da Farmanguinhos.

Todas as soluções foram preparadas utilizando água deionizada pelo sistema de osmose reversa, através do equipamento Q842, da Quimis. A solução estoque do tampão Tetraborato de Sódio (TBS, Na₂B₄O₇) foi preparada, em balão volumétrico de 200 mL, pela pesagem e dissolução de uma massa referente à concentração de 100 mmol.L⁻¹; como se trata de um soluto de difícil solubilização o balão contendo a solução foi deixado em banho ultrassônico por 4 horas. A solução estoque do tampão TRIS-ABE foi preparada, em balão volumétrico de 200 mL, pela pesagem e dissolução de massa correspondente a 50 mmol.L⁻¹ de TRIS e 25 mmol.L⁻¹ de ABE, gerando um tampão de TRIS-ABE 50 mmol.L⁻¹.

A solução referente ao eletrólito de corrida consistiu da mistura das soluções tampão de TBS e TRIS-ABE, através das diluições dos estoques para uma concentração de 20 mmol.L⁻¹ para TBS e 15 mmol.L⁻¹ para TRIS-ABE, resultando em uma solução de pH em torno de 9,15. Vale ressaltar que o eletrólito de corrida sempre foi preparado no dia da análise.

As soluções estoque dos analitos foram preparadas pela pesagem e dissolução de massa referente às concentrações 2500 mg.L⁻¹, 2000 mg.L⁻¹, 2000 mg.L⁻¹ e 1500 mg.L⁻¹ de ASP, CIC, ACSK e SAC, respectivamente. Para construção da curva de calibração, os estoques foram diluídos a fim de que as concentrações dos níveis estabelecidos fossem encontradas. As soluções estoques foram armazenadas sob refrigeração em uma geladeira comum.

As amostras analisadas nesse trabalho foram todas adquiridas em supermercados locais, sendo estas: Adoçantes de mesa (líquidos e em pó), sucos light e diet (em pó), gelatinas (em pó) e achocolatado em pó. Os primeiros testes foram realizados com as amostras de sucos e estas foram pesadas e dissolvidas em água deionizada. Os balões volumétricos contendo as soluções foram levados ao banho ultrassônico por 30 minutos e então os mesmos tiveram seu volume acertado até a marca de aferição. Até então nenhum tipo de pré-processamento nas amostras foi realizado, isto é, após a dissolução em água, as soluções de amostra foram injetadas no equipamento. As soluções de amostra foram estocadas sob congelamento em um freezer de geladeira comum.

3.3.2 CROMATOGRAFIA LÍQUIDA COM DETECÇÃO POR ESPECTROMETRIA DE MASSAS (sigla do inglês, LC-MS)

As análises por LC-MS ocorreram após os primeiros testes e resultados obtidos por CE, o que será melhor discutido nos itens 3.4.1 e 3.4.2 dos resultados. Essas análises foram realizadas na Universidade Federal de Santa Catarina (UFSC) pela doutoranda Maressa Dolzan, integrante do Laboratório de Eletroforese Capilar e Técnicas Cromatográficas. O material enviado da UFJF consistiu em cerca de 100 mg de cada amostra de trabalho, além dos respectivos padrões; mencionados no item anterior.

Para as análises das amostras de trabalho no LC-MS, foram pesadas aproximadamente 10 mg de cada amostra e transferidos para frascos eppendorffs, aos quais foi adicionado 1,0 mL de água deionizada. Os sistemas foram agitados por vortex e em seguida permaneceram em ultrason por 24 minutos, sendo novamente agitados por vortex a cada 8 minutos. Por fim foram centrifugados por 1 min a 12000 rpm, diluídos adequadamente em água e injetados automaticamente no sistema cromatográfico.

O método consistiu na utilização de uma coluna C18 (150 mm x 4.6 mm, 5µm tamanho de partícula), fase móvel com mistura de soluções, sendo a primeira composta da mistura de água deionizada com acetato de amônio na concentração de 5 mmol L⁻¹ (A) e a segunda, preparada pela mistura de metanol e acetato de amônio na concentração de 5 mmol.L⁻¹(B). O gradiente foi: 0 min) 5% B; 5 min) 20% B; 6 min) 95% B; 9 min) 95% B; 9.06 min) 5% B; 14 min) 5% B, com fluxo de 0,8mL min⁻¹. A temperatura da coluna foi mantida a 40°C e a injeção foi de 20 µL. As principais transições monitoradas, no modo MRM, foram:

	QIT	CIT
ACE	162 > 82	162 > 82
SAC	182 > 102	182 > 62
CIC	178 > 80	178 > 64
ASP	293 > 200	293 > 261

3.3.3 ESPECTROSCOPIA RAMAN NIR

Para aquisição dos espectros foi utilizado o espectrofotômetro FT-Raman da Bruker com radiação excitante em 1064 nm, modelo RFS-100/S, equipado com detector de Germânio e resfriado com Nitrogênio líquido, que se encontra disponível nas instalações da UFJF.

Os padrões utilizados para se obter os espectros puros dos analitos (dos quatro edulcorantes) foram os mesmos para o trabalho desenvolvido na eletroforese capilar, ou seja,

todos adquiridos da Sigma-Aldrich. As amostras comerciais também foram as mesmas analisadas via eletroforese capilar.

O procedimento de aquisição dos espectros foi o mais simples possível e sem pré-tratamento das amostras. Para os padrões puros dos quatro edulcorantes, a aquisição dos espectros consistiu apenas na transferência de uma pequena quantidade do sólido para o porta-amostra do equipamento. Para esses espectros não se importou com a quantidade de amostra nem com o tipo de porta-amostra, visto que tratava-se apenas de uma análise qualitativa dos espectros dos padrões. Esses espectros foram então adquiridos pela irradiação excitante em 1604 nm, com potência fixada em 100 mW, visto que não deteriorava a amostra; foram feitas 128 varreduras com resolução de 4 cm^{-1} .

Para análises quantitativas, é necessário que se tenha repetibilidade nas análises. Para essa finalidade o porta-amostras para sólidos disponível no equipamento não é a melhor opção, pois existem vários deles com diferentes tamanhos de diâmetro e profundidade do orifício onde se introduz a amostra. Sendo assim, foi utilizado um porta-amostra próprio para o equipamento de NIR, mas que foi adaptável para o espectrofotômetro FT-Raman, permitindo uma análise mais confiável e reprodutível. Com a finalidade de validar o uso desse porta-amostra, foram adquiridos espectros em ambos porta-amostras e percebeu-se que não há diferenças significativas entre os espectros, o que nos permitiu prosseguir com os trabalhos da maneira pretendida.

No caso de aquisição dos espectros referentes às amostras, a maioria delas se apresenta como um sólido de alta granulometria, sendo necessária uma maceração prévia, que foi realizada em um almofariz de ágata. Após essa etapa o sólido foi transferido para o porta-amostras de NIR e a leitura dos espectros realizada nas mesmas condições instrumentais para os padrões de edulcorantes.

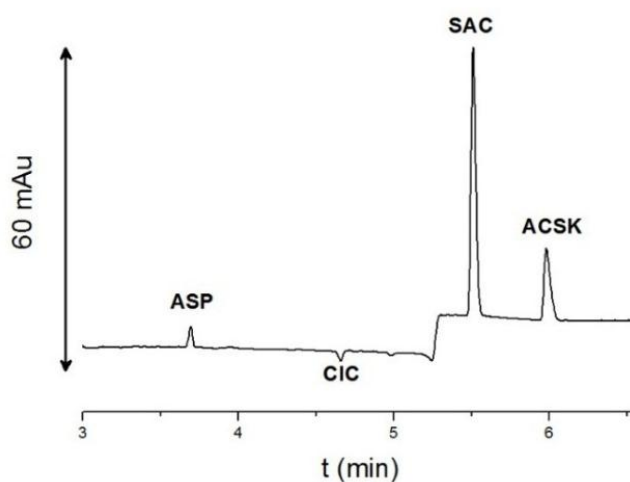
Logo em que terminou-se a aquisição dos espectros no equipamento FT-Raman, as amostras foram levadas para análise de NIR. O conteúdo sólido foi transferido para o frasco específico de NIR que permite a leitura por reflectância, modo pelo qual os espectros foram adquiridos. Após todas as amostras de trabalho estarem nos respectivos frascos (porta-amostras), as leituras foram realizadas no equipamento da Bruker modelo FT-NIR MPA; foram feitas 32 varreduras com resolução de 4 cm^{-1} .

3.4 RESULTADOS

3.4.1 ELETROFORESE CAPILAR – REPRODUÇÃO E APLICAÇÃO DA METODOLOGIA

Como o método para análise simultânea de edulcorantes já havia sido desenvolvido por Fernandes e colaboradores (Fernandes et al. 2013), a primeira tentativa foi a reprodução do mesmo. Para tal, foi realizado um experimento que consistiu no preparo e injeção de uma mistura de padrões, contendo os quatro analitos de interesse na concentração de 100 mg L^{-1} (ASP, CIC, SAC e ACSK). As condições de análise otimizadas no trabalho de Fernandes foram mantidas. O perfil do primeiro eletroferograma obtido pode ser visto abaixo:

Figura 38: Eletroferograma para mistura de padrões dos analitos em estudo



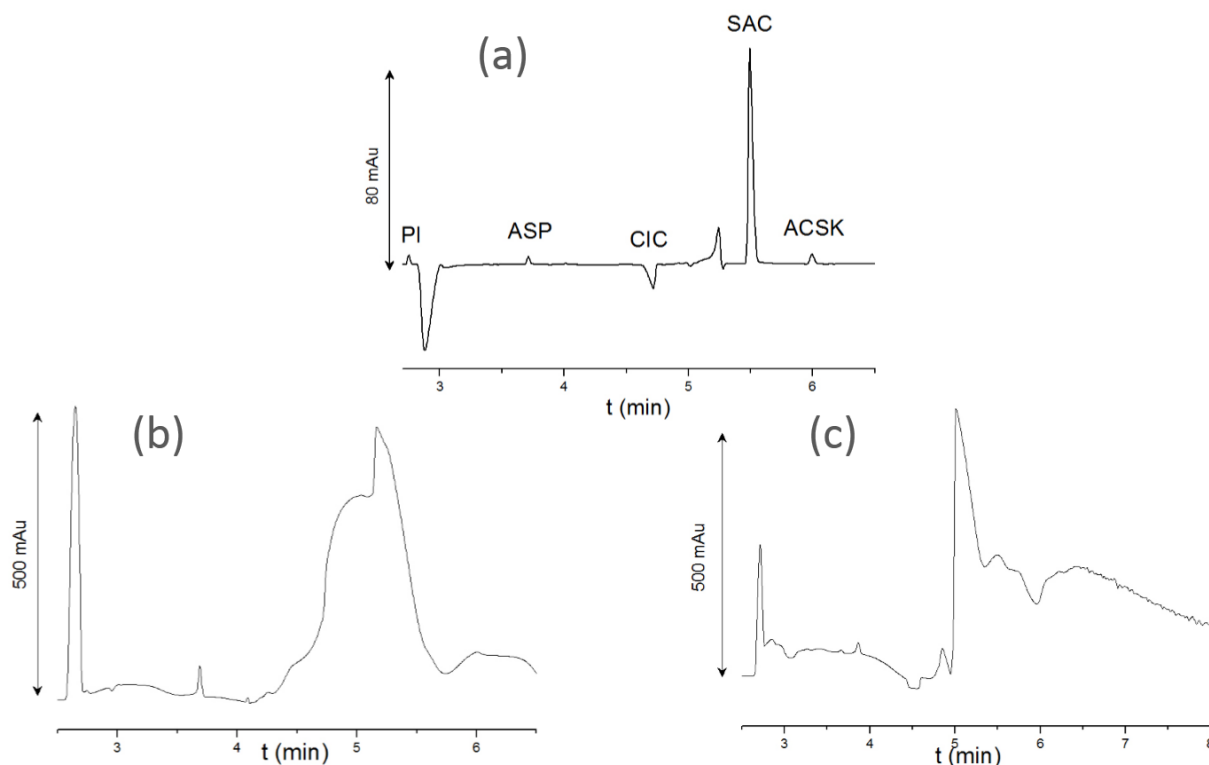
Fonte: ELABORADA PELO PRÓPRIO AUTOR

Por meio da Figura 38 pode-se afirmar que o eletroferograma obtido nessa corrida está em concordância com o obtido por Fernandes e colaboradores (Fernandes et al. 2013). Tanto os tempos de migração para os quatro analitos, quanto o perfil dos picos são muito semelhantes. Conforme discutido no item 1.6.4 a ordem de eluição dos edulcorantes obedece ao mecanismo CZE, sendo essa: ASP, CIC, SAC e ACSK.

Sabendo que o método foi reprodutível, o passo seguinte foi o teste em amostras comerciais. Algumas amostras de adoçantes de mesa foram as primeiras testadas, sendo 3 nesses primeiros testes. Duas dessas amostras eram líquidas (Figura 39 (b) e (c)) e estas foram transferidas para o vial e injetadas sem nenhum tipo de pré-tratamento e os resultados não foram satisfatórios, visto que os picos referentes aos edulcorantes estavam encobertos pelas bandas

largas e intensas oriundas da matriz. A amostra sólida, nomeada AM1, teve seu conteúdo dissolvido em água deionizada, seguida de sua injeção no equipamento de CE e o perfil do eletroferograma se apresentou em concordância com o obtido para a mistura de padrões, como pode ser visto na Figura 39 (a). Nesse eletroferograma, também pode-se perceber a presença de um padrão interno (PI), nesse caso a Levofloxacin (LEVO), que tem por objetivo minimizar as incertezas nas injeções, bem como, variações instrumentais (Skoog et al. 2005)

Figura 39: Eletroferograma das amostras em que foram aplicadas a metodologia desenvolvida por Fernandes. Uma amostra sólida (a), nomeada AM1 e duas líquidas, (b) e (c)



Fonte: ELABORADA PELO PRÓPRIO AUTOR

A amostra AM1 é composta pelos quatro edulcorantes artificiais aqui estudados, sendo esses os minoritários, além de Lactose (majoritário) e Dióxido de Silício; o rótulo do produto informa que de 800 mg, 750 mg são de Lactose e os 50 mg restantes, referem-se ao somatório dos quatro edulcorantes. Para essa amostra, foi observado em seu eletroferograma, um pico de sistema (PS), originário da matriz, próximo a 2,8 min de corrida. Esse PS compromete a integração do pico referente ao PI, que seria utilizado como fator de normalização para as respostas dos analitos.

A fim de verificar se a resposta do PI influencia negativamente como fator de normalização, foi realizada a quantificação dos analitos na amostra AM1, por meio de curvas

de adição de padrão (também chamada de curva interna). Para tanto, a curva foi construída em cinco níveis com triplicatas genuínas em cada nível, sendo que para o primeiro ponto nenhuma quantidade de analito foi adicionada, contendo apenas amostra e PI na concentração de 50 mg L^{-1} . Nos demais níveis, além de amostra e PI, foram adicionados padrões de cada analito em níveis crescentes de concentração. Através da extrapolação da curva, a concentração de cada analito foi calculada na amostra (onde $y = 0$). Os primeiros cálculos foram realizados levando em consideração o padrão interno, ou seja, a curva interna foi plotada através de uma normalização pelo padrão interno, tanto para resposta instrumental (área dos picos) quanto para as concentrações; em outras palavras, a curva é plotada pela razão das áreas (ordenada) versus a razão das concentrações (abscissa), sempre do analito com relação ao padrão interno. A segunda forma de cálculo não levou em consideração a contribuição do padrão interno. O resultado dos cálculos, para ambas as formas, é apresentado na Tabela 10:

Tabela 10: Cálculo das massas dos edulcorantes na amostra, AM1, via curvas de adição de padrão (a) e (b).

	CAP(a)		CAP(b)	
	R ²	m (mg)	R ²	m (mg)
ASP	0.949	1.88	0.992	2.40
CIC	0.812	22.34	0.940	30.69
SAC	0.663	36.04	0.835	1.62
ACSK	0.969	1.30	0.997	13.13
Σ_{massa}		61.55		47.84

C_{AP(a)} - Curva de Adição de Padrão utilizando o PI como fator de normalização

C_{AP(b)} - Curva de Adição de Padrão sem PI

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Em termos da correlação dos dados (R^2), houve um ganho considerável sem PI. Outro ponto importante é que os valores de massa obtidos para a amostra real, através dos cálculos realizados sem a utilização do PI, apresentaram melhores resultados quando se compara o somatório dos edulcorantes com o informado pelo fabricante do produto, como dito no parágrafo anterior, 50 mg. Logo, foi possível perceber que para essa amostra o PS próximo ao pico do PI comprometeu a integração do mesmo, o que acarretou em uma deterioração da modelagem. Sendo assim, a utilização do PI como fator de normalização passou a ser desconsiderada.

3.4.2 ESTRATÉGIA DE AÇÃO

Visto que para a amostra AM1 a técnica de eletroforese capilar respondeu aparentemente bem, o foco do trabalho foi direcionado para amostras que continham a mesma matriz, isto é, Lactose, Dióxido de Silício (excipientes) e os quatro edulcorantes de interesse, ASP, CIC, SAC e ACSK. Além da amostra AM1, outras amostras de marcas diferentes, contendo a mesma composição, são encontradas disponíveis para o consumo dos brasileiros.

Partiu-se de 30 sachês de adoçantes de mesa sólidos da marca AM1, para a formação do primeiro universo amostral, denominado *blend*. Esse *blend* consiste da massa total agrupada dos 30 sachês macerados, de 5 em 5 por 5 minutos, em almofariz de ágata. Esse aglomerado de adoçante de em pó macerado, refere-se ao ponto de partida para o doseamento e construção das demais amostras de trabalho.

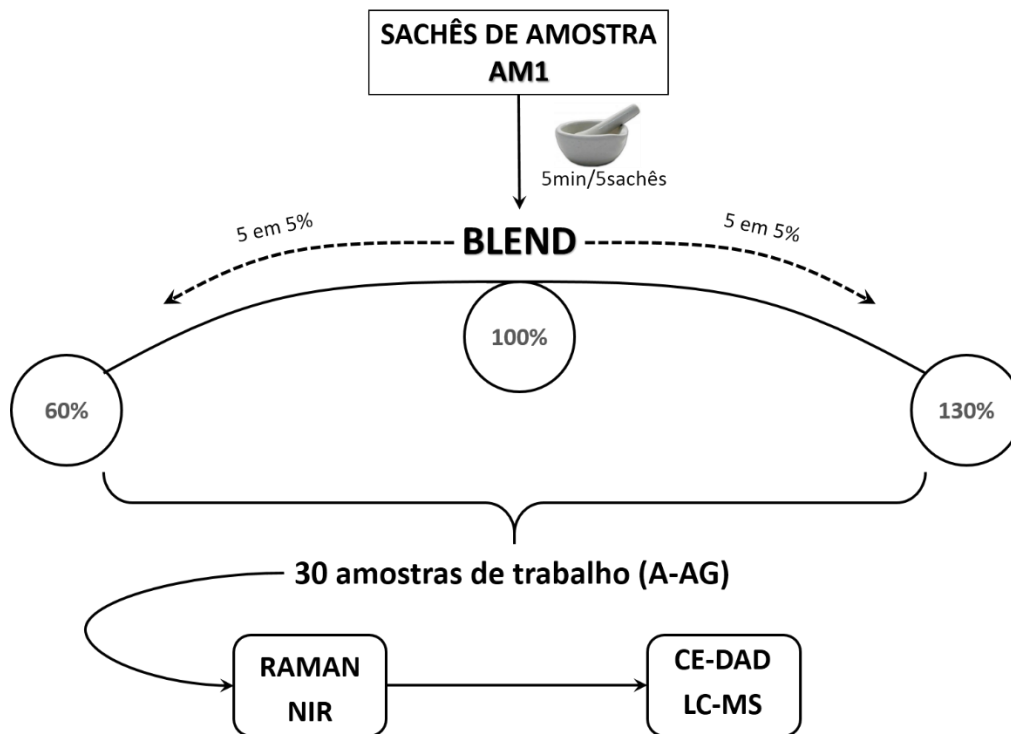
O doseamento consiste na diluição e concentração do *blend* a partir da adição de excipiente (lactose) e mistura de padrões (MP), respectivamente; onde essa MP contém todos os analitos na mesma proporção em que foram encontrados na amostra AM1. Por meio do doseamento foi possível construir um novo conjunto amostral, esse que foi analisado pelas técnicas espectroscópicas, Raman e NIR e, por conseguinte, pelas técnicas de referência, CE e LC-MS. Logo, é importante frisar que as novas amostras foram construídas com sobreposição à matriz da amostra AM1, ocorrendo somente diluição ou concentração dos analitos em questão.

O *blend* refere-se ao 100% de concentração dos analitos, pois é completamente fiel a amostra original, isto é, não sendo adicionada lactose ou MP. As amostras com concentração abaixo de 100%, foram preparadas pela diluição do *blend* com lactose e esses novos níveis decresceram, de 5 em 5% em termos de concentração, até alcançar a marca de 60% da concentração inicial; para tanto a massa das amostras com níveis baixos foram fixadas em 400 mg. Para as amostras que compreendiam os níveis superiores a 100%, até 130% ou 30% a mais que a amostra original, foram adicionados aos 400 mg do BLEND uma determinada massa referente às frações de 5 em 5%. É importante frisar, que se trata da preparação de soluções sólidas, então a cada adição de excipiente ou MP, as novas amostras foram agitadas em vortex e então maceradas novamente, uma a uma, com o intuito de garantir a homogeneidade.

Uma vez preparado esse novo conjunto amostral, sendo agora o conjunto de trabalho, as amostras foram submetidas às espectroscopias, seguida das análises de referência. Essa sequência foi estabelecida pelo fato da radiação utilizada na espectroscopia não destruir as

amostras, podendo ainda utilizá-las para posteriores avaliações. O esquema a seguir tem o objetivo de tornar a estratégia de ação mais clara:

Figura 40: Esquema para a estratégia de ação adotada nessa etapa do trabalho



Fonte: ELABORADA PELO PRÓPRIO AUTOR

3.4.3 ESPECTROSCOPIAS

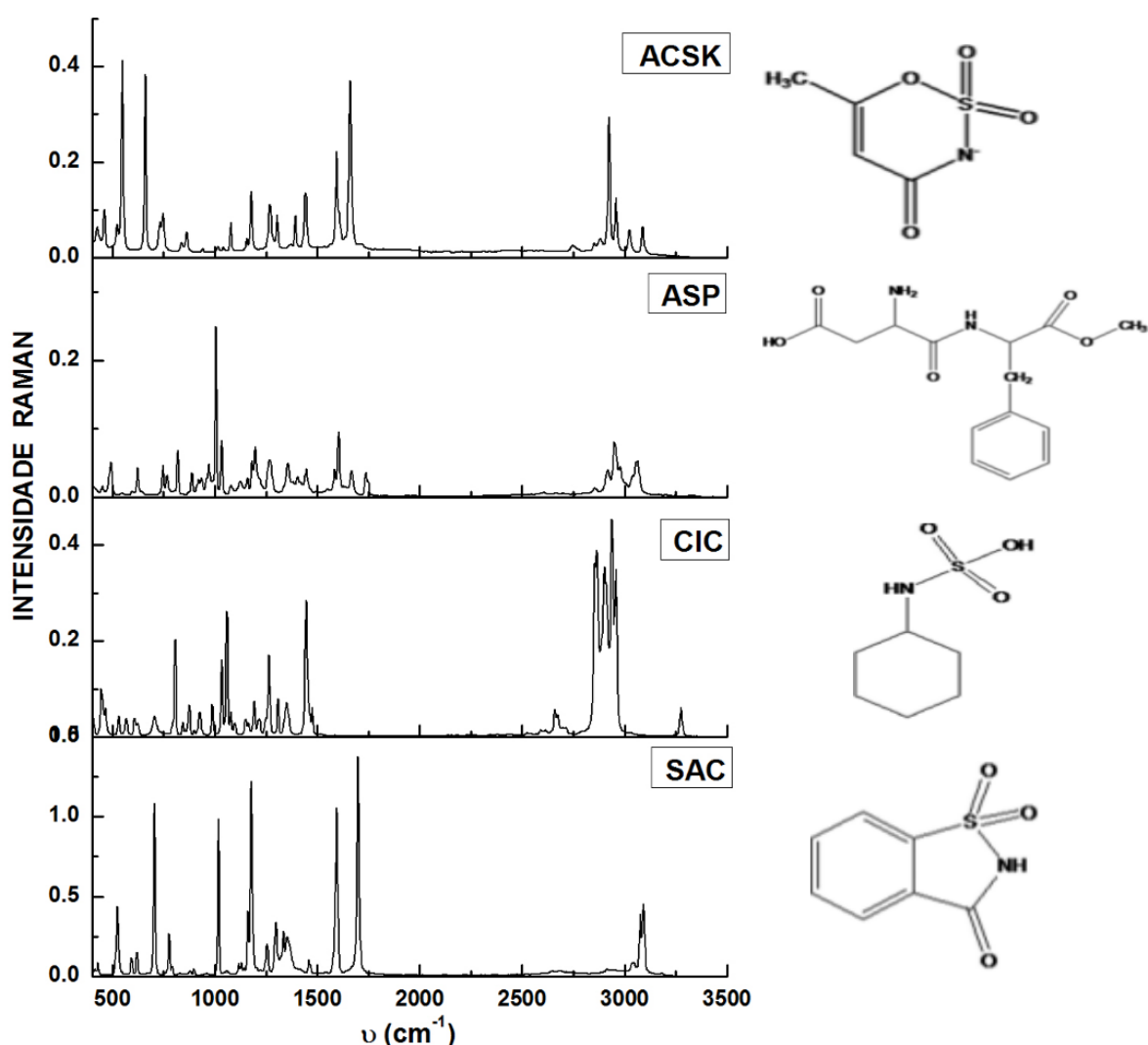
De posse do conjunto amostral de trabalho, composto por 30 amostras variando a concentração dos analitos de 60 a 130% da real (apontada para a amostra AM1), adquiriu-se os espectros para cada uma delas e então obteve-se as concentrações pelo método de referência. Anteriormente à análise quantitativa pelo método de referência, realizou-se a parte espectroscópica pelo fato de se tratar de técnicas não destrutivas.

Foram aplicadas duas diferentes técnicas espectroscópicas por dois principais motivos. O primeiro deles era avaliar o desempenho de cada uma delas no que diz respeito à análise quantitativa, onde se esperava melhor desempenho da técnica de absorção de radiação, no caso o NIR, frente a uma técnica de espalhamento de luz, como o Raman. O segundo ponto interessante é que o Raman, além de possibilitar a informação quantitativa (quando associado a ferramentas quimiométricas), pode fornecer a informação qualitativa de grande qualidade, principalmente para avaliação da modificação da matriz após a diluição ou fortificação, por

meio de alguma banda duvidosa que possa aparecer, seja essa oriunda de contaminação ao longo da manipulação da amostra ou possíveis oxidações.

Primeiramente foi feita a aquisição dos espectros raman para os quatro edulcorantes artificiais e para o constituinte majoritário da amostra trabalhada, a lactose. Os espectros para os analitos podem ser vistos na Figura 41.

Figura 41: Espectro FT-Raman para os padrões puros de ACSK, ASP, CIC e SAC

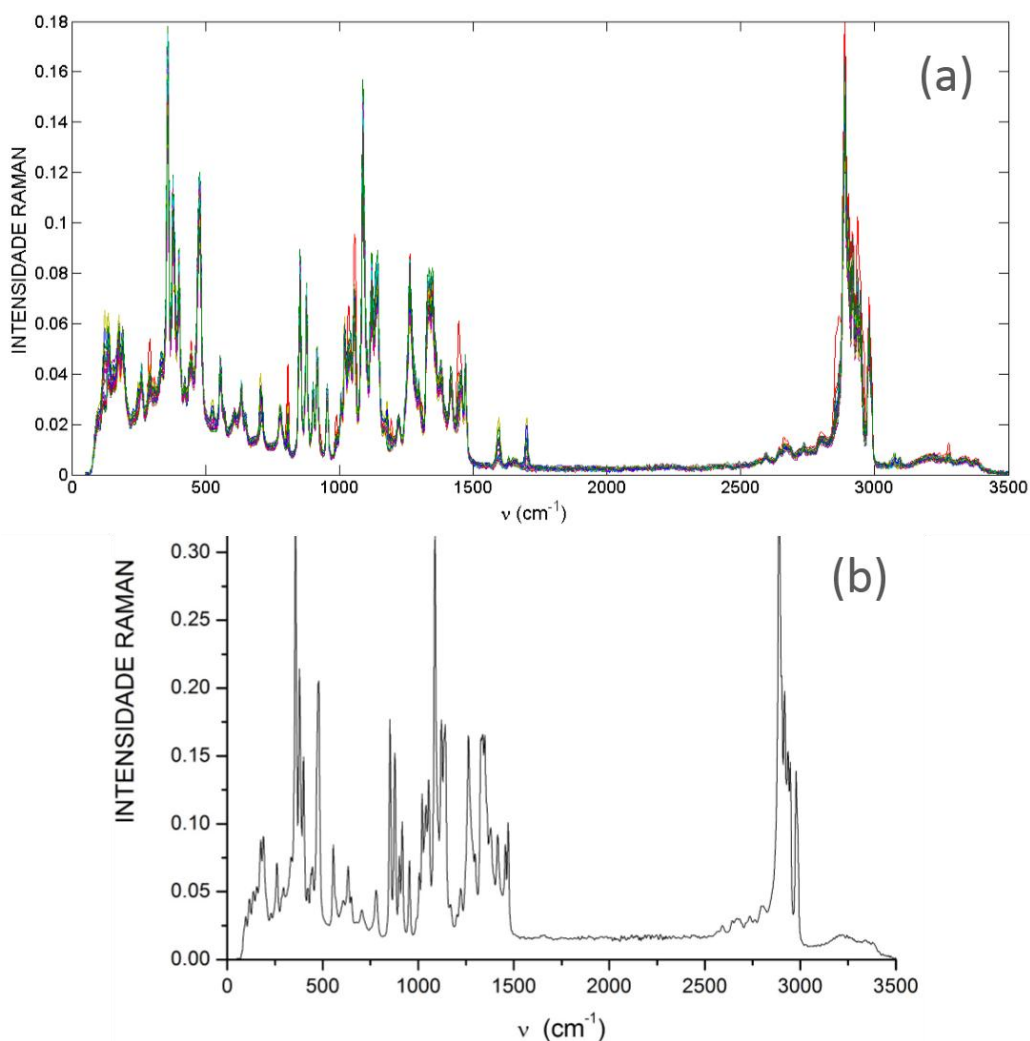


Fonte: ELABORADA PELO PRÓPRIO AUTOR

Como a lactose é o componente majoritário dessa formulação o espectro das amostras de trabalho apresentam um perfil muito similar ao dessa substância o que pode ser percebido

pela análise da Figura 42, onde é mostrado o conjunto de espectros referentes a todas amostras de trabalho (a) e o espectro que diz respeito à lactose (b); os espectros para cada amostra de trabalho foram adquiridos no mesmo dia para que houvesse o mínimo de variação instrumental. A diferença mais perceptível entre as duas figuras aparece na região compreendida entre 1550 e 1750 cm^{-1} , onde a lactose não apresenta bandas, enquanto que para o conjunto de amostras é possível observar duas. A primeira, com máximo de intensidade em 1580 cm^{-1} pode-se referir aos estiramentos simétricos $C = C$, presentes nas estruturas do ASP, SAC e ACSK, já a segunda, em 1700 cm^{-1} , refere-se ao estiramento simétrico da carbonila ($C = O$), que pode estar associados também aos analitos ASP, SAC e ACSK.

Figura 42: Espectros Raman agrupados para as 30 amostras de trabalho (a) e espectro Raman para a lactose (b)

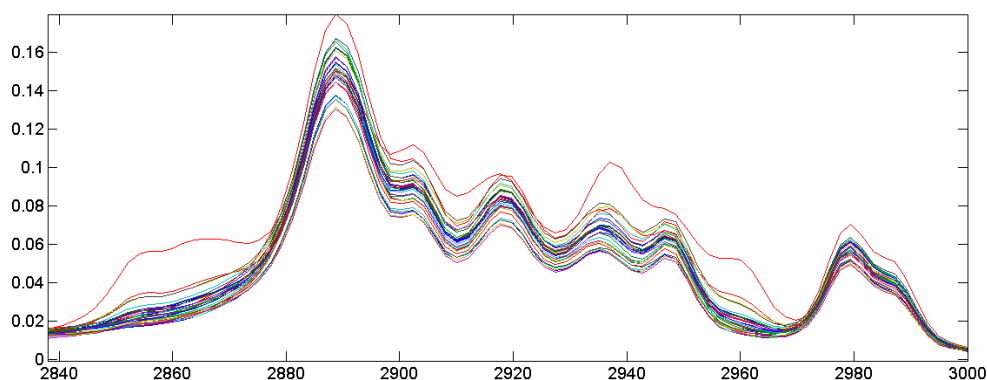


Fonte: ELABORADA PELO PRÓPRIO AUTOR

A Figura 42 (a) dá a impressão de que para as 30 amostras existe uma grande similaridade nos espectros, no entanto, para aplicação em análises quantitativas é muito

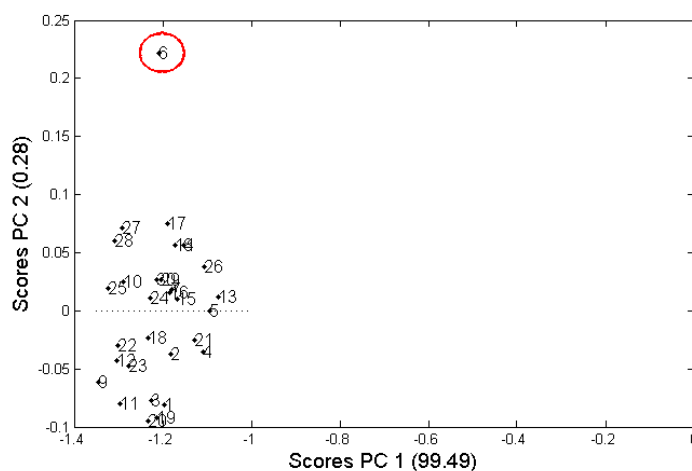
importante que uma análise mais minuciosa no conjunto de espectros seja feita. Visualmente, é possível perceber que o espectro de uma amostra, destacado em coloração vermelha, aparenta estar em discordância com os demais, o que pode ser melhor visualizado pela aproximação feita na região compreendida entre 2840 e 3000 cm^{-1} (Figura 43). Trata-se da amostra F (ou número 6), na qual os analitos estão com 70% da concentração original, pelo fato desta ter sido diluída. Entretanto, seu espectro apresenta-se com altas intensidades para a maioria das bandas em comparação com os demais espectros e isso seria incoerente, visto que essa é uma amostra diluída. Sendo assim, os espectros das 30 amostras foram submetidos à PCA, para uma melhor classificação do conjunto (Figura 44).

Figura 43: Aproximação feita na região espectral compreendida entre 2840 e 3000 cm^{-1} , para melhor visualização do possível *outlier*



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 44: PCA para o conjunto de espectros Raman, referente às 30 amostras de trabalho



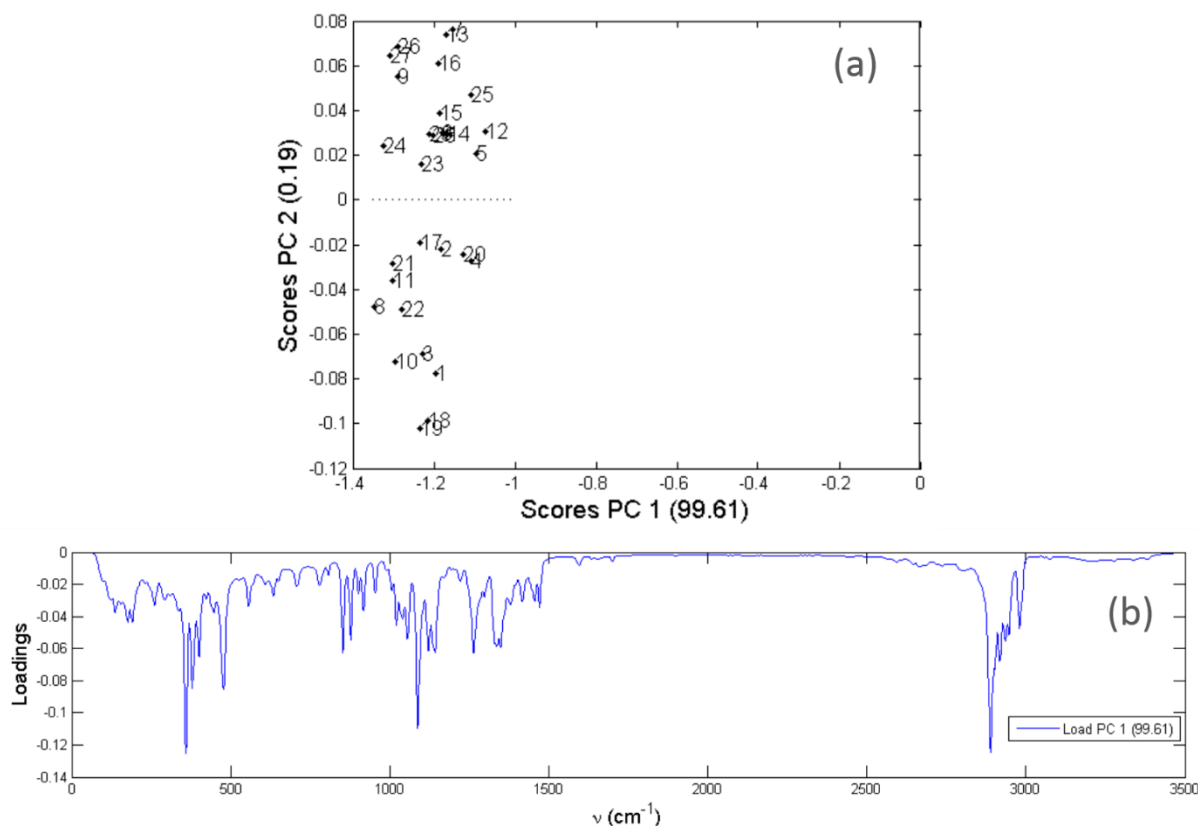
Fonte: ELABORADA PELO PRÓPRIO AUTOR

Pela PCA mostrada anteriormente percebe-se que a amostra de número 6, ou F, encontra-se distante do principal grupo de amostras e visto que já se desconfiava desta, por ser

uma potencial *outlier*, a mesma foi retirada do conjunto e para as modelagens com dados de Raman, foram utilizadas 29 amostras ao invés de 30. Após a retirada dessa amostra, uma nova PCA foi realizada, tornando possível a confirmação de que essa era, de fato, a amostra problema (Figura 45(a)). Na primeira PCA, assim como na segunda, nota-se que por meio de uma PC é possível descrever mais de 99% das variáveis originais, o que ainda tem um ligeiro aumento após a retirada do *outlier*. Outra forma interessante de olhar, é por meio do gráfico de *loadings* para a PC1, que descreve perfeitamente o espectro Raman para as amostras de trabalho, dando ênfase nas principais regiões onde existe relevância química e apresentando quase nenhum ruído (Figura 45(b)).

Para o conjunto espectral utilizado, não foi realizado nenhum tipo de pré-processamento, pois as amostras foram todas construídas com sobreposição a uma mesma matriz, os espectros todos adquiridos no mesmo dia (minimizando a variação instrumental) e por tratar-se de uma matriz mais simples, contendo apenas um excipiente. Foram realizados alguns testes para os pré-processamentos, como a aplicação de SNV e MSC, mas o conjunto espectral não apresentou diferenças significativas, logo, optou-se por utilizar os espectros como adquiridos (brutos), a fim de que não houvesse manipulação matemática, o que tornaria os dados ainda mais fiéis à realidade química das amostras.

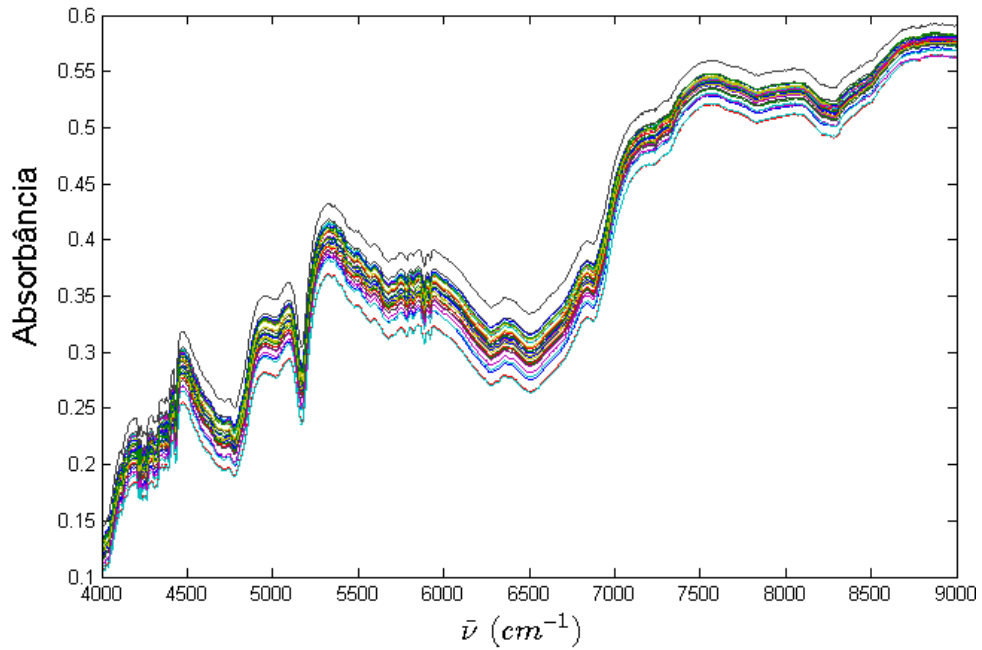
Figura 45: PCA para o novo conjunto de espectros Raman, referente as 29 amostras de trabalho (a) e o gráfico de *loading* mostrando que a primeira PC descreveu 99,61% da variabilidade dos dados originais (b)



Fonte: ELABORADA PELO PRÓPRIO AUTOR

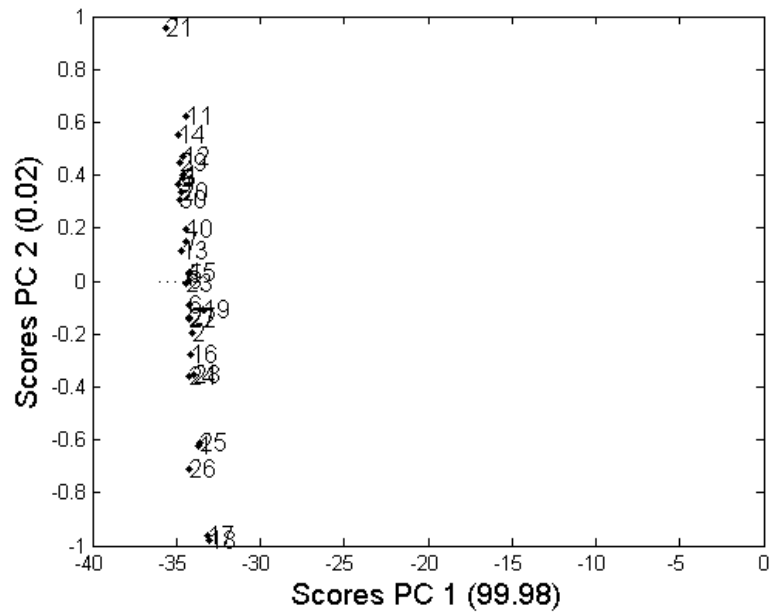
Os espectros de NIR agrupados para formar o conjunto amostral foram adquiridos na faixa de 4000 a 9000 cm^{-1} , tendo como informação principal os primeiros e segundos sobretons, além de bandas de combinação. Na Figura 46 são mostrados os espectros de NIR agrupados para as 30 amostras referentes ao conjunto de amostras de trabalho. Para esses dados também foi realizada uma PCA, a qual, por meio de essencialmente uma PC, descreveu quase 100% das variáveis originais e mostrou que apesar do conjunto variar ao longo da PC1, não há evidências de *outlier* (Figura 47). O gráfico de loadings para a PC1 é mostrado na Figura 48.

Figura 46: Espectros NIR agrupados para as 30 amostras de trabalho



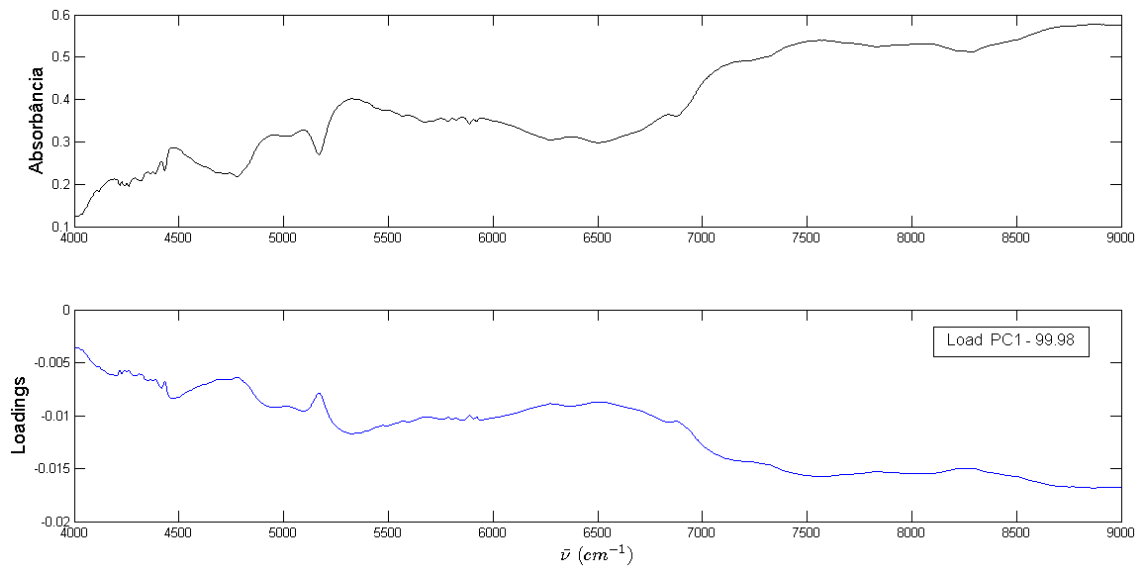
Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 47: PCA para o conjunto de espectros NIR, referente as 30 amostras de trabalho



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 48: Espectro médio para as 30 amostras (superior) e gráfico de loadings para a PC1 (inferior), que descreveu 99,98% das variáveis originais



Fonte: ELABORADA PELO PRÓPRIO AUTOR

3.4.4 ADQUIRINDO AS CONCENTRAÇÕES PELO MÉTODO DE REFERÊNCIA: COMPARAÇÃO ENTRE CE-DAD E LC-MS

Após a realização das espectroscopias, a matriz de espectros X foi montada, entretanto os modelos por PLS necessitam da resposta do parâmetro de base estudado, oriunda de um método de referência, para montar o vetor de resposta Y, conforme descrito no item 1.4.2. A priori, as respostas de concentração viriam da CE, então as análises das 30 amostras foram realizadas e os cálculos de concentração se deram por duas formas a fim de fazer uma comparação e conseqüentemente ter maior confiabilidade dos resultados finais.

Primeiro, pensou-se no método de calibração externa, pois com as curvas validadas para cada analito, bastava fazer a leitura das amostras no equipamento de CE e então substituir as respostas instrumentais nas equações das curvas analíticas para realizar a inferência estatística e obter o valor de concentração. O outro método de quantificação se deu por meio da calibração interna, por meio de uma curva de dois pontos (Bader 1980). Nesse método de padronização interna, utilizam-se dois balões ou a preparação é feita diretamente no vial, sendo que ao primeiro balão ou vial não é adicionado nenhuma quantidade de padrão referente aos analitos, mas somente amostra e PI, caso haja. Ao segundo balão, ocorre a adição de padrão com concentração conhecida. A concentração do analito na amostra é conseguida por meio da equação:

$$C_A = \frac{A_1 \times C_{PEtq} \times V_{PEtq}}{V_A(A_2 - A_1)}$$

onde C_A refere-se à concentração do analito no vial, sendo necessário levar em consideração a diluição para encontrar o valor real de concentração na amostra, C_{PEtq} a concentração do padrão referente ao analito na solução estoque, V_{PEtq} o volume da alíquota de padrão retirada do estoque, V_A o volume de amostra adicionado ao balão ou vial, A_1 é a resposta instrumental referente ao pico do analito na amostra e A_2 à área do analito na amostra acrescida de padrão, isto é, a resposta instrumental para a segunda preparação.

Como ao primeiro balão para os cálculos via curva interna, não são adicionados nenhuma quantidade de padrão, temos somente a resposta dos analitos já presentes na amostra e essa informação pode também ser utilizada na curva externa para a quantificação. Logo essas duas metodologias foram realizadas e os resultados comparados. Entretanto, um passo prévio à inferência estatística por meio de uma curva externa, é a validação da mesma por meio da avaliação de alguns pressupostos estatísticos. Para avaliação da qualidade do ajuste do modelo, o exame dos resíduos é de suma importância. Sendo assim, além dos erros de observação serem

independentes e estarem normalmente distribuídos, estes devem ter variância comum (homocedásticos), ou seja, as variâncias devem ser homogêneas (Neto et al. 2010; Anjos 2005); a homocedasticidade foi avaliada pelo teste de Cochran. Para todas as curvas tivemos comportamento homocedástico, pois os valores calculados para o teste foram menores que os tabelados (ou críticos), sendo assim a regressão por mínimos quadrados ordinários pôde ser utilizada, caso contrário, a condição de contorno seria a utilização da ponderada. O último passo de validação para a utilização das curvas, foi a submissão das mesmas ao teste de falta de ajuste, para verificação de indícios de desvios da linearidade; também para nenhuma foi encontrada, visto que o valor de F calculado foi menor que o tabelado, dentro do nível de significância de 0,05. Uma tabela resumo, com os principais parâmetros de avaliação das curvas externas é mostrada na tabela a seguir:

Tabela 11: Tabela resumo da avaliação dos pressupostos para as curvas externas referentes aos quatro analitos

	Extensão (mg L ⁻¹)	R ²	Homocedasticidade (Cochran)	Regressão	Linearidade	Falta de Ajuste
ASP	150-2000	0,9951	^a C _{calc} =0,46 < C _{tab} =0,68	^b MMO	F _{calc} =0,14 < ^c F _(3.10) =3,71	Não há
CIC	150-1300	0,991	^a C _{calc} =0,48 < C _{tab} =0,68	^b MMO	F _{calc} =0,68 < ^c F _(3.10) =3,71	Não há
SAC	35-420	0,9913	^a C _{calc} =0,55 < C _{tab} =0,68	^b MMO	F _{calc} =0,06 < ^c F _(3.10) =3,71	Não há
ACSK	35-380	0,9929	^a C _{calc} =0,51 < C _{tab} =0,68	^b MMO	F _{calc} =0,82 < ^c F _(3.10) =3,71	Não há

$\alpha = 0.05$; a - valor para o teste de Cochran calculado; b - Mínimos Quadrados ordinários (MMO);
c - F(m-p,n-m), onde: n - total de medidas; m - número de níveis; p - parâmetros do modelo

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Uma vez validadas as curvas com calibração externa, estas estão aptas para estimar a concentração dos analitos. Os cálculos para a quantificação dos analitos foram feitos nas 30 amostras de trabalho, por meio das curvas com calibração externa e interna e a comparação entre as duas metodologias é mostrada na tabela Tabela 12. No corpo do texto será mostrada a comparação para as cinco primeiras amostras (A-E), para dois dos quatro analitos, estando a tabela completa para os 4 analitos nas 30 amostras no apêndice B.

Tabela 12: Quantificação e comparação da concentração de dois analitos, ASP e CIC, nas 5 primeiras amostras de trabalho, por meio de curvas externas e internas com dois pontos

	[ASP] _{CEX} (mg.L ⁻¹)	[ASP] _{CI} (mg.L ⁻¹)	Er	[CIC] _{CEX} (mg.L ⁻¹)	[CIC] _{CI} (mg.L ⁻¹)	Er
A	119,83	80,55	48,76	1276,53	835,42	52,80
B	116,21	78,27	48,47	1243,23	908,12	36,90
C	123,69	82,42	50,07	1267,40	753,86	68,12
D	131,17	84,16	55,86	1312,51	920,18	42,64
E	131,17	89,88	45,95	1298,55	643,09	10,92

CEX - Curva Externa | CI - Curva Interna | |Er| - módulo do erro absoluto

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Para construir um modelo de calibração multivariada, uma das premissas é que o método de referência seja confiável, pois os espectros refletirão os teores dos componentes na amostra e o que teremos é uma diferença na intensidade dos picos dos analitos, o que é diretamente proporcional à concentração dos mesmos. Como as quantificações pelos dois métodos de calibração apresentaram diferenças significativas, para todos os analitos em todas as amostras (Tabela 13 e Apêndice C), passou-se a desconfiar das respostas fornecidas pela CE. A saída foi a realização das quantificações por um método mais robusto e confiável, a saber, a LC-MS.

As quantificações foram feitas por LC-MS e os resultados foram comparados com os de CE. No corpo do texto será mostrada a seguir, uma tabela com os testes t para comparação de duas metodologias e no Apêndice C, encontram-se os valores de concentração comparados pelas duas metodologias, para cada analito em cada amostra.

Tabela 13: Comparação estatística entre as metodologias, CE e LC-MS, para quantificação dos analitos nas amostras

	Teste-F (95%)	Variância	Teste-t (95%)	Metodologias
ASP	p-valor = 0.324	Equivalente	p-valor = 0.147	Equivalentes
CIC	p-valor = 3.4×10^{-9}	Diferente	p-valor = 3.7×10^{-3}	Diferentes
SAC	p-valor = 5.65×10^{-8}	Diferente	p-valor = 0.03	Diferentes
ACSK	p-valor = 0.118	Equivalente	p-valor = 0.02	Diferentes

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Primeiramente foi realizado um teste-F para avaliar se as variâncias entre os dois conjuntos de dados, obtidos pelas duas metodologias, eram equivalentes ou não, para então decidir qual teste-t aplicar. Para ASP e ACSK as variâncias foram equivalentes e para CIC e SAC diferentes, aplicando os devidos testes-t verificou-se que ao nível de confiança de 95%, apenas para o ASP as duas metodologias, CE e LC-MS, são equivalentes. Diante desse quadro,

foi tomada a decisão de utilizar as respostas oriundas do LC-MS para alimentar os modelos de calibração multivariada. Uma das possibilidades pela qual a metodologia desenvolvida por Fernandes e colaboradores não se apresentou de maneira eficaz, para a quantificação dessas amostras, é com relação ao PI. Para essa matriz, o PS próximo ao pico do PI compromete a integração do mesmo e então os cálculos foram todos realizados sem a normalização pelo PI, o que deixa o método mais vulnerável às variações instrumentais e, conseqüentemente, menos robusto.

3.4.5 MODELAGEM MULTIVARIADA PARA DETERMINAÇÃO DOS EDULCORANTES ARTIFICIAIS

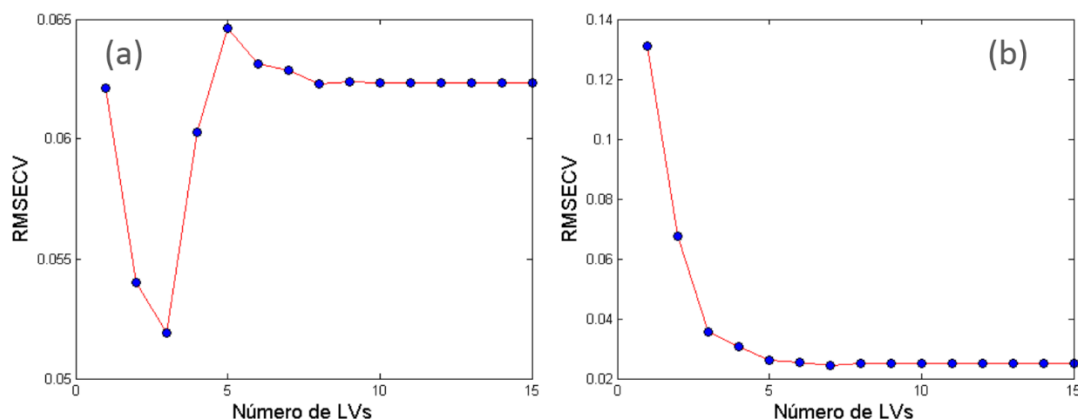
Com os espectros de todas as amostras de trabalho, bem como a resposta do método de referência, LC-MS, para as concentrações dos analitos, foi possível construir os modelos de calibração multivariada. Como foram realizados dois tipos de espectroscopia, obtivemos diferentes modelos com as informações de Raman e NIR, ambos alimentados pela resposta do LC-MS. Serão apresentados a seguir, os resultados para Raman e NIR, de cada analito separadamente.

3.4.5.1 ASPARTAME

Dois modelos preditivos para o edulcorante artificial ASP foram construídos. O modelo de PLS com os espectros Raman foi construído a partir de 29 amostras, pois uma foi identificada como *outlier* e retirada, conforme discutido no item 3.4.3. Para o modelo de PLS com NIR, foram utilizadas 30 amostras, pois não houve evidência da presença de *outliers*. Para ambos os casos, não houve pré-tratamento nos espectros, muito menos alinhamento, como foi feito para a parte do petróleo. Sendo assim, por meio do algoritmo Kennard-Stone, a matriz X de espectros, bem como o vetor Y de respostas oriundas do LC-MS, foram separadas em dois conjuntos, a saber, 70% para calibração e os 30% restantes, para a validação.

Para a etapa de validação interna do conjunto de calibração, foi escolhido o número ótimo de variáveis latentes, em que ocorria a minimização dos erros de validação cruzada, RMSECV, pelo método iterativo *leave one out*, conforme discutido no item 1.5. Os gráficos de otimização das LVs, tanto para PLS-Raman, quanto para PLS-NIR, são mostrados na Figura 49.

Figura 49: Otimização do número de variáveis latentes para construção de modelo de calibração, com dados de Raman (a) e NIR (b), para o analito Aspartame

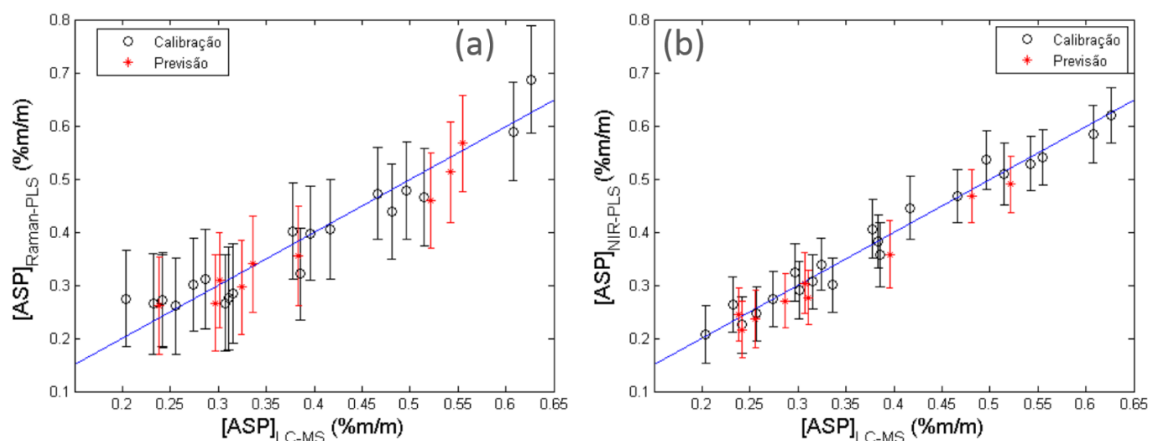


Fonte: ELABORADA PELO PRÓPRIO AUTOR

Para o modelo com Raman o número de LVs foi otimizado para 3 e para o de NIR, 4 LVs. Entretanto, pode-se perceber que o comportamento do gráfico para modelagem com NIR, (b), apresenta um melhor comportamento, visto que o RMSECV fica praticamente constante para um maior número de LVs, além de ser mais baixo que o apresentado pelo modelo de Raman, o que provavelmente levará o modelo de NIR a apresentar melhor desempenho, no que diz respeito a robustez e a capacidade preditiva.

Otimizados esses valores, ocorreu a validação por meio do conjunto que não participou da etapa de calibração. Os gráficos de correlação entre os valores medidos pelo método de referência e os previstos pela modelagem, para ambas as técnicas espectroscópicas, são mostrados na Figura 50.

Figura 50: Correlação entre os valores de referência e os previstos pelos modelos PLS, com dados de Raman (a) e NIR, para o analito ASP

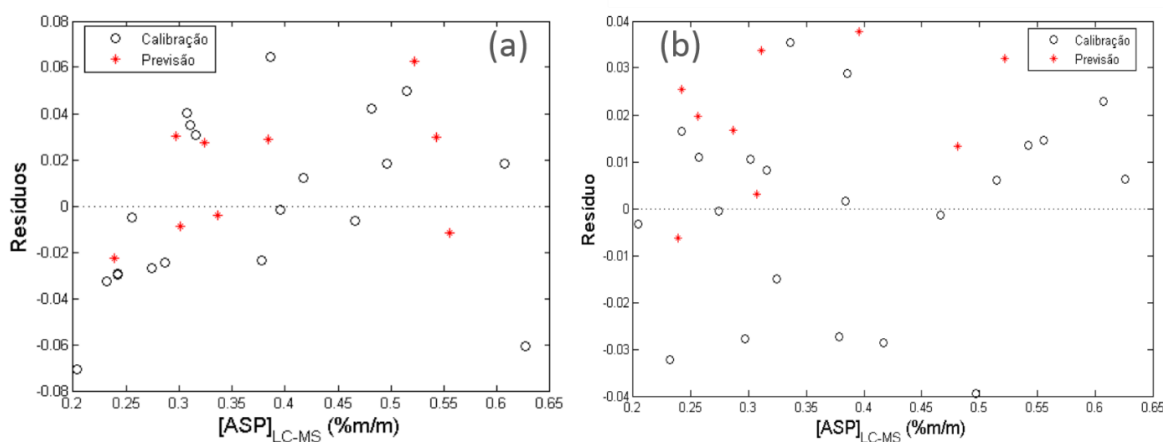


Fonte: ELABORADA PELO PRÓPRIO AUTOR

Ambos os modelos apresentaram boas correlações entre os dados de previsão, sendo o R^2_p igual a 0,948 para o Raman e 0,982 para o NIR e uma maior diferença para a correlação da validação cruzada, sendo o R^2_{vc} iguais a 0,844 e 0,941 para Raman e NIR, respectivamente. A exatidão ficou principalmente a cargo da análise do RMSEP, sendo 0,03 para o Raman e 0,02 para o NIR.

Os resíduos deixados pelos modelos, Figura 51, também foram avaliados e submetidos aos testes para avaliação de erros sistemáticos e de tendências. Analisando Figura 51 (b), gráfico de resíduos para o modelo NIR-PLS, percebe-se que os resíduos de previsão, com exceção de um, se encontram na parte positiva, evidenciando que todos os valores previstos, por mais próximos que fossem, foram menores que os medidos, o que implica em um erro sistemático, consequentemente acusado no teste de bias ao nível de significância de 0,05 ($t_{cal} = 4,00 > t_{tab(\alpha=0,05,n-1)} = 2,31$). Já para o modelo Raman-PLS, não há evidências de erros sistemáticos nos resíduos ($t_{cal} = 1,59 < t_{tab(\alpha=0,05,n-1)} = 2,31$)

Figura 51: Gráfico dos resíduos dos modelos PLS, com dados de Raman (a) e NIR (b), para o analito ASP



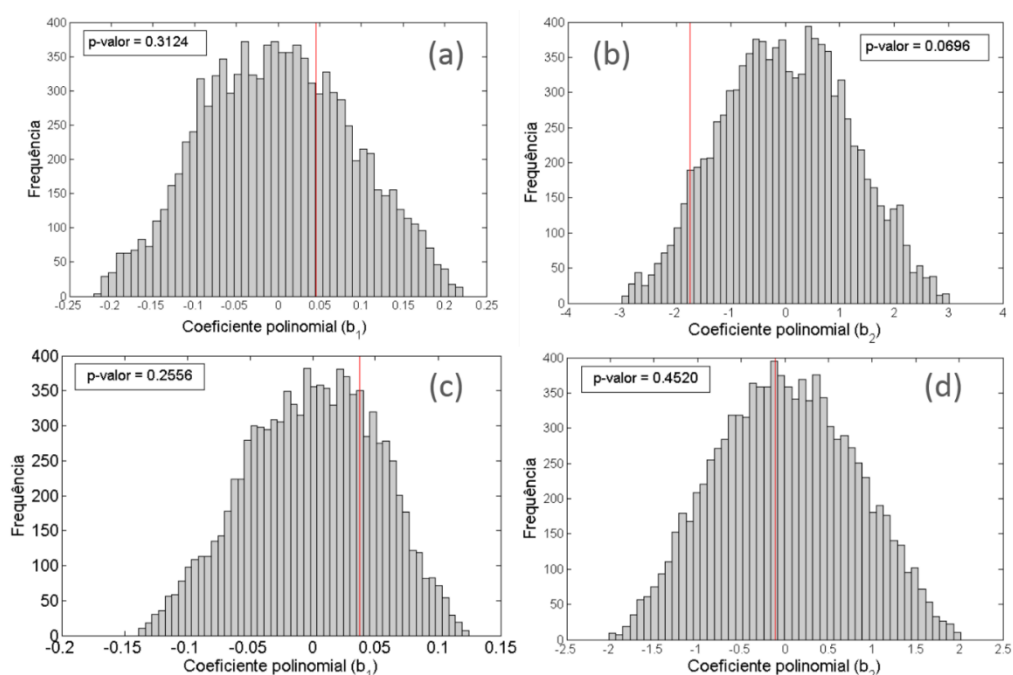
Fonte: ELABORADA PELO PRÓPRIO AUTOR

O teste de permutação não paramétrico também foi aplicado aos resíduos, por meio de funções lineares e quadráticas para ambos modelos, mostrando que não há evidência de tendência, visto que os p-valores foram superiores ao nível de significância de 0,05. A Figura 52 mostra os histogramas de distribuição para os coeficientes dos modelos.

Fazendo ainda uma análise da importância que o modelo deu às variáveis originais para a construção dos modelos, para ambos os casos pode-se perceber (Figura 53) que foram consideradas as principais regiões que trazem informação química relevante. É interessante

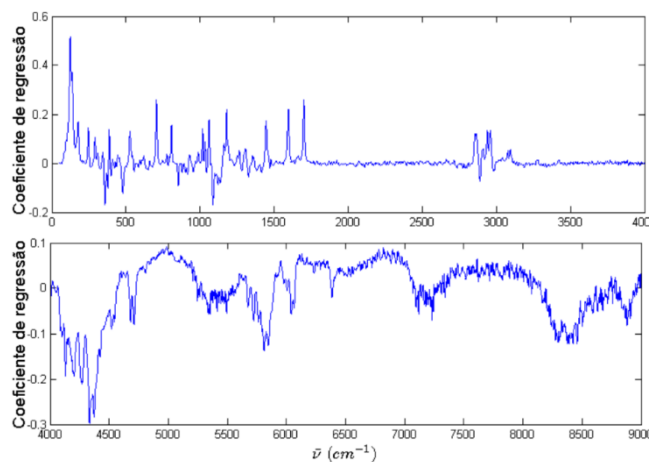
comentar o caso do modelo NIR-PLS, onde a Figura 49 (b), aponta como sendo 5 o número de LVs que ocorre a minimização do RMSECV, entretanto, analisando a Figura 53 (inferior) é possível perceber que a partir de 7000 cm^{-1} perfil do espectro começa a ficar ruidoso e caso utilizássemos 5 LVs, como era apontado, isso ainda seria agravado, ocorrendo o chamado sobre-ajuste ou “overfitting”. Um modelo sobre-ajustado pode até se apresentar com alta exatidão, porém, tal modelo não é uma boa representação da realidade, devendo ser evitado.

Figura 52: Histograma de permutação utilizando uma função linear, (a) e (c) e, uma quadrática, (b) e (d), para descrever os resíduos dos modelos construídos, para o analito ASP. Os gráficos superiores referem-se ao modelo de raman e os inferiores ao modelo de NIR



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 53: Coeficiente de regressão para os modelos de Raman-PLS (superior) com 3LVs e NIR-PLS (inferior) com 4LVs



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Por fim, duas tabelas serão apresentadas. A primeira, contém um resumo dos principais parâmetros de modelagem e a segunda contendo os valores medidos e previstos, dentro dos intervalos de confiança, para ambas as modelagens. Com esses resultados, pode-se afirmar que o ASP é modelável por ambas as metodologias, apresentando resultados com enorme correlação entre os valores medidos e previstos.

Tabela 14: Principais parâmetros de modelagem para o analito ASP

	RAMAN-PLS	NIR-PLS
R^2_{cv}	0.8441	0.941
R^2_p	0.9482	0.9819
RMSECV	0.05	0.03
RMSEP	0.03	0.02
LVs	3	4
E_{sist}	$t_{cal} = 1.59 < t_{tab} = 2.31$	$t_{cal} = 4.00 < t_{tab} = 2.08$
$E_{aleat}(b_1)$	p-valor = 0.31 > 0.05	p-valor = 0.25 > 0.05
$E_{aleat}(b_2)$	p-valor = 0.07 > 0.05	p-valor = 0.45 > 0.05

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Tabela 15: Valores medidos e preditos e intervalo de confiança para o analito ASP

Raman-PLS				NIR-PLS			
[ASP] _{med} (%m/m)	[ASP] _{prev} (%m/m)	IC		[ASP] _{med} (%m/m)	[ASP] _{prev} (%m/m)	IC	
0.24	0.26	0.17	0.35	0.24	0.22	0.16	0.27
0.30	0.31	0.22	0.40	0.26	0.24	0.18	0.29
0.34	0.34	0.25	0.43	0.24	0.24	0.19	0.30
0.32	0.30	0.21	0.38	0.29	0.27	0.22	0.32
0.30	0.27	0.18	0.36	0.31	0.28	0.23	0.33
0.38	0.36	0.26	0.45	0.31	0.30	0.25	0.36
0.52	0.46	0.37	0.55	0.40	0.36	0.29	0.42
0.54	0.51	0.42	0.61	0.52	0.49	0.44	0.54
0.56	0.57	0.48	0.66	0.48	0.47	0.42	0.52

Fonte: ELABORADA PELO PRÓPRIO AUTOR

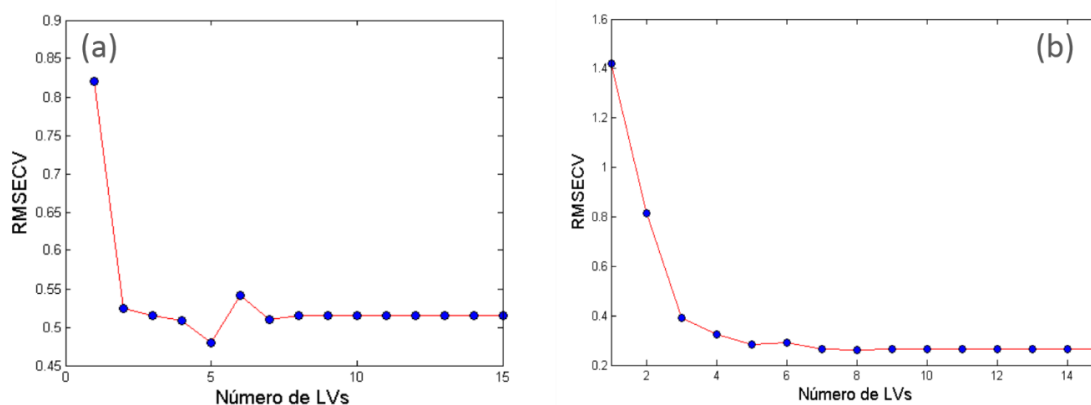
3.4.5.2 CICLAMATO

Assim como feito para o ASP, para os demais analitos também foram construídos dois modelos PLS, um para Raman e outro para NIR. O modelo de PLS-Raman foi construído a partir de 29 amostras e o modelo de PLS-NIR com 30 amostras. Para ambos os casos, não houveram pré-tratamento nos espectros. Sendo assim, por meio do algoritmo Kennard-Stone, a

matriz X de espectros, bem como o vetor Y de respostas oriundas do LC-MS, foram separadas em dois conjuntos, a saber, 70% para calibração e os 30% restantes, para a validação.

Para a etapa de validação interna do conjunto de calibração, foi escolhido o número ótimo de variáveis latentes, em que ocorria a minimização dos erros de validação cruzada, RMSECV, pelo método iterativo *leave one out*, conforme discutido no item 1.5. Os gráficos de otimização das LVs, tanto para PLS-Raman, quanto para PLS-NIR, são mostrados na figura abaixo:

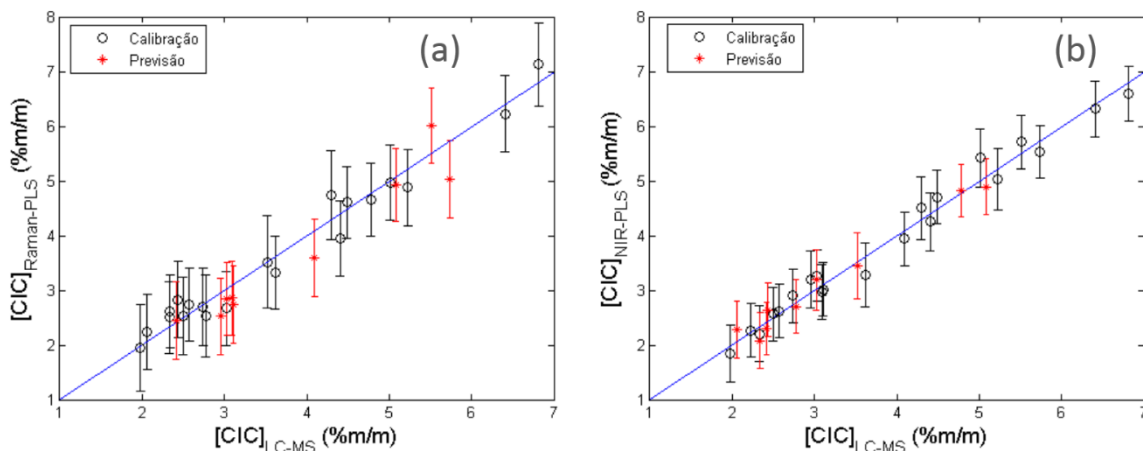
Figura 54: Otimização do número de variáveis latentes para construção de modelo de calibração, com dados de Raman (a) e NIR (b), para o analito Ciclamato



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Para o modelo com Raman o número de LVs foi otimizado para 5 e para o de NIR, 4 LVs. Assim como para as modelagens anteriores, o gráfico de otimização das LVs para NIR, (b), apresenta um melhor comportamento em relação ao de Raman, (a), o que repercutirá no modelo final. Otimizados esses valores, passou-se para a etapa de validação com o conjunto não participante da calibração. Os gráficos de correlação entre os valores medidos pelo método de referência e os previstos pela modelagem, para ambas as técnicas espectroscópicas, são mostrados na Figura 55.

Figura 55: Correlação entre os valores de referência e os previstos pelos modelos PLS, com dados de Raman (a) e NIR, para o analito CIC

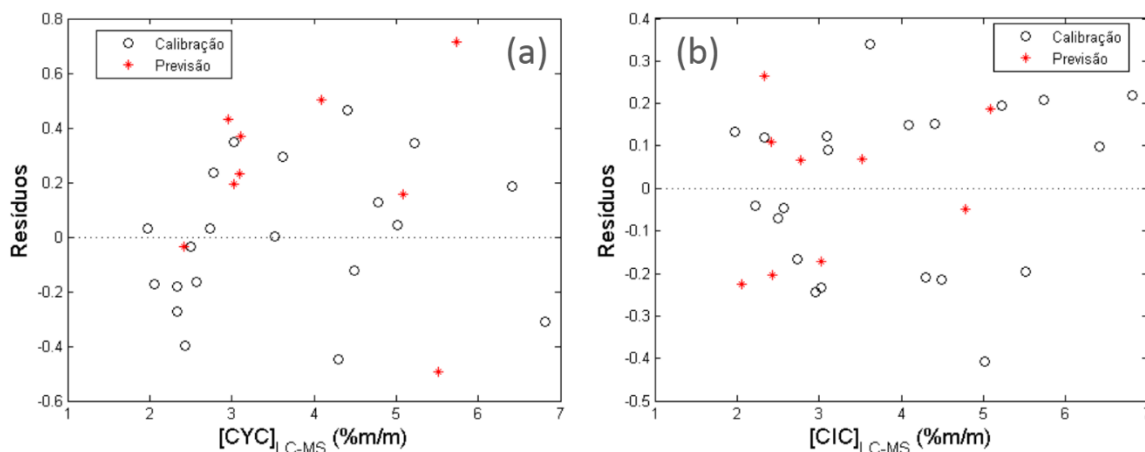


Fonte: ELABORADA PELO PRÓPRIO AUTOR

O modelo NIR-PLS apresentou melhores valores de R^2_p e R^2_{vc} quando comparado ao modelo de Raman-PLS, sendo estes 0,974 e 0,948 respectivamente, para NIR-PLS e 0,930 e 0,892, para Raman-PLS. A exatidão também foi mais significativa para o modelo NIR-PLS, apresentando valor de 0,17 para o RMSEP, enquanto no Raman-PLS, o valor foi de 0,40 para o mesmo parâmetro.

Os resíduos deixados pelos modelos, Figura 56, foram avaliados e submetidos aos testes para avaliação de erros sistemáticos e de tendências. Aplicando o teste de bias para ambos resíduos, pode-se afirmar que não há evidências de erros sistemáticos para nenhum deles ao nível de significância de 0,05, pois para os dois o valor do teste calculado foi menor que o crítico (Raman-PLS- $t_{cal} = 1,97 < t_{tab(\alpha=0,05,n-1)} = 2,31$ e NIR-PLS - $t_{cal} = 0,09 < t_{tab(\alpha=0,05,n-1)} = 2,31$).

Figura 56: Gráfico dos resíduos dos modelos PLS, com dados de Raman (a) e NIR (b), para o analito CIC

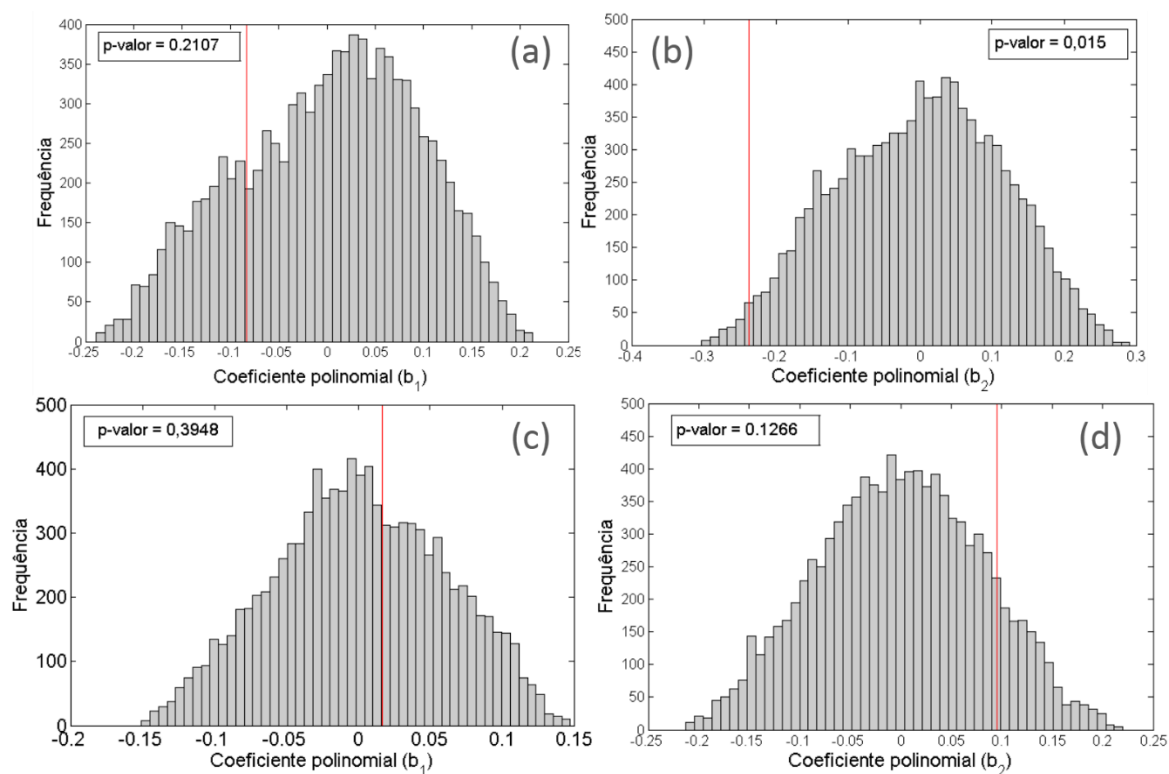


Aplicando o teste de permutação não paramétrico aos resíduos, por meio de funções lineares e quadráticas para ambos modelos, pode-se afirmar que não há evidência de tendências para o modelo NIR-PLS, pois o p-valor é superior ao nível de significância para os dois tipos de funções. Em contrapartida, para o modelo Raman-PLS, seus resíduos apresentam caráter quadrático em um número considerável de vezes, de acordo com a significância do teste, quando realizadas 10.000 permutações. Visto que os p-valores foram superiores ao nível de significância de 0,05. A Figura 57 mostra os histogramas de distribuição para os coeficientes dos modelos.

A análise de importância do coeficiente de regressão dos modelos pode ser vista pela Figura 58, onde se pode perceber que para o modelo NIR-PLS com 4 LVs, as principais regiões com informação química do sistema estudado foi levado em consideração, apresentando um aspecto ruidoso a partir de 7000 cm^{-1} como mostrado na modelagem para o ASP. Entretanto, para o modelo Raman-PLS (Figura 58 superior) um perfil muito ruidoso foi observado para esse gráfico, o que foi ocasionado pelo aumento do número de LVs para tentar descrever as variáveis originais, incorrendo na deterioração do modelo. Comparando esse gráfico com o obtido para o analito ASP, onde foram utilizadas 3 LVs para modelar, pode-se perceber o aumento considerável de ruído. Porém, esse foi o número de LVs ótimos para que se conseguisse o melhor modelo para Raman-PLS, no que diz respeito à correlação e exatidão.

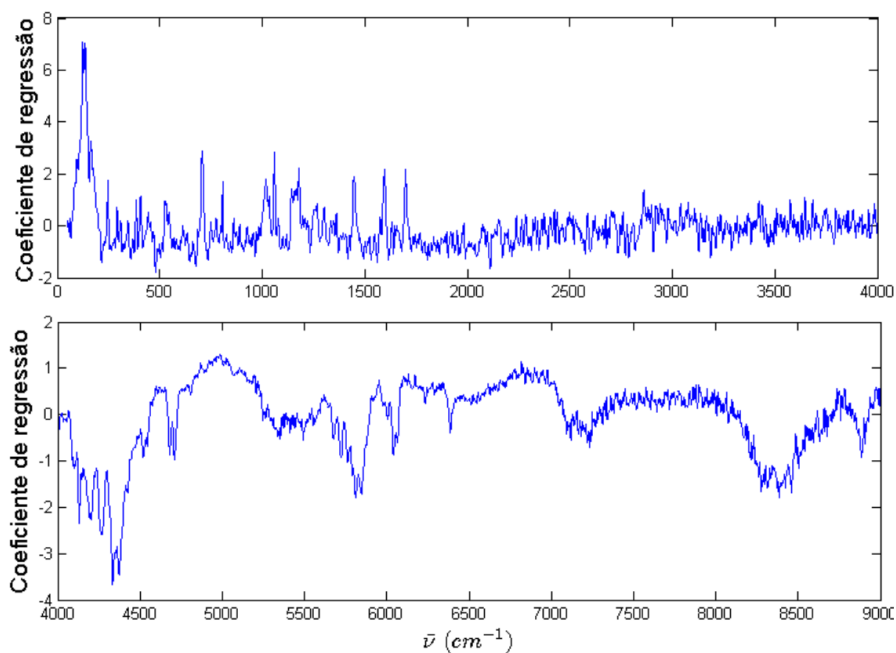
A Tabela 16 traz um resumo dos principais parâmetros de modelagem e a subsequente os valores medidos e previstos, dentro dos intervalos de confiança, para ambas as modelagens. Os dois modelos cumprem o papel de previsão para o analito CIC, sendo que o NIR se comporta um pouco melhor no que diz respeito à exatidão e correlação, além dos resíduos deixados pelo modelo não apresentarem tendência, porém, o modelo de Raman também prevê com grande exatidão a concentração desses analito, uma vez que os valores medidos e previstos são próximos, como mostrado na última tabela.

Figura 57: Histograma de permutação utilizando uma função linear, (a) e (c) e, uma quadrática, (b) e (d), para descrever os resíduos dos modelos construídos, para o analito CIC. Os gráficos superiores referem-se ao modelo de Raman e os inferiores ao modelo de NIR



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 58: Coeficiente de regressão para os modelos de Raman-PLS (superior) com 5LVs e NIR-PLS (inferior) com 4LVs



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Tabela 16: Principais parâmetro de modelagem para o analito CIC

	RAMAN	NIR
R^2_{cv}	0.8922	0.9477
R^2_p	0.9305	0.974
RMSECV	0.48	0.32
RMSEP	0.40	0.17
LVs	5	4
E_{sist}	$t_{cal} = 1.97 < t_{tab} = 2.31$	$t_{cal} = 0.09 < t_{tab} = 2.31$
$E_{aleat}(b_1)$	p-valor = 0.21 > 0.05	p-valor = 0.39 > 0.05
$E_{aleat}(b_2)$	p-valor = 0.01 < 0.05	p-valor = 0.12 > 0.05

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Tabela 17: Valores medidos e preditos e intervalo de confiança para o analito CIC

Raman-PLS			NIR-PLS			
[CIC] _{med} (%m/m)	[CIC] _{prev} (%m/m)	IC	[CIC] _{med} (%m/m)	[CIC] _{prev} (%m/m)	IC	
2.42	2.45	1.75 3.15	2.34	2.08	1.56 2.59	
3.09	2.86	2.18 3.53	2.06	2.29	1.77 2.80	
3.11	2.74	2.04 3.44	2.42	2.31	1.82 2.79	
3.03	2.84	2.17 3.51	2.44	2.64	2.15 3.13	
2.95	2.52	1.83 3.21	2.78	2.71	2.22 3.19	
4.09	3.59	2.89 4.30	3.02	3.19	2.64 3.74	
5.08	4.93	4.26 5.60	3.52	3.45	2.84 4.06	
5.74	5.02	4.32 5.73	5.08	4.90	4.39 5.40	
5.51	6.01	5.32 6.69	4.78	4.83	4.35 5.30	

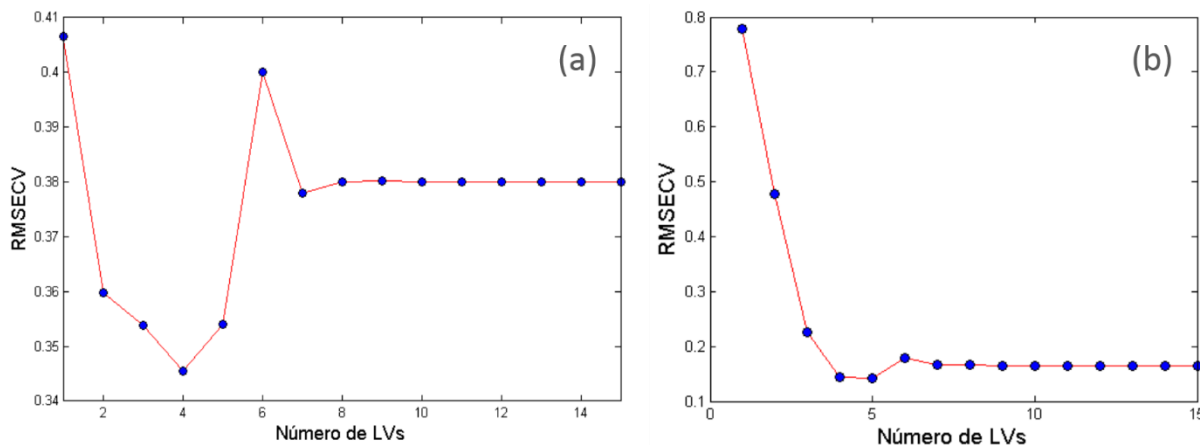
Fonte: ELABORADA PELO PRÓPRIO AUTOR

3.4.5.3 SACARINA

Para a SAC o modelo de PLS-Raman foi construído a partir de 29 amostras e o modelo de PLS-NIR com 30 amostras. Para ambos os casos, não houve pré-tratamento nos espectro. Os conjuntos de calibração (70%) e validação (30%) foram separados pela aplicação do algoritmo Kennard-Stone e assim formada a matriz X de espectros e o vetor Y de respostas.

A validação interna do conjunto de calibração, foi feita pelo método iterativo *leave one out* e então otimizado o número de LVs pela análise do gráfico de RMSECV vs LVs. Tais gráficos, para os modelos PLS-Raman e PLS-NIR, são mostrados na Figura 59.

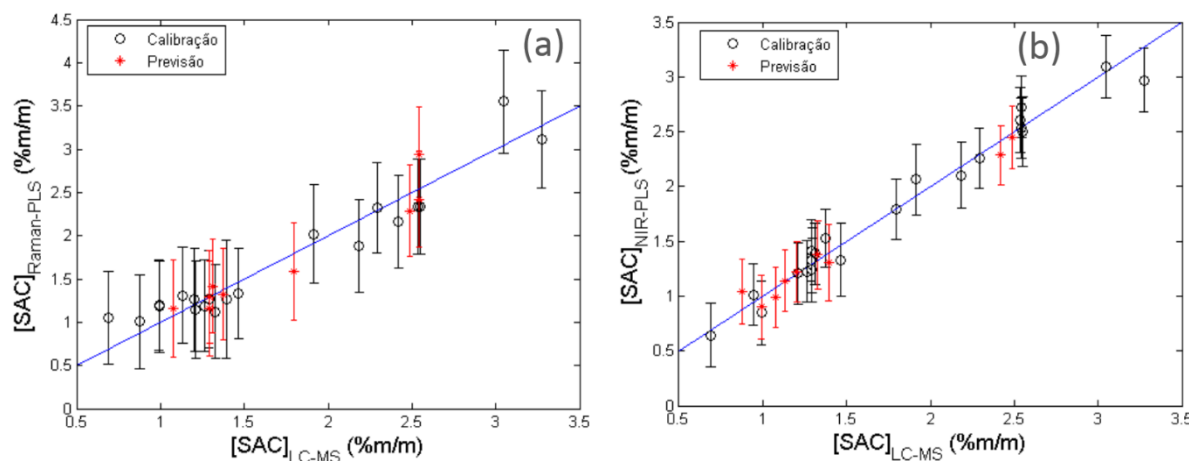
Figura 59: Otimização do número de variáveis latentes para construção de modelo de calibração, com dados de Raman (a) e NIR (b), para o analito Sacarina



Fonte: ELABORADA PELO PRÓPRIO AUTOR

O modelo com Raman teve o número de LVs otimizado para 5, enquanto o de NIR foi otimizado para 4. Assim como para as modelagens anteriores, o gráfico de otimização das LVs para NIR, (b), apresenta um melhor comportamento em relação ao de Raman, (a), o que repercutirá do modelo final. Otimizados esses valores, passou-se para a etapa de validação com o conjunto não participante da calibração. Os gráficos de correlação entre os valores medidos pelo método de referência e os previstos pela modelagem, para ambas as técnicas espectroscópicas, são mostrados a seguir:

Figura 60: Correlação entre os valores de referência e os previstos pelos modelos PLS, com dados de Raman (a) e NIR (b), para o analito SAC

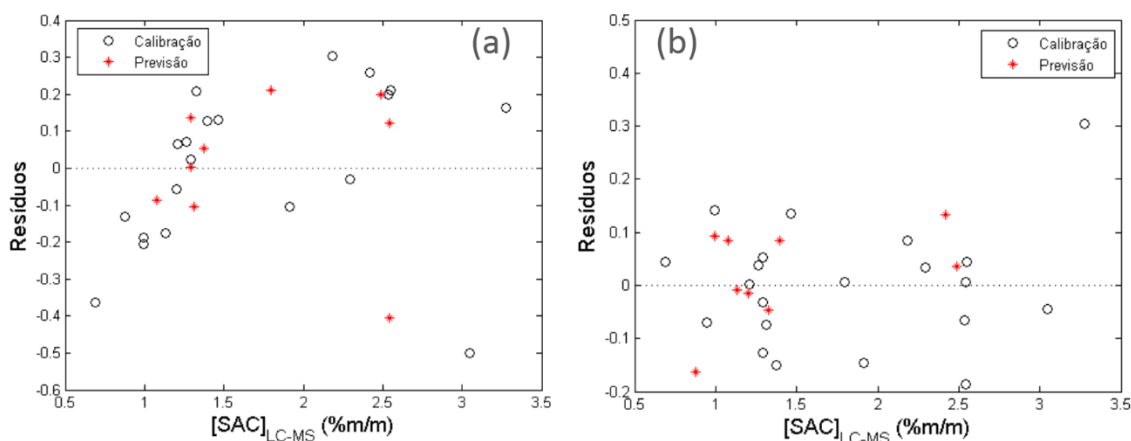


Fonte: ELABORADA PELO PRÓPRIO AUTOR

O modelo NIR-PLS apresentou melhor correlação para validação cruzada e previsão que o modelo Raman-PLS, além de melhor exatidão. Os valores de R^2_p e R^2_{vc} para o modelo de Raman foram 0,8097 e 0,9121, respectivamente, enquanto para o de NIR iguais a 0,9606 e 0,9788. A exatidão também foi mais significativa para o modelo NIR-PLS, apresentando valor de 0,09 para o RMSEP, enquanto no Raman-PLS, o valor foi de 0,18 para o mesmo parâmetro.

Avaliando os resíduos do modelo, Figura 61, pode-se afirmar que não há evidências de erros sistemáticos para ambos os modelos, ao nível de significância de 0,05, quando aplicado o teste de bias (Raman-PLS- $t_{cal} = 0,22 < t_{tab(\alpha=0,05,n-1)} = 2,31$ e NIR-PLS - $t_{cal} = 1,70 < t_{tab(\alpha=0,05,n-1)} = 2,31$).

Figura 61: Gráfico dos resíduos dos modelos PLS, com dados de Raman (a) e NIR (b), para SAC



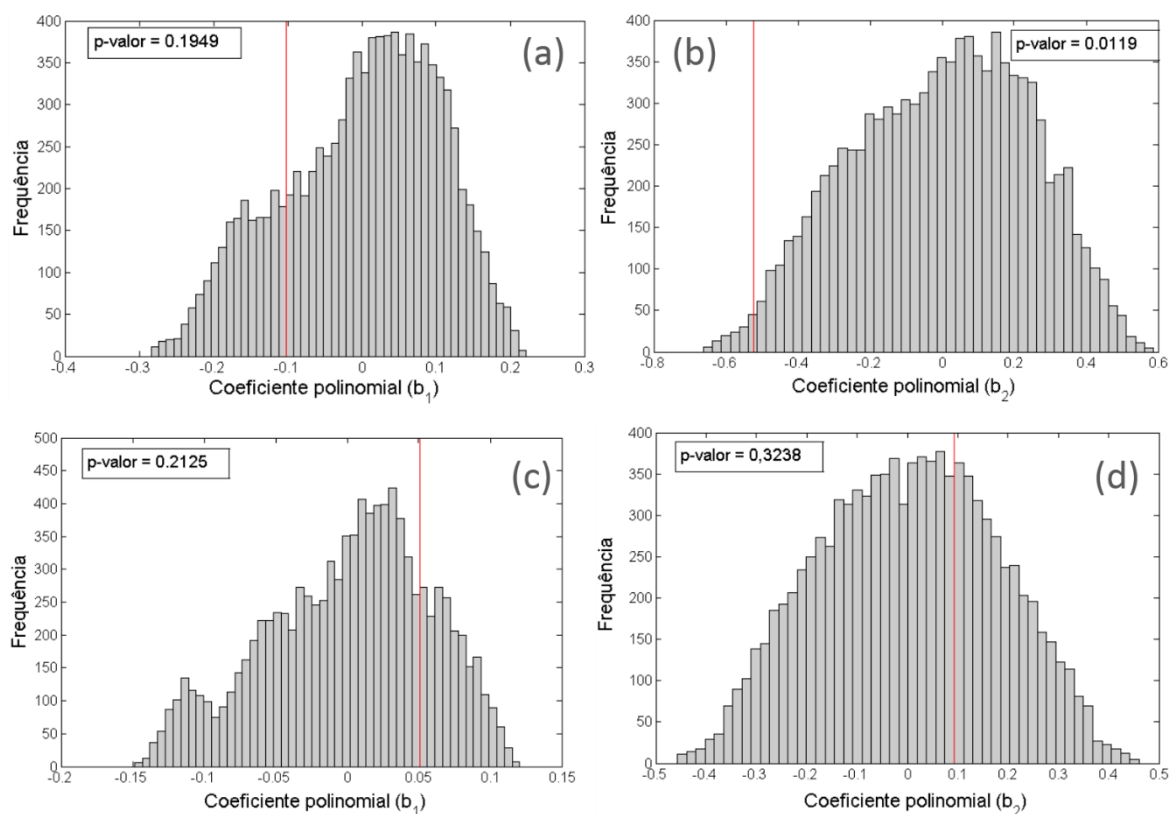
Fonte: ELABORADA PELO PRÓPRIO AUTOR

O teste de permutação indica que os resíduos para o modelo Raman-PLS se comportam com tendência quadrática, visto que o p-valor foi inferior a 0,05 (Figura 62 (b)). Analisando o gráfico de resíduos do modelo NIR-PLS, percebe-se que os mesmos se comportam aleatoriamente em torno de zero, o que é confirmado pelos testes de permutação para dois tipos de função, linear e quadrática. A Figura 62 mostra os histogramas de distribuição para os coeficientes dos modelos.

Os gráficos dos coeficientes de regressão para ambos os modelos, são mostrados na Figura 63. Para o modelo de NIR-PLS foram utilizadas o mesmo número de LVs e o perfil do gráfico é o mesmo para as modelagens anteriores. Já para o modelo de Raman-PLS, foi utilizado um número menor de LVs, quando comparado à modelagem anterior (Figura 58 superior) o que proporcionou uma diminuição do ruído fazendo com que a modelagem fosse mais fiel à realidade da matriz.

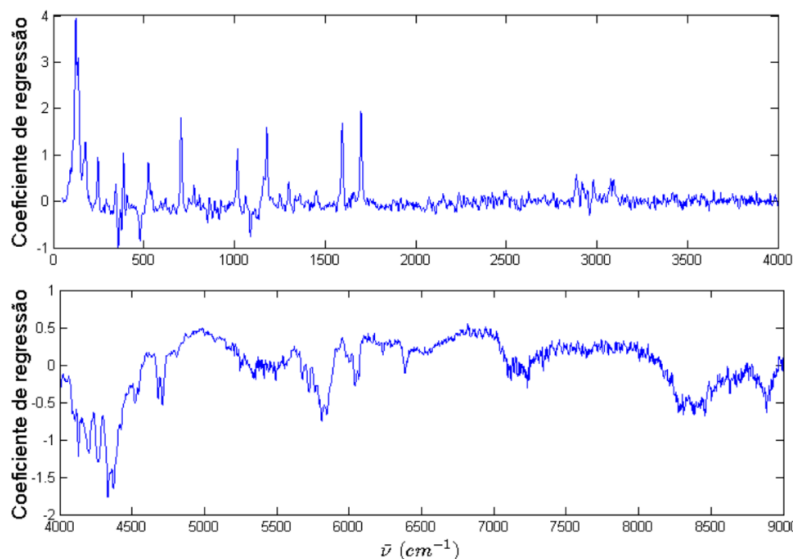
As tabelas 18 e 19 trazem os principais parâmetros de modelagem e a comparação os valores medidos e previstos, respectivamente, para ambas modelagens. Para a SAC, os dois modelos respondem bem, apresentando valores medidos e previstos bem próximo e dentro dos intervalos de confiança calculados, no entanto, o modelo de NIR-PLS se mostra com maior capacidade preditiva e mais robusto, no que diz respeito à análise dos principais parâmetros, além de ter sido aprovado em todos os testes para análise dos resíduos.

Figura 62: Histograma de permutação utilizando uma função linear, (a) e (c) e, uma quadrática, (b) e (d), para descrever os resíduos dos modelos construídos, para o analito SAC. Os gráficos superiores referem-se ao modelo de Raman e os inferiores ao modelo de NIR



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 63: Coeficiente de regressão para os modelos de Raman-PLS (superior) com 4LVs e NIR-PLS (inferior) com 4LVs



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Tabela 18: Principais parâmetros de modelagem para o analito SAC

	RAMAN	NIR
R^2_{cv}	0.8097	0.9606
R^2_p	0.9121	0.9788
RMSECV	0.345	0.14
RMSEP	0.183	0.09
LVs	4	4
E_{sist}	$t_{cal} = 0,22 < t_{tab} = 2,31$	$t_{cal} = 1,70 < t_{tab} = 2,31$
$E_{aleat}(b_1)$	p-valor = 0.19 > 0.05	p-valor = 0.33 > 0.05
$E_{aleat}(b_2)$	p-valor = 0.01 < 0.05	p-valor = 0.02 < 0.05

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Tabela 19: Valores medidos e preditos e intervalo de confiança para o analito SAC

Raman-PLS				NIR-PLS			
[SAC] _{med} (%m/m)	[SAC] _{prev} (%m/m)	IC		[SAC] _{med} (%m/m)	[SAC] _{prev} (%m/m)	IC	
1.08	1.16	0.60	1.72	0.99	0.90	0.61	1.19
1.29	1.29	0.76	1.82	0.88	1.04	0.75	1.34
1.31	1.42	0.87	1.96	1.08	0.99	0.72	1.27
1.38	1.32	0.79	1.85	1.13	1.14	0.86	1.42
1.29	1.16	0.61	1.70	1.20	1.22	0.94	1.50
1.79	1.58	1.02	2.15	1.33	1.38	1.06	1.69
2.49	2.29	1.76	2.82	1.39	1.31	0.96	1.65
2.54	2.42	1.86	2.98	2.49	2.45	2.16	2.74
2.54	2.95	2.40	3.49	2.42	2.29	2.02	2.56

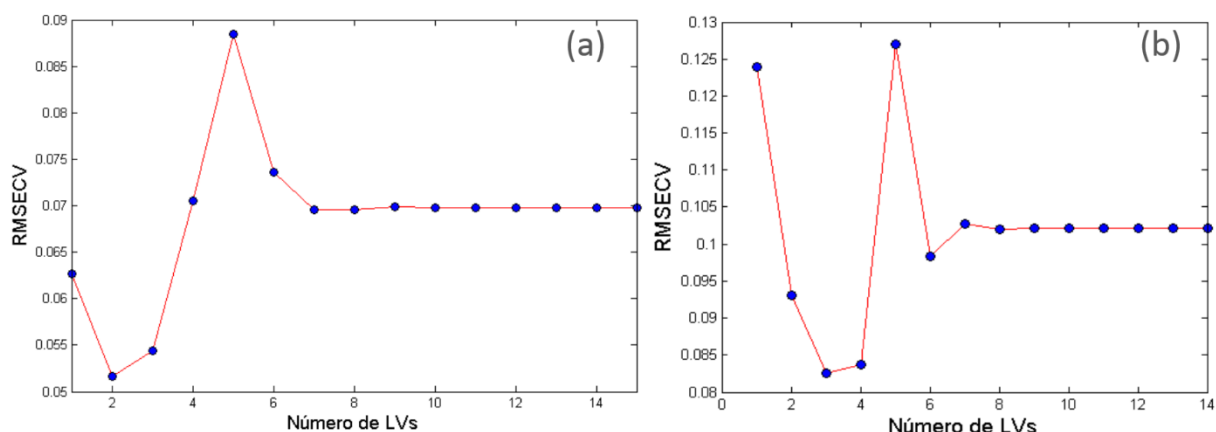
Fonte: ELABORADA PELO PRÓPRIO AUTOR

3.4.5.4 ACESULFAME-K

Para o ACSK o modelo de PLS-Raman foi construído a partir de 29 amostras e o modelo de PLS-NIR a partir de 27 amostras. Primeiramente o modelo de PLS-NIR foi construído como havia sendo feito para os analitos anteriores, porém, na etapa de separação do conjunto em amostras de calibração e validação, o algoritmo selecionou três amostras de validação antes do primeiro ponto de calibração, o que não faz sentido, pois estas amostras não seriam contempladas pela extensão da calibração; o gráfico que exemplifica esse fato esta em anexo. Então, essas três primeiras amostras foram retiradas e o modelo de NIR-PLS foi construído a partir de 27 e não 30. Aplicando novamente o algoritmo para separação do conjunto amostral, os dados de calibração e validação foram separados na proporção de 70 e 30%, respectivamente.

A otimização das LVs, feita pelo método iterativo *leave one out*, para os modelos de Raman-PLS e NIR-PLS, podem ser vistos por meio dos gráficos a seguir:

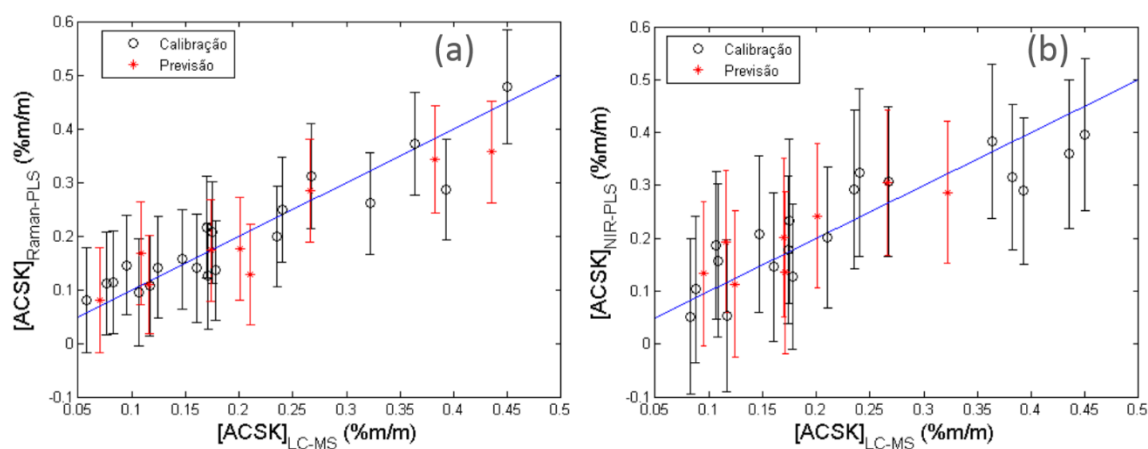
Figura 64: Otimização do número de variáveis latentes para construção de modelo de calibração, com dados de Raman (a) e NIR (b), para ACSK



Fonte: ELABORADA PELO PRÓPRIO AUTOR

O número de LVs foi otimizado para ambos os modelos, Raman-PLS e NIR-PLS, sendo 2 e 3 os valores, respectivamente. Diferentemente das modelagens para os outros analitos, o gráfico de LVs para o modelo NIR-PLS se comportou de maneira mais “confusa” e adquiriu a constância somente após a sétima LV, no entanto, o menor valor de RMSECV ocorreu com um número baixo de LVs. Otimizados esses valores, os modelos foram validados pelos conjuntos não participantes da calibração. Os gráficos de correlação entre os valores medidos pelo método de referência e os previstos pela modelagem, para ambas as técnicas espectroscópicas, são mostrados na Figura 65.

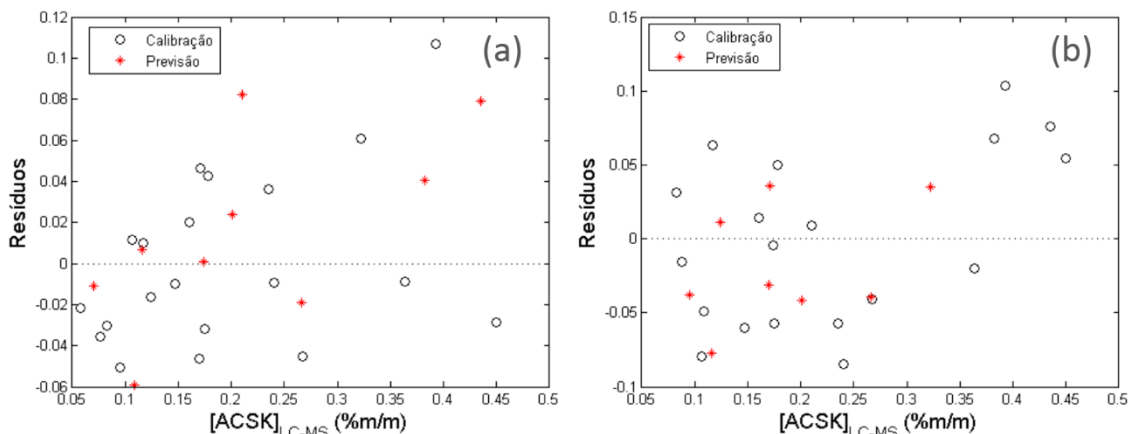
Figura 65: Correlação entre os valores de referência e os previstos pelos modelos PLS, com dados de Raman (a) e NIR (b), para o analito ACSK



Fonte: ELABORADA PELO PRÓPRIO AUTOR

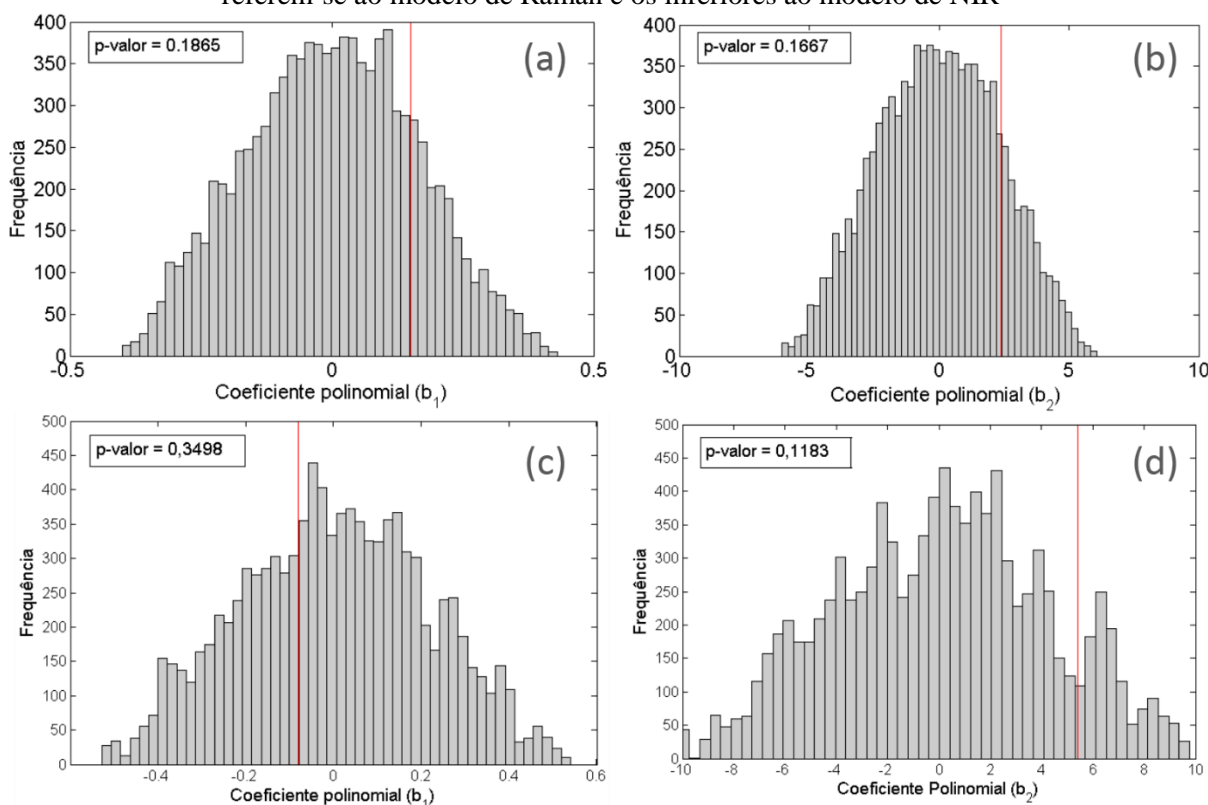
Para esse caso, o modelo Raman-PLS apresentou melhores correlações quando comparado ao modelo de NIR-PLS. Os valores de R^2_p e R^2_{vc} , para o modelo de Raman foram 0.878 e 0.781 respectivamente e para o modelo de NIR, 0.7285 e 0.5832. Entretanto, a exatidão para o modelo NIR-PLS foi maior que para o Raman-PLS, a saber, os valores de RMSEP foram 0,04 e 0,05, respectivamente. Os resíduos dos modelos foram avaliados e submetidos ao teste de bias. Ao nível de significância de 0,05, pode-se afirmar que não há evidências de erros sistemáticos para ambos modelos (Raman-PLS- $t_{cal} = 0,22 < t_{tab(\alpha=0,05,n-1)} = 2,31$ e NIR-PLS - $t_{cal} = 1,70 < t_{tab(\alpha=0,05,n-1)} = 2,31$). Os resíduos deixados pelos modelos, Figura 66, foram avaliados e submetidos aos testes para avaliação de erros sistemáticos e de tendências. Aplicando o teste de bias para ambos resíduos, pode-se afirmar que não há evidências de erros sistemáticos para nenhum deles ao nível de significância de 0,05, pois para os dois o valor do teste calculado foi menor que o crítico (Raman-PLS- $t_{cal} = 1,04 < t_{tab(\alpha=0,05,n-1)} = 2,31$ e NIR-PLS - $t_{cal} = 1,25 < t_{tab(\alpha=0,05,n-1)} = 2,31$). O teste de permutação indica que para ambos os modelos, os resíduos se distribuem aleatoriamente em torno de zero, visto que os p-valores para função linear e quadrática, são superiores ao nível de significância. A Figura 67 mostra os histogramas de distribuição para os coeficientes dos modelos.

Figura 66: Gráfico dos resíduos dos modelos PLS, com dados de Raman (a) e NIR (b), para o analito ACSK



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Figura 67: Histograma de permutação utilizando uma função linear, (a) e (c) e, uma quadrática, (b) e (d), para descrever os resíduos dos modelos construídos, para o analito ACSK. Os gráficos superiores referem-se ao modelo de Raman e os inferiores ao modelo de NIR



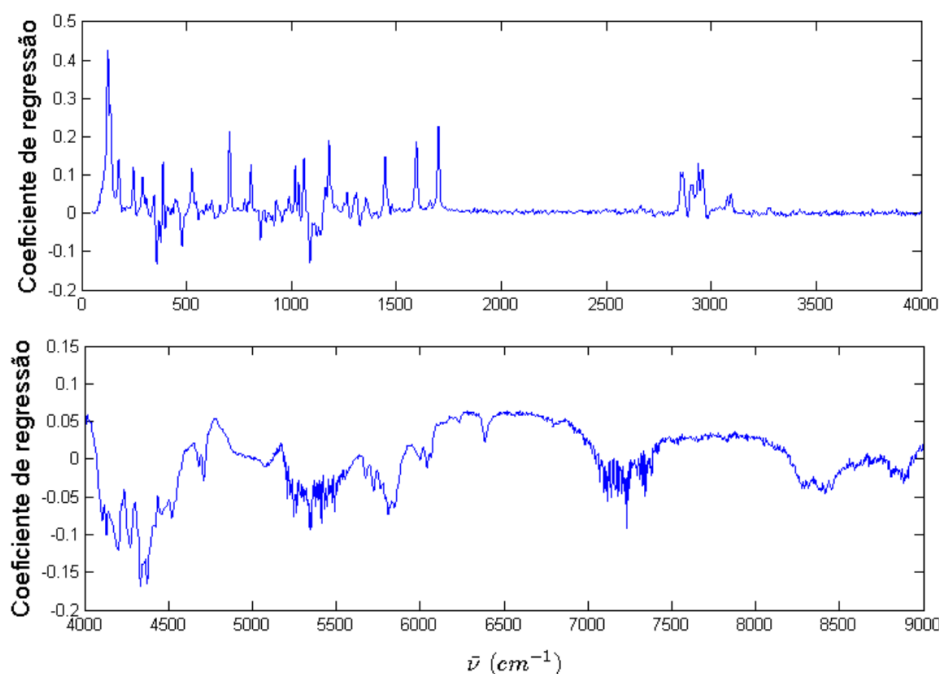
Fonte: ELABORADA PELO PRÓPRIO AUTOR

Os gráficos dos coeficientes de regressão para ambos os modelos, são mostrados na Figura 68. Para o modelo de NIR-PLS foram utilizadas uma LV a menos que para as outras modelagens, entretanto regiões importantes do espectro ficaram mais ruidosas, como por

exemplo entre 5000 e 6000 cm^{-1} , além da região entre 7000 e 7500 cm^{-1} . Para o modelo de Raman-PLS, como foi utilizado um número menor de LVs, pode-se perceber que as principais regiões do espectros ficam mais evidenciadas no gráfico do coeficiente de regressão.

As tabelas 20 e 21 mostram os principais parâmetros de modelagem e a comparação dos valores medidos e previstos, respectivamente, para ambas as modelagens. Para o ACSK, os dois modelos apresentam boa previsão, para as concentrações medidas pelo método de referência, entretanto, o modelo Raman-PLS se mostra mais robusto, uma vez que os principais parâmetros de modelagem foram melhores.

Figura 68: Coeficiente de regressão para os modelos de Raman-PLS (superior) com 2LVs e NIR-PLS (inferior) com 3LVs



Fonte: ELABORADA PELO PRÓPRIO AUTOR

Tabela 20: Principais parâmetros de modelagem para o analito ACSK

	RAMAN	NIR
R^2_{cv}	0.7813	0.5832
R^2_p	0.8776	0.7285
RMSECV	0.05	0.08
RMSEP	0.05	0.04
LVs	2	3
E_{sist}	$t_{cal} = 1.04 < t_{tab} = 2.31$	$t_{cal} = 1.25 < t_{tab} = 2.31$
$E_{aleat}(b_1)$	p-valor = 0.19 > 0.05	p-valor = 0.35 > 0.05
$E_{aleat}(b_2)$	p-valor = 0.17 > 0.05	p-valor = 0.11 > 0.05

Fonte: ELABORADA PELO PRÓPRIO AUTOR

Tabela 21: Valores medidos e preditos e intervalo de confiança para o analito ACSK

Raman-PLS				NIR-PLS			
[ACSK] _{med} (%m/m)	[ACSK] _{prev} (%m/m)	IC		[ACSK] _{med} (%m/m)	[ACSK] _{prev} (%m/m)	IC	
0.07	0.08	-0.02	0.18	0.12	0.11	-0.02	0.25
0.11	0.17	0.07	0.26	0.10	0.13	0.00	0.27
0.17	0.17	0.08	0.27	0.17	0.14	-0.02	0.29
0.21	0.13	0.04	0.22	0.12	0.19	0.06	0.33
0.12	0.11	0.02	0.20	0.17	0.20	0.05	0.35
0.20	0.18	0.08	0.27	0.20	0.24	0.11	0.38
0.27	0.28	0.19	0.38	0.27	0.31	0.17	0.44
0.38	0.34	0.24	0.44	0.32	0.29	0.15	0.42
0.44	0.36	0.26	0.45	-	-	-	-

Fonte: ELABORADA PELO PRÓPRIO AUTOR

3.5 CONCLUSÕES

O método desenvolvido por Fernandes e colaboradores é reproduzível, entretanto os cálculos realizados por normalização são comprometidos quando aplicado a matrizes contendo lactose. É interessante a revisão dessa metodologia, para ampliar a aplicação em amostras com matrizes distintas.

Visto que a quantificação dos analitos por CE-DAD gerou dúvidas, as amostras foram submetidas à análise por LC-MS, que é uma técnica amplamente difundida e já estabelecida para esse tipo de amostra. Logo, as respostas de concentração oriundas do LC-MS, para os quatro analitos, serviram como referência para a construção de modelos multivariados.

Foram construídos modelos multivariados por espectroscopias Raman e NIR, as quais se mostraram eficazes para inferências quantitativas, quando associadas à ferramenta PLS. Os modelos apresentaram boas correlações para os dados de previsão e alta exatidão. O modelo de NIR-PLS se mostrou mais eficaz para os analitos ASP, CIC e SAC, enquanto para o analito ACSK, a modelagem via Raman-PLS apresentou melhores resultados.

As metodologias desenvolvidas, por NIR e Raman, se apresentam como alternativas viáveis e de grande potencialidade para análises de aspartame, ciclamato, sacarina e acesulfame-K em formulações de adoçantes de mesa em pó. As principais vantagens dessas metodologias são: mínimo preparo de amostras, não-destrutivas, rápidas, de baixo custo e com nenhuma implicação ambiental, visto que não se utiliza solventes orgânicos, o que está de acordo com as premissas da química verde.

3.6 PERSPECTIVAS FUTURAS

- Pretende-se aplicar da metodologia multivariada desenvolvida a outras amostras de adoçantes sólidos em sachês que contenham a mesma matriz. A partir do espectro Raman e NIR, para essas amostras, espera-se estimar as concentrações dos quatro edulcorantes artificiais através dos respectivos modelos;
- Aplicação de algoritmo de seleção de variáveis para otimizar a aplicação do algoritmo de regressão, como por exemplo Algoritmo Genético (GA) e Redes Neurais (ANN);
- Testes com aplicação de método de regressão não-lineares como as Máquinas de Vetor de Suporte (SVR);
- Revisitação da metodologia desenvolvida por Fernandes e colaboradores;
- Pretende-se também a publicação desses resultados em revistas indexadas da área.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abbas, O. et al., 2012. **PLS regression on spectroscopic data for the prediction of crude oil quality: API gravity and aliphatic/aromatic ratio.** *Fuel*, 98, pp.5–14.
- Anh, T. & Magi, S., 2009. **Principal Component Analysis.** National Cheng Kung University.
- Anjos, A., 2005. **Curso Planejamento de Experimentos I** 1st ed. UFPR.
- ANVISA, 2014. **Regulamentação de aditivos alimentares e coadjuvantes de tecnologia no Brasil.** , p.<http://portal.anvisa.gov.br>
- ASTM, 2012. ASTM internacional, **ASTM Standard: Standard Practices for Infrared Multivariate Quantitative Analysis.** West Conshohocken, PA. Reaprovada em 2012
- ASTM, 2011. ASTM internacional, **ASTM Standard: Standard Test Method for Determination of Carcon Residue (Micro Method).** West Conshohocken, PA.
- Armenta, S., Garrigues, S. & de la Guardia, M., 2004a. **FTIR determination of Aspartame and Acesulfame-K in tabletop sweeteners.** *Journal of agricultural and food chemistry*, 52(26), pp.7798–803.
- Armenta, S., Garrigues, S. & de la Guardia, M., 2004b. **Sweeteners determination in tabletop formulations using FT-Raman spectrometry and chemometric analysis.** *Analytica Chimica Acta*, 521(2), pp.149–155. ASTM, E.-05, 2012. Standard Practices for Infrared Multivariate Quantitative Analysis 1. , pp.1–29.
- Bader, M., 1980. **A systematic approach to standard addition methods in instrumental analysis.** *J. Chem. Educ.*, 57(10), p.703.
- Barnes, R.J., Dhanoa, M.S. & Lister, S.J., 1989. **Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra.** *Appl. Spectrosc.*, 43, pp.772–777.
- Barros, R.R., 2008. **Consumo de alimentos industrializados e fatores associados em adultos e idosos residentes no Município de São Paulo.**
- Brereton, R.G., 2007. **Applied Chemometrics for Scientists.** John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England
- Brereton, R.G., 2003. **Data Analysis for the Laboratory and Chemical Plant** 1st ed. A. J. Wiley, ed., England.
- Brereton, R.G., 2000. **Introduction to multivariate calibration in analytical chemistry.** *The Analyst*, 125(11), pp.2125–2154.
- Bro, R., 2003. **Multivariate calibration.** *Analytica Chimica Acta*, 500(1-2), pp.185–194.
- Burns, D.A. & Ciurcsak, E.W., 2008. **Handbook of Near-Infrared Analysis** 3rd ed.

- Cantarelli, M. a et al., 2008. **Simultaneous determination of saccharin and aspartame in commercial noncaloric sweeteners using the PLS-2 multivariate calibration method and validation by capillary electrophoresis.** *Journal of agricultural and food chemistry*, 56(20), pp.9345–9.
- Cantarelli, M. a. et al., 2009. **Simultaneous determination of aspartame and acesulfame-K by molecular absorption spectrophotometry using multivariate calibration and validation by high performance liquid chromatography.** *Food Chemistry*, 115(3), pp.1128–1132.
- CODEX, 2013. **Codex General Standard for Food Additives.** *Codex Stan 192-1995*.
- Contreras, J., 2011. **A MODERNIDADE ALIMENTAR: ENTRE A SUPERABUNDÂNCIA E A INSEGURANÇA.** *História: Questões e Debates*, 54, pp.19–45.
- Dalmaschio, G.P., 2012. **Caracterização de Compostos Polares no Petróleo por Espectrometria de Massas de Altíssima Resolução e Exatidão – ESI(±)-FT-ICR MS.** Dissertação de Mestrado, UFES.
- Farah, M.A., 2012. **Petróleo e seus derivados.** Rio de Janeiro: LTC, 2012.
- Fatibello-filho, O. et al., 1996. **Adoçantes Artificiais.** *Química Nova*, 19(3).
- Fernandes, V.N.O. et al., 2013. **Simultaneous analysis of aspartame, cyclamate, saccharin and acesulfame-K by CZE under UV detection.** *Analytical Methods*, 5(6), p.1524.
- Ferreira, A. de A. & Neto, F.R. de A., 2005. **A DESTILAÇÃO SIMULADA NA INDÚSTRIA DO PETRÓLEO.** *Química Nova*, 28(3), pp.478–482.
- Ferreira, M.M.C. et al., 1999. **Quimiometria I: Calibração Multivariada, um tutorial.** *Química Nova*, 22(5), pp.1–8.
- Filgueiras, P.R., 2011. **Determinação da composição de Blends de Petróleos utilizando FTIR-ATR e Calibração Multivariada.** Dissertação de Mestrado, UFES
- Filgueiras, P.R. et al., 2014. **Evaluation of trends in residuals of multivariate calibration models by permutation test.** *Chemometrics and Intelligent Laboratory Systems*, 133, pp.33–41.
- Filgueiras, P.R., 2014. **REGRESSÃO POR VETORES DE SUPORTE APLICADO NA DETERMINAÇÃO DE PROPRIEDADES FÍSICO-QUÍMICAS DE PETRÓLEO E BIOCOMBUSTÍVEIS.** Tese de Doutorado, UNICAMP
- Geladi, P. & Kowalski, B.R., 1986. **Partial least-squares regression: a tutorial.** *Analytica Chimica Acta*, 185, pp.1–17.
- Gomes, E.A.D.S., 2009. **Estudo da Cristalização de Parafinas em Sistemas Solventes/Tensoativos/Água.** Tese de Doutorado, UFRN.

- He, Y.-C., Fang, S. & Xu, X.-J., 2012. **SIMULTANEOUS DETERMINATION OF ACESULFAME-K, ASPARTAME AND STEVIOSIDE IN SWEETENER BLENDS BY ULTRAVIOLET SPECTROSCOPY WITH VARIABLE SELECTION BY SIPLS ALGORITHM.** *Journal of Macedonian, Engineering Chemical*, 31(1), pp.17–28.
- Hotteling, H., 1933. **Analysis of a complex of statistical variables into principal components.** *Journal of Educational Psychology*, 24, pp.417–441 and 498–520.
- Isaksson, T. & Naes, T., 1988. **The Effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy.** , 42(7), pp.1273–1284.
- Kalivas, J.H., 2009. 3.01 **Calibration Methodologies.** In *Comprehensive Chemometrics, Chemical and Biochemical Data Analysis*. pp. 1–32.
- Khurana, H.K. et al., 2008. **Application of multibounce attenuated total reflectance fourier transform infrared spectroscopy and chemometrics for determination of aspartame in soft drinks.** *Journal of agricultural and food chemistry*, 56(3), pp.778–83.
- Lanças, F.M., 2008. **Avanços Recentes e Tendências Futuras das Técnicas de Separação : uma visão pessoal.** *Scientia Chromatographica*, 0(0), pp.17–44.
- Llamas, N.E. et al., 2008. **Direct Determination of Saccharin and Acesulfame-K in Sweeteners and Fruit Juices Powders.** *Food Analytical Methods*, 1(1), pp.43–48.
- Masili, A. et al., 2012. **Prediction of physical-chemical properties of crude oils by ¹H NMR analysis of neat samples and chemometrics.** *Magnetic resonance in chemistry : MRC*, 50(11), pp.729–38.
- Massart, D.L. et al., 1991. **Handbook of Chemometrics and Qualimetrics: Part A.**
- Nascimento, D.R. et al., 2014. **Determination of Carbon Microresidue at Mixtures of Diesel/Biodiesel.** *Revista Virtual de Química*, 6(4). Neto, B.D.B., Scarminio, I.S. & Bruns, R.E., 2006. Assuntos Gerais. *Química Nova*, 29(6), pp.1401–1406.
- Neto, B.D.B., Scarminio, I.S. & Bruns, R.E., 2010. **Como fazer experimentos 4^o ed.** Bookman, ed., Porto Alegre.
- De Noord, O.E., 1994. **The influence of data preprocessing on the robustness and parsimony of multivariate calibration models.** *Chemometrics and Intelligent Laboratory Systems*, 23(1), pp.65–70.
- Oliveira, L.F.C. De, 2001. **Espectroscopia Molecular. Química Nova na Escola, N^o4(Cadernos Temáticos), pp.1–7.**
- Oliveira, V.N., 2010. **Desenvolvimento e Otimização de Metodologia para Análise de Edulcorantes por Eletroforese Capilar.** Dissertação de Mestrado, UFJF

- Olivieri, A.C. et al., 2006. **Uncertainty estimation and figures of merit for multivariate calibration (IUPAC Technical Report)**. *Pure and Applied Chemistry*, 78(3), pp.633–661.
- Parisotto, G., 2007. **Determinação do número de acidez total em resíduo de destilação atmosférica e de vácuo do petróleo empregando a espectroscopia no infravermelho (ATR-FTIR) e calibração multivariada**. Departamento de química, UFSM.
- Pasquini, C., 2003. **Review Near Infrared Spectroscopy : Fundamentals , Practical Aspects and Analytical Applications**. *J. Braz. Chem. Soc*, 14(2), pp.198–219.
- Pearson, K., 1901. **On lines and planes of closest fit to systems of points in space**. F.R.S, *University College*, London, pp.559–572.
- De Peinder, P. et al., 2009. **Partial least squares modeling of combined infrared, ¹H NMR and ¹³C NMR spectra to predict long residue properties of crude oils**. *Vibrational Spectroscopy*, 51(2), pp.205–212.
- Pierini, G.D. et al., 2013. **Simultaneous determination of acesulfame-K and aspartame using linear sweep voltammetry and multivariate calibration**. *Microchemical Journal*, 106, pp.347–350.
- Raman, C. V & Krishnan, K., 1928. **A new class of spectra due to secondary radiation Part I**. *Indian J. Phys*, 2, pp.399–419.
- Ribeiro, D.B.M., 2009. **Desenvolvimento de metodologias analíticas multivariadas empregando espectroscopia Raman de baixa resolução amplificada por superfície**. Tese de doutorado, UNICAMP.
- Sala, O., 2008a. **Fundamentos da Espectroscopia Raman e no Infravermelho 2° ed**. UNESP, ed., São Paulo.
- Sala, O., 2008b. **I 2 - UMA MOLÉCULA DIDÁTICA**. *Química Nova*, 31(4), pp.914–920.
- Savitzky, A. & Golay, M.J.E., 1964. **Smoothing and differentiation by simplified least squares procedures**. *Anal. Chem.*, 36, pp.1627–1632.
- Savorani, F., Tomasi, G. & Engelsen, S.B., 2010. **icoshift: A versatile tool for the rapid alignment of 1D NMR spectra**. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 202(2), pp.190–202.
- Silverstein, R.M., Webster, F.X. & Kiemle, D.J., 2007. **Identificação Espectrométrica de Compostos Orgânicos 7th ed**. LTC, ed.
- Skoog et al., 2005. **Fundamentos de Química Analítica - tradução da 8° edição Norte-Americana**.
- Speight, A.J., 2002. **Handbook of Petroleum Product Analysis**, John Wiley & Sons, Inc., Hoboken, New Jersey.

- Tavares, M.F.M., 1996. **Eletroforese Capilar: Conceitos Básicos**. *Química Nova*, 19(2), pp.1–9.
- Tavares, M.F.M., 1997. **Métodos de separação em eletroforese Capilar**. *Química Nova*, 20(5), pp.493–511.
- Turak, F. & Ozgur, M.U., 2013. **Validated Spectrophotometric Methods for Simultaneous Determination of Food Colorants and Sweeteners**. *Journal of Chemistry*, 2013, pp.19.
- Valderrama, P., Braga, J.W.B. & Poppi, R., 2009. **Estado da arte de figuras de mérito em calibração multivariada**. *Quim. Nova*, , 32(5), pp.1278–1287.
- Wold, S., Sjostrom, M. & Eriksson, L., 2001. **PLS-regression : a basic tool of chemometrics**. *Chemometrics and Intelligent Laboratory Systems*, 58, pp.109–130.
- Zygler, A., Wasik, A. & Namieśnik, J., 2009. **Analytical methodologies for determination of artificial sweeteners in foodstuffs**. *Trends in Analytical Chemistry*, 28(9), pp.1082–1102.

APÊNDICES

A. INTERVALOS DE CONFIANÇA PARA OS MODELOS CONSTRUÍDOS

Tabela 22: Valores medidos e preditos e intervalo de confiança para as propriedades modeladas (API, TIAC e RC)

API				TIAC				RC			
Medido	Predito	Intervalo		Medido	Predito	Intervalo		Medido	Predito	Intervalo	
26.800	27.367	25.308	29.426	41.80	42.31	35.74	48.88	7.320	6.288	4.942	7.634
30.900	31.201	29.099	33.304	37.42	41.21	34.61	47.82	4.350	4.637	3.380	5.893
31.000	30.033	27.963	32.104	38.32	42.83	36.18	49.48	5.700	5.836	4.559	7.112
30.200	30.097	28.076	32.119	35.55	38.38	31.81	44.95	3.150	3.223	1.952	4.494
28.000	29.701	27.621	31.781	37.58	39.97	33.51	46.43	2.950	3.109	1.802	4.416
30.100	29.749	27.722	31.775	41.58	39.83	33.16	46.50	2.800	3.530	2.293	4.767
27.800	26.926	24.858	28.995	31.87	32.61	26.03	39.18	3.150	4.280	3.023	5.537
33.000	33.693	31.600	35.785	37.42	39.21	32.69	45.74	4.090	4.300	3.038	5.561
28.400	26.762	24.699	28.825	41.02	37.18	30.51	43.85	3.100	3.183	1.884	4.481
30.100	29.598	27.511	31.685	35.03	34.16	27.38	40.94	3.500	3.323	2.056	4.590
29.400	29.857	27.798	31.916	38.67	38.96	32.46	45.46	4.250	4.095	2.790	5.400
29.500	28.325	26.289	30.361	36.99	38.56	32.05	45.07	3.060	3.983	2.735	5.231
26.800	27.423	25.392	29.455	45.82	44.52	38.00	51.03	4.350	3.984	2.741	5.227
18.100	18.358	16.222	20.494	39.18	37.40	30.92	43.88	7.630	6.864	5.536	8.192
29.000	29.326	27.287	31.364	16.22	19.19	12.63	25.75	4.850	4.092	2.858	5.326
30.800	30.622	28.583	32.661	38.47	33.57	27.06	40.08	5.270	4.851	3.577	6.125
28.800	29.734	27.686	31.782	36.60	38.49	31.94	45.04	3.800	3.833	2.576	5.089
29.600	30.004	27.968	32.039	36.36	41.57	35.10	48.04	3.490	3.404	2.157	4.651
26.800	26.639	24.617	28.661	35.49	36.17	29.69	42.65	3.770	3.367	2.111	4.623
22.100	21.403	19.326	23.480	34.23	36.12	29.63	42.61	2.440	3.048	1.804	4.291
28.800	28.243	26.199	30.288	13.63	17.90	11.29	24.50	2.930	3.412	2.166	4.658
32.300	31.454	29.423	33.485	13.630	16.798	16.80	16.80	3.720	3.748	2.459	5.038
30.500	29.119	27.079	31.159	-	-	-	-	4.110	5.095	3.815	6.375
29.000	27.997	25.858	30.137	-	-	-	-	-	-	-	-
22.200	22.848	20.769	24.928	-	-	-	-	-	-	-	-

Fonte: ELABORADA PELO PRÓPRIO AUTOR

B. QUANTIFICAÇÃO DOS ANALITOS NAS AMOSTRAS POR CURVA EXTERNA E INTERNA

Tabela 23: Quantificação e comparação da concentração dos analitos nas 30 amostras de trabalho, por meio de curvas externas e internas com dois pontos

	[ANALITO] mg.L ⁻¹											
	ASP _{CE}	ASP _{CI}	Er	CIC _{CE}	CIC _{CI}	Er	SAC _{CE}	SAC _{CI}	Er	ACSK _{CE}	ACSK _{CI}	Er
A	119.83	80.55	48.76	1276.53	835.42	52.80	479.66	334.96	43.20	65.04	50.29	29.31
B	116.21	78.27	48.47	1243.23	908.12	36.90	498.32	292.36	70.45	70.05	51.88	35.02
C	123.69	82.42	50.07	1267.40	753.86	68.12	490.48	154.41	217.64	58.07	38.58	50.53
D	131.17	84.16	55.86	1312.51	920.18	42.64	524.46	274.55	91.02	60.02	43.77	37.13
E	131.17	89.88	45.95	1298.55	643.09	101.92	615.15	241.63	154.58	55.30	35.22	57.03
F	133.89	91.52	46.29	1414.55	848.53	66.71	570.40	309.43	84.34	68.26	44.74	52.58
G	140.47	95.31	47.39	1653.54	1205.04	37.22	657.40	352.33	86.58	99.18	67.87	46.12
H	165.41	113.87	45.27	1697.58	1268.90	33.78	669.42	469.51	42.58	85.33	62.27	37.04
I	139.11	93.59	48.64	1452.15	1117.88	29.90	654.48	574.74	13.87	82.56	59.43	38.91
J	154.30	109.79	40.54	1585.87	1239.89	27.90	636.38	466.09	36.54	61.59	47.30	30.23
K	169.04	116.30	45.34	1823.79	1399.13	30.35	733.17	309.08	137.21	96.56	67.34	43.38
L	143.87	90.00	59.86	1582.65	1255.65	26.04	591.31	564.75	4.70	65.19	48.06	35.64
M	195.12	142.50	36.93	2150.86	1668.22	28.93	861.88	612.93	40.62	103.45	76.90	34.52
N	178.34	130.04	37.14	1881.25	1532.44	22.76	699.54	351.69	98.91	88.10	65.04	35.44
O	175.39	149.37	17.42	1779.75	1975.33	9.90	739.87	507.07	45.91	102.77	76.92	33.60
P	167.91	122.82	36.71	1926.90	1753.78	9.87	746.77	548.49	36.15	87.80	63.12	39.10
Q	199.65	135.81	47.01	2025.18	1996.78	1.42	769.29	565.00	36.16	78.22	58.24	34.29
R	202.83	157.05	29.15	2294.25	2890.69	20.63	882.44	764.98	15.35	119.10	90.96	30.93
S	193.53	146.21	32.37	2482.22	2198.71	12.89	962.27	666.04	44.48	130.55	97.77	33.53
T	198.29	145.33	36.44	2470.94	1758.12	40.54	907.03	671.07	35.16	136.02	97.99	38.80
U	211.90	163.05	29.96	2766.32	3363.76	17.76	1068.78	979.81	9.08	134.82	96.09	40.30
V	210.76	147.52	42.87	2752.36	2901.90	5.15	1153.79	1286.99	10.35	159.38	123.02	29.56
X	239.11	191.53	24.84	3304.99	2865.36	15.34	1366.78	1346.97	1.47	207.97	155.94	33.36
Z	200.33	156.65	27.89	2550.43	1934.97	31.81	1085.13	901.04	20.43	158.25	116.40	35.96
AB	238.43	169.88	40.35	3381.25	5064.00	33.23	1279.72	1772.35	27.80	194.12	150.78	28.74
AC	275.84	197.47	39.69	4093.39	5056.05	19.04	1583.07	2219.83	28.69	206.32	134.13	53.83
AD	308.72	222.28	38.89	4294.25	4311.94	0.41	1739.43	2034.35	14.50	285.09	215.58	32.24
AE	276.52	198.01	39.65	3895.22	3381.10	15.21	1449.89	1187.12	22.13	220.10	153.80	43.11
AF	291.72	217.61	34.06	4030.56	6335.20	36.38	1640.35	2629.68	37.62	209.02	152.80	36.79
AG	282.19	234.20	20.49	3953.22	4652.40	15.03	1562.18	2438.93	35.95	263.75	189.34	39.30

CE - Curva Externa | CI - Curva Interna | |Er| - módulo do erro absoluto

Fonte: ELABORADA PELO PRÓPRIO AUTOR

C. COMPARAÇÃO ENTRE CE E LC-MS

Tabela 24: Comparação entre CE e LC-MS para a concentração dos analitos nas 30 amostras de trabalho

	ANALITO %m/m											
	ASP			CIC			SAC			ACSK		
	ASP _{CE}	LC-MS	Er	CIC _{CE}	LC-MS	Er	SAC _{CE}	LC-MS	Er	ACSK _{CE}	LC-MS	Er
A	0.25	0.20	24.51	2.64	1.98	33.06	1.06	0.69	53.18	0.16	0.08	91.39
B	0.25	0.24	3.17	2.89	2.33	24.23	0.93	0.99	6.06	0.17	0.12	41.83
C	0.27	0.24	9.91	2.43	2.34	3.89	0.50	0.99	49.84	0.12	0.08	62.47
D	0.26	0.26	1.50	2.84	2.06	37.84	0.85	0.88	3.32	0.14	0.06	133.31
E	0.28	0.24	17.66	2.01	2.42	16.79	0.76	1.08	29.88	0.11	0.07	57.45
F	0.29	0.26	11.33	2.65	2.23	19.15	0.97	0.94	2.61	0.14	0.09	59.95
G	0.30	0.27	9.69	3.80	2.58	47.57	1.11	1.26	11.96	0.21	0.18	20.29
H	0.33	0.29	14.50	3.66	2.44	50.07	1.35	1.13	19.46	0.18	0.12	44.01
I	0.29	0.23	26.17	3.49	2.50	39.99	1.80	1.21	48.35	0.19	0.16	15.94
J	0.35	0.31	11.09	3.90	2.78	40.47	1.47	1.20	21.73	0.15	0.10	56.23
K	0.35	0.31	14.07	4.21	3.02	39.60	0.93	1.33	29.91	0.20	0.17	18.42
L	0.28	0.30	6.04	3.95	3.09	27.76	1.78	1.29	37.23	0.15	0.11	39.55
M	0.38	0.32	20.20	4.44	2.74	62.19	1.63	1.29	26.41	0.20	0.11	91.50
N	0.41	0.34	23.30	4.88	3.11	57.07	1.12	1.31	14.64	0.21	0.17	19.19
O	0.47	0.32	44.52	6.19	3.03	104.29	1.59	1.38	15.57	0.24	0.21	14.50
P	0.39	0.30	29.72	5.50	2.95	86.09	1.72	1.29	32.78	0.20	0.12	70.72
Q	0.42	0.40	6.21	6.18	3.52	75.79	1.75	1.39	25.68	0.18	0.17	6.26
R	0.47	0.39	21.91	8.65	3.62	139.31	2.29	1.47	56.22	0.27	0.15	85.51
S	0.46	0.42	9.94	6.89	4.29	60.59	2.09	1.92	8.95	0.31	0.17	75.33
T	0.46	0.38	18.70	5.51	4.09	34.67	2.10	1.79	17.21	0.31	0.20	53.13
U	0.52	0.38	37.77	10.75	4.41	143.94	3.13	2.19	43.23	0.31	0.24	30.42
V	0.45	0.52	13.06	8.93	5.08	75.68	3.96	2.49	59.26	0.38	0.27	42.35
X	0.55	0.48	13.77	8.19	4.78	71.33	3.85	2.42	59.01	0.45	0.32	38.50
Z	0.50	0.47	7.33	6.18	4.49	37.78	2.88	2.29	25.67	0.37	0.39	5.40
AB	0.54	0.50	8.30	16.03	5.01	219.90	5.61	2.54	121.11	0.48	0.27	78.84
AC	0.53	0.52	2.51	13.52	5.23	158.52	5.94	2.55	132.77	0.36	0.24	49.49
AD	0.65	0.63	4.08	12.64	6.81	85.56	5.97	3.05	95.73	0.63	0.45	40.49
AE	0.56	0.54	3.11	9.55	5.74	66.49	3.35	2.54	31.84	0.43	0.38	13.52
AF	0.68	0.61	11.26	19.67	6.41	206.83	8.17	3.27	149.49	0.47	0.36	30.54
AG	0.73	0.56	31.00	14.45	5.51	162.08	7.57	2.54	197.96	0.59	0.44	35.11

CE - Eletroforese Capilar | LC-MS - Cromatografia Líquida com detecção por Espectrometria de Massas

|Er| - módulo do erro absoluto

Fonte: ELABORADA PELO PRÓPRIO AUTOR

D. EVIDÊNCIA DE MÁ SEPARAÇÃO DO CONJUNTO AMOSTRAL FEITO PELO KENNARD-STONE

Essa figura tem o objetivo de mostrar que os três primeiros pontos em vermelho referem-se às amostras de validação, sendo que essas não se apresentam ao longo da extensão das amostras de calibração.

Figura 69: Correlação entre os valores de referência e os previstos pelos modelos PLS, com dados de NIR, para o analito ACSK. Evidenciando que três amostras de validação aparecem antes do primeiro ponto de calibração

