



**UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE QUÍMICA
DEPARTAMENTO DE QUÍMICA ANALÍTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

**APLICAÇÃO DA ESPECTROSCOPIA NIR, ASSOCIADA À TÉCNICAS
DE CALIBRAÇÃO MULTIVARIADA, NA DETERMINAÇÃO DA
COMPOSIÇÃO QUÍMICA DO SOLVENTE UTILIZADO NA
PRODUÇÃO DE ISOPRENO**

DENISE SOUSA BRANDÃO

Salvador, Bahia

2012

**UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE QUÍMICA
DEPARTAMENTO DE QUÍMICA ANALÍTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

**APLICAÇÃO DA ESPECTROSCOPIA NIR, ASSOCIADA À TÉCNICAS
DE CALIBRAÇÃO MULTIVARIADA, NA DETERMINAÇÃO DA
COMPOSIÇÃO QUÍMICA DO SOLVENTE UTILIZADO NA
PRODUÇÃO DE ISOPRENO**

DENISE SOUSA BRANDÃO

Dissertação apresentada ao Programa de Pós – Graduação, Instituto de Química, Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Mestre.

Orientador: Prof. Dr. Leonardo S. Gomes Teixeira
Co-orientador: Dr. Márcio das Virgens Rebouças

Salvador, Bahia

2012

Sistema de Bibliotecas/IQ - UFBA

Brandão, Denise Sousa.

Aplicação da espectroscopia Nir, associada à técnicas de calibração multivariada, na determinação da composição química do solvente utilizado na produção de isopreno / Denise Sousa Brandão. – 2012.

105 f. : il.

Orientador: Prof. Dr. Leonardo Sena Gomes Teixeira.

Co - orientador: Márcio das Virgens Rebouças.

Dissertação (mestrado) - Universidade Federal da Bahia, Instituto de Química, Salvador, 2012.

1. Espectroscopia de infravermelho. 2. Solventes orgânicos. 3. Isopreno.

I. Teixeira, Leonardo Sena Gomes. II. Universidade Federal da Bahia. Instituto de Química.

III. Título.

CDD – 543.57

CDU – 543.42:547.315.2

TERMO DE APROVAÇÃO

Prof. Dr. Leonardo Sena Gomes Teixeira

(Orientador - IQ/UFBA)

Dr. Márcio das Virgens Rebouças

(Co-Orientador – Braskem S.A.)

Prof^a. Dr^a. Maria Fernanda Pimentel

(DEQ/UFPE)

Prof^a. Dr^a. Lícia Passos Santos Cruz

(IQ– UFBA)

Homologada pelo Colegiado do Programa de Pós-Graduação em Química.

Em __/__/__.

“...Cada um de nós compõe a sua história.
Cada ser em si carrega o dom de ser capaz.
E ser feliz...”

Almir Sater

Dedico este trabalho à minha mãe
pelo amor incondicional e
por me fazer acreditar na concretização dos meus ideais.

À Márcio Rebouças pela confiança,
incentivo e apoio na realização deste trabalho.

AGRADECIMENTOS

À Força Superior, a qual chamo de Deus, por sempre guiar meus passos sobre caminhos de luz. Por me dar uma família valiosa e por colocar em meu convívio pessoas tão especiais.

Aos meus pais, Fátima e Fernando, pela educação sobre uma base sólida e por serem exemplo na construção dos meus valores e princípios.

Ao meu irmão, David, pela amizade e pelo carinho.

À todos os professores do Instituto de Química da UFBA, pela contribuição acadêmica e ensinamentos fundamentais para a minha formação.

Ao professor Leonardo Teixeira pela confiança, incentivo e colaborações.

À professora Maria Fernanda Pimentel pelas contribuições e importantes discussões científicas.

À Alex Trindade, Juliane Brito e Daniele Pereira pelo grande apoio na preparação das misturas, aquisição dos espectros e realização das análises cromatográficas.

À equipe da cromatografia da Braskem pelo agradável convívio no ambiente de trabalho, pela guarda das amostras e análises cromatográficas.

Aos amigos da graduação: Ana Cecília, Chico, Dani, Drika, Kaio e Viviane. Obrigada pelos incontáveis momentos de estudo, aprendizado e pelas alegrias vividas.

Às minhas queridas amigas (de infância): Cila, Ia, Karol, Mila e Nanda. Obrigada pela amizade sempre presente, pelas trocas e pelas reflexões compartilhadas.

Às amigas e colegas de trabalho: Consuelo, Daniela Domingos, Jaciara e Nélia pelo estímulo e pelos inúmeros momentos de muita descontração.

À Rogério Machado, meu amor, por me encorajar e apoiar. Pela compreensão, amor e carinho.

À Fernanda Seabra por ser a líder mais querida, amiga, humana, competente e talentosa.

À BRASKEM por ter acreditado em meu potencial e tornado possível a realização deste projeto.

SUMÁRIO

LISTA DE FIGURAS	iii
LISTA DE TABELAS	vii
LISTA DE ABREVIATURAS E SIGLAS	ix
RESUMO	ix
ABSTRACT	xi
1. INTRODUÇÃO	1
1.1. Objetivo.....	4
1.1.1. Objetivos específicos.....	4
2. FUNDAMENTAÇÃO TEÓRICA	5
2.1. A produção de isopreno.....	5
2.2. A espectroscopia na região do infravermelho.....	9
2.2.1. Histórico.....	9
2.2.2. Fundamentos da técnica NIR.....	11
2.2.3. Características espectrais na região do NIR [24,25].....	15
2.2.4. Técnicas quimimométricas.....	18
2.3. Calibração multivariada.....	26
2.3.1. Métodos de regressão multivariada.....	26
2.4. Planejamento experimental.....	28
2.4.1. Matriz Doehlert.....	29
2.5. A função desejabilidade e os experimentos com multirespostas [47].....	32
2.6. Aplicações da espectroscopia NIR na indústria petroquímica.....	34
3. EXPERIMENTAL	40
3.1. Amostras.....	40
3.2. Efeito da temperatura da amostra no espectro NIR.....	46
3.3. Reagentes utilizados.....	46

3.4.	Instrumentação e condições de operação	47
3.5.	Programas estatísticos.....	47
3.6.	Construção dos modelos: Calibração multivariada	48
3.6.1.	Detecção de amostras anômalas (<i>outliers</i>)	49
3.6.2.	Seleção dos conjuntos de amostras de calibração e validação.....	50
3.6.3.	Seleção das faixas espectrais	50
3.6.4.	Planejamento de experimentos empregando a matriz Doehlert.....	50
3.6.5.	Otimização das variáveis de resposta: Função desejabilidade	52
3.6.6.	Validação estatística externa dos modelos.....	53
4.	RESULTADOS E DISCUSSÃO	54
4.1.	Características espectrais do solvente na região NIR.....	54
4.2.	Efeito da temperatura da amostra no espectro NIR.....	59
4.3.	Construção dos modelos: Análise multivariada	61
4.3.1.	Avaliação do preparo das misturas na variabilidade do conjunto de dados	61
4.3.2.	Detecção de <i>outliers</i> no conjunto de dados.....	62
4.3.3.	Definição dos conjuntos de calibração e validação	63
4.3.4.	Seleção das faixas espectrais	66
4.3.5.	Resultados do planejamento experimental e otimização dos modelos de calibração multivariada	73
4.3.6.	Refinamento dos modelos	83
4.3.7.	Validação estatística dos modelos	91
5.	CONCLUSÕES	98
	REFERÊNCIAS BIBLIOGRÁFICAS	99

LISTA DE FIGURAS

Figura 1. Origem da corrente C5.....	5
Figura 2. Esquema simplificado da unidade de produção de isopreno.	7
Figura 3. Diagrama de energia potencial para os modelos harmônico (A) e anarmônico (B) de um sistema diatômico. d_e é a distância interatômica de equilíbrio.	13
Figura 4. Gráfico do modelo λ PLS: número de intervalos vs RMSECV.	24
Figura 5. Gráfico do modelo λ PLS: faixa escolhida para a construção do modelo PLS.	25
Figura 6. Representação geométrica do planejamento bidimensional Doehlert para duas variáveis: um hexágono regular.....	29
Figura 7. Representação geométrica do planejamento tridimensional Doehlert para três variáveis: a) cubooctaedro centrado sobre uma face triangular, b) cubooctaedro centrado sobre uma face quadrada. Fonte: Rebouças, M. V.; Determinação de Arsênio em nafta e condensado de petróleo por espectrometria de absorção atômica com atomização eletrotérmica utilizando lantânio como modificador, Universidade Federal da Bahia, 2004.	30
Figura 8. Três projeções planas possíveis para o planejamento Doehlert com três variáveis: a) sobre uma face triangular, b) sobre uma face quadrada, c) sobre um vértice. Fonte: Rebouças, M. V.; Determinação de Arsênio em nafta e condensado de petróleo por espectrometria de absorção atômica com atomização eletrotérmica utilizando lantânio como modificador, Universidade Federal da Bahia, 2004.	31
Figura 9. Gráficos da função desejabilidade.	33
Figura 10. Principais etapas desenvolvidas para a obtenção do melhor modelo de calibração.....	49
Figura 11: Espectro NIR de uma amostra típica do solvente na região entre 10000 e 3700 cm^{-1}	55

Figura 12. Espectro NIR da acetona pura na região entre 10000 e 3700 cm^{-1}	55
Figura 13. Espectro NIR da acetonitrila pura na região entre 10000 e 3700 cm^{-1}	56
Figura 14. Espectro NIR da água pura na região entre 10000 e 3700 cm^{-1}	56
Figura 15. Espectro NIR do n-propanol puro na região entre 10000 e 3700 cm^{-1}	57
Figura 16. Espectro NIR do i-propanol puro na região entre 10000 e 3700 cm^{-1}	57
Figura 17. Espectro NIR do t-butanol puro na região entre 10000 e 3700 cm^{-1}	58
Figura 18. Espectro NIR do álcool alílico puro na região entre 10000 e 3700 cm^{-1} . 58	
Figura 19. Espectro NIR do DCPD puro na região entre 10000 e 3700 cm^{-1}	59
Figura 20. Diferença entre os espectros NIR da amostra adquiridos nos extremos dos testes de temperatura (4,1 e 50,9 $^{\circ}\text{C}$) e a temperatura ambiente (20 $^{\circ}\text{C}$).	60
Figura 21. Gráfico dos dados de concentração para a propriedade acetona.....	62
Figura 22. Gráfico de <i>scores</i> PC2 vs PC1 com resultado do teste <i>Hotelling T²</i>	63
Figura 23. Gráfico de resíduo vs <i>leverage</i> para a propriedade água.	65
Figura 24. Gráfico do modelo <i>i</i> PLS para a propriedade acetona.	69
Figura 25. Gráfico do modelo <i>i</i> PLS para a propriedade acetonitrila.	69
Figura 26. Gráfico do modelo <i>i</i> PLS para a propriedade água.	70
Figura 27. Gráfico do modelo <i>i</i> PLS para a propriedade álcool alílico.....	70
Figura 28. Gráfico do modelo <i>i</i> PLS para a propriedade n-propanol.	71
Figura 29. Gráfico do modelo <i>i</i> PLS para a propriedade i-propanol.	71
Figura 30. Gráfico do modelo <i>i</i> PLS para a propriedade t-butanol.	72
Figura 31. Gráfico do modelo <i>i</i> PLS para a propriedade DCPD.	72
Figura 32. Gráfico do modelo <i>i</i> PLS para a propriedade outros.	73

Figura 33. Gráficos de desejabilidade para a propriedade acetona.	85
Figura 34. Gráficos de desejabilidade para a propriedade acetonitrila.	85
Figura 35. Gráficos de desejabilidade para a propriedade água.....	86
Figura 36. Gráficos de desejabilidade para a propriedade n-propanol.....	86
Figura 37. Gráficos de desejabilidade para a propriedade i-propanol.....	87
Figura 38. Gráficos de desejabilidade para a propriedade t-butanol.....	87
Figura 39. Gráficos de desejabilidade para a propriedade álcool alílico.	88
Figura 40. Gráficos de desejabilidade para a propriedade DCPD.....	88
Figura 41. Gráficos de desejabilidade para a propriedade outros.....	89
Figura 42. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade acetona.	93
Figura 43. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade acetonitrila.....	94
Figura 44. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade água.....	94
Figura 45. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade n-propanol.....	95
Figura 46. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade i-propanol.	95
Figura 47. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade t-butanol.	96
Figura 48. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade álcool alílico.	96

Figura 49. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade DCPD.....	97
Figura 50. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade Outros.	97

LISTA DE TABELAS

Tabela 1. Principais características espectrais dos grupos funcionais estudados neste trabalho na região NIR.	18
Tabela 2. Valores codificados das coordenadas planas do planejamento Doehlert para três variáveis. As colunas A, B e C correspondem às projeções da Figura 8 [52].	31
Tabela 3. Composição típica das amostras do solvente.	41
Tabela 4. Níveis experimentais para o planejamento das misturas.	41
Tabela 5. Planejamento experimental das misturas.	42
Tabela 6. Planejamento experimental das misturas incluindo o componente DCPD.	44
Tabela 7. Planejamento experimental das 25 misturas adicionais.	45
Tabela 8. Matriz Doehlert para as variáveis: comprimento de onda, pré-processamento e seleção de variáveis, com os valores reais e suas respectivas coordenadas.	51
Tabela 9. Códigos referentes aos pré-processamentos utilizados.	52
Tabela 10. Número de amostras do conjunto de calibração para cada propriedade.	65
Tabela 11. Faixas espectrais selecionadas a partir do <i>i</i> PLS para cada propriedade.	67
Tabela 12. Resultados do planejamento experimental da propriedade acetona (n=109).	74
Tabela 13. Resultados do planejamento experimental da propriedade acetonitrila (n=106).	75

Tabela 14. Resultados do planejamento experimental da propriedade água (n=105).	76
Tabela 15. Resultados do planejamento experimental da propriedade n-propanol (n=108).....	77
Tabela 16. Resultados do planejamento experimental da propriedade i-propanol (n=108).....	78
Tabela 17. Resultados do planejamento experimental da propriedade t-butanol (n=105).....	79
Tabela 18. Resultados do planejamento experimental da propriedade álcool alílico (n=107).....	80
Tabela 19. Resultados do planejamento experimental da propriedade DCPD (n=108).....	81
Tabela 20: Resultados do planejamento experimental da propriedade outros (n=107).....	82
Tabela 21. Condições experimentais para obtenção dos modelos otimizados.	89
Tabela 22. Figuras de mérito dos modelos de calibração.	90
Tabela 23. Resultados da validação estatística dos modelos PLS.	93

LISTA DE ABREVIATURAS E SIGLAS

DCPD: Diciclopentadieno1,3.

SPXY: Algoritmo SPXY (*Sample set Partitioning based on joint x-y distances*).

CG: Cromatografia Gasosa.

NIRS: Espectroscopia na região do infravermelho próximo (*Near Infrared Spectroscopy*).

PLS: Regressão dos Mínimos Quadrados Parciais (*Partial Least-Squares Regression*).

NIR: Infravermelho próximo (*Near Infrared*).

GAP: Gasolina de Pirólise.

GLP: Gás Liquefeito de Petróleo.

PGH: Gasolina de Pirólise Hidrogenada (*Pyrolysis Gasoline Hydrogenation*).

CPD: Ciclopentadieno.

KS: Algoritmo Kennard-Stone.

GA: Algoritmo Genético (*Genetic Algorithm*).

\bar{r} PLS: Mínimos Quadrados Parciais por Intervalos (*Interval Partial Least Square regression*)

SPA: Algoritmo das Projeções Sucessivas (*Sucessive Projections Algorithm*).

JK: Algoritmo Jack Knife.

RMSECV: Erro Médio Quadrático da Validação Cruzada (*root mean squared error of cross-validation*).

VL: Variável Latente.

PC: Componente Principal (*Principal Component*).

MLR: Regressão Linear Múltipla (*Multiple Linear Regression*).

PCR: Regressão em Componentes Principais (*Principal Component Regression*).

PCA: Análise de Componentes Principais (*Principal Componente Analysis*).

RMSEP: Erro Médio Quadrático de Previsão (*root mean square error prediction*).

MIR: Infravermelho Médio (Mid Infrared).

PONA: Parafinas, Olefinas, Naftênicos e Aromáticos.

RESUMO

A espectroscopia na região do infravermelho próximo (NIRS), associada a técnicas quimiométricas, foi utilizada no desenvolvimento de modelos de calibração multivariada para a predição da composição química de um solvente orgânico-aquoso utilizado na produção de isopreno. Os modelos foram construídos para a predição da concentração dos componentes presentes no solvente: acetona, acetonitrila, água, n-propanol, i-propanol, álcool alílico, t-butanol e DCPD (diciclopentadieno-1,3). Uma análise exploratória dos dados foi realizada para detecção de amostras anômalas, através da análise de componentes principais (PCA). Para seleção dos conjuntos de calibração e validação, o algoritmo SPXY (*Sample set Partitioning based on joint x-y distances*) foi utilizado. As condições experimentais para a construção dos modelos foram escolhidas através da realização de um planejamento experimental para três variáveis baseado na matriz Doehlert, em que foram estudadas 1) técnica de pré-processamento, 2) faixa de comprimento de onda, obtidas através do algoritmo de seleção de variáveis por regressão dos mínimos quadrados parciais por intervalo (iPLS), e 3) seleção de variáveis espectrais, com a utilização do algoritmo Jack Knife. A técnica de regressão por mínimos quadrados parciais (PLS) foi utilizada para a calibração multivariada. As variáveis de resposta, PC, RMSECV, r^2_{regr} , RMSEP, r^2_{valid} , *Slope Valid* e p_{value} , obtidas a partir dos modelos definidos no planejamento experimental, foram otimizadas simultaneamente, através da função de desejabilidade. Novos modelos de calibração foram construídos a partir dos valores ótimos encontrados para cada variável de entrada. Pelo menos 85 amostras de calibração foram usadas em cada modelo, sendo necessários entre 1 e 10 fatores para assegurar o melhor desempenho de cada modelo. Os coeficientes de determinação de validação cruzada (r^2_{regr}) foram maiores que 0,92 para todas as propriedades. Os valores para os RMSEPs variaram de 0,39 a 1,66 (% m/m). Portanto, o método desenvolvido mostrou-se viável para determinação da composição química do solvente e auxiliar no controle do processo da unidade de produção de isopreno.

Palavras-chave: infravermelho próximo, calibração multivariada, matriz Doehlert, solvente orgânico, isopreno, função desejabilidade.

ABSTRACT

Near-infrared spectroscopy (NIRS) combined with chemometrics techniques was applied in the development of multivariate calibration models to predict the chemical composition of an organic-aqueous solvent used in an isoprene production plant. The models were developed to predict the concentration of the components present in the solvent, namely acetone, acetonitrile, water, n-propanol, i-propanol, allyl alcohol, t-butanol and DCPD (1,3-dicyclopentadiene). Principal component analysis (PCA) was used for exploratory data analysis to search for outliers in the data set. The calibration and validation sets were selected by the SPXY (*Sample set Partitioning based on joint x-y distances*) algorithm. The most appropriate experimental conditions for modeling were chosen using an Doehlert experimental design with three input variables: 1) preprocessing technique; 2) wavenumber range, which was obtained using the algorithm based on the interval partial least squares regression (iPLS), and 3) selection of spectral variables with the Jackknife algorithm. Partial Least Squares Regression (PLS) was the technique used in the multivariate calibration. The response variables, such as the number of PLS factors, RMSECV, r^2_{regr} , RMSEP, r^2_{valid} , *Slope Valid* and p_{value} , were simultaneously assessed using a desirability function to find the optimum conditions for PLS modeling. From the optimum setting for each input variable, new multivariate models were developed. At least 85 samples were used in each model and 1 – 10 PLS factors were needed to ensure the best performance. The cross validation determination coefficient was higher than 0,92 for all models and the standard error of prediction (SEP) ranged from 0,39 to 1,66 (%w/w). Therefore, the feasibility of performing the chemical composition analysis required for process control of the isoprene unit solvent has been demonstrated.

Keywords: near infrared, multivariate calibration, Doehlert matrix, organic solvent, isoprene, desirability function.

1. INTRODUÇÃO

A indústria petroquímica é o setor mais expressivo e dinâmico da diversificada indústria química nacional, a qual é considerada como o maior segmento da indústria de transformação brasileira. A petroquímica brasileira Braskem S.A. é a maior produtora de resinas termoplásticas das Américas, sendo a primeira empresa petroquímica a integrar operações da primeira e segunda gerações da cadeia produtiva do plástico.

A nafta é a principal matéria-prima da cadeia produtiva da petroquímica e do plástico no Brasil, sendo obtida através do fracionamento do petróleo e transformada em diversos petroquímicos básicos, tais como eteno, propeno, para-xileno, benzeno, tolueno e isopreno (2-metil-butadieno-1,3).

O isopreno é um monômero de alta importância para a indústria de transformação, cujo mercado apresenta alta atratividade, com forte crescimento de demanda, mundialmente e no Brasil. O isopreno é largamente utilizado na fabricação da borracha poliisopreno, equivalente sintético da borracha natural que possui qualidade superior à mesma. Entretanto, a sua principal aplicação é o copolímero estireno-isopreno-estireno, sendo empregado na produção de adesivos para rótulos e aqueles por pressão como, por exemplo, utilizados em fraldas e *post-it*. É esperado que o crescimento do consumo deste copolímero se intensifique, impulsionado pela popularização de suas aplicações fora do mercado americano.

Existem diversos processos para a produção do isopreno, como através da reação de isobuteno e formaldeído, processo utilizado comercialmente na Rússia e no Japão. Outro processo é a desidrogenação do isopentano, também usado comercialmente na Rússia. Contudo, o processo mais atrativo e utilizado atualmente pela indústria mundial é a destilação fracionada extrativa, acompanhada de destilação fracionada simples. Nesse processo, são utilizadas como matérias-primas correntes concentradas em compostos C5 olefínicos vindo de unidades de cocraqueamento de nafta.

A única produtora de isopreno no Brasil é a Braskem, cujo processo produtivo consiste basicamente na destilação extrativa do isopreno de uma corrente C5, por intermédio de um solvente misto, contendo essencialmente acetonitrila e água. Entretanto, outros compostos orgânicos podem ser acrescentados à sua composição à medida que este é continuamente utilizado em circuito fechado no processo. Espécies como cetonas, alcoóis e hidrocarbonetos se acumulam gradativamente no solvente como resultado da sua extração da corrente C5 (carga da unidade), degradação do solvente e reações paralelas.

No solvente usado na extração de isopreno, a água tem um papel fundamental, pois é utilizada para melhorar a eficiência de separação do solvente, tornando-se um parâmetro importante a ser controlado. Acetona e álcool alílico, assim como terc-butanol, iso-propanol e n-propanol estão presentes na carga da unidade, em concentrações mínimas, e se acumulam no solvente durante a campanha da unidade. Um acúmulo excessivo destes compostos pode afetar a seletividade do solvente e, por conseguinte, a eficiência da destilação extrativa. Reações paralelas favorecem ainda a formação de DCPD (diclopentadieno-1,3), o qual compromete a campanha da unidade. Sendo assim, um controle rigoroso da qualidade e da composição do solvente é essencial para aumentar a eficiência da extração e garantir a sustentabilidade do processo.

A técnica de cromatografia tem exercido um papel fundamental no controle do processo na indústria química utilizando métodos analíticos baseados em cromatografia à gás (CG) *on-line* e *at-line*. Neste contexto, o controle de qualidade do solvente utilizado na unidade de produção de isopreno é geralmente projetado para ser realizado por analisadores *on-line* baseados em cromatografia à gás. Entretanto, a disponibilidade e confiabilidade analítica destes analisadores são comprometidas devido às características inerentes do processo e à qualidade das amostras, uma vez que a presença de particulados e outros sólidos suspensos nas amostras do solvente demandam limpezas frequentes no sistema de filtração de amostra. Portanto, com frequência faz-se necessária a utilização de cromatógrafos de bancada a fim de determinar a composição química do solvente. Porém, a presença de particulados na amostra também é um ponto de atenção, mesmo após o processo de filtração, já que colunas capilares são utilizadas neste processo de

caracterização. A presença de compostos pesados, resultantes da degradação do solvente, também pode contribuir para uma diminuição do desempenho do método cromatográfico.

Diversos estudos mostram que a espectroscopia na região do infravermelho próximo – NIRS (*Near Infrared Spectroscopy*) é eficiente para aplicações petroquímicas, fornecendo a base para a predição de propriedades químicas e físicas a partir de um único espectro de uma amostra (MACHO; LARRECHI, 2002; CHUNG, 2007). Por meio de poderosas técnicas de análise multivariada como a Regressão dos Mínimos Quadrados Parciais – PLS (*Partial Least-Squares Regression*), cada propriedade de interesse pode ser correlacionada com os espectros de absorção NIR, a partir de um modelo matemático. Muitos trabalhos utilizam a técnica NIR para a predição de propriedades em gasolinas (HONORATO *et al.*, 2008; KELLI; CALLIS, 1990; KELLY *et al.*, 1989; CÔCCO; YAMAMOTO; VON MEIEN, 2005; BOHÁCS; OVÁDI, 1998; CHOQUETTE; CHESLER; DUEWER, 1996), combustíveis (PARISI; NOGUEIRAS; PRIETO, 1990), nafta (REBOUÇAS; SANTOS; VIEIRA, 2007; LEE; CHUNG; KIM, 2006), petróleo (FALLA *et al.*, 2006; PASQUINI; BUENO, 2007), hidrocarbonetos (REBOUÇAS *et al.*, 2010; REBOUÇAS; BARROS NETO, 2001) e misturas de hidrocarbonetos preparadas em laboratório (CHUNG; LEE; KU, 1998).

Poucos artigos relatam o emprego da técnica na predição de compostos com grupos funcionais similares aos presentes no solvente da unidade de isopreno. A maioria dos trabalhos apresenta resultados para a determinação de alcoóis, particularmente o etanol presente em misturas de gasolina (FERNANDES *et al.*, 2008; MENDES *et al.*, 2003) e a determinação de água em diferentes matrizes (GARRIGUES; GALLIGNANI; DE LA GUARDIA, 1993). Nenhuma referência a respeito da aplicação da espectroscopia NIR na predição de cetonas e compostos nitrílicos foi encontrada na literatura.

Diante do exposto, no presente trabalho propõe-se o desenvolvimento de método analítico através da aplicação da espectroscopia NIR, associada às técnicas de calibração multivariada, para obtenção de modelos capazes de prever de forma satisfatória a composição química do solvente utilizado em uma unidade de

produção de isopreno. Com isso, pretende-se desenvolver uma alternativa à determinação por cromatografia, cujo método analítico requer longo tempo de análise para uma separação adequada e cuidado especial para remoção de particulados e resíduos presentes na amostra.

1.1. Objetivo

Desenvolver método analítico através da aplicação da espectroscopia NIR, associada às técnicas de calibração multivariada, para a predição da composição do solvente utilizado na unidade de produção de isopreno.

1.1.1. Objetivos específicos

- Construir modelos matemáticos utilizando a técnica de calibração multivariada a partir dos espectros NIR e dos valores de composição do solvente, obtidos por análises cromatográficas;
- Otimizar o processo de construção dos modelos através de planejamento experimental;
- Predizer os teores de acetona, acetonitrila, água, n-propanol, i-propanol, terc-butanol, álcool alílico e DCPD presentes no solvente a partir dos modelos matemáticos desenvolvidos.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. A produção de isopreno

A unidade de produção de isopreno tem como objetivo produzir isopreno para indústria de adesivos, borracha sintética e borracha termoplástica SIS, a partir de um corte C5 originado nas unidades de dehidrodessulfurização de gasolina de pirólise (GAP). A composição da corrente C5 é função da qualidade da nafta e da severidade dos fornos de pirólise. Na Figura 1 é mostrada a origem da corrente C5 que é carga da unidade de isopreno.

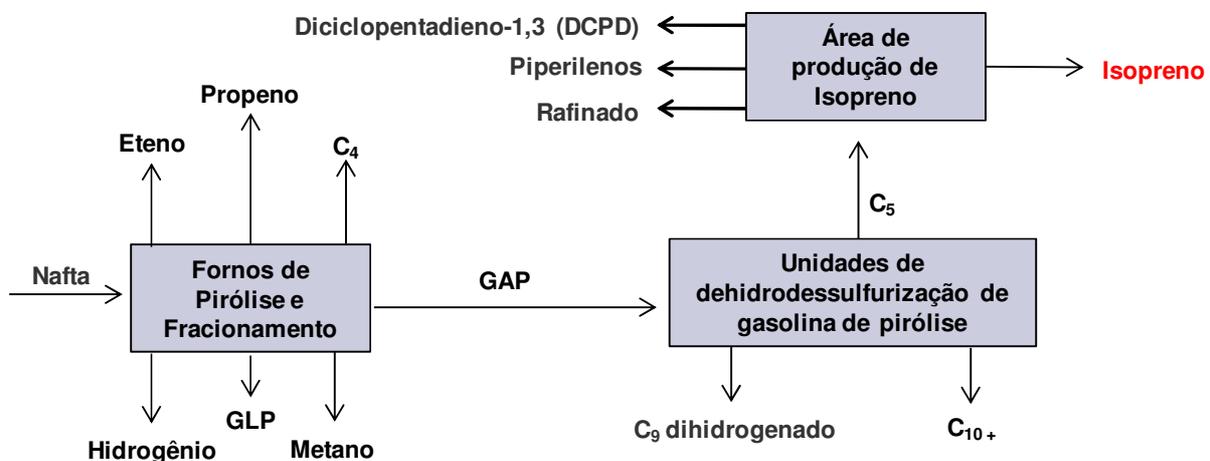


Figura 1. Origem da corrente C5.

A nafta é uma fração leve do petróleo enriquecida por hidrocarbonetos contendo de 2 a 12 átomos de carbono, principalmente parafinas (alcanos), isoparafinas (alcanos de cadeia ramificada) e naftênicos (cicloalcanos). Ela é submetida ao processo de craqueamento nos fornos de pirólise, o qual consiste na “quebra” das moléculas maiores, sob condições controladas, objetivando a produção de eteno e propeno, obtendo como co-produtos hidrogênio, metano, gás liquefeito de petróleo (GLP), correntes ricas em hidrocarbonetos contendo 4 átomos de carbono (corrente C4) e GAP.

GAP é uma corrente rica em hidrocarbonetos C5-C12 que alimenta a unidade de tratamento de gasolina. O objetivo desta unidade é a produção de uma corrente rica

em hidrocarbonetos contendo entre 6 a 8 átomos de carbono, chamada de gasolina de pirólise hidrogenada – PGH (*Pyrolysis Gasoline Hydrogenation*), destinada à unidade de extração de aromáticos. Como co-produtos, são separados as frações leves (corrente C5, matéria-prima para a produção de isopreno) e correntes pesadas (correntes C10+ e C9 dihidrogenado, adequadas à produção de combustíveis).

A separação do isopreno da corrente de C5 produzida nas unidades de pirólise é realizada a partir do processo de destilação extrativa, em que um solvente, a base de acetonitrila e água, é utilizado para mudar a volatilidade relativa dos componentes. O principal objetivo desta unidade é a obtenção de isopreno de alta pureza (99,5 %m/m), embora também sejam produzidas corrente de DCPD (com pureza 75 a 82 %m/m, podendo ser comercializada para a produção de resina C10), corrente de piperilenos (trans e cis-pentadieno-1,3, que pode ser comercializada para a produção de resina C5) e corrente de refinado C5 (rica em n-pentano e isopentano, a qual é encaminhada para co-craqueamento nos fornos de pirólise ou para a produção de gasolina automotiva).

Cinco sistemas principais compõem a unidade de produção de isopreno: 1) Preparação de Carga, 2) Extração, 3) Recuperação do Solvente, 4) Remoção de Enxofre e 5) Fracionamento. Um esquema simplificado da unidade de produção de isopreno está ilustrado na Figura 2.

O sistema de preparação de carga tem por finalidade: separar, da fração C5 da gasolina de pirólise, uma corrente rica em isopreno com teor mínimo de 23 %m/m, um teor máximo de 4,5 %m/m de ciclopentadieno (CPD) e especificar os co-produtos piperilenos e DCPD.

A seção de extração tem como objetivo separar o isopreno dos demais C5's existentes na corrente proveniente da seção de preparação de carga, por intermédio do processo de destilação utilizando um solvente. A destilação é o processo de separação mais utilizado industrialmente, sendo baseado na diferença de volatilidade dos compostos a serem separados. Quando a volatilidade relativa entre os compostos é muito baixa, como na separação por destilação do isopreno dos

demais compostos C5, adiciona-se um composto externo para modificar as volatilidades entre os componentes. Normalmente este composto adicionado é mais pesado que os demais da mistura, e devido à diferença entre as forças intermoleculares deste composto externo e os demais existentes na mistura, as forças de atração e repulsão entre todas as moléculas são alteradas, elevando a volatilidade relativa dos elementos que se deseja separar. Neste processo, o componente externo é chamado de solvente e quando este não forma duas fases líquidas com a mistura a ser separada, o processo passa a ser chamado de destilação extrativa.

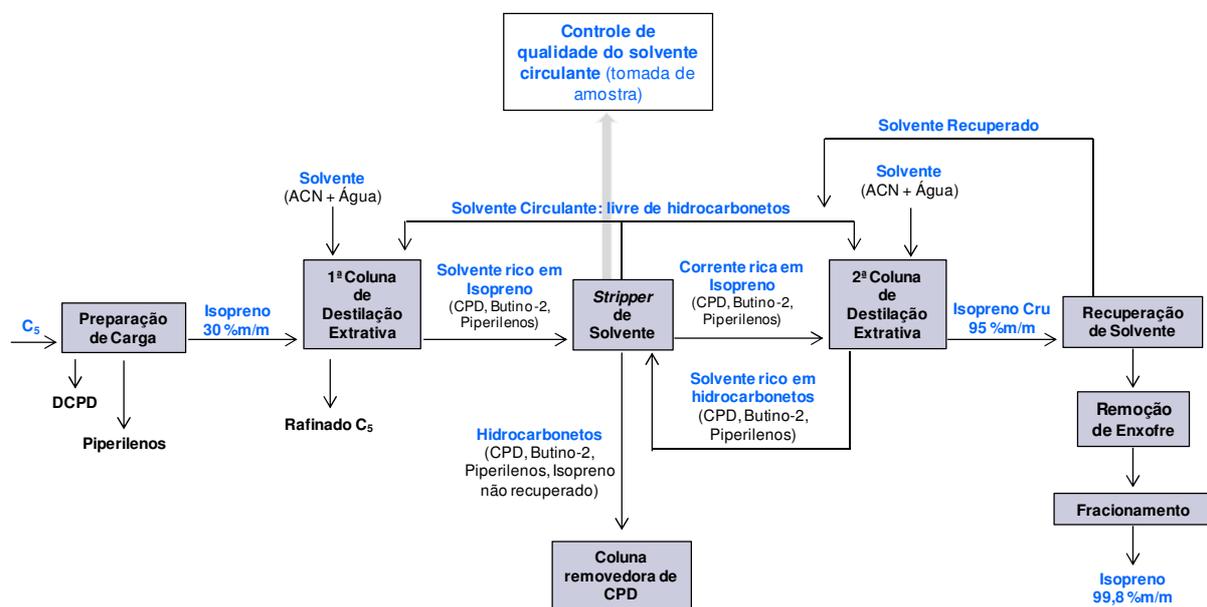


Figura 2. Esquema simplificado da unidade de produção de isopreno.

O solvente usado para promover a separação do isopreno é uma mistura de acetonitrila e água. Entretanto, outros componentes, como álcool alílico, acetona, terc-butanol, isopropanol estão presentes na carga da unidade, em concentrações mínimas, e se acumulam gradativamente no solvente durante a campanha da unidade, como resultado da sua extração da carga C5 e reações paralelas.

A seção extrativa é composta por duas colunas de destilação extrativa, uma seção chamada de “*stripper*” de solvente e uma coluna removedora de CPD. A primeira coluna gera como produto de topo um refinado C5, contendo compostos como

pentanos, pentenos e pentadieno-1,4. O solvente rico contendo isopreno e compostos como os dienos e acetilenos (CPD, piperilenos e butino-2) são obtidos no fundo e alimentam a coluna “*stripper*” de solvente. Esta coluna tem por objetivo separar o solvente dos demais compostos C5 presentes na mistura. A corrente do topo da coluna “*stripper*” de solvente, rica em isopreno, alimenta a segunda coluna de destilação extrativa, a qual gera como produto de topo o isopreno cru (95 %m/m) com um pouco de acetonitrila e água. A corrente rica em solvente, contendo CPD, butino-2, piperilenos e uma pequena parte do isopreno não recuperado, obtida no fundo da 2ª coluna de destilação extrativa, alimenta também a coluna “*stripper*” de solvente, onde todo o solvente usado nas duas colunas de destilação extrativa é recuperado pelo fundo e retorna como solvente circulante para alimentar as colunas de destilação extrativa. Nessa etapa, uma amostra do solvente é coletada e encaminhada para o laboratório para acompanhamento da composição do solvente circulante (Figura 2). Os hidrocarbonetos (CPD, acetilenos, piperilenos) e o próprio isopreno não recuperado são concentrados em um estágio intermediário desta coluna e transferidos para a coluna de remoção de CPD. Esta coluna separa e retira do sistema praticamente todo o CPD contido na corrente de carga da seção de extração, grande parte do butino-2 e piperilenos.

Diversos parâmetros são controlados durante o processo. O teor de água no solvente circulante, por exemplo, deve ser mantido entre 10 e 15 %m/m. Concentração de água muito elevada no solvente pode causar a formação de mais uma fase líquida e levar a perda completa de produção. Valores muito baixos podem causar diminuição da eficiência de separação e aumento de temperatura na coluna de destilação extrativa. Assim como a água, o álcool alílico, acima de uma determinada concentração, afeta a seletividade do solvente, prejudicando a separação nas colunas de destilação extrativa, aumentando a perda e contaminando o isopreno cru com alfa e beta olefinas, além de aumentar a taxa de corrosão em todo o sistema. Também são verificadas as presenças de DCPD, o qual é formado através da reação de dimerização do CPD, além de codímeros, que são formados, principalmente, pela reação de adição do CPD com o isopreno ou com os piperilenos. Os teores de DCPD, dímeros e codímeros no solvente circulante devem ser mantidos em no máximo 1 %m/m.

A seção de recuperação de solvente tem como objetivo não só recuperar o solvente e especificar o isopreno produto em relação ao teor de acetonitrila, como também de manter o solvente circulante dentro das especificações requeridas no processo de destilação extrativa. A concentração desejada para o solvente é: acetonitrila 55-70 %m/m, água 10-15 %m/m, álcool alílico 20-25 %m/m e acetona 3-5 %m/m.

A seção de remoção de enxofre tem por objetivo especificar o isopreno em no máximo 5 ppm de enxofre usando leitos de adsorção.

Finalmente, a seção de fracionamento recebe o isopreno cru com pureza de cerca de 97% m/m, proveniente da segunda coluna de destilação extrativa, e realiza dois processos de separação utilizando duas superfracionadoras para levar o isopreno produto à especificação de mercado internacional.

A seleção do tema deste trabalho fundamentou-se na necessidade de um controle rigoroso da qualidade e composição do solvente para melhorar a eficiência de extração e garantir a eficiência do processo. Neste sentido, a combinação da espectroscopia na região do infravermelho com técnicas quimiométricas fornece uma ferramenta poderosa para monitorar uma variedade de processos, tendo em vista o fato de que as amostras requerem mínimo ou nenhum pré-tratamento e diversos parâmetros de qualidade podem ser estimados simultaneamente a partir de um único espectro NIR.

2.2. A espectroscopia na região do infravermelho

2.2.1. Histórico

A radiação na região do infravermelho próximo foi descoberta pelo astrônomo inglês, Frederick William Herschel. Ao filtrar a luz do sol com vidros escuros de diferentes cores, ele observou que a iluminação e o aquecimento dos objetos não eram igualmente distribuídos entre os vários raios coloridos (HERSCHEL, 1800).

De modo a confirmar este poder das diferentes cores de aquecer corpos de maneira desigual, Herschel realizou um experimento no qual empregou um prisma de vidro, transparente à radiação, para decompor a luz branca do sol em seus comprimentos de onda visíveis e, com o auxílio de termômetros de bulbo enegrecido, constatou que a dilatação do líquido do termômetro variava para cada cor, sendo a variação do vermelho maior do que as cores verde e violeta (HERSCHEL, 1800).

“[...] the heating power of the prismatic colours, is very far from being equally divided, and that the red rays are chiefly eminent in that respect”.

(HERSCHEL, 1800, p. 261)

“[...] o poder de aquecimento das cores decompostas pelo prisma está longe de ser igualmente distribuído, e que os raios vermelhos se destacam nesse aspecto”.

(Tradução livre)

Ao contrário do senso comum, Herschel continuou investigando o que acontecia com a temperatura, colocando o termômetro além da região vermelha do espectro. Surpreendentemente, ele observou que o poder de aquecimento máximo está entre os raios invisíveis (HERSCHEL, 1800). Naquele momento, ele não podia imaginar o impacto que a sua descoberta exerceria sobre a ciência. A região do espectro eletromagnético estudada oferecia as primeiras ferramentas para o desenvolvimento de mais uma das mais poderosas técnicas espectroscópicas. Herschel publicou suas descobertas se referenciando à região dos *raios caloríficos*, que mais tarde foi denominada infravermelho.

Até o ano de 1950, pouco avanço foi registrado na espectroscopia NIR. Embora a região do espectro eletromagnético correspondente à região do infravermelho tenha sido descoberta em 1800, a espectroscopia NIR foi negligenciada durante muito tempo devido à grande dificuldade em se extrair informações fornecidas por estes espectros, principalmente por apresentarem bandas largas, superpostas e de fraca absorção.

Foi na agricultura que a espectroscopia NIR ganhou impulso e teve suas primeiras aplicações práticas, tornando-se bastante difundida nesta área. Grande parte dos estudos, os quais visavam determinações dos teores de umidade, proteínas e óleos em produtos agrícolas, estava baseado nos trabalhos desenvolvidos pelo Professor Karl Norris (HART; NORRIS; GOLUMBIC, 1962). Norris revolucionou a química analítica e tornou-se reconhecido como pioneiro na utilização da espectroscopia NIR. Em seguida, outras áreas passaram a experimentar intensamente o uso desta técnica, sendo a indústria farmacêutica e petroquímica as áreas em que a espectroscopia na região do infravermelho tem encontrado larga aplicabilidade.

A técnica NIR tornou-se então mais acessível, e passou a ser interpretada não apenas por químicos ou espectroscopistas, mas por todos os profissionais interessados nessa importante técnica analítica.

2.2.2. Fundamentos da técnica NIR

A espectroscopia do infravermelho próximo é um tipo de espectroscopia vibracional correspondendo à região delimitada pelos comprimentos de onda entre 780 e 2500 nm ($12800 - 4000 \text{ cm}^{-1}$) (PASQUINI, 2003). Os fótons empregados na espectroscopia NIR possuem energia entre $2,65 \times 10^{-19}$ a $7,96 \times 10^{-20}$ J, a qual é suficiente para gerar movimentos que alteram os comprimentos e os ângulos de ligação das moléculas, promovendo-as de seu estado vibracional fundamental para outros estados excitados.

Entretanto, para que uma molécula absorva radiação NIR, precisa receber radiação numa frequência capaz de fornecer exatamente a energia necessária para promover a transição entre dois níveis energéticos ou para a ocorrência de sobretons (transições de mais de um nível energético) ou combinações de duas ou mais vibrações. Além disso, o deslocamento dos átomos em vibração precisa produzir uma alteração do momento dipolar da molécula ou do grupo de átomos em vibração. Apenas nessas circunstâncias, a radiação eletromagnética pode interagir com a molécula e causar variações na amplitude de um de seus movimentos

(PASQUINI, 2003). O momento dipolar é determinado pela magnitude da diferença de carga e a distância entre os dois centros de carga. Isso implica que moléculas diatômicas homonucleares não absorvem radiação no intervalo do espectro infravermelho, visto que suas vibrações não produzem alterações no momento dipolar.

As vibrações moleculares podem ser classificadas em estiramentos e deformações angulares. As vibrações de estiramento envolvem uma mudança contínua na distância interatômica ao longo do eixo da ligação entre dois átomos, podendo ser simétricas ou assimétricas. Tais vibrações ocorrem a frequências mais altas (comprimentos de onda menores) do que as vibrações de deformação angular que, por sua vez, são caracterizadas pela mudança do ângulo entre duas ligações e podem acontecer no plano ou fora do plano da molécula (WORKMAN; WEYER, 2008).

Considerando que as vibrações acontecem de forma isolada em uma molécula, estas podem ser representadas por um modelo mecânico simples, conhecido como oscilador harmônico. Neste modelo, uma molécula diatômica pode ser representada por duas esferas (m_1 e m_2) conectadas por uma mola com uma constante de força k (WORKMAN; WEYER, 2008; WORKMAN, 2001). Quando uma molécula absorve radiação infravermelha, seus átomos vibram com maior amplitude e se deslocam numa determinada frequência, a qual depende da força de ligação e das massas envolvidas, definida pela equação a seguir:

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

em que, μ é a massa reduzida:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (2)$$

A energia potencial (U), descrita como função do deslocamento dos átomos (x), obedece à lei de Hooke e é dada por:

$$U = \frac{1}{2} k x^2 \quad (3)$$

A figura 3A mostra o comportamento da energia potencial como função da amplitude de vibração (deslocamento), a partir da posição de equilíbrio (PASQUINI, 2003). A energia potencial assume a forma de uma parábola e tem um comportamento simétrico em torno da distância internuclear de equilíbrio d_e , sofrendo um aumento quando a distância entre os átomos é máxima e quando esta é mínima.

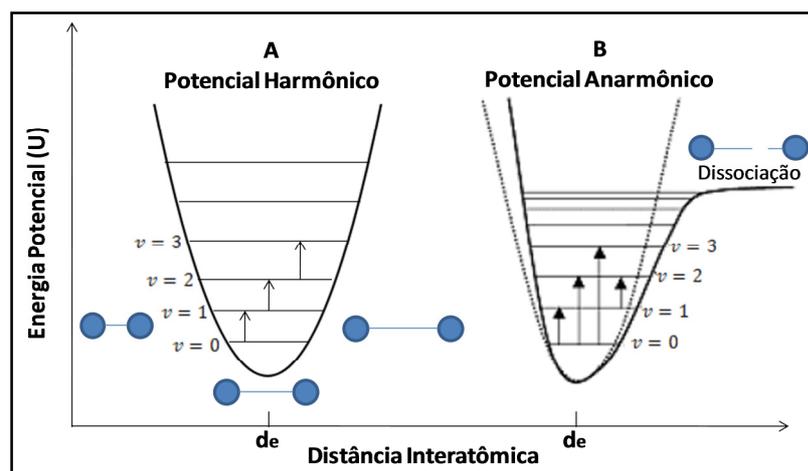


Figura 3. Diagrama de energia potencial para os modelos harmônico (A) e anarmônico (B) de um sistema diatômico. d_e é a distância interatômica de equilíbrio (PASQUINI, 2003).

Embora o modelo do oscilador harmônico seja utilizado para elucidar o movimento de vibração de um sistema diatômico, ele apresenta falhas quando aplicado a moléculas reais. Assim, o completo entendimento da espectroscopia vibracional só é possível no contexto da mecânica quântica. A teoria da mecânica quântica mostra que, em um sistema molecular real, a energia de vibração entre os átomos de uma molécula não pode apresentar um perfil de energia contínuo, como o apresentado na Figura 3A, mas níveis de energia discretos, ou quantizados (PASQUINI, 2003; WORKMAN; WEYER, 2008; WORKMAN, 2001), os quais são definidos por:

$$E_v = \left(v + \frac{1}{2} \right) h\nu, v = 0, 1, 2, 3, \dots \quad (4)$$

em que, ν é o número quântico vibracional, h é a constante de Planck e ν_0 é a frequência do oscilador harmônico clássico, prevista pela lei de Hooke.

Outra restrição imposta pela mecânica quântica ao modelo do oscilador harmônico está no fato de que as únicas transições energéticas possíveis somente ocorrem entre níveis de energia adjacentes, $\Delta\nu = \pm 1$, e a diferença de energia entre dois níveis energéticos adjacentes é sempre a mesma. Neste caso, as transições entre mais de um nível de energia são proibidas e, assim, o aparecimento de sobretons, bandas correspondentes às transições entre níveis energéticos com $\Delta\nu \geq \pm 2$, não deveriam existir. Além disso, como as vibrações presentes no modelo harmônico são independentes, as bandas correspondentes às suas combinações não existiriam. Contudo, tanto os sobretons como as bandas de combinação são observados na região do infravermelho próximo (PASQUINI, 2003).

Na prática, os níveis de energia quantizados não são igualmente espaçados, pois as moléculas desviam da idealidade e comportam-se como um oscilador anarmônico. Diferente do modelo ideal, o oscilador anarmônico compreende alguns comportamentos não-ideais. Ele leva em consideração as forças de repulsão entre as nuvens eletrônicas, as quais tem origem quando núcleos atômicos se aproximam, e a variação da força de ligação quando os átomos se afastam. Em uma molécula real, um afastamento interatômico significativo iria provocar o rompimento da ligação molecular com consequente dissociação dos átomos (WORKMAN, 2001). A Figura 3B apresenta a forma do potencial anarmônico para um sistema diatômico.

As ocorrências espectrais na região NIR são dominadas por sobretons e combinações harmônicas das vibrações (estiramentos e deformações angulares) fundamentais de ligações C-H, O-H e N-H, provocadas pela grande anarmonicidade dessas vibrações. Essas bandas são de fraca intensidade, de cerca de 10 vezes, para bandas de combinação, a 1000 vezes menos intensas, para sucessivos sobretons, do que as observadas na região do infravermelho médio (PASQUINI, 2003). Além disso, são também bem mais sobrepostas, tornando o espectro NIR complexo e de difícil interpretação.

As bandas de combinação e os sobretons são fortemente influenciados pelas pontes de hidrogênio, as quais afetam a anarmonicidade das moléculas. Essa influência resulta no aparecimento de bandas mais largas, de fraca intensidade e em deslocamentos da frequência de absorção, uma vez que moléculas ligadas às moléculas mais próximas possuem anarmonicidade menor. Assim, deslocamentos das bandas para frequências mais baixas (comprimentos de onda maiores) podem ser observados com a formação de pontes de hidrogênio. Deslocamentos das bandas para frequências mais altas (comprimentos de onda menores) ocorrem com o enfraquecimento das pontes de hidrogênio devido a diluições e aumento de temperatura (WORKMAN, 2001).

2.2.3. Características espectrais na região do NIR (WORKMAN; WEYER, 2008; WORKMAN, 2001)

2.2.3.1. Alcanos

A primeira consideração a ser feita a respeito da interpretação espectral dos alcanos inclui a vibração C-H de estiramento e deformações angulares dos grupos metil (-CH₃) e metileno (-CH₂-). Rotações internas e sobreposições das bandas tornam a interpretação dos espectros deste grupo de hidrocarbonetos complexa. De forma geral, os primeiros sobretons de estiramento da ligação C-H ocorrem entre 5555 e 5882 cm⁻¹, os segundos sobretons entre 8264 e 8696 e os terceiros sobretons entre 11364 e 10929 cm⁻¹. As regiões de combinação mais importantes ocorrem entre 6666 e 7690 cm⁻¹ e 4545 e 4500 cm⁻¹.

2.2.3.2. Alcenos

Assim como no infravermelho médio, a posição da vibração de estiramento da ligação C-H pode ser influenciada pelos átomos adjacentes. Assim, o primeiro sobreton de estiramento da ligação C-H situada próxima a uma dupla ligação ocorre a números de onda maiores (comprimentos de onda menores) do que

absorções de estiramento da ligação C-H envolvendo apenas carbonos saturados. Em algumas estruturas, este pico é nítido e forte, particularmente o grupo metileno de ligações duplas terminais. Entretanto, na maioria dos casos, o primeiro sobretudo é fraco e de difícil localização, especialmente na presença de grupos metílicos.

2.2.3.3. Cetonas

Compostos contendo grupos carbonílicos (C=O) exibem bandas de absorção na região entre $5679 - 4043 \text{ cm}^{-1}$ e $7500 - 5166 \text{ cm}^{-1}$, as quais correspondem ao 2º e 3º sobretudo de estiramento da ligação C=O respectivamente. Na acetona, essas bandas estão localizadas a 5100 cm^{-1} e a 6897 cm^{-1} . Já as bandas de combinação envolvendo a ligação C=O e C-H nas cetonas são fracas, podendo até não existir, uma vez que os dois grupos não compartilham o mesmo átomo de carbono.

2.2.3.4. Alcoóis

O primeiro sobretudo de um grupo hidroxila livre em solução diluída de tetracloreto de carbono (CCl_4) ou em um gás de baixa densidade ocorre a 7090 cm^{-1} . Este pico está localizado em diferentes posições do espectro para alcoóis primários, secundários e terciários. Em alcoóis cujas hidroxilas estão ligadas por pontes de hidrogênio, o primeiro sobretudo ocorre sob a forma de um pico largo na região entre $6850 - 6240 \text{ cm}^{-1}$. Por outro lado, o primeiro sobretudo de hidroxilas livres é forte e pode ser facilmente visualizado. O segundo e terceiro sobretudo de estiramento da ligação O-H de hidroxilas livres ocorrem a 10400 e 13500 cm^{-1} , respectivamente. A região de combinação dos alcoóis está situada entre 4550 e 5550 cm^{-1} .

2.2.3.5. Água

A molécula de água apresenta um pico forte a 5150 cm^{-1} o qual é atribuído à combinação do estiramento assimétrico e deformação angular da molécula. A água

no estado líquido exibe ainda bandas de combinação largas e fracas, as quais ocorrem a 5620 cm^{-1} , 8310 cm^{-1} , e um segundo conjunto de picos a 11800 e 13000 cm^{-1} . As bandas localizadas a 6900 e 10300 cm^{-1} indicam que estas são bandas de combinação envolvendo os modos de estiramento simétrico e assimétrico da molécula.

2.2.3.6. Nitrilas

Pouca informação foi encontrada a respeito das absorções na região do infravermelho dos compostos nitrílicos. Entretanto, com relação à acetonitrila, a literatura relata que o 1º sobreton de estiramento da ligação $\text{C}\equiv\text{N}$ aparece a $4444,44\text{ cm}^{-1}$ [25]. Similarmente à acetona, bandas de absorção relativas ao 3º, 2º e 1º sobretons de estiramento da ligação C-H de metil podem ser observadas na região NIR. Estas bandas aparecem a 11428 , 8771 e entre 5988 e $5882,4\text{ cm}^{-1}$, respectivamente.

As principais atribuições espectrais na região do infravermelho próximo dos compostos estudados neste trabalho estão resumidas na Tabela 1.

O espectro NIR está longe de apresentar uma interpretação rápida e direta, visto que as bandas de absorção nessa região são de fraca intensidade e bem mais sobrepostas do que àquelas que ocorrem na região do infravermelho médio, o que dificulta a obtenção de uma boa correlação entre grupos de átomos presentes numa molécula e a sua composição nas amostras. Sendo assim, somente com a ajuda dos métodos quimiométricos de análise, a obtenção de informações qualitativa ou quantitativa desses espectros torna-se possível. Com o desenvolvimento das técnicas matemáticas (algoritmos para pré-processamento dos dados e regressão multivariada) e a grande disponibilidade de softwares, a espectroscopia NIR alcançou sucesso, consagrando-se como uma ferramenta para análises rápidas e precisas [26].

Tabela 1. Principais características espectrais dos grupos funcionais estudados neste trabalho na região NIR.

Região Espectral (cm ⁻¹)	Álcoois e Água	Cetonas	Nitrilas	Hidrocarbonetos
12821 - 10000	Combinação de O-H (2v) e δC-H de metil (9386 cm ⁻¹), R-C-OH; O-H (3v) (10040 cm ⁻¹), alcoóis primários;			C-H (4v), (11052 - 10953 cm ⁻¹) metil;
10000 - 7500	O-H (3v) (9960 cm ⁻¹ , alcoóis secundários e 9940 cm ⁻¹ , alcoóis terciários); O-H (3v) com ponte de hidrogênio (9550 cm ⁻¹), álcool alquil de cadeia curta;			C-H (3v) (8368 cm ⁻¹), •C-H ₃ ; Combinação C-H (3vCH ₃ e δCH ₃) (9795 cm ⁻¹); Combinação C-H (3vCH ₂ e δCH ₂) (9600 cm ⁻¹);
7500 - 6500	O-H (2v) com ponte de hidrogênio (6850 - 6240 cm ⁻¹), R-C-OH; O-H (2v) (6887 cm ⁻¹ , alcoóis primários; 6831 cm ⁻¹ , alcoóis secundários e 6812 cm ⁻¹ , alcoóis terciários);	C=O (4v) (6897 cm ⁻¹), acetona		Combinação C-H (7353 cm ⁻¹), •C-H ₃ ; Combinação C-H (7067 cm ⁻¹), •C-H ₂ ; Combinação C-H (2vCH ₃ e δCH ₃) (7194 cm ⁻¹); Combinação C-H (2vCH ₂ e δCH ₂) (7085 cm ⁻¹);
6500 - 5500	O-H (2v) com ponte de hidrogênio (6330 cm ⁻¹), R-C-OH; Combinação de estiramento O-H e deformação angular H-O-H (5150 cm ⁻¹), água;			C-H (2v) (5797 cm ⁻¹), •CH ₂ ; C-H (2v) (5900 cm ⁻¹),•CH ₃ ;
5500 - 4600	Combinação de estiramento simétrico e assimétrico (5587 cm ⁻¹), água; Combinação de νO-H e δO-H (5550 - 4550 cm ⁻¹); O-H (3δ) da banda do infravermelho médio ocorrendo a 1200 cm ⁻¹ (4770 cm ⁻¹), alcoóis e água;	C=O (3v) (5100 cm ⁻¹), acetona		Combinação de estiramento C-H e deformação C-H (4608 cm ⁻¹), (•HC=CH);
4600 - 4000			C≡N (2v) (4444 cm ⁻¹), acetonitrila	Combinação de estiramento C-H e deformação angular C-H (2vCH ₂ estiramento assimétrico e δCH ₂) (4314 cm ⁻¹); C-H (3δCH ₃) (4068 cm ⁻¹);

2v, 3v, 4v: primeiro, segundo e terceiro sobretons das bandas de vibrações fundamentais de estiramento.

2δ e 3δ: primeiro e segundo sobretons das bandas de vibrações fundamentais de deformações angulares.

2.2.4. Técnicas quimimométricas

A quimiometria é o ramo da ciência que faz uso de ferramentas matemáticas e estatísticas a fim de obter informações quimicamente relevantes a partir de dados analíticos, como por exemplo, dados espectrais na região do infravermelho (PASQUINI, 2003; WORKMAN; WEYER, 2008; WORKMAN, 2001; BOKOBZA, 1998; WOLD, 1995). Considera-se ser uma de suas grandes contribuições ter

fornecido um modelo de pensar multivariado aos químicos, que até então aplicavam à química analítica instrumental, eminentemente multivariada, um modo de pensar univariado. A análise multivariada fornece maior compreensão de um conjunto de dados, permitindo um tratamento dos efeitos sinérgicos ou antagônicos entre as variáveis que estão sendo investigadas (WOLD, 1995).

Na quimiometria existem várias técnicas extremamente úteis, as quais são utilizadas na calibração multivariada na região NIR, dentre elas estão: as técnicas de pré-processamento (BEEBE; RANDY; SEASHOLTZ, 1998; ZEAITER; ROGER; BELLON-MAUREL, 2005), seleção de variáveis [OLIVEIRA *et al.*, 2004; LEARDI; GONZÁLEZ, 1998; NORGAARD *et al.*, 2000; PONTES *et al.*, 2005; ARAÚJO *et al.*, 2001) e de regressão multivariada (BEEBE; RANDY; SEASHOLTZ, 1998; LEARDI; GONZÁLEZ, 1998; FERREIRA *et al.*, 1999; DARDENNE; SINNAEVE; BAETEN, 2000). As técnicas de pré-processamento são utilizadas para reduzir ou remover fontes de variação irrelevantes, como por exemplo, ruído aleatório, deslocamentos da linha de base, espalhamentos de luz, enquanto que as técnicas de seleção de variáveis permitem obter modelos de calibração mais simples e de fácil interpretação. Após pré-tratamentos dos dados, pode-se obter modelos de calibração multivariada com ferramentas que em geral supõem uma correlação linear entre os dados espectrais e a concentração ou valores de propriedades a serem determinados.

2.2.4.1. Pré-processamento de dados

No contexto da calibração multivariada, pré-processamento pode ser definido como qualquer manipulação matemática dos dados que possibilite reduzir ou remover fontes de variação não informativas, sejam elas sistemáticas ou aleatórias, antes da modelagem multivariada propriamente dita (BEEBE; RANDY; SEASHOLTZ, 1998). O pré-processamento pode ser aplicado às amostras ou às variáveis.

Normalização, atribuição de pesos, suavização ou correção da linha de base podem ser utilizados para pré-processar as amostras. A normalização é realizada dividindo-

se cada variável de um vetor amostra por uma constante. A atribuição de pesos a uma amostra consiste na multiplicação de cada elemento de um vetor amostra por uma constante. Nesse sentido, a influência que uma amostra exerce sobre um modelo matemático pode ser manipulada e, com isso, o peso atribuído só deve ser aplicado quando informações confiáveis, a respeito da importância de algumas amostras em relação a outras, estão disponíveis. Ambas as técnicas são utilizadas para remover variações sistemáticas (BEEBE; RANDY; SEASHOLTZ, 1998).

A técnica de suavização é empregada para reduzir o ruído aleatório (ruídos de alta frequência) a fim de aumentar a relação sinal/ruído, porém não é capaz de remover os ruídos de baixa frequência sem que algumas informações químicas de interesse também sejam removidas. As correções de linha de base são capazes de reduzir esse tipo de variação sistemática, capaz de provocar mudanças no sinal de interesse e dominar a análise.

Um dos pré-processamentos mais utilizados para corrigir a linha base é a derivada (BEEBE; RANDY; SEASHOLTZ, 1998). A primeira derivada remove deslocamentos constantes (*offsets*) da linha de base e a segunda derivada uma variação linear da linha de base. O cálculo das derivadas é obtido através das diferenças dos valores de pontos adjacentes, o que torna a relação sinal/ruído pior. Uma forma de contornar esse efeito é fazer uso das diferenças das médias, em que um número fixo de pontos (largura da janela) é selecionado, ao invés de se utilizar as diferenças entre pontos individuais (BEEBE; RANDY; SEASHOLTZ, 1998). Além disso, recomenda-se aplicar aos dados algum tipo de suavização antes da diferenciação (BOKOBZA, 1998).

As variáveis podem ser pré-processadas através dos métodos (BEEBE; RANDY; SEASHOLTZ, 1998): 1) centralização na média, o qual consiste em subtrair o valor de cada elemento (absorvância no mesmo comprimento de onda em diferentes amostras) do vetor variável pelo valor médio dos elementos, fazendo com que as diferenças relativas das intensidades sejam discriminadas mais facilmente, e 2) atribuição de pesos, no qual está incluída a operação de escalonar na variância, cujo objetivo desta é colocar todas as variáveis em uma mesma base, e corresponde à divisão de cada elemento do vetor variável pelo desvio padrão da

variável. O autoescanolamento é a denominação dada quando as técnicas de centralização na média e o escalonamento são aplicadas simultaneamente (BEEBE; RANDY; SEASHOLTZ, 1998).

2.2.4.2. Seleção de amostras

Na calibração multivariada, a seleção de um subconjunto representativo de calibração e validação, a partir de um conjunto grande de amostras, é fundamental. O conjunto de calibração deve ser selecionado de forma a reproduzir a variabilidade dos parâmetros em estudo de amostras reais. O conjunto de validação deve ser escolhido de forma a assegurar a qualidade do modelo (DANTAS FILHO *et al.*, 2004; GALVÃO *et al.*, 1999). Alguns métodos podem ser aplicados para realizar esta separação, e dentre esses pode-se citar, o algoritmo Kennard-Stone (KS) e o algoritmo SPXY (GALVÃO *et al.*, 1999).

2.2.4.2.1. Algoritmo Kennard-Stone (KS) (GALVÃO *et al.*, 1999)

O algoritmo clássico Kennard-Stone visa realizar a seleção dos subconjuntos, assegurando a uniformidade de distribuição de cada subconjunto ao longo do espaço de dados X (respostas instrumentais). Sendo assim, KS segue o procedimento por etapas, em que novas seleções são feitas nas regiões do espaço longe das amostras já selecionadas. Neste contexto, o algoritmo emprega a distância Euclidiana $d_x(a,b)$ entre os vetores X de cada par (a,b) de amostras, como mostrado na Equação 5.

$$d_x(a, b) = \sqrt{\sum_{j=1}^J [Xa(j) - Xb(j)]^2}; a e b \in [1, N] \quad (5)$$

Para dados espectrais, $Xa(j)$ e $Xb(j)$ são as respostas instrumentais no j -ésimo comprimento de onda para as amostras a e b , respectivamente, e J denota o número de comprimentos de onda no espectro.

A seleção se inicia tomando o par de amostras (a_1, b_1) em que a distância Euclidiana $d_x(a_1, b_1)$ é maior. A cada subsequente iteração, o algoritmo seleciona a amostra que exibe a maior distância em relação a uma amostra já selecionada. Este procedimento é repetido até que o número de amostras requerido seja alcançado.

2.2.4.2.2. Algoritmo SPXY (*Sample set Partitioning based on joint x-y distances*) (GALVÃO *et al.*, 1999)

Neste trabalho, a separação dos conjuntos de dados entre os conjuntos de calibração e validação foi realizada através do algoritmo SPXY. O algoritmo SPXY é uma extensão do algoritmo KS que consiste no aumento da distância definida pela Equação 5, pois leva em consideração tanto as diferenças de X (Equação 5) quanto as de Y no cálculo das distâncias inter-amostras. A distância $d_y(a, b)$ pode ser calculada através da Equação 6.

$$d_y(a, b) = \sqrt{(Ya - Yb)^2} = |Ya - Yb|; a, b \in [1, N] \quad (6)$$

De forma a atribuir igual importância às distribuições das amostras no espaço (x, y) , este algoritmo calcula cada distância $d_x(a, b)$ (Equação 5) e cada distância $d_y(a, b)$ (Equação 6), soma as distâncias de X com Y e normaliza dividindo $d_x(a, b)$ e $d_y(a, b)$ pelos seus valores máximos no conjunto de dados. Assim, uma distância normalizada xy é calculada, de acordo com a Equação 7.

$$d_{xy}(a, b) = \frac{d_x(a, b)}{\max_{a, b \in [1, N]} d_x(a, b)} + \frac{d_y(a, b)}{\max_{a, b \in [1, N]} d_y(a, b)}; a e b \in [1, N] \quad (7)$$

2.2.4.3. Seleção de variáveis

A escolha da região espectral envolve a busca de variáveis (um comprimento de onda ou um conjunto de comprimentos de onda), contendo informações relevantes para a propriedade de interesse, a fim de basear a modelagem apenas nessas variáveis, permitindo que os modelos de calibração construídos sejam mais robustos e de simples interpretação, além de minimizar os erros de predição (PASQUINI, 2003; OLIVEIRA *et al.*, 2004).

A escolha da região espectral a ser modelada pode ser realizada com base na experiência do analista e do seu conhecimento a respeito do sistema sob investigação. No entanto, diversas técnicas tem sido empregadas para auxiliar nesse propósito, dentre as quais pode-se destacar: o algoritmo genético – GA (*Genetic Algorithm*) (LEARDI; GONZÁLEZ, 1998), o método dos mínimos quadrados parciais por intervalos – *i*PLS (*Interval Partial Least Square regression*) (NORGAARD *et al.*, 2000), o algoritmo das projeções sucessivas – SPA (*Successive Projections Algorithm*) (PONTES, 2005; ARAÚJO *et al.*, 2001) e o algoritmo JackKnife (JK) (HONORATO *et al.*, 2007; MARTENS, H.; MARTENS, M., 2000; WOLD; SJOSTROM; ERIKSSON, 2001; WEHRENS; PUTTER; BUYDENS, 2000). O algoritmo *i*PLS e JackKnife foram utilizados neste trabalho e serão apresentados a seguir.

2.2.4.3.1. Algoritmo *i*PLS (NORGAARD *et al.*, 2000)

O método dos mínimos quadrados parciais por intervalos é uma extensão interativa desenvolvida para o PLS, em que modelos locais PLS são desenvolvidos em subintervalos iguais de toda a região do espectro. O principal objetivo deste método é identificar as informações mais relevantes do espectro através das diferentes subdivisões espectrais, focando nas regiões espectrais mais importantes e removendo as regiões cujas variáveis se apresentam como supostamente de menor relevância e/ou relacionadas a interferentes.

Os modelos *i*PLS para cada região do espectro subdividido são comparados em um gráfico com o modelo PLS global (desenvolvido com todo o espectro). A comparação está baseada, principalmente, no parâmetro de validação RMSECV (erro médio quadrático da validação cruzada, do inglês *root mean squared error of cross-validation*). O espectro pode ser subdividido em quantas partes se desejar, até que, através de tentativa e erro, chega-se a uma divisão ótima, ou seja, obtêm-se regiões do espectro com menor RMSECV do que o modelo global. Outros parâmetros como r^2 (coeficiente quadrático de correlação) da reta de correlação

entre os valores reais e previstos pelo modelo também podem ser avaliados. Amostras e/ou medidas anômalas detectadas pelo PLS devem ser removidas antes da aplicação do *i*PLS. As regiões espectrais com valores de RMSECV menor que o modelo global são selecionadas para a construção dos modelos PLS. Gráficos do *i*PLS são apresentados nas Figuras 4 e 5.

Na figura 4, a divisão dos dados é apresentada em 10 subgrupos. Para cada subgrupo é desenvolvido um modelo PLS, e os resultados de RMSECV são mostrados como barras na figura. Os números no interior das faixas no gráfico são os números de variáveis latentes (VL's ou PC's) indicados pelo modelo *i*PLS para serem usados na construção dos modelos com as variáveis selecionadas em cada intervalo. A linha horizontal tracejada indica o valor de RMSECV para o modelo global (utilizando todas as variáveis). Nesse exemplo, é possível perceber que o 2º e o 5º intervalo apresentam valor de RMSECV menor que o modelo global e, por isso, são indicados para o desenvolvimento do modelo de calibração.

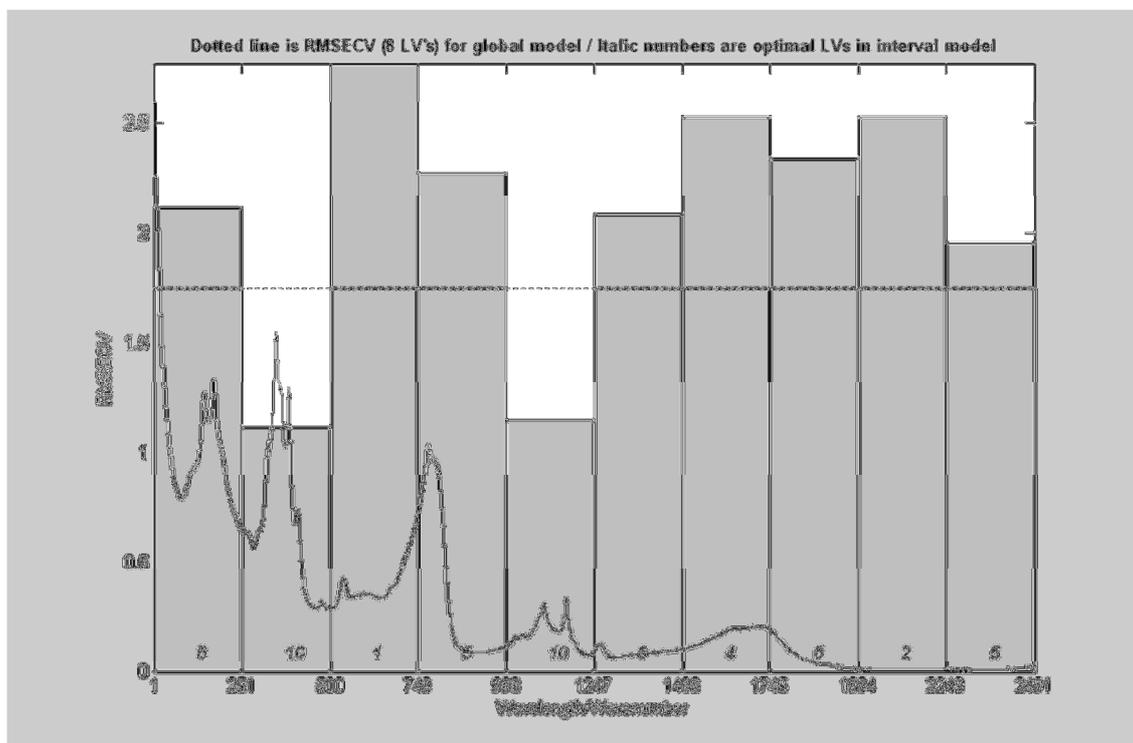


Figura 4. Gráfico do modelo *i*PLS: número de intervalos vs RMSECV.

Na Figura 5, a faixa das variáveis escolhidas (251-499) é apresentada. Portanto, no exemplo mencionado, o modelo PLS pode ser construído a partir de 248 variáveis do espectro original (variáveis 251-499), ao invés das 2491 variáveis iniciais.

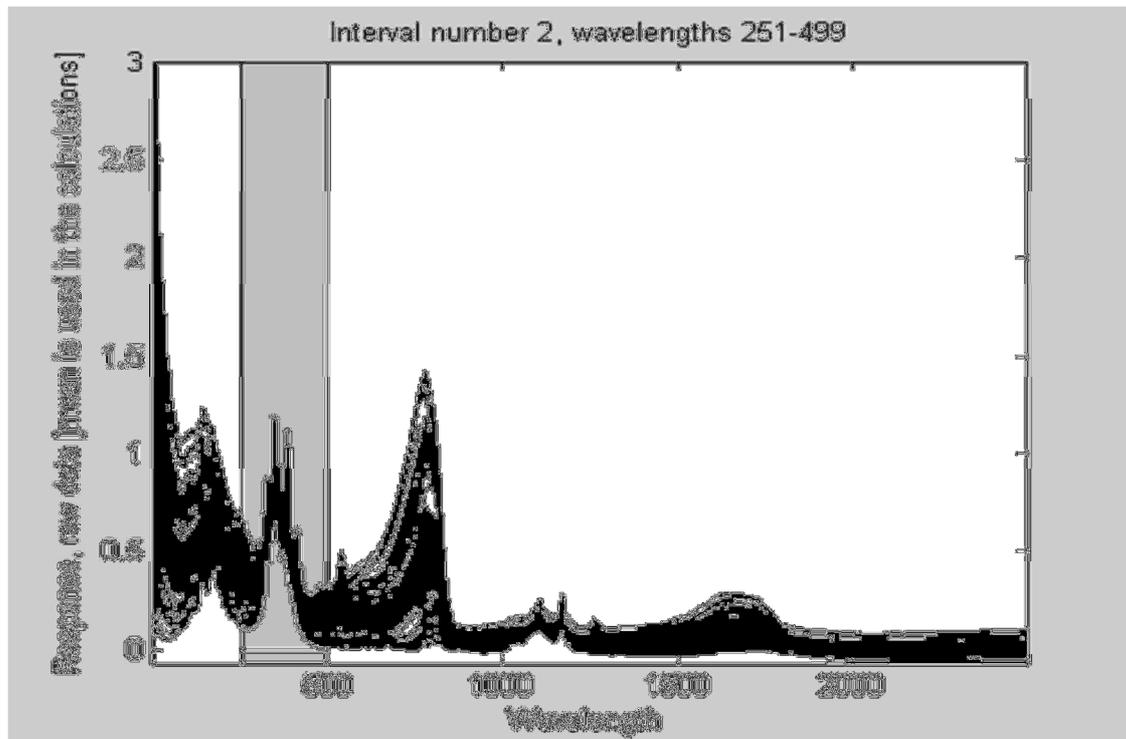


Figura 5. Gráfico do modelo PLS: faixa escolhida para a construção do modelo PLS.

2.2.4.3.2. Algoritmo Jackknife (MARTENS, H.; MARTENS, M., 2000; WOLD; SJOSTROM; ERIKSSON, 2001)

O algoritmo JK pode ser utilizado para selecionar as variáveis mais informativas na construção dos modelos de regressão, incluindo-se o PLS. Nesse algoritmo, os erros padrão das estimativas dos coeficientes de regressão obtidos pelo modelo são avaliados utilizando-se uma série de subconjuntos de dados. A partir daí, pode-se calcular os intervalos de confiança para os coeficientes de regressão e avaliar a significância de cada variável utilizada no modelo.

2.3. Calibração multivariada

A calibração multivariada é provavelmente a área da quimiometria que tem atraído mais interesse em aplicações na espectroscopia na região do infravermelho. O propósito da calibração é estabelecer uma relação matemática entre os dados espectrais e o parâmetro físico ou químico de interesse previamente determinado por uma técnica independente. Na construção de modelos de calibração multivariada, os dados espectrais de natureza multivariada são representados em uma matriz \mathbf{X} , de dimensão $(n \times m)$, em que n representa o número de amostras cujos espectros foram obtidos e m representa o número de variáveis (comprimentos de onda) em que as absorbâncias foram medidas. Os parâmetros de interesse são representados pela matriz \mathbf{Y} de dimensão $(n \times p)$, em que p são os parâmetros a serem estimados pelos modelos construídos. As técnicas de calibração tem por finalidade construir uma relação matemática entre as duas matrizes, e otimizá-la de forma a obter uma melhor descrição da propriedade de interesse. Após a construção de um modelo satisfatório, deve ser possível, a partir do espectro de qualquer amostra, prever os valores dos parâmetros calibrados (BEEBE; RANDY; SEASHOLTZ, 1998; BRERETON, 2000).

2.3.1. Métodos de regressão multivariada

Existem muitas ferramentas quimiométricas para fazer uso das informações espectroscópicas para fins de calibração, sendo as mais comuns, a Regressão Linear Múltipla – MLR (*Multiple Linear Regression*) (WOLD, 1995; BRERETON, 2000), a Regressão em Componentes Principais – PCR (*Principal Component Regression*) (BEEBE; RANDY; SEASHOLTZ, 1998; DARDENNE; SINNAEVE; BAETEN, 2000) e a Regressão por Mínimos Quadrados Parciais – PLS (*Partial Least Squares Regression*) (BEEBE; RANDY; SEASHOLTZ, 1998; BRERETON, 2000, FERREIRA *et al.*, 1999; DARDENNE; SINNAEVE; BAETEN, 2000; HELLAND, 2001; GELADI; KOWALSKI, 1986; LARRECHI; CALLAO, 2006), sendo que PCR e PLS são as mais utilizadas. Tendo em vista que o método dos mínimos quadrados parciais foi utilizado no presente trabalho, este será abordado de forma destacada.

A base fundamental para o PCR e PLS é a Análise de Componentes Principais – PCA (*Principal Componente Analysis*). Esta consiste em uma manipulação matemática na matriz de dados cujo objetivo é representar as variações presentes em muitas variáveis através de um número menor de fatores. Um novo sistema de eixos denominados fatores ou componentes principais (PC`s) são obtidos, no qual a natureza multivariada dos dados pode ser visualizada em poucas dimensões (BEEBE; RANDY; SEASHOLTZ, 1998). Em termos matemáticos, o principal objetivo da PCA é descrever a variação ou a dispersão entre os pontos (amostras) utilizando o menor número de eixos que se alinham com os dados. As novas coordenadas no novo sistema de eixos das componentes principais são denominadas escores. Estes definem as similaridades e/ou diferenças entre as amostras. As contribuições de cada variável original para uma determinada componente são denominadas *loadings*, as quais são determinadas através dos cossenos dos ângulos entre os eixos das variáveis e a componente principal.

O número de PC`s pode ser determinado utilizando-se o método da validação cruzada (*cross validation*). A validação cruzada consiste em remover um subconjunto de amostras do conjunto de calibração e aplicar o procedimento de modelagem com as amostras remanescentes. Em seguida, os resíduos das amostras excluídas do modelo são calculados. Este procedimento é repetido para diferentes subconjuntos de amostras até que cada amostra tenha sido excluída do modelo uma vez. Na prática, os resíduos de validação são resumidos em um único número denominado RMSECV, o qual pode ser calculado através da Equação 8. O melhor número de PC`s será aquele capaz de produzir o menor RMSECV (BEEBE; RANDY; SEASHOLTZ, 1998).

$$RMSECV = \sqrt{\frac{\sum_{i=1}^N (y_{ref} - y_{cal})^2}{n}} \quad (8)$$

em que, y_{ref} é o valor de referência da concentração ou da propriedade, y_{cal} é o valor estimado e n corresponde ao número de amostras de calibração.

Depois de construído, o modelo deve passar por um processo de validação externa antes da sua aplicação. Na validação, o modelo será testado, prevendo a

concentração ou propriedades das amostras que não participaram da modelagem. O RMSEP (erro médio quadrático de previsão, do inglês *root mean square error prediction*) é então calculado de forma semelhante à Equação 8. Como critério de aceitação do modelo, este valor normalmente é comparado com o valor da reprodutibilidade intralaboratorial ou da reprodutibilidade do método convencional (LARRECHI; CALLAO, 2003).

2.4. Planejamento experimental

Na otimização de métodos analíticos, o objetivo principal é determinar a influência de uma ou mais variáveis sobre uma outra variável de interesse. Na maioria dos procedimentos analíticos, os experimentos são otimizados de forma univariada, alterando uma variável de cada vez, enquanto as demais permanecem constantes. Suas principais desvantagens são o tempo gasto, pois requer um número elevado de experimentos, e a falta de informação a respeito das interações entre as variáveis que afetam a resposta analítica. Técnicas envolvendo otimização multivariada, na qual todas as variáveis são avaliadas simultaneamente, é a forma mais eficiente de otimização por permitir a avaliação de efeitos de interação entre as variáveis com uma menor quantidade de experimentos. Estas metodologias baseiam-se em técnicas estatísticas de planejamento experimental e análise de dados (BARROS NETO; SCARMINIO; BRUNS, 2001).

Os planejamentos experimentais podem ser de primeira ordem, dos quais se pode ajustar modelos lineares e podem ser descritos da seguinte forma: $R = a_0 + a_1X + a_2Y + a_{12}XY$, ou de segunda ordem, que permitem ajustar modelos quadráticos e seguem a equação do tipo $R = a_0 + a_1X + a_2Y + a_{11}X^2 + a_{22}Y^2 + a_{12}XY$. Para os dois polinômios descritos anteriormente, a_0 é o termo independente, a_1 e a_2 são coeficientes dos termos lineares, a_{11} e a_{22} são os coeficientes dos termos quadráticos e a_{12} é o coeficiente do termo de interação.

O planejamento fatorial é uma estratégia multivariada bastante empregada quando se deseja estudar os efeitos de uma ou mais variáveis sobre a variável resposta

escolhida. Essa estratégia permite uma combinação de todas as variáveis em todos os níveis, obtendo-se uma análise de uma variável sujeita a todas as combinações das demais. O planejamento fatorial é geralmente utilizado para análise exploratória do sistema sob investigação, sendo indicado para determinação das variáveis que exercem maior influência no desempenho de um dado processo analítico (BARROS NETO; SCARMINIO; BRUNS, 2001).

Nos planejamentos experimentais onde as variáveis são exploradas em dois níveis costuma-se identificar os níveis superior e inferior com os sinais (+) e (-), respectivamente. A atribuição é feita de forma arbitrária e não interfere na interpretação dos resultados (BARROS NETO; SCARMINIO; BRUNS, 2001).

2.4.1. Matriz Doehlert

O planejamento Doehlert ou matriz Doehlert (TEÓFILO; FERREIRA, 2006; FERREIRA *et al.*, 2004; BOSQUE-SENDRA, 1995) foi apresentado como sendo uma alternativa bastante útil e atrativa como planejamento experimental de 2ª ordem. Este planejamento descreve uma distribuição uniforme dos pontos experimentais ao redor de círculos centrados ou esferas. Para duas variáveis, a matriz Doehlert consiste de um ponto central e mais seis pontos adicionais formando um hexágono regular, conforme Figura 6.

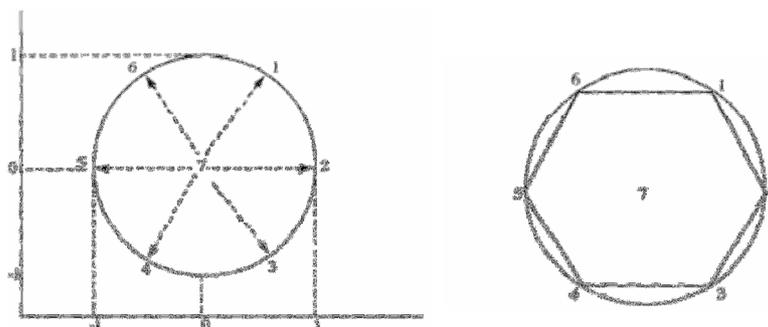


Figura 6. Representação geométrica do planejamento bidimensional Doehlert para duas variáveis: um hexágono regular (TEÓFILO; FERREIRA, 2006).

O número total de pontos experimentais no planejamento é igual a $k^2 + k + 1$, em que k é o número de fatores (TEÓFILO; FERREIRA, 2006). Na otimização de três variáveis, a distribuição dos pontos experimentais para o planejamento depende da escolha da projeção de um cubooctaedro (Figura 7) sobre um plano (FERREIRA *et al.*, 2004). Três projeções são possíveis para esse planejamento (Figura 8). As projeções podem ser obtidas sobre uma face triangular, uma face quadrada ou ainda sobre um vértice da figura tridimensional (REBOUÇAS, 2004). O número de experimentos requeridos pela matriz Doehlert de três variáveis é dado por $k^2 + k + 1$, ou seja, treze experimentos (REBOUÇAS, 2004).

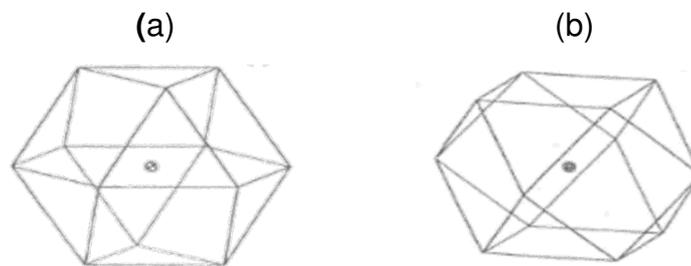


Figura 7. Representação geométrica tridimensional do planejamento Doehlert para três variáveis: a) cubooctaedro centrado sobre uma face triangular, b) cubooctaedro centrado sobre uma face quadrada (REBOUÇAS, 2004).

Uma importante propriedade do planejamento Doehlert diz respeito ao número de níveis que cada variável possui. Em um planejamento para duas variáveis, enquanto uma variável é estudada com cinco níveis, a outra o é somente com três níveis. Para a otimização de três variáveis, três planejamentos experimentais são possíveis: o primeiro, e mais utilizado, é aquele em que as variáveis assumem 3, 5 e 7 níveis (Tabela 2). Para os outros dois, as variáveis assumem somente 3 e 5 níveis. As projeções planas do planejamento geram uma matriz teórica onde os valores experimentais encontram-se codificados pelas coordenadas de cada projeção (REBOUÇAS, 2004).

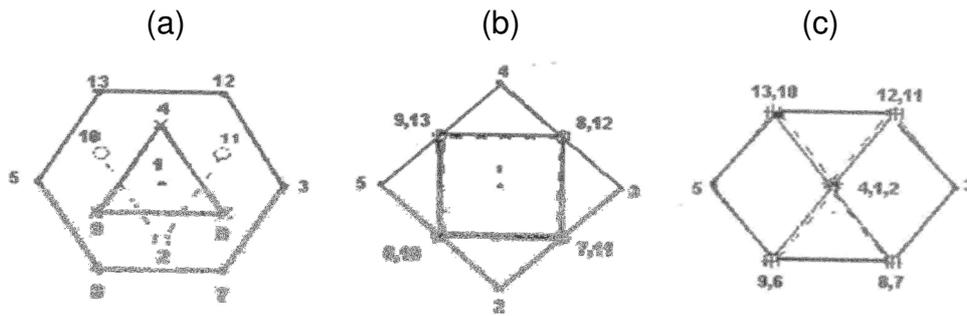


Figura 8. Três projeções planas possíveis para o planejamento Doehlert com três variáveis: a) sobre uma face triangular, b) sobre uma face quadrada, c) sobre um vértice (Rebouças, 2004).

Tabela 2. Valores codificados das coordenadas planas do planejamento Doehlert para três variáveis. As colunas A, B e C correspondem às projeções da Figura 8 (Doehlert, 1970).

Experimento	A	B	C
1	(0;0;0)	(0;0;0)	(0;0;0)
2	(0;-0,577;-0,816)	(0;-1;0)	(0;0;-1)
3	(1;0;0)	(1;0;0)	(1;0;0)
4	(0;0,577;-0,816)	(0;1;0)	(0;0;1)
5	(-1;0;0)	(-1;0;0)	(-1;0;0)
6	(-0,5;-0,866;0)	(-0,5;-0,5;0,707)	(-0,5;-0,707;-0,5)
7	(0,5;-0,866;0)	(0,5;-0,5;0,707)	(0,5;-0,707;-0,5)
8	(0,5;0,289;0,816)	(0,5;0,5;0,707)	(0,5;-0,707;0,5)
9	(-0,5;-0,289;-0,816)	(-0,5;0,5;0,707)	(-0,5;-0,707;0,5)
10	(-0,5;0,289;0,816)	(-0,5;-0,5;-0,707)	(-0,5;0,707;-0,5)
11	(0,5;-0,289;-0,816)	(0,5;-0,5;-0,707)	(0,5;0,707;0,5)
12	(0,5;0,866;0)	(0,5;0,5;-0,707)	(0,5;0,707;-0,5)
13	(-0,5;0,866;0)	(-0,5;0,5;-0,707)	(-0,5;0,707;0,5)

2.5. A função desejabilidade e os experimentos com multirespostas (BARROS NETO; SCARMINIO; BRUNS, 2001)

Geralmente, os planejamentos experimentais são realizados com uma única variável resposta. Entretanto, na maioria dos processos analíticos, muitas variáveis estão envolvidas na resposta de uma outra variável de interesse, sendo necessário um método de otimização simultânea, o qual é bastante útil quando se deseja encontrar as condições ótimas de várias respostas. Este método se baseia na definição de uma função de desejabilidade para cada resposta, com valores restritos ao intervalo $[0,1]$, em que zero significa um valor inaceitável e um, o valor mais desejável.

Suponhamos que a resposta tenha um valor alvo A , situado em algum ponto dentro de uma faixa de aceitação, cujos limites inferior e superior são representados por LI e LS , respectivamente. Desta forma, a função de desejabilidade será dada por (equação 9, 10 e 11) (BARROS NETO; SCARMINIO; BRUNS, 2001):

$$d = \left(\frac{\hat{y} - LI}{A - LI} \right)^s, \text{ para } LI \leq \hat{y} \leq A \quad (9)$$

$$d = \left(\frac{\hat{y} - LS}{A - LS} \right)^t, \text{ para } A \leq \hat{y} \leq LS \quad (10)$$

$$d = 0, \text{ para } \hat{y} \text{ fora do intervalo } (LI, LS) \quad (11)$$

em que d é a desejabilidade individual de uma dada resposta e \hat{y} é o valor predito para a resposta. A taxa de variação da função desejabilidade é definida pelos expoentes s e t .

Uma desejabilidade igual a 1 será obtida se a resposta coincidir exatamente com o valor alvo e tornar os numeradores das frações iguais aos respectivos denominadores. À medida que o valor de y se afasta do alvo A , o valor da desejabilidade irá caindo, tornando-se zero quando um dos limites da faixa de aceitação for alcançado.

Uma vez que as funções de desejabilidade tenham sido especificadas para todas as respostas, deve-se combiná-las numa desejabilidade global (Equação 12), dada pela média geométrica das m desejabilidades individuais. Dessa forma, a otimização simultânea das várias respostas se reduz à maximização de um único valor, a desejabilidade global, e o problema em questão transforma-se em descobrir os níveis dos fatores que maximizem o valor de D . Uma vantagem desta função é que o uso da média geométrica faz com que a desejabilidade global se anule sempre que uma das respostas tiver um valor inaceitável, não importando quão satisfatórios sejam os valores das outras respostas.

$$D = \sqrt[m]{d_1 d_2 \dots d_m} \quad (12)$$

A Figura 9 exemplifica a aplicação do método mostrando os resultados da otimização.

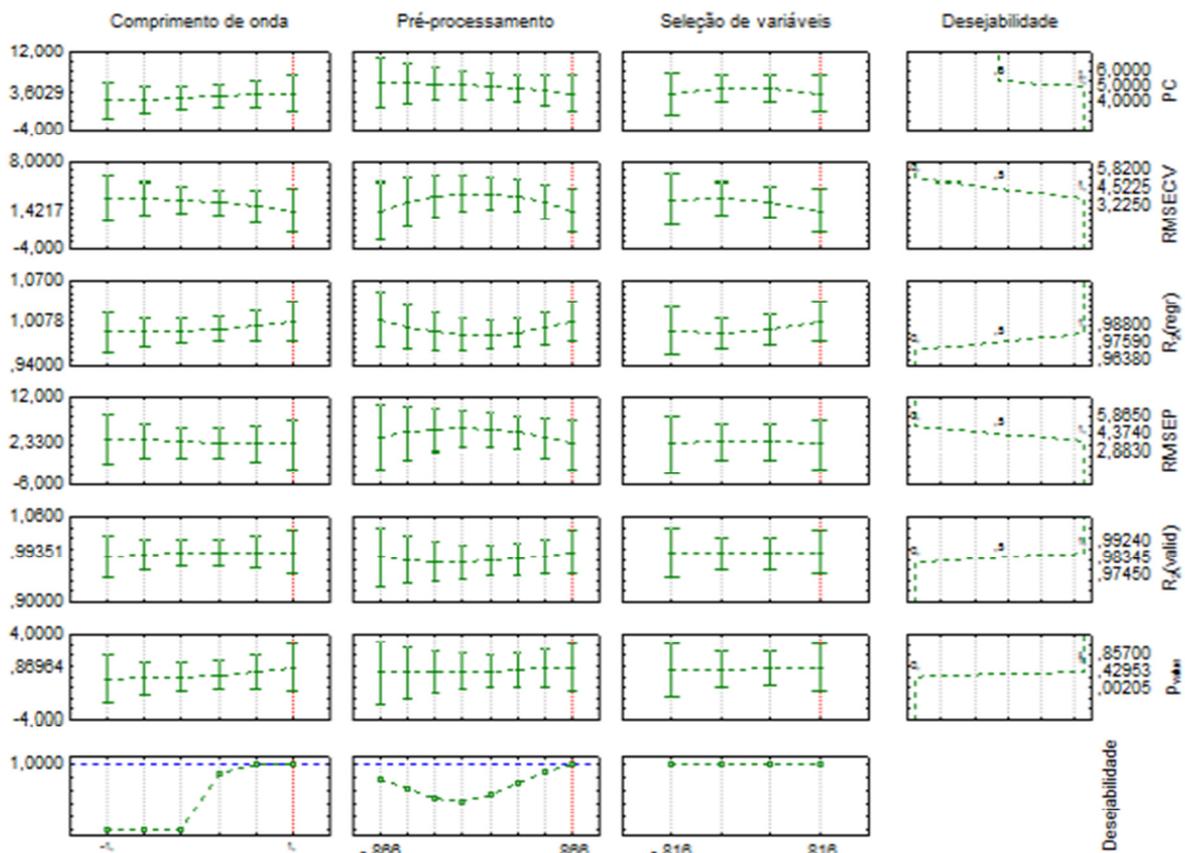


Figura 9. Gráficos da função desejabilidade.

As curvas mostram como as respostas variam com cada fator, mantidos fixos os níveis dos outros fatores nos valores especificados. Os fatores estão apresentados nas primeiras três colunas e as respostas são representadas por cada linha. Portanto, o primeiro gráfico, localizado no canto esquerdo superior representa a influência do comprimento de onda, quando o nível é variado de (-1) a (1), na resposta PC. Os valores alvos para as respostas são mostrados na quarta coluna. Os três últimos gráficos apresentados na última linha mostram a influência dos fatores na desejabilidade global, considerando a combinação das desejabilidades individuais de cada resposta. As linhas tracejadas verticais sinalizam as condições de máxima desejabilidade global, que nesse caso representa 1,0.

2.6. Aplicações da espectroscopia NIR na indústria petroquímica

A espectroscopia NIR, em associação a técnicas de regressão multivariada como, por exemplo, PLS, tem sido utilizada com sucesso para diversos fins na indústria petroquímica. A aplicabilidade do uso desta técnica como método alternativo para o controle da qualidade de produtos petroquímicos, como a gasolina automotiva, já foi suficientemente demonstrada através de inúmeros trabalhos. Este combustível, provavelmente, foi um dos primeiros derivados do petróleo a ser analisado através da espectroscopia NIR. Nestes trabalhos, foram previstas propriedades como octanagem, densidade e teor de aromáticos, etanol e saturados, para aplicações tanto em laboratório como em processo.

Honorato *et al.* (2008, p.3706) desenvolveram modelos de calibração multivariada baseados em dados espectrais nas regiões NIR e MIR (Infravermelho médio, do inglês *Mid Infrared*) para prever as principais propriedades de gasolinas comercializadas na região nordeste do Brasil. Cento e sessenta amostras de gasolina foram coletadas e os modelos de calibração foram construídos considerando-se dados espectrais nas regiões NIR e MIR. Dois algoritmos de calibração (PLS e MLR) e diferentes pré-processamentos (derivadas, alisamentos e seleção de variáveis com o algoritmo genético ou o algoritmo de projeções sucessivas) foram utilizados. Os erros médios quadráticos relativos de previsão para

todos os modelos foram analisados e observou-se que todas as propriedades envolvidas (densidade, curva de destilação - ponto de ebulição inicial, 5%, 10%, 30%, 50%, 70%, 90%, 95% evaporado e ponto final de ebulição – número de octanas – MON e RON – e composição química – parafinas, isoparafinas, aromáticos, naftênicos e olefinas) puderam ser preditas de forma satisfatória a partir do espectro NIR na faixa 1600 a 2500 nm, com calibração por MLR e seleção de variáveis pelo algoritmo genético, com qualquer dos pré-processamentos utilizados.

Parisi *et al.* (1990, p.95) utilizaram a espectroscopia NIR com fibra óptica, em combinação com calibração multivariada, no monitoramento remoto de algumas propriedades da gasolina (MON, RON e composição PONA – parafinas, olefinas, naftênicos e aromáticos) e do índice de cetano no óleo diesel. Para construção do modelo de calibração, 28 amostras foram utilizadas. Para a determinação de MON e RON, os erros absolutos obtidos foram comparáveis aos desvios determinados nos métodos de referência. Para determinação da composição PONA, os erros relativos foram sempre menores que 4%. Para o índice de cetano em óleo diesel, o desvio padrão médio obtido foi comparável à tolerância do método de referência.

Rebouças *et al.* (2007, p.191) utilizaram modelos de calibração multivariada associados à espectroscopia NIR para prever propriedades físicas e químicas de amostras de nafta média em uma unidade de fracionamento de nafta. Propriedades como densidade relativa, curva de destilação (ponto de ebulição inicial, 5%, 10%, 30%, 50%, 70%, 90%, 95% evaporado e ponto final de ebulição) e composição química (parafinas, isoparafinas, naftênicos e aromáticos) foram estimadas a partir de espectros NIR nas regiões entre 1660 a 1800 e 2200 a 2500 nm. PCA foi utilizada para análise exploratória dos dados. Os modelos foram desenvolvidos usando a técnica de regressão PLS. Os erros de previsão obtidos foram próximos aos valores de reprodutibilidade intralaboratorial do método de referência e abaixo dos limites definidos pelo método ASTM. O trabalho demonstrou a viabilidade da realização de diversas análises requeridas pelo controle de qualidade da unidade de fracionamento de nafta baseadas em um único espectro NIR, fornecendo meios para um controle de processo rápido e eficaz.

Fernandes *et al.* (2008, p.804) desenvolveram modelos PLS para determinação simultânea dos teores de metanol e etanol em gasolina. Um conjunto de 120 amostras foi preparado em laboratório, sendo 80 amostras utilizadas na construção dos modelos de calibração multivariada, através da técnica de regressão PLS, e 40 amostras utilizadas para avaliar a capacidade preditiva dos modelos, os quais foram desenvolvidos usando a faixa espectral de 1105 a 1682 nm. Os valores obtidos para o RMSEP foram 0,28% (v/v) para etanol para ambos os modelos PLS1 (*One-block PLS*) e PLS2 (*Two-block PLS*) e 0,31 e 0,32% (v/v) para metanol para os modelos PLS1 e PLS2, respectivamente. O número máximo de 4 PCs foi utilizado nos modelos construídos. Os autores realizaram também um estudo do efeito da variação da composição da gasolina na predição do metanol e etanol, sendo verificado que alguns solventes, como tolueno e o-xileno, interferem na previsão do teor de etanol, enquanto iso-octano, o-xileno, m-xileno e p-xileno interferem na previsão do teor de metanol. Os autores associaram este fato à sobreposição das bandas de absorção. Para resolver esse problema realizaram seleção de variáveis ou faixas espectrais, obtendo modelos mais robustos, demonstrando que é possível determinar simultaneamente metanol e etanol em gasolina.

Rebouças *et al.* (2010, p.97) demonstraram a possibilidade de determinar, de forma rápida e exata, a composição química de diferentes amostras de uma série complexa de diversos processos petroquímicos para a produção de compostos aromáticos, como benzeno, tolueno, o-xileno, p-xileno e xilenos mistos, através de modelos de calibração multivariada associados à espectroscopia NIR. Os modelos foram desenvolvidos visando à predição dos parâmetros químicos aromáticos e não-aromáticos em diferentes correntes de uma unidade de reforma catalítica até a produção final de o-xileno e p-xileno. A faixa espectral utilizada para estimar os teores de não-aromáticos, benzeno, tolueno, etilbenzeno, m-xileno, o-xileno, p-xileno, cumeno, aromáticos C8 e aromáticos C9 foi de 2130 a 2500 nm (4600 a 4000 cm^{-1}). A análise de componentes principais foi utilizada para avaliação exploratória dos dados. A técnica PLS foi empregada para calibração, utilizando-se cerca de 200 amostras reais do processo. Os erros médios quadráticos de predição (RMSEP) obtidos para os modelos de calibração foram comparáveis à reprodutibilidade intralaboratorial do método de referência. O método se mostrou

capaz de diferenciar características espectroscópicas dos isômeros o-xileno, p-xileno e m-xileno.

Rebouças e Barros Neto (2001, p.263) utilizaram modelos de calibração multivariada associados à espectroscopia NIR para prever propriedades físicas de misturas de hidrocarbonetos aromáticos, principalmente C8 – C10. Duas regiões espectrais que se estendem entre 1670 a 1800 nm e 2100 a 2500 nm foram utilizadas para prever parâmetros de destilação (10%, 50%, 90% e 100% evaporado) e a densidade relativa. A técnica PCA foi utilizada para avaliação exploratória dos dados. Os modelos foram calculados utilizando-se PLS1 e PLS2 cujos erros médios de predição foram próximos dos valores de reprodutibilidade intralaboratorial de referência e dos limites definidos pelos métodos ASTM: densidade relativa = 0,0005 mg/L, 10% evaporado = 0,489 °C, 50% evaporado = 0,350 °C, 90% evaporado = 0,579 °C, 100% evaporado = 3,276 °C. O trabalho demonstrou ainda a capacidade das técnicas quimiométricas em diferenciar espectros de compostos aromáticos similares, principalmente devido à natureza e localização dos grupos substituintes no anel aromático.

Chung *et al.* (1998, p.885) desenvolveram um método para determinar simultaneamente a concentração de compostos, tais como o-xileno, m-xileno, p-xileno, etilbenzeno, tolueno, hidrocarbonetos alifáticos e hidrocarbonetos aromáticos C9-C10 no processo de produção de p-xileno através da espectroscopia NIR. Sessenta misturas sintéticas foram preparadas de forma a simular a concentração real destes compostos no processo, sendo quarenta selecionadas para calibração e vinte para os testes de validação. Cinco diferentes faixas espectrais, 1100 a 1270, 1270 a 1650, 1800 a 2100, 1100 a 1650, e a combinação 1100 a 1650 e 1800 a 2100 nm foram examinadas. Mesmo com características espectrais muito semelhantes, os teores dos isômeros o-xileno, p-xileno e m-xileno, e outros compostos aromáticos, foram preditos utilizando o algoritmo PLS, o qual se mostrou preciso e exato para a determinação proposta, exibindo resultados concordantes com os resultados obtidos por cromatografia à gás.

Jun *et al.* (2010, p.79) desenvolveram um método para determinar o teor de olefinas em amostras de gasolina através da espectroscopia NIR. A técnica PLS foi

empregada para calibração, utilizando-se cerca de trinta amostras de gasolina. A faixa empregada para o desenvolvimento do modelo foi de 1000 a 2000 nm. Os valores obtidos para o RMSECV e o coeficiente de determinação de validação cruzada (r^2_{regr}) foram 0,2 %m/m e 0,9881, respectivamente. O valor para o RMSEP foi de 0,39 %m/m. Estes resultados foram concordantes com os resultados obtidos por cromatografia, demonstrando a capacidade do método em prever o teor de olefinas em gasolina.

A maioria dos trabalhos publicados até o momento, os quais abordam o uso da técnica de regressão PLS para calibração multivariada como um método quantitativo de análise, não fornece detalhes a respeito dos procedimentos utilizados para a construção e refinamento dos modelos de calibração. Costuma-se considerar que o desenvolvimento dos modelos de calibração é feito com uma abordagem univariada, em que as variáveis de entrada, tais como, a faixa de comprimento de onda e as técnicas de pré-processamento são sucessivamente variadas. Além disso, na maioria das vezes, a única variável de resposta utilizada como guia para avaliar o desempenho do modelo é o RMSECV (ou RMSEP).

O uso de técnicas de planejamento de experimentos pode fornecer uma sistemática promissora para conduzir os experimentos necessários para o desenvolvimento de modelos PLS. Caneca *et al.* (2006, p.344) empregaram um planejamento fatorial 2^3 para avaliar o procedimento de pré-processamento e seleção de variáveis na capacidade preditiva de modelos MLR, PLS e PCR utilizados no monitoramento de óleos lubrificantes. Uma vez que o planejamento fatorial 2^n foi utilizado, cada variável de entrada foi testada em apenas dois níveis. Os fatores estudados foram: 2ª derivada Savitsky-Golay, suavização Savitsky-Golay (filtro = 5) e seleção de variáveis, através dos algoritmos GA e SPA. Os experimentos foram planejados de forma a se obter o RMSEP como variável de resposta. Esta abordagem para construção dos modelos de calibração multivariada não teve como finalidade fornecer um procedimento de otimização dos modelos, entretanto, proveu informações bastante úteis sobre os efeitos das variáveis estudadas.

O trabalho desenvolvido por Rebouças *et al.* (2011, p.185) propõe uma nova abordagem a respeito do desenvolvimento de modelos de calibração multivariada,

utilizando a técnica de regressão PLS, através de um planejamento de experimentos baseado na matriz Doehlert. Os modelos foram construídos para prever os teores de compostos aromáticos, saturados, olefinas e benzeno, além da densidade relativa a 20/4 °C, em amostras de gasolina. Três variáveis de entrada foram consideradas: faixa de comprimento de onda, técnica de pré-processamento e técnica de regressão, as quais foram variadas em cinco, sete e três níveis, respectivamente. Além dos valores de RMSECV, RMSEP e r^2_{regr} (coeficiente de determinação da validação cruzada), as variáveis de resposta *slope valid* (inclinação da curva de correlação entre os valores preditos e medidos), r^2_{valid} (coeficiente de determinação da validação externa) e p_{value} (obtido no teste t_{pareado}), obtidas através da validação externa, bem como os fatores obtidos com os modelos PLS, foram simultaneamente avaliadas com a finalidade de se encontrar as melhores condições para construção dos modelos PLS. Os pontos ótimos de cada variável de entrada foram simultaneamente definidos por meio de uma abordagem multivariada através da função desejabilidade. Os resultados obtidos foram bastante satisfatórios. A faixa espectral utilizada foi de 7300 a 3900 cm^{-1} . Os valores para o RMSECV variavam de 0,04 a 0,63 %m/m para os teores de aromáticos, saturados, olefinas e benzeno e 0,0008 para a densidade relativa a 20/4 °C. Cada modelo PLS otimizado foi obtido com menos de 40 experimentos, o que demonstra a eficiência da metodologia proposta.

Do exposto, percebe-se que a literatura existente que aborda a utilização da espectroscopia NIR, associada à quimiometria, não traz aplicações desta técnica para a predição da composição de amostras contendo compostos com diferentes grupos funcionais, como por exemplo, hidrocarbonetos, alcoóis, acetona e acetonitrila. Portanto, o desenvolvimento deste trabalho é de fundamental importância, pois além de prover informações até então inexistentes na literatura atual, as quais são de extrema relevância para indústria petroquímica, irá fornecer uma metodologia para a construção e refinamento de modelos de calibração, uma questão até então pouco discutida.

3. EXPERIMENTAL

3.1. Amostras

Para o presente estudo, 75 amostras do solvente foram coletadas, entre agosto de 2009 e dezembro de 2010, no ponto de amostragem codificado como S-52041, na Unidade de Produção de Isopreno da Braskem, localizada no site de Camaçari, Bahia. Todas as amostras foram coletadas em cilindros de aço inox e transferidas para frascos de vidros limpos e secos. Em seguida, as amostras foram armazenadas em geladeira a 4 °C para minimizar as perdas de componentes voláteis, o que alteraria as suas composições e, portanto, seus espectros. Antes da análise, as amostras foram retiradas da geladeira e analisadas somente após atingir o equilíbrio com a temperatura ambiente (22 ± 3 °C).

A composição química típica da amostra do solvente é apresentada na Tabela 3. “Outros” correspondem a outros compostos orgânicos presentes na amostra em baixa concentração, principalmente, hidrocarbonetos e oxigenados. A sua concentração é determinada a partir da diferença dos componentes majoritários.

Uma vez que as amostras reais coletadas no processo não apresentaram variação significativa em sua composição, foram preparadas misturas em laboratório visando atribuir maior variabilidade ao conjunto de dados. Essas misturas visam simular situações de descontrole operacional, atribuindo maior robustez ao conjunto de dados (BARROS NETO; SCARMINIO; BRUNS, 2001). Os níveis inferiores e superiores utilizados como limites para as concentrações dos componentes são mostrados na Tabela 4 e foram baseados na composição atual da corrente do processo para cada componente.

Desta forma, 60 misturas foram preparadas de acordo com as proporções definidas na Tabela 5, obtidas a partir de um planejamento de mistura em rede simplex pseudo-componente (*simplex lattice design pseudo-component*). Uma matriz com 159 misturas foi obtida, em sequência aleatória. Entretanto, somente as 60 primeiras misturas foram selecionadas, a partir do planejamento original em ordem aleatória, no intuito de minimizar o esforço experimental.

Tabela 3. Composição típica das amostras do solvente.

Componente (% m/m)	Valores típicos do processo*		
	Média	Desvio Padrão	Variação
Água	16,5	7,2	6,1 - 42,3
Acetonitrila	42,6	6,4	23,9 - 52,5
Acetona	4,1	1,1	2,2 - 7,7
n-Propanol	1,89	0,22	1,36 - 2,42
i-Propanol	2,16	0,41	1,12 - 2,88
t-Butanol	0,11	0,05	0,01 - 0,24
Ácool Alílico	28,2	2,1	22,6 - 34,1
DCPD** + codímeros	1,36	0,82	0,07 - 3,95
Outros	3,8	1,4	0,8 - 8,0

* Baseados em 134 amostras, por um período de 12 meses (Fevereiro de 2009 a Janeiro de 2010).

** DCPD = dicitlopentadieno-1,3.

Tabela 4. Níveis experimentais para o planejamento das misturas.

Componente (% m/m)	Planejamento Experimental	
	Nível Inferior (-)	Nível Superior (+)
Água	0	46
Acetonitrila	8	90
Acetona	0	20
n-Propanol	0	10
i-Propanol	0	10
t-Butanol	0	20
Ácool Alílico	0	52

Tabela 5. Planejamento experimental das misturas.

Mistura	Água (% m/m)	Acetonitrila (% m/m)	Acetona (% m/m)	n-Propanol (% m/m)	i-Propanol (% m/m)	t-Butanol (% m/m)	Álcool Alílico (% m/m)
1	27,64	19,01	4,29	6,89	2,19	4,28	35,69
2	0,00	89,97	0,00	0,00	0,00	10,03	0,00
3	11,88	9,33	0,00	0,00	9,33	18,74	50,72
4	0,00	9,71	19,52	9,93	9,70	19,49	31,65
5	0,00	9,63	19,35	9,84	0,00	19,33	41,85
6	10,66	19,73	13,95	7,16	2,27	9,18	37,05
7	45,87	27,03	0,00	0,00	9,01	18,09	0,00
8	33,71	28,34	4,32	2,26	6,79	13,53	11,05
9	10,71	19,83	14,02	2,34	7,03	14,00	32,08
10	0,00	28,69	0,00	9,78	9,56	0,00	51,97
11	45,78	17,98	0,00	9,20	8,99	18,05	0,00
12	10,88	49,10	4,54	7,31	2,32	14,22	11,62
13	33,69	37,49	4,32	6,94	2,21	4,31	11,04
14	10,67	33,97	4,46	2,33	7,01	4,45	37,12
15	33,06	18,80	4,24	2,21	2,16	4,23	35,29
16	10,89	53,99	4,55	2,38	2,32	14,24	11,63
17	44,81	8,80	17,70	0,00	0,00	0,00	28,70
18	45,16	8,87	0,00	0,00	8,87	17,81	19,29
19	10,89	44,31	14,25	2,37	2,32	14,23	11,62
20	0,00	90,00	0,00	0,00	10,00	0,00	0,00
21	10,66	19,74	9,20	7,16	2,27	13,93	37,05
22	0,00	59,80	20,04	10,19	9,96	0,00	0,00
23	45,06	8,85	17,80	9,05	0,00	0,00	19,24
24	0,00	69,77	20,04	10,19	0,00	0,00	0,00
25	35,16	9,21	18,52	9,42	9,21	18,49	0,00
26	44,17	8,68	0,00	0,00	0,00	0,00	47,15
27	10,67	29,22	13,97	2,33	2,27	4,45	37,10
28	33,66	19,14	13,53	6,94	2,20	13,51	11,03
29	28,18	19,38	13,70	7,03	6,87	13,68	11,17
30	45,78	26,98	0,00	9,20	0,00	18,05	0,00

31	0,00	28,72	19,25	0,00	0,00	0,00	52,03
32	21,09	29,71	8,80	4,60	4,50	8,79	22,51
33	0,00	9,55	9,60	9,77	0,00	19,17	51,91
34	33,19	18,87	4,25	2,22	6,69	4,25	30,52
35	33,67	23,73	13,53	6,94	6,79	4,31	11,03
36	10,66	24,46	13,95	7,16	2,27	4,44	37,06
37	10,66	19,74	13,95	7,16	7,00	4,44	37,06
38	10,90	58,83	9,41	2,38	2,32	4,54	11,63
39	45,76	26,97	18,07	9,19	0,00	0,00	0,00
40	11,85	9,31	18,72	9,52	0,00	0,00	50,60
41	33,71	32,93	4,32	2,26	2,21	13,53	11,05
42	33,41	19,00	4,28	6,89	2,19	13,41	20,83
43	33,69	32,91	4,32	6,94	6,79	4,31	11,04
44	22,14	19,26	4,34	2,27	2,22	13,60	36,17
45	10,66	33,93	4,45	7,16	2,27	4,45	37,08
46	45,73	8,98	18,06	9,19	0,00	18,03	0,00
47	0,00	19,15	0,00	0,00	9,57	19,23	52,05
48	33,66	19,14	13,53	6,94	6,78	8,91	11,03
49	33,31	18,94	13,38	2,23	2,18	4,26	25,70
50	10,88	39,43	14,24	7,30	2,32	14,22	11,61
51	10,90	58,83	4,55	2,38	2,32	9,39	11,63
52	0,00	79,94	0,00	0,00	0,00	20,06	0,00
53	33,54	19,07	13,48	6,91	6,76	4,29	15,95
54	33,93	8,89	0,00	0,00	8,88	0,00	48,30
55	11,85	9,31	0,00	9,52	0,00	18,70	50,61
56	0,00	9,57	19,24	0,00	0,00	19,21	51,99
57	0,00	19,15	19,25	0,00	9,57	0,00	52,03
58	10,89	49,16	4,55	2,38	7,15	14,24	11,63
59	10,89	39,48	14,25	2,37	7,15	14,23	11,62
60	16,46	19,49	13,77	7,07	2,24	4,39	36,59

Um conjunto de 15 misturas adicionais foi planejado para obter variação na composição do componente DCPD + codímeros (Tabela 6). Inicialmente, esse parâmetro estaria incluído no resultado do constituinte 'Outros' e, por isso, não foi

previsto no planejamento inicial das misturas. Entretanto, ao longo do estudo, percebeu-se que seria uma boa oportunidade fornecer os resultados para esse parâmetro já que o mesmo é utilizado no controle do processo.

Para obtenção das misturas contendo o componente DCPD foram utilizadas como base as proporções definidas pelas misturas 45, 50 e 60, as quais foram escolhidas aleatoriamente. As concentrações do componente DCPD foram fixadas em cinco níveis: 2, 4, 6, 8 e 10 %v/v, de forma a se incluir a variação natural do processo conforme apresentado na Tabela 3. Para cada uma das 3 misturas selecionadas foi calculado o volume necessário de DCPD para obter uma mistura final com a concentração requerida desse componente, obtendo-se também a concentração final dos demais componentes após adição do DCPD. Os resultados para este planejamento estão apresentados na Tabela 6.

Tabela 6. Planejamento experimental das misturas incluindo o componente DCPD.

Mistura	Água (% m/m)	Acetonitrila (% m/m)	Acetona (% m/m)	n-Propanol (% m/m)	i-Propanol (% m/m)	t-Butanol (% m/m)	Álcool Alílico (% m/m)	DCPD (% m/m)
1	10,30	33,12	4,44	7,15	2,21	4,43	36,00	2,34
2	10,06	32,33	4,33	6,98	2,15	4,33	35,14	4,67
3	9,82	31,55	4,23	6,81	2,10	4,22	34,29	6,98
4	9,57	30,77	4,12	6,64	2,05	4,12	33,45	9,27
5	9,33	30,00	4,02	6,48	2,00	4,02	32,61	11,55
6	10,50	38,26	13,95	7,29	2,25	13,93	11,42	2,39
7	10,25	37,34	13,62	7,11	2,20	13,60	11,14	4,76
8	10,00	36,41	13,28	6,94	2,14	13,26	10,87	7,11
9	9,75	35,50	12,95	6,76	2,09	12,93	10,59	9,44
10	9,50	34,59	12,62	6,59	2,03	12,60	10,32	11,75
11	16,17	18,88	13,51	7,06	2,18	4,37	35,52	2,31
12	15,79	18,44	13,19	6,89	2,13	4,27	34,68	4,61
13	15,42	18,00	12,87	6,72	2,08	4,17	33,86	6,89
14	15,04	17,56	12,56	6,56	2,03	4,07	33,03	9,16
15	14,67	17,12	12,25	6,40	1,98	3,97	32,21	11,41

Após o preparo das misturas, gráficos de concentração das propriedades, incluindo amostras reais do processo e misturas, foram construídos de forma a avaliar se o conjunto de dados se distribuía uniformemente em todo o domínio experimental (Tabela 3). Com isso, 25 misturas adicionais foram preparadas de forma a garantir que toda a faixa de trabalho fosse preenchida. As proporções de cada uma das 25 misturas estão definidas na Tabela 7.

Tabela 7. Planejamento experimental das 25 misturas adicionais.

Mistura	Água (% m/m)	Acetonitrila (% m/m)	Acetona (% m/m)	n-Propanol (% m/m)	i-Propanol (% m/m)	t-Butanol (% m/m)	Álcool Alílico (% m/m)	DCPD (% m/m)
1	1,26	86,67	1,00	1,01	0,99	1,99	6,46	0,62
2	2,52	84,18	2,99	3,54	3,47	0,99	1,08	1,24
3	3,74	80,92	0,00	5,52	0,00	2,95	3,20	3,67
4	4,97	75,59	5,88	0,00	5,36	0,00	2,12	6,08
5	6,10	50,83	7,23	8,34	6,23	5,29	5,21	10,77
6	7,15	15,91	8,00	9,09	7,95	6,11	37,63	8,17
7	0,00	72,03	0,00	0,00	9,44	10,97	7,56	0,00
8	8,74	73,53	10,84	0,00	0,00	6,89	0,00	0,00
9	0,00	62,20	15,88	0,00	0,00	8,42	12,88	0,62
10	30,68	60,26	0,00	0,00	0,00	0,00	9,07	0,00
11	18,85	59,71	17,29	0,00	0,00	0,00	4,15	0,00
12	0,00	52,05	0,00	0,00	0,00	2,90	45,05	0,00
13	22,11	11,43	4,14	0,00	4,11	11,01	47,20	0,00
14	28,72	13,54	0,00	0,00	0,00	13,59	44,15	0,00
15	12,27	77,14	0,00	0,00	0,00	0,00	9,75	0,84
16	0,00	23,13	16,47	1,97	0,00	15,48	42,95	0,00
17	25,57	44,39	17,85	0,00	0,00	0,00	12,19	0,00
18	35,29	64,71	0,00	0,00	0,00	0,00	0,00	0,00
19	42,77	25,54	0,00	0,00	0,00	17,09	14,61	0,00
20	0,00	87,60	3,00	0,00	1,99	2,00	5,41	0,00
21	3,08	49,82	9,43	1,48	1,45	3,88	30,49	0,36
22	0,00	55,07	1,96	2,49	7,31	3,42	29,14	0,61

23	37,59	12,98	0,00	0,00	8,05	14,82	24,32	2,23
24	1,21	10,48	16,77	7,80	2,86	10,05	50,24	0,59
25	0,00	52,87	3,44	3,50	2,94	14,74	21,28	1,22

Assim, o conjunto de dados foi constituído de 175 amostras, sendo 75 amostras reais do processo e 100 misturas preparadas em laboratório.

3.2. Efeito da temperatura da amostra no espectro NIR

Um estudo do efeito da variação da temperatura da amostra no espectro NIR foi realizado em sete níveis de temperatura: 4, 6, 10, 20, 30, 40 e 50 °C. Um banho de gelo e uma placa aquecedora foram utilizados para resfriar e aquecer a amostra, respectivamente, até que a temperatura desejada fosse atingida. A temperatura da amostra foi controlada através de um termômetro. Gráficos das diferenças entre os espectros obtidos nos níveis: 4, 6, 10, 30, 40 e 50 °C e o espectro obtido a 20 °C foram construídos e as diferenças espectrais foram analisadas.

3.3. Reagentes utilizados

- Clorofórmio p.a. para limpeza da célula do FT-NIR (J. T. Baker, 99,99 %m/m);
- Acetona (Merck, 99,8 %m/m);
- Acetonitrila (J. T. Baker, 99,99 %m/m);
- n-propanol (Merck, 99,5 %m/m);
- i-propanol (Merck, 99,8 %m/m)
- t-butanol (merck, 99,5 %m/m);
- Álcool Alílico (Merck, 99,0 %m/m);
- DCPD (Produto produzido pela Braskem, 83,71 %m/m).

3.4. Instrumentação e condições de operação

A composição das amostras foi determinada por cromatografia em fase gasosa através da injeção de 0,3 µL de amostra em um cromatógrafo, modelo 7890A fabricado pela Agilent Technologies (Estados Unidos), equipado com detector de condutividade térmica e coluna capilar Carbowax (polietilenoglicol) Agilent Technologies (Estados Unidos), com 30 m de comprimento, 0,53 mm de diâmetro interno e 1,3 µm de espessura de filme. A temperatura de operação do forno variou de 40 a 200 °C, de acordo com a seguinte programação: 40 °C/ 5 min; 120 °C/ 0 min e 200 °C/ 7 min. A taxa de aquecimento foi de 20 °C/min. Hidrogênio 4.5 fornecido pela White Martins (Brasil) foi utilizado como gás de arraste. As temperaturas do injetor e do detector foram mantidas a 200 °C e a pressão na coluna a 2 psi.

Os espectros de transmitância no infravermelho próximo foram registrados na região de 15797,77 a 3,86 cm⁻¹, e convertido automaticamente em unidades de absorvância. Os espectros foram obtidos como média de 32 varreduras em resolução de 4 cm⁻¹, utilizando o espectro do ar ambiente como referência. As regiões abaixo de 3700 cm⁻¹ e acima de 10000 cm⁻¹ foram eliminadas por conter alto ruído e não conter informações espectroscópicas relevantes. A obtenção dos espectros foi realizada utilizando-se espectrômetro NIR com transformada de Fourier série MB-160 (Bomem/ Hartmann & Braun; Canadá). Utilizou-se sistema de detecção de arseneto de índio (InAs), sendo a amostra acondicionada em célula de Fluoreto de Cálcio (CaF₂) modelo SL-4 com caminho ótico de 0,5mm (International Crystal Labs, Nova Jersey, Estados Unidos).

3.5. Programas estatísticos

O programa STATISTICA, versão 6.0 (Stasoft Inc., Tulsa, EUA), foi utilizado na obtenção do planejamento das misturas, no tratamento dos resultados do planejamento de experimentos da matriz Doehlert e na aplicação da função desejabilidade.

O programa MATLAB®, versão 6.5, foi utilizado para executar o algoritmo SPXY para a seleção de amostras do conjunto de calibração e validação. Para seleção de variáveis espectrais foi utilizado o pacote PLS-Toolbox 2.1 baseado no método PLS1.

Para realização da análise exploratória dos dados e a regressão por mínimos quadrados parciais, utilizou-se o programa UNSCRAMBLER®, versão 9.7 (CAMO Process A/S).

3.6. Construção dos modelos: Calibração multivariada

Antes da construção dos modelos, foi realizada a análise exploratória dos dados para detecção de amostras anômalas, utilizando a Análise de Componentes Principais (PCA). Utilizando o algoritmo SPXY, o conjunto de amostras foi dividido nos conjuntos de calibração e validação. Posteriormente, as faixas espectrais relevantes para cada constituinte em estudo foram selecionadas através da aplicação do algoritmo PLS. Para otimização do processo de calibração multivariada e definição da melhor condição de modelagem foi realizado um planejamento experimental utilizando a matriz Doehlert. As variáveis selecionadas para o planejamento foram estudadas de acordo com os níveis previamente escolhidos. Desta forma, foram realizados os experimentos definidos pelo planejamento experimental, obtendo as respostas para os experimentos planejados. Para avaliar essas respostas, foram construídos gráficos de desejabilidade, com objetivo de obter os pontos ótimos para cada variável estudada. Em seguida, foram obtidos os modelos de regressão PLS nessas condições. Posteriormente, os modelos foram refinados, excluindo-se amostras com alto resíduo de concentração e espectral. O modelo refinado foi então submetido à etapa de validação estatística externa, em que o modelo de calibração construído foi comparado com o método de referência. Na Figura 10 está ilustrado o fluxograma das etapas utilizadas para o desenvolvimento dos modelos de calibração e o detalhamento de cada uma das etapas é descrito nas seções seguintes.

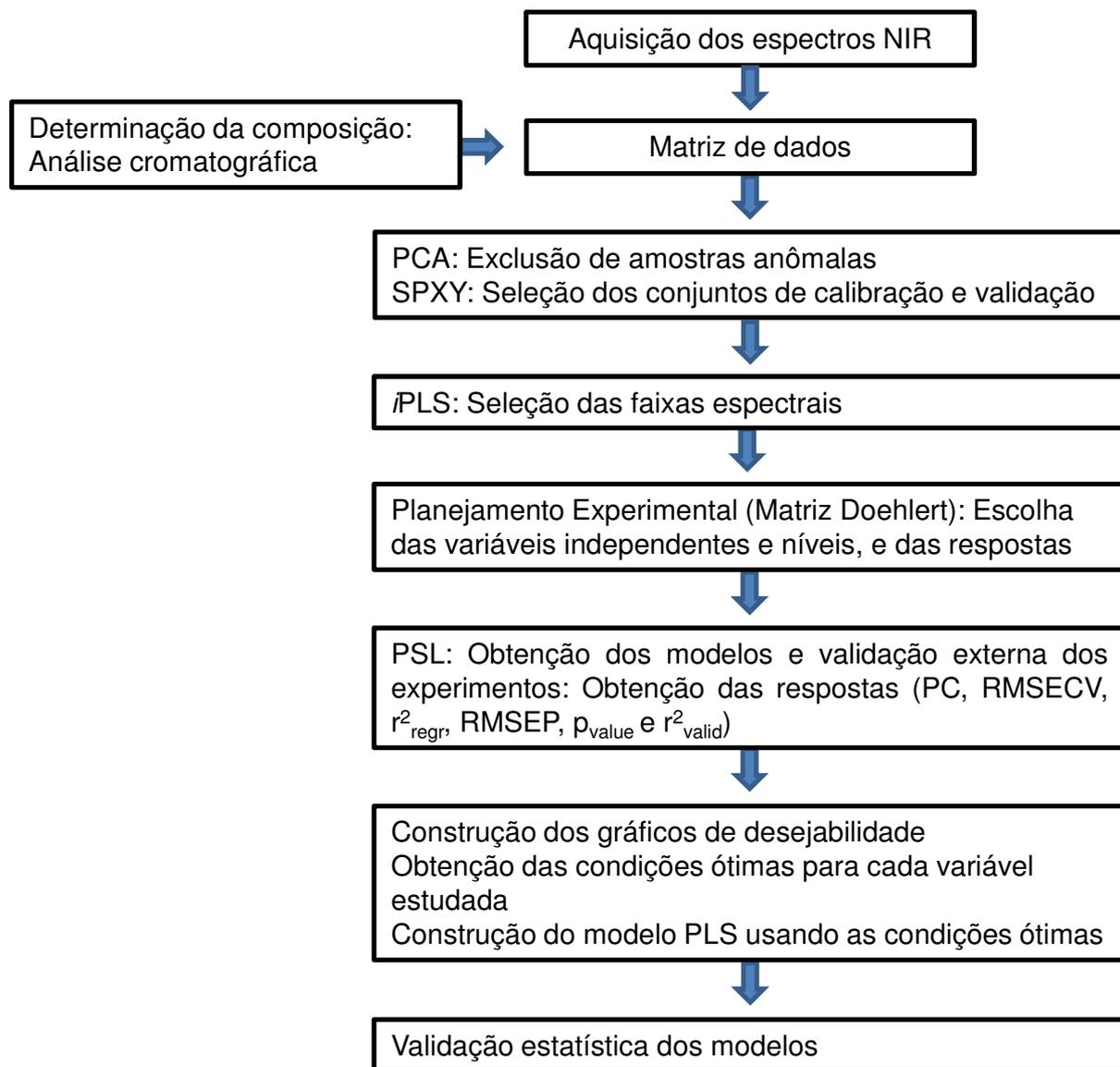


Figura 10. Principais etapas desenvolvidas para a obtenção do melhor modelo de calibração.

3.6.1. Detecção de amostras anômalas (*outliers*)

Durante o processo de construção dos modelos de calibração, ferramentas estatísticas foram empregadas para identificar amostras que possuíam um comportamento atípico, não característico do conjunto de dados, as quais podem influenciar negativamente os modelos de calibração.

Neste trabalho, a identificação de *outliers* no conjunto de dados foi realizada através da PCA (BEEBE; RANDY; SEASHOLTZ, 1998) utilizando as ferramentas: 1)

gráficos dos *scores* para as duas primeiras componentes da PCA (PC2 vs PC1), 2) Teste *Hotelling T²* através da elipse de confiança (95%) e 3) gráficos de resíduo vs *leverage*. Amostras que se encontraram muito distantes das demais no gráfico de *scores*, fora da elipse de confiança e com alto resíduo foram descartadas.

3.6.2. Seleção dos conjuntos de amostras de calibração e validação

A seleção do conjunto de dados para calibração e validação foi realizada através do algoritmo SPXY (GALVÃO *et al.*, 1999), sendo definidas as proporções de 65% do total de amostras para o conjunto de calibração e 35% para o conjunto de validação. A análise gráfica da distribuição das amostras na faixa de concentração também foi utilizada para garantir que os conjuntos de calibração e validação estivessem homogeneamente distribuídos no domínio experimental, sendo realizado o ajuste dos conjuntos quando necessário.

3.6.3. Seleção das faixas espectrais

Para a seleção das faixas espectrais, foi utilizado o algoritmo *i*PLS através do software MATLAB® 6.5, pacote PLS-Toolbox 2.1. Para isso, o espectro derivativo foi dividido em 10 intervalos equidistantes e modelos PLS foram construídos nesses intervalos. As faixas espectrais que resultaram em modelos PLS com menor RMSECV foram selecionadas para definição dos níveis no planejamento experimental.

3.6.4. Planejamento de experimentos empregando a matriz Doehlert

O processo de construção dos modelos de calibração foi otimizado através do planejamento experimental baseado na matriz Doehlert, em que as melhores condições de modelagem PLS para cada propriedade foram definidas. As variáveis de entrada estudadas foram: 1) Técnica de pré-processamento; 2) Faixa de

comprimento de onda e 3) Seleção de variáveis espectrais através do algoritmo JackKnife. A faixa de comprimento de onda e a técnica de pré-processamento de dados foram consideradas as variáveis mais importantes e, por isso, foram definidas na matriz Doehlert com o número máximo de níveis, cinco e sete, respectivamente. A aplicação da técnica de seleção das variáveis espectrais com o algoritmo JackKnife foi estudada em três níveis. Para o planejamento experimental Doehlert, com três variáveis independentes, foram necessários 13 experimentos. A ordem da realização dos experimentos foi aleatória. Na Tabela 8 encontra-se a matriz Doehlert aplicada com os valores reais das variáveis independentes e entre parênteses suas respectivas coordenadas.

Tabela 8. Matriz Doehlert para as variáveis: comprimento de onda, pré-processamento e seleção de variáveis, com os valores reais e suas respectivas coordenadas.

Experimento	Variáveis Independentes		
	Faixa de comprimento de onda	Técnica de pré-processamento ^a	Seleção de variáveis ^b
1	Selecionada iPLS 1 (0)	PRE4 (0)	Com 1 Jack Knife (0)
2	Selecionada iPLS 1 (0)	PRE2 (-0,577)	Sem Seleção (-0,816)
3	Selecionada iPLS 2 (1)	PRE4 (0)	Com 1 Jack Knife (0)
4	Selecionada iPLS 1 (0)	PRE6 (0,577)	Sem Seleção (-0,816)
5	Selecionada iPLS 3 (-1)	PRE4 (0)	Com 1 Jack Knife (0)
6	Selecionada iPLS 4 (-0,5)	PRE1 (-0,866)	Com 1 Jack Knife (0)
7	Selecionada iPLS 5 (0,5)	PRE1 (-0,866)	Com 1 Jack Knife (0)
8	Selecionada iPLS 5 (0,5)	PRE5 (0,289)	Com 2 Jack Knife (0,816)
9	Selecionada iPLS 4 (-0,5)	PRE3 (-0,289)	Sem Seleção (-0,816)
10	Selecionada iPLS 4 (-0,5)	PRE5 (0,289)	Com 2 Jack Knife (0,816)
11	Selecionada iPLS 5 (0,5)	PRE3 (-0,289)	Sem Seleção (-0,816)
12	Selecionada iPLS 5 (0,5)	PRE7 (0,866)	Com 1 Jack Knife (0)
13	Selecionada iPLS 4 (-0,5)	PRE7 (0,866)	Com 1 Jack Knife (0)

^aDetalhes dos códigos de pré-processamento de dados na Tabela 9.

^bQuantidade de rodadas do algoritmo JackKnife para seleção de variáveis.

As técnicas de pré-processamento utilizadas estão resumidas na Tabela 9. Os níveis utilizados para o algoritmo Jack Knife corresponderam a: não utilização da técnica, utilização da técnica uma vez e utilização da técnica duas vezes consecutivas.

Tabela 9. Códigos referentes aos pré-processamentos utilizados.

Código	Pré-processamento ^a
PRE1	Nenhum
PRE2	SNV (normalização padrão)
PRE3	Suavização Savitsky-Golay (filtro = 3)
PRE4	Normalização (área)
PRE5	Normalização + Correção da linha de base
PRE6	1ª derivada Savitsky-Golay (número de pontos da janela = 21)
PRE7	2ª derivada Savitsky-Golay (número de pontos da janela = 21)

^aEm todos os experimentos foi aplicado o autoescalamento.

3.6.5. Otimização das variáveis de resposta: Função desejabilidade

As variáveis de resposta do planejamento experimental foram estabelecidas com base nas duas etapas de validação do modelo (interna e externa). Os parâmetros utilizados para monitorar o desempenho do modelo na etapa de validação interna foram o número de componentes principais (PC's), RMSECV e r^2_{regr} (coeficiente de determinação da validação cruzada). Já os parâmetros RMSEP, r^2_{valid} (coeficiente de determinação da validação externa), *slope valid* (inclinação da curva de correlação entre os valores preditos e medidos) e p_{value} (obtido no teste $t_{pareado}$) foram utilizados para avaliação do desempenho da validação externa.

Para cada propriedade modelada, os parâmetros de resposta foram avaliados simultaneamente através da função de desejabilidade, utilizando-se o *software* STATISTICA. Dentre os valores obtidos para as variáveis de resposta, através dos experimentos realizados para criação dos modelos de calibração, buscou-se um valor alvo mínimo para as variáveis PC, RMSECV e RMSEP. Para as variáveis r^2_{regr}

e r^2_{valid} , buscou-se um valor máximo. Para a variável p_{value} , buscou-se valores maiores que 0,05. O valor máximo possível para a função desejabilidade é 1.

Uma vez que as funções de desejabilidade foram especificadas para todas as respostas, a desejabilidade global foi obtida e as condições ideais de cada variável de entrada foram estabelecidas. Com isso, um novo modelo de calibração foi obtido nessas condições e sucessivamente refinado, buscando obter um melhor desempenho do modelo.

3.6.6. Validação estatística externa dos modelos

Após refinamento, o modelo foi submetido ao procedimento de validação estatística, em que o conjunto de dados de validação (item 3.9) foi usado para avaliar a exatidão dos modelos. O desempenho dos modelos foi avaliado através da comparação entre os valores preditos e medidos usando o teste t_{pareado} . O erro de predição (RMSEP) foi comparado com a precisão intermediária do método de referência.

4. RESULTADOS E DISCUSSÃO

4.1. Características espectrais do solvente na região NIR

O espectro NIR na região entre 10000 - 3700 cm^{-1} de uma amostra típica do solvente é mostrado na Figura 11. A região de 10000 a 7500 cm^{-1} não contém informação espectroscópica relevante nas condições utilizadas, com bandas fracas atribuídas ao 2º sobretom de vibrações fundamentais de estiramento O-H e C-H. A banda larga observada na região entre 7500 – 6500 cm^{-1} está relacionada, principalmente, ao 1º sobretom de bandas fundamentais de estiramento O-H, com contribuições do 3º sobretom da banda fundamental de estiramento C=O e combinação C-H, decorrentes dos componentes água, alcoóis e acetona presentes na amostra. Na região de 6500 a 5500 cm^{-1} , bandas de absorção relativas ao 1º sobretom de vibrações fundamentais de estiramento O-H e C-H, incluindo a ligação C-H do grupo metil adjacente à carbonila, podem ser observadas. A região entre 5500 e 4600 cm^{-1} mostra uma banda de absorção intensa atribuída à combinação de estiramento O-H e deformação angular H-O-H da água. O 2º sobretom da banda fundamental de estiramento C=O também está presente nesta região. As bandas de combinação C-H dominam a região de 4600 a 3900 cm^{-1} , associadas aos hidrocarbonetos, bem como aos grupos alquila presentes nas estruturas dos componentes da amostra. Essa região também engloba o 1º sobretom da vibração fundamental de estiramento da ligação C≡N, relacionada à presença da acetonitrila.

Os espectros NIR dos componentes presentes na amostra podem ser visualizados nas Figuras 12 a 19. É possível notar diferenças espectrais acentuadas entre os componentes, até mesmo quando compostos com o mesmo grupo funcional, por exemplo, alcoóis, são comparados. Com isso, estabelece-se a base para que a regressão PLS nos espectros NIR seja bem sucedida. A região espectral entre 7500 e 3700 cm^{-1} foi utilizada como ponto de partida para calibração, uma vez que fornece as informações espectrais mais relevantes dos compostos presentes na amostra.

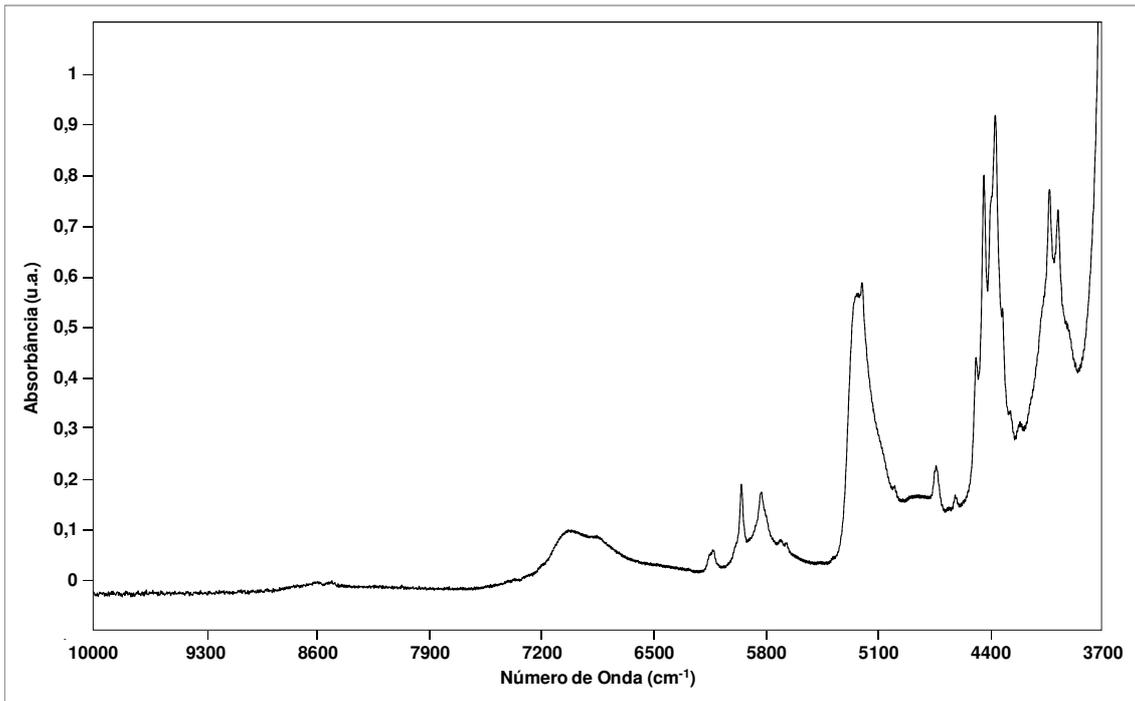


Figura 11: Espectro NIR de uma amostra típica do solvente na região entre 10000 e 3700 cm⁻¹.

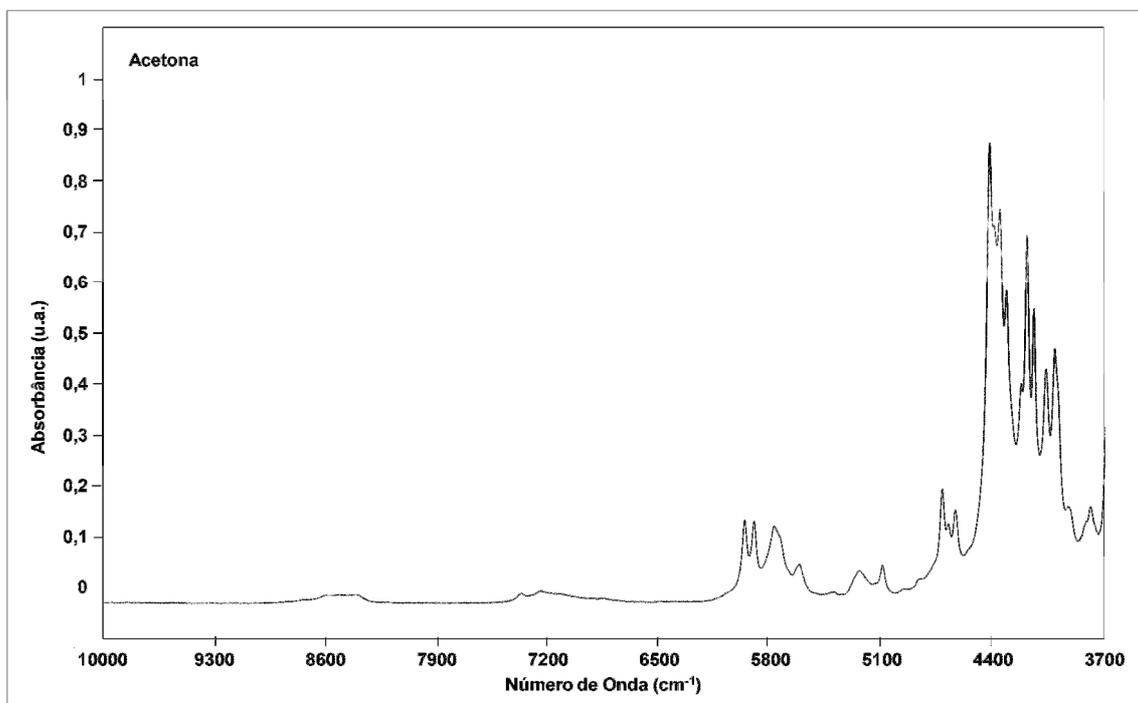


Figura 12. Espectro NIR da acetona pura na região entre 10000 e 3700 cm⁻¹.

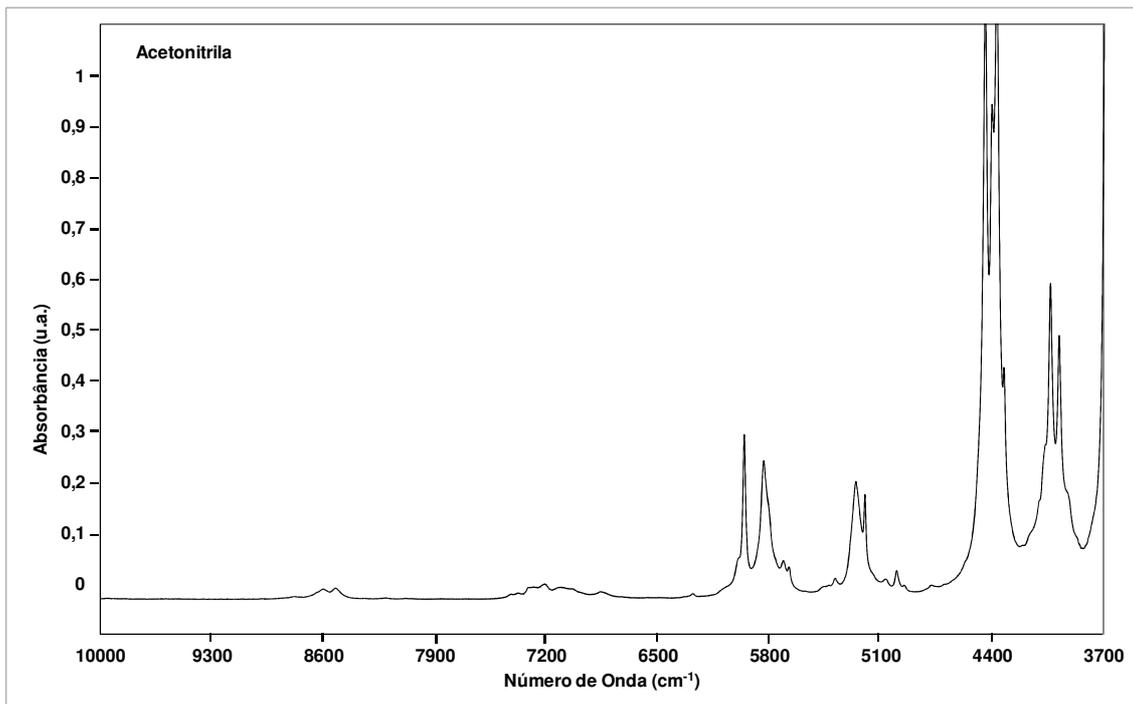


Figura 13. Espectro NIR da acetonitrila pura na região entre 10000 e 3700 cm⁻¹.

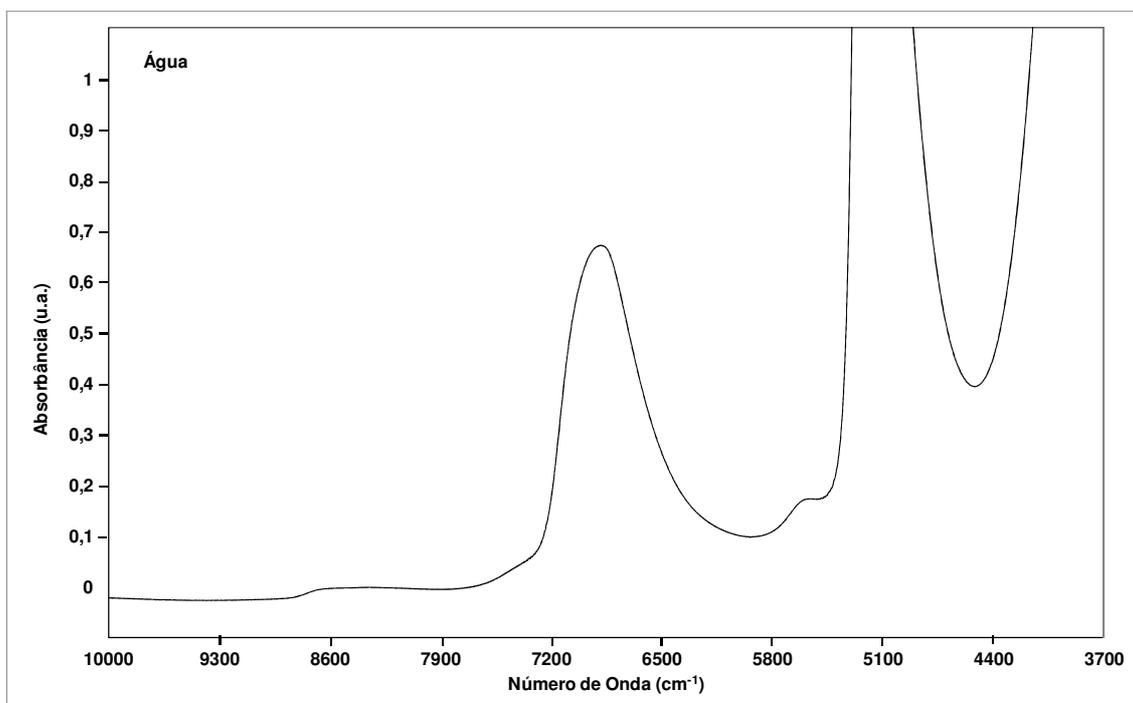


Figura 14. Espectro NIR da água pura na região entre 10000 e 3700 cm⁻¹.

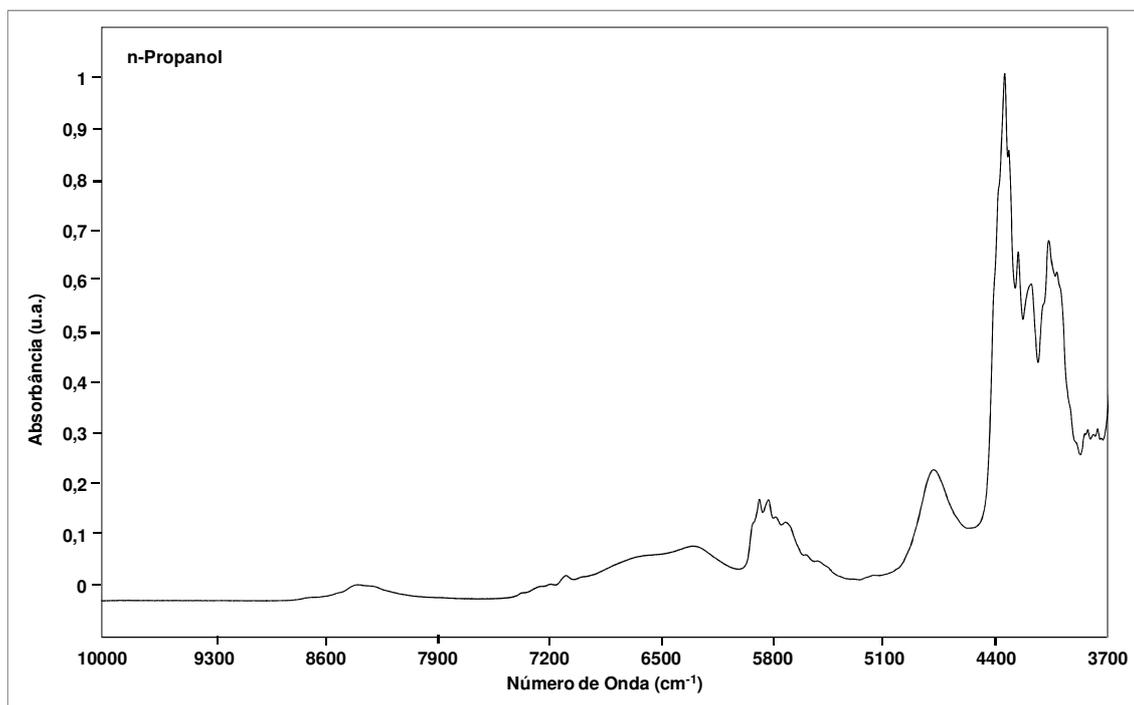


Figura 15. Espectro NIR do n-propanol puro na região entre 10000 e 3700 cm⁻¹.

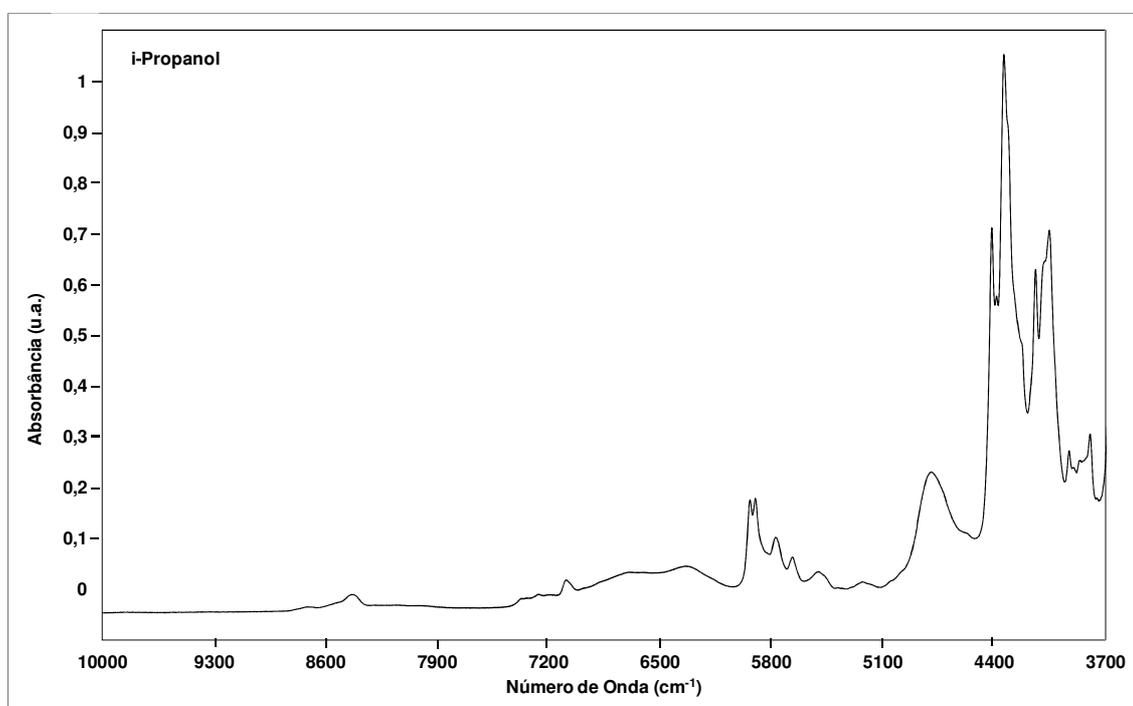


Figura 16. Espectro NIR do i-propanol puro na região entre 10000 e 3700 cm⁻¹.

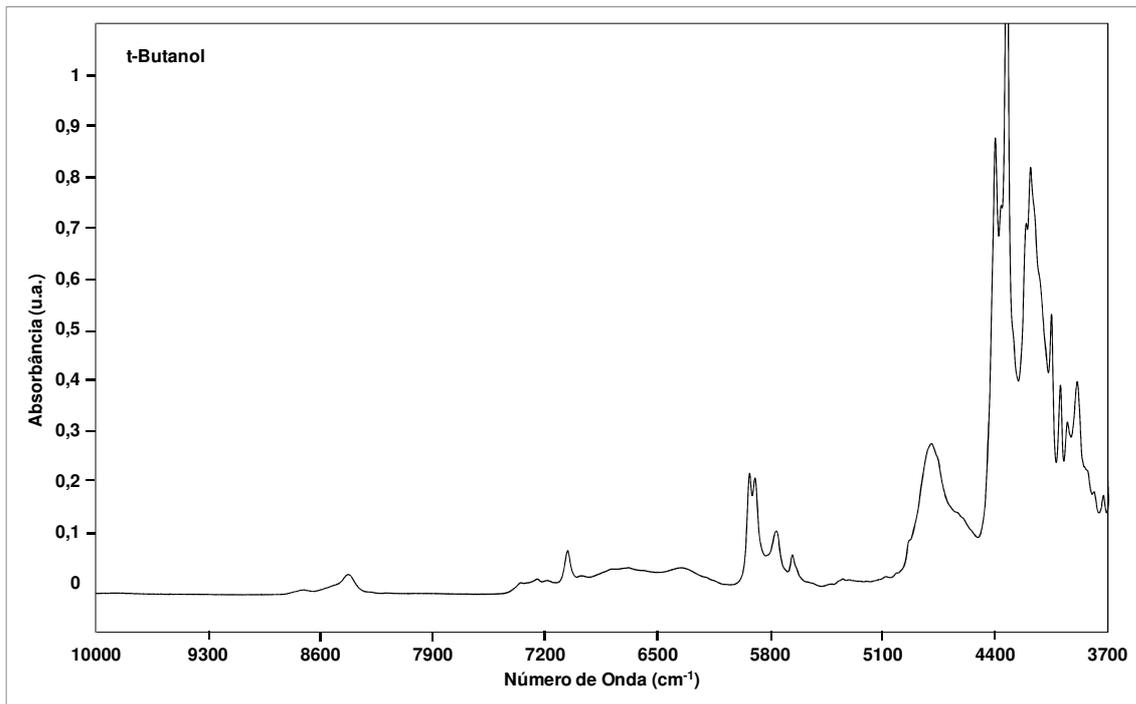


Figura 17. Espectro NIR do t-butanol puro na região entre 10000 e 3700 cm⁻¹.

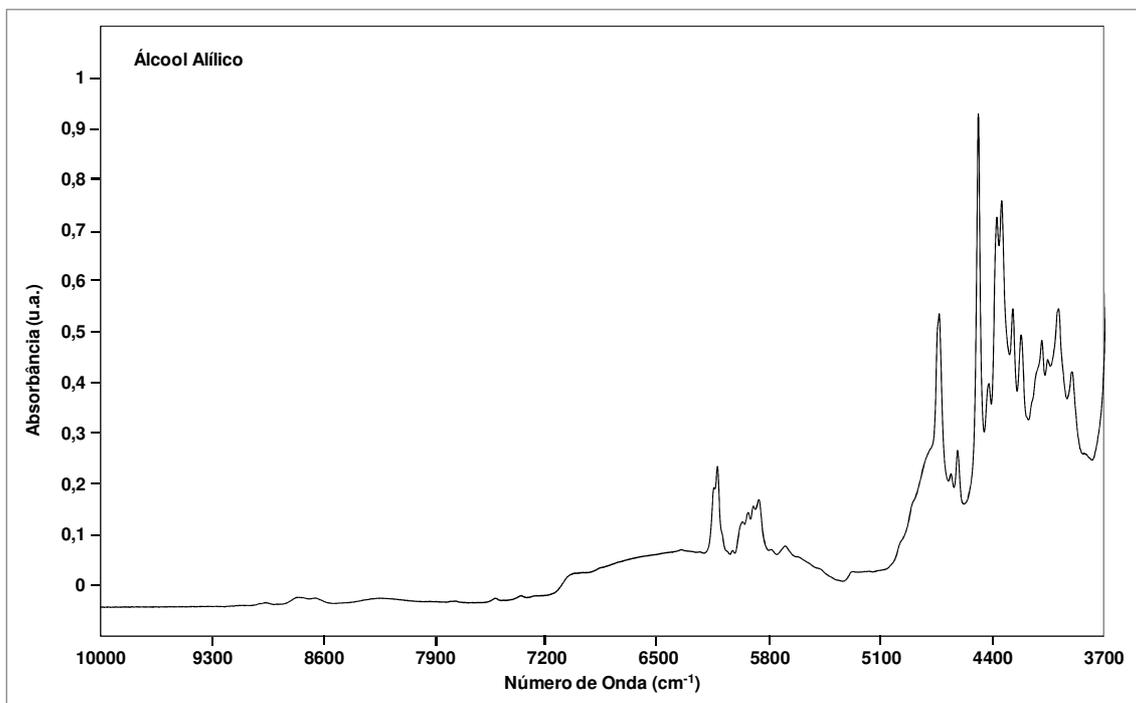


Figura 18. Espectro NIR do álcool alílico puro na região entre 10000 e 3700 cm⁻¹.

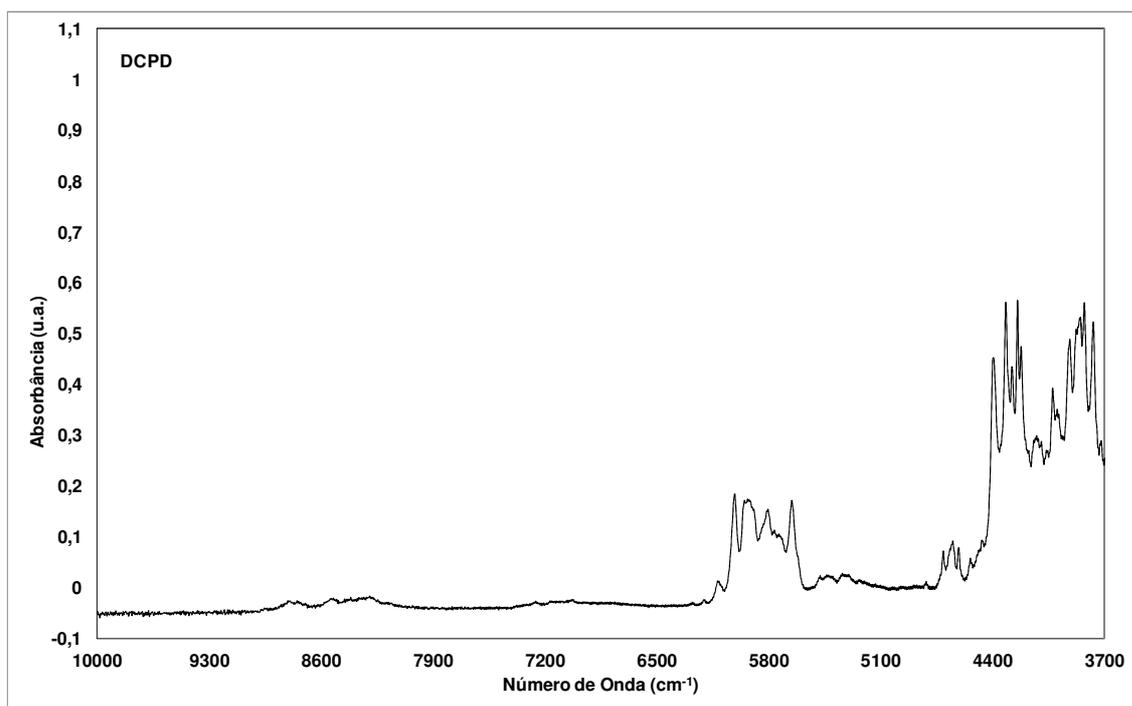


Figura 19. Espectro NIR do DCPD puro na região entre 10000 e 3700 cm⁻¹.

4.2. Efeito da temperatura da amostra no espectro NIR

O efeito da temperatura no espectro NIR da amostra foi avaliado nos níveis 4, 6, 10, 20, 30, 40 e 50 °C. Os resultados das diferenças entre os espectros coletados a 4 e 50 °C e o espectros coletado na temperatura ambiente são apresentados na Figura 20. Os resultados para as outras temperaturas foram omitidos, uma vez que nenhuma variação de absorbância fora da faixa de $\pm 0,01$ foi observada em toda faixa espectral. As intensidades positivas representam áreas no espectro NIR que aumentam em intensidade com o aumento da temperatura, enquanto que as intensidades negativas correspondem a áreas no espectro NIR que possuem correlação negativa com a temperatura (LIBNAU *et al.*, 1994, p.243). O impacto mais significativo foi observado na região entre 5500 – 4000 cm⁻¹, quando o espectro da amostra é obtido a 50,9 °C, particularmente a 5300, 5200, 4400 e 4350 cm⁻¹. A variação observada a 5300 cm⁻¹ pode ser explicada pelo deslocamento da banda de combinação das vibrações fundamentais de estiramento O-H e deformação angular O-H com o aumento da temperatura (MAEDA *et al.*, 1999,

p.121; CZARNECKI; WOJTKÓW, 2008, p.203). Esta banda ocorre a 5200 cm^{-1} na região NIR (WORKMAN, 2001, p.164), entretanto, com aumento da temperatura, o deslocamento das bandas para números de onda maiores é favorecido (LIBNAU *et al.*, 1994, p.243). A intensidade positiva da variação a 5300 cm^{-1} sugere que espécies OH ligadas por pontes de hidrogênio se dissociam com o aumento da temperatura (MAEDA *et al.*, 1999, p.121). Na região de 10000 a 5500 cm^{-1} , nenhum efeito significativo foi observado, exceto pela variação evidente a 7100 cm^{-1} , região na qual ocorre o 1º sobretom de estiramento de grupos OH livres (MAEDA *et al.*, 1999, p.121). Tais observações demonstram a influência do aumento da temperatura sobre as absorções na região NIR, particularmente quando espécies contendo o grupo OH estão presentes.

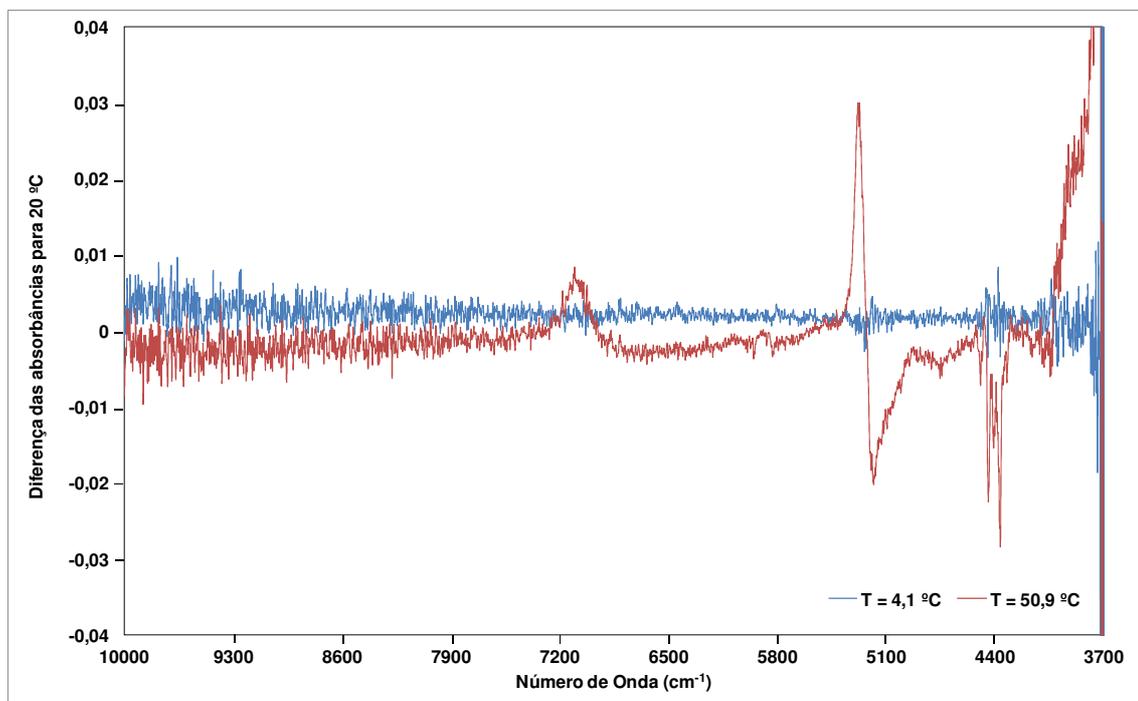


Figura 20. Diferença entre os espectros NIR da amostra adquiridos nos extremos dos testes de temperatura ($4,1$ e $50,9\text{ }^{\circ}\text{C}$) e a temperatura ambiente ($20\text{ }^{\circ}\text{C}$).

É importante ressaltar que as maiores variações na absorvância só aconteceram com as mais altas variações de temperatura. Portanto, para esse estudo, a variação natural do ambiente climatizado do laboratório não influenciou significativamente nos espectros, não havendo necessidade de controle adicional de temperatura da

amostra. Assim, o cuidado requerido para efetuar a aquisição dos espectros, foi apenas aguardar a amostra chegar à temperatura ambiente.

4.3. Construção dos modelos: Análise multivariada

A seguir, serão discutidos os parâmetros avaliados e os resultados obtidos com o desenvolvimento dos modelos de calibração para determinação da composição do solvente utilizado na área de produção de isopreno.

4.3.1. Avaliação do preparo das misturas na variabilidade do conjunto de dados

Após preparo das 75 misturas iniciais, a variabilidade do conjunto de dados foi avaliada através de gráficos de concentração das propriedades. A Figura 21, por exemplo, ilustra o gráfico de concentração para a propriedade acetona. Os pontos em azul representam as concentrações das 75 misturas preparadas no laboratório e das amostras reais do solvente coletadas no processo. Os pontos em vermelho representam as 25 misturas adicionais, preparadas para tornar mais robusto o conjunto de dados. É possível perceber que os espaços, anteriormente vazios, foram preenchidos tornando a distribuição dos dados mais homogênea em todo o domínio experimental. Efeito similar foi obtido para os demais componentes.

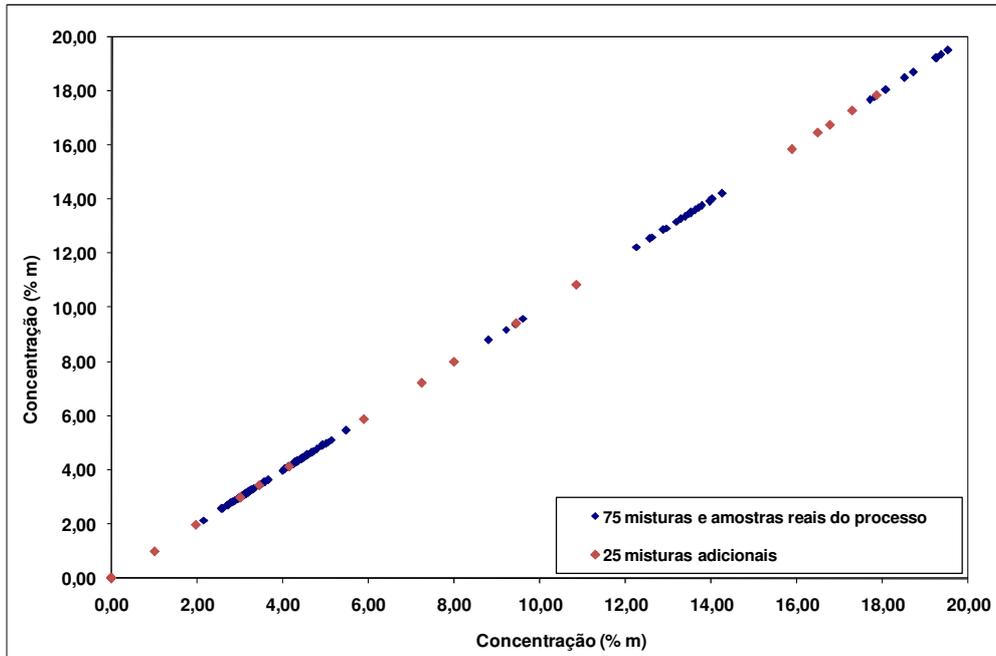


Figura 21. Gráfico dos dados de concentração para a propriedade acetona.

4.3.2. Detecção de *outliers* no conjunto de dados

A matriz de dados **X** (espectros) e **Y** (propriedades) contendo 175 amostras foi analisada para detecção de amostras anômalas através da PCA, que apresentou 82% de variância explicada pelas duas primeiras componentes principais. Nesta etapa, o gráfico dos *scores* (PC2 vs PC1) e o teste *Hotelling T²* foram empregados e estão ilustrados na Figura 22. As amostras fora da elipse foram eliminadas do conjunto de dados. Com isso, 11 amostras foram excluídas da matriz de dados, que passou a ser constituída por 164 amostras.

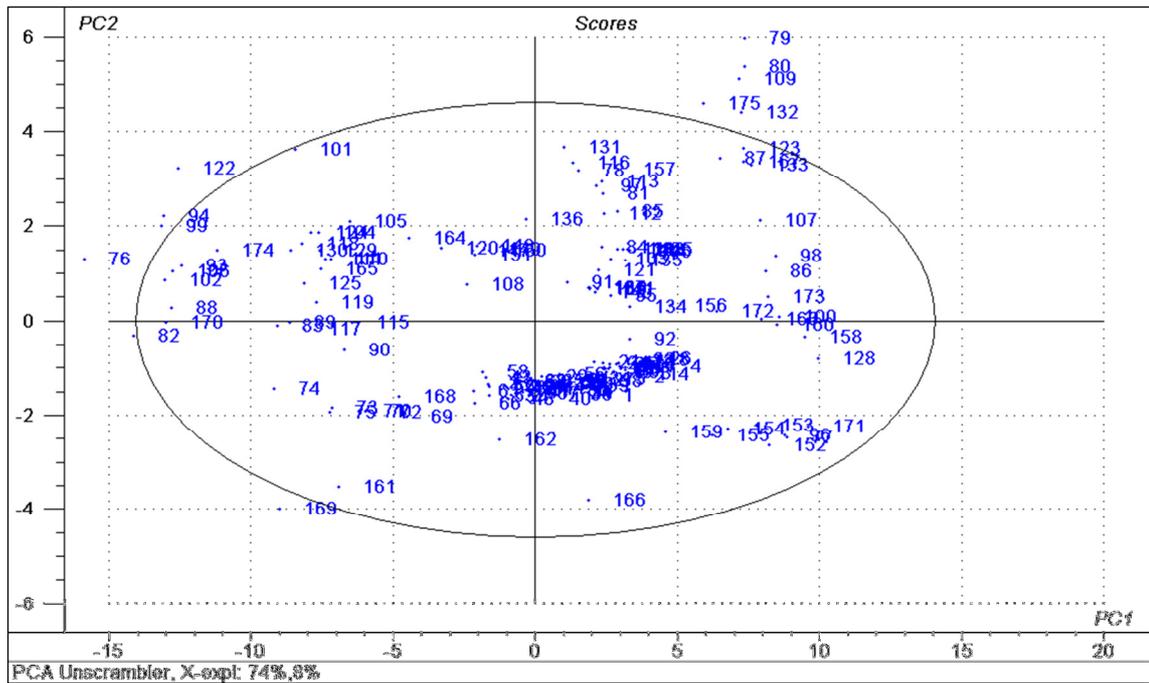


Figura 22. Gráfico de *scores* PC2 vs PC1 com resultado do teste *Hotelling T²*.

4.3.3. Definição dos conjuntos de calibração e validação

4.3.3.1. Seleção das amostras de calibração e validação

Após exclusão de *outliers* presentes no conjunto de dados, realizou-se a seleção do conjunto de dados para calibração e validação, através do algoritmo SPXY. A análise gráfica dos dados de concentração de cada propriedade a ser modelada (item 4.3.1) também auxiliou na seleção das amostras para calibração e validação, garantindo a uniformidade de distribuição de cada subconjunto ao longo do espaço experimental. Foram selecionadas 109 amostras para calibração e 55 amostras para a validação.

4.3.3.2. Detecção de *outliers* no conjunto de calibração por componente

Antes da realização dos experimentos previstos pelo planejamento experimental, um modelo PLS para cada componente em estudo foi construído com o objetivo de identificar amostras do conjunto de calibração com comportamentos anômalos para cada componente a ser modelado. Nesta etapa, os gráficos de resíduo *vs leverage* foram utilizados. Amostras com alto resíduo de concentração foram removidas do conjunto de calibração desse componente. Nessa etapa, para construção dos modelos, foi aplicado apenas o autoescalamento dos dados e a faixa espectral empregada foi de 8503 a 3700 cm^{-1} .

Na Figura 23, por exemplo, é ilustrado o gráfico de resíduo *vs leverage* para a propriedade água. As amostras circuladas em vermelho foram excluídas do conjunto de calibração para a modelagem da propriedade água por apresentarem resíduo elevado. Na Tabela 11 é mostrado o número de amostras que compõem o conjunto de calibração de cada propriedade, após exclusão de amostras com elevado resíduo.

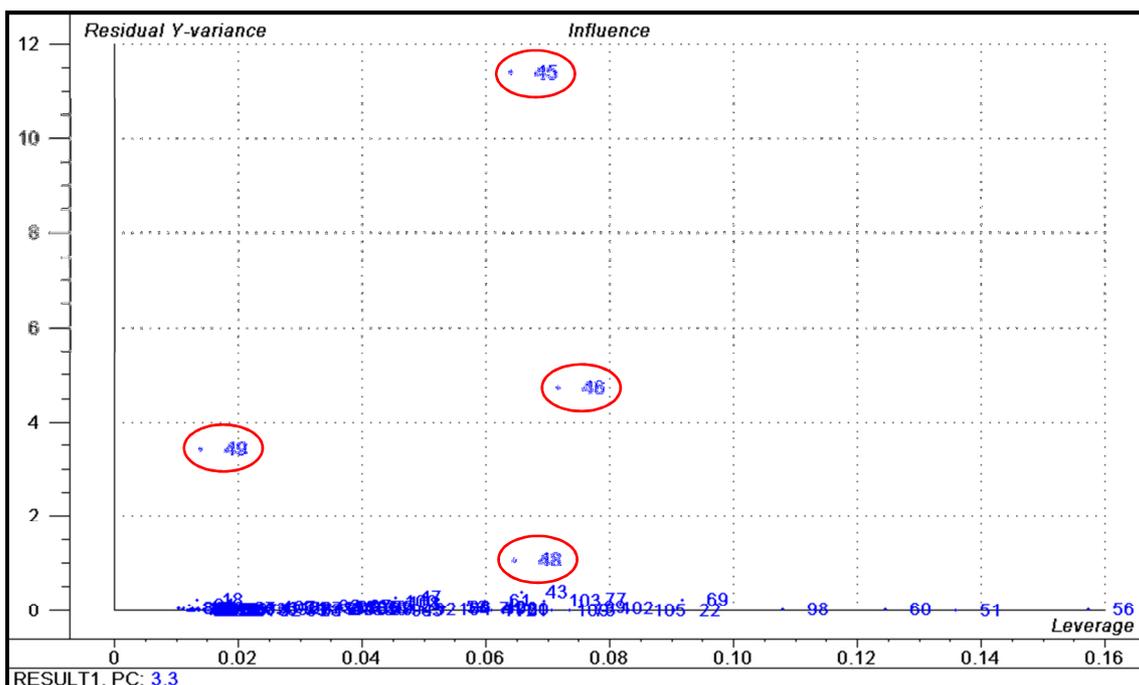


Figura 23. Gráfico de resíduo vs leverage para a propriedade água.

Tabela 10. Número de amostras do conjunto de calibração para cada propriedade.

Propriedade	Outliers	n (número de amostras do conjunto de calibração)
Acetona	0	109
Acetonitrila	3	106
Água	4	105
n-Propanol	1	108
iso-Propanol	1	108
t-Butanol	4	105
Ácool Alílico	2	107
DCPD	1	108
Outros	2	107

4.3.4. Seleção das faixas espectrais

As faixas espectrais selecionadas para definição dos níveis do planejamento experimental foram obtidas de acordo com o algoritmo *i*PLS, conforme descrito na Seção 2.3.2.2. É importante ressaltar que o intervalo indicado pelo *i*PLS corresponde a uma faixa espectral muito estreita, podendo levar a uma limitação da informação necessária para a construção de um modelo robusto. Assim, faixas mais largas, incluindo o intervalo indicado pelo algoritmo, também foram utilizadas, além de combinações das melhores regiões que apresentaram os menores valores de RMSECV e as regiões com informações espectroscópicas importantes, conforme discutido no item 4.1.

As faixas espectrais selecionadas para o planejamento experimental de cada propriedade estão resumidas na Tabela 11. Os resultados do *i*PLS obtidos para os parâmetros modelados são mostrados nas Figuras 24 a 32. Em cada uma das figuras é também mostrado um espectro da amostra para uma melhor interpretação dos gráficos.

Através dos resultados do *i*PLS é possível observar que, para as propriedades acetona, n-propanol, i-propanol, DCPD e Outros, o algoritmo conseguiu identificar regiões espectrais com RMSECV menor do que o modelo global, as quais foram selecionadas para o planejamento experimental. Entretanto, para as propriedades acetonitrila, água, álcool alílico e t-butanol, o *i*PLS não conseguiu indicar faixas espectrais que levariam a um melhor modelo do que o modelo global, o qual utiliza toda a região espectral. Nestes casos, priorizou-se a seleção de regiões em que os valores obtidos para o RMSECV apresentaram os menores valores. Na maioria dos casos, a região acima de 6102 cm^{-1} (intervalo *i*PLS 1247 a 2491) foi descartada por se tratar de uma região com valores de RMSECV bastante elevados.

É válido salientar que as principais bandas de absorção de todos os componentes em estudo, como por exemplo, o 2º sobretom da banda fundamental de estiramento C=O da acetona que ocorre a 5100 cm^{-1} , a banda de combinação das vibrações fundamentais de estiramento O-H e deformação angular O-H da água que aparece a 5200 cm^{-1} e o 1º sobretom da banda de estiramento C≡N da acetonitrila que

ocorre a 4444 cm^{-1} , estão presentes nas faixas espectrais selecionadas para o planejamento experimental.

Tabela 11. Faixas espectrais selecionadas a partir do $\bar{\nu}$ PLS para cada propriedade.

Componentes	Intervalos $\bar{\nu}$ PLS selecionados para o planejamento experimental	Região espectral correspondente ao intervalo indicado (cm^{-1})
Acetona	251 - 500	4181 - 4661
	1 - 500	3700 - 4661
	1 - 1247	3700 - 6102
	500 - 1247	4661 - 6102
	251 - 500 e 998 - 1247	4181 - 4661 e 5622 - 6102
Acetonitrila	251 - 500	4181 - 4661
	998 - 1247	5622 - 6102
	1 - 1247	3700 - 6102
	500 - 998	4661 - 5622
	251 - 500 e 998 - 1247	4181 - 4661 e 5622 - 6102
Água	1 - 500	3700 - 4661
	1 - 1994	3700 - 7543
	500 - 1247	4661 - 6102
	1496 - 1994	6582 - 7543
	1 - 998 e 1496 - 1994	3700 - 5622 e 6582 - 7543
n-propanol	251 - 500	4181 - 4661
	998 - 1247	5622 - 6102
	1 - 1247	3700 - 6102
	1 - 500	3700 - 4661
	251 - 500 e 998 - 1247	4181 - 4661 e 5622 - 6102

i-propanol	1 - 251	3700 - 4181
	251 - 500	4181 - 4661
	1 - 500	3700 - 4661
	1 - 1247	3700 - 6102
	500 - 1247	4661 - 6102
t-butanol	1 - 500	3700 - 4661
	1 - 1247	3700 - 6102
	998 - 1247	5622 - 6102
	749 - 1247	5141 - 6102
	1 - 500 e 998 - 1247	3700 - 4661 e 5622 - 6102
Álcool alílico	500 - 1247	4661 - 6102
	1 - 500	3700 - 4661
	1 - 1247	3700 - 6102
	1 - 251	3700 - 4181
	251 - 500	4181 - 4661
DCPD	251 - 500	4181 - 4661
	998 - 1247	5622 - 6102
	1 - 500 e 749 - 1247	3700 - 4661 e 5141 - 6102
	749 - 1247	5141 - 6102
	251 - 500 e 998 - 1247	4181 - 4661 e 5622 - 6102
Outros	1 - 500	3700 - 4661
	251 - 500	4181 - 4661
	500 - 1247	4661 - 6102
	1 - 1994	3700 - 7543
	251 - 500 e 1496 - 1994	4181 - 4661 e 6582 - 7543

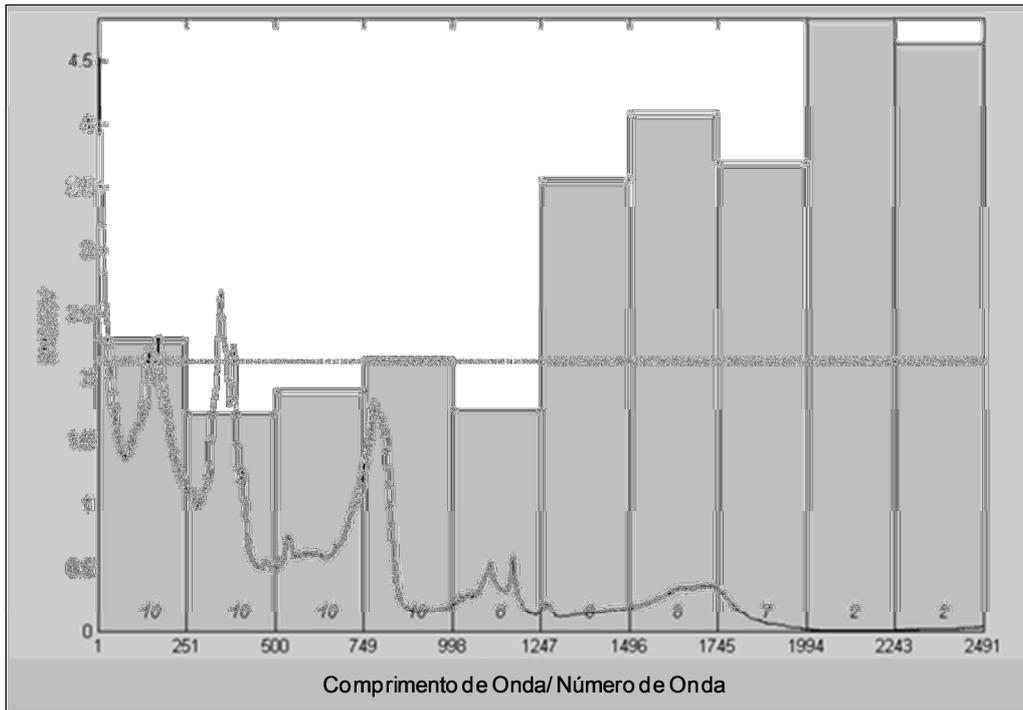


Figura 24. Gráfico do modelo iPLS para a propriedade acetona.

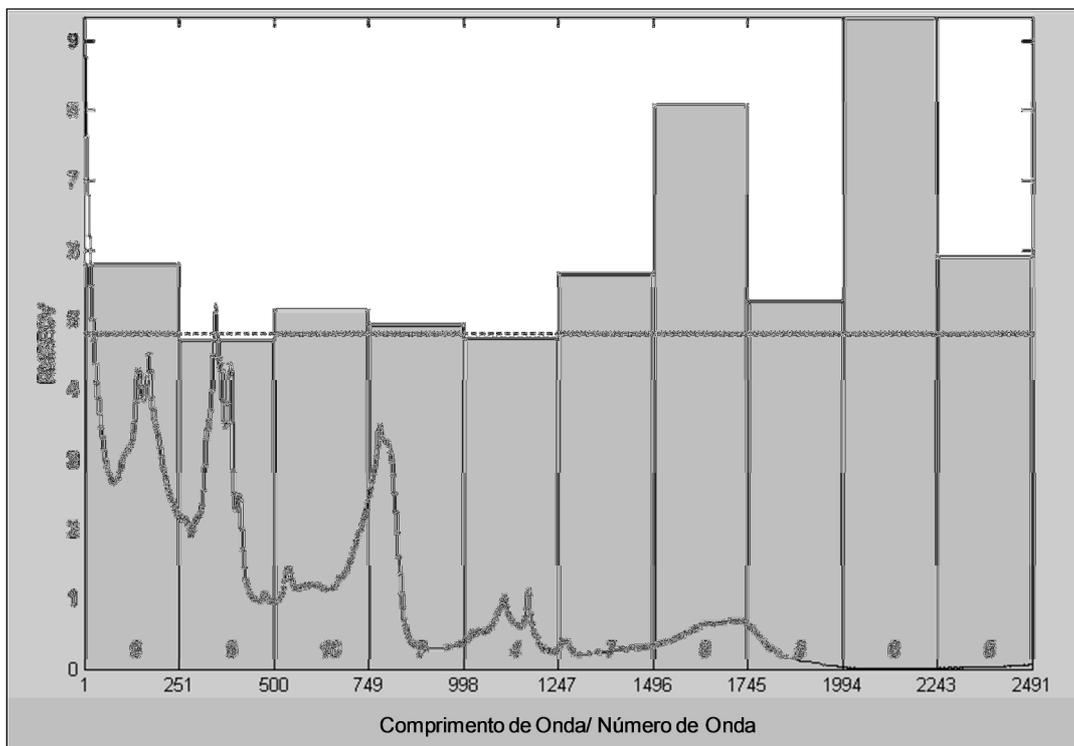


Figura 25. Gráfico do modelo iPLS para a propriedade acetonitrila.

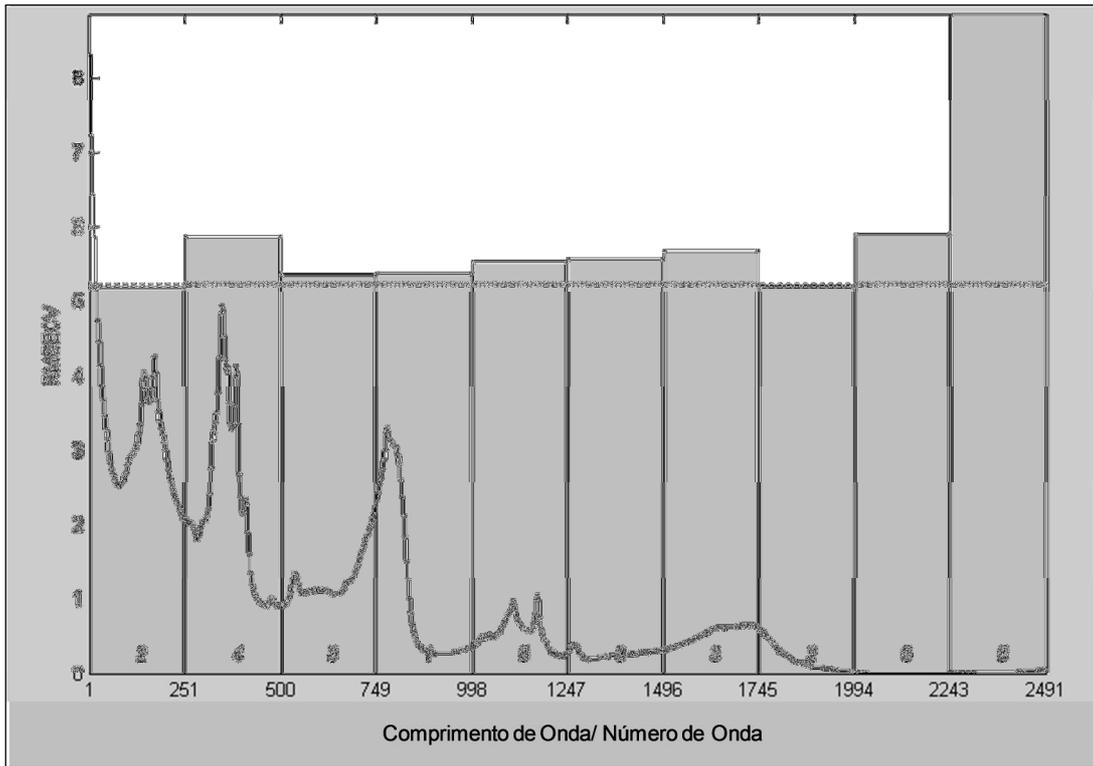


Figura 26. Gráfico do modelo PLS para a propriedade água.

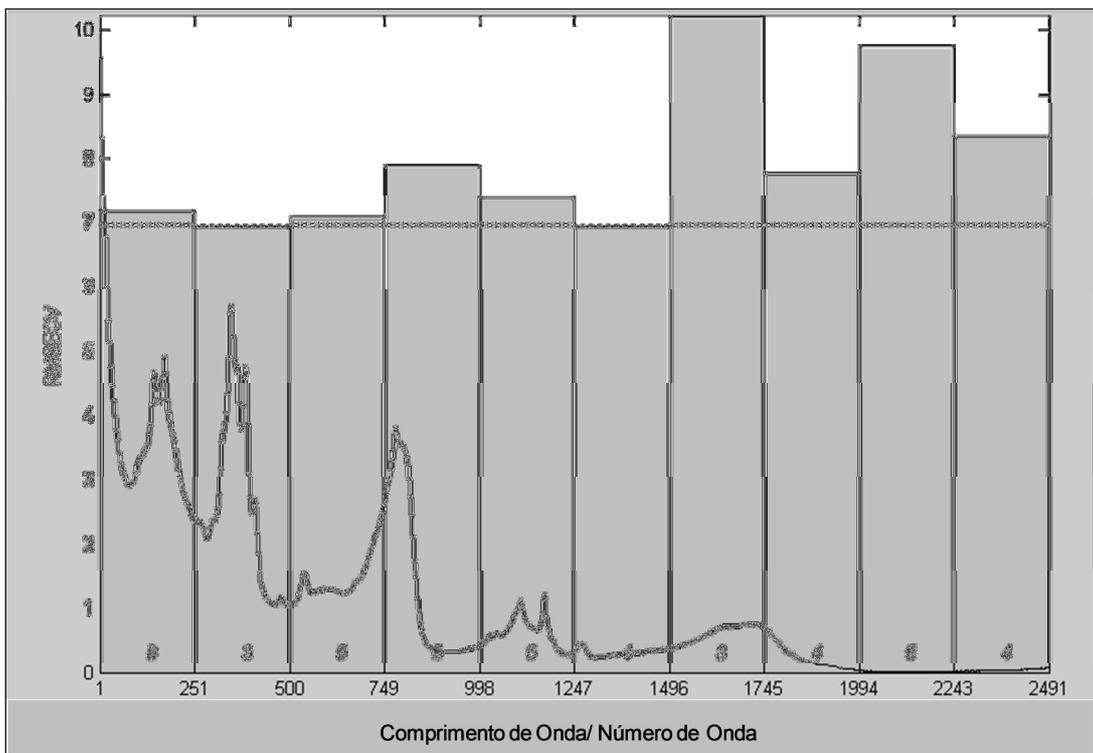


Figura 27. Gráfico do modelo PLS para a propriedade álcool alílico.

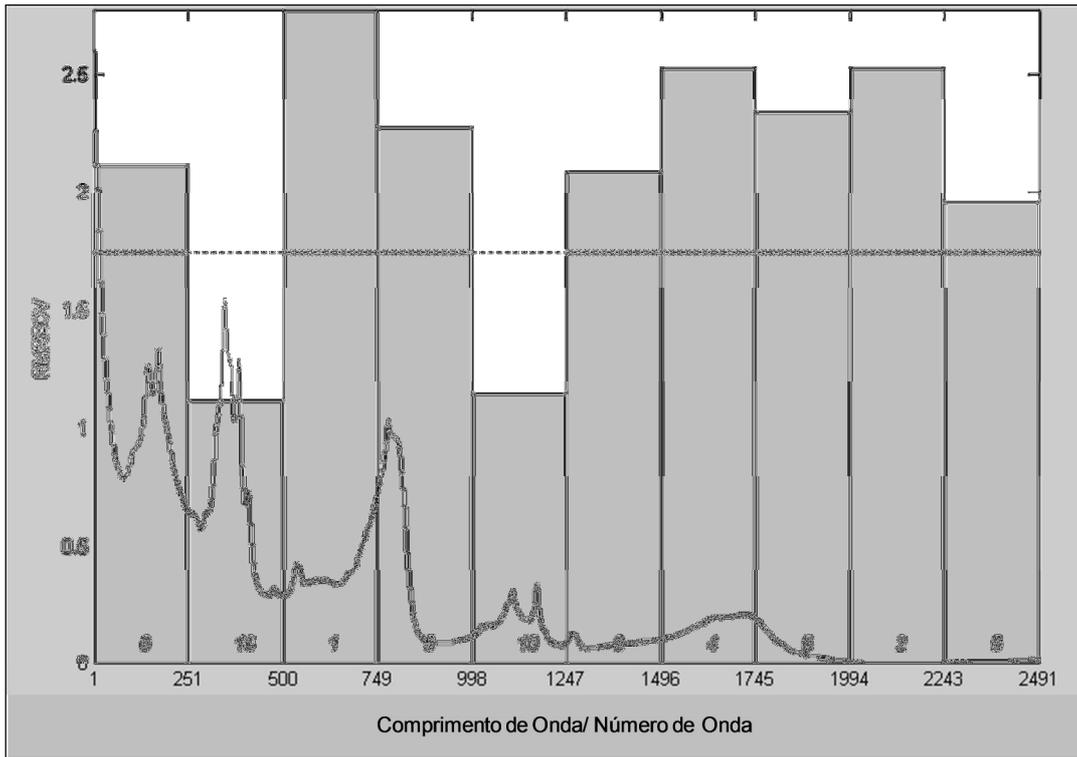


Figura 28. Gráfico do modelo *i*PLS para a propriedade n-propanol.

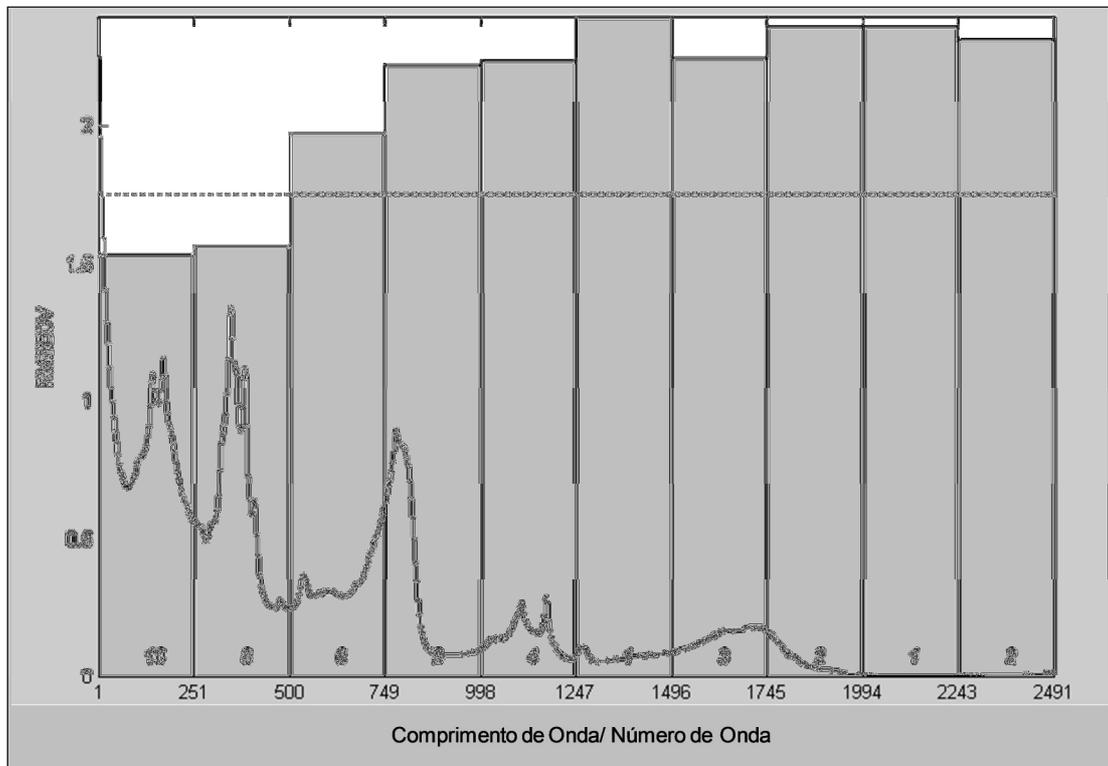


Figura 29. Gráfico do modelo *i*PLS para a propriedade i-propanol.

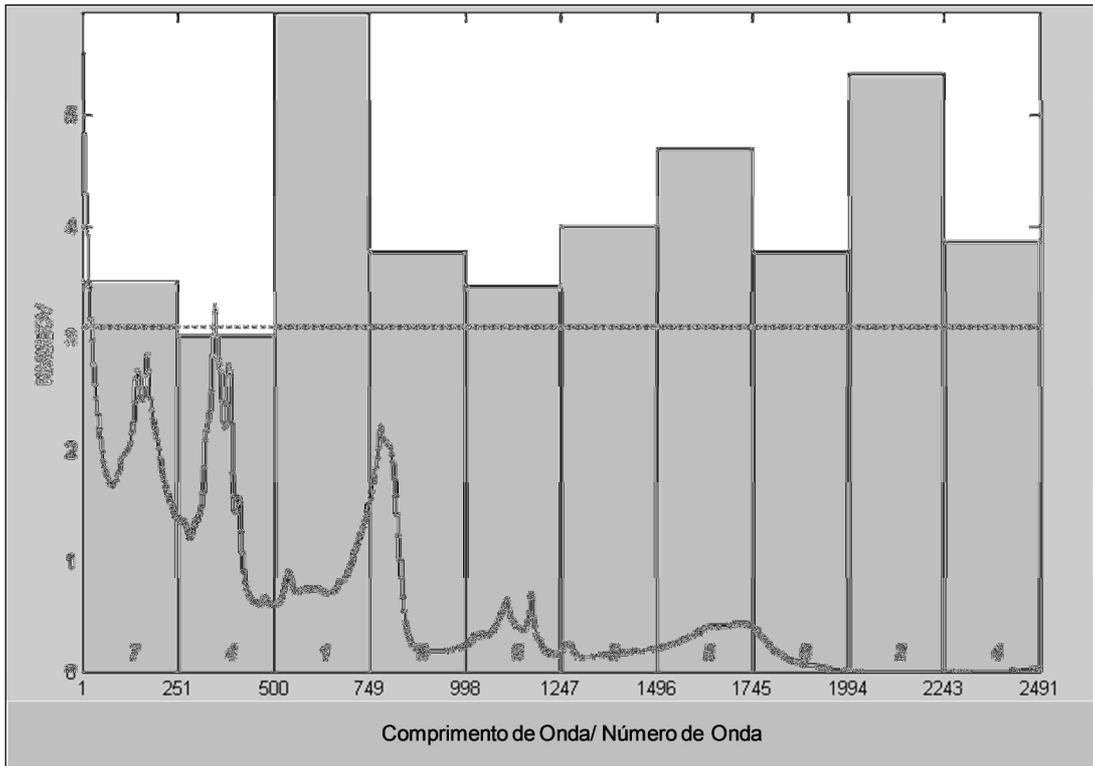


Figura 30. Gráfico do modelo PLS para a propriedade t-butanol.

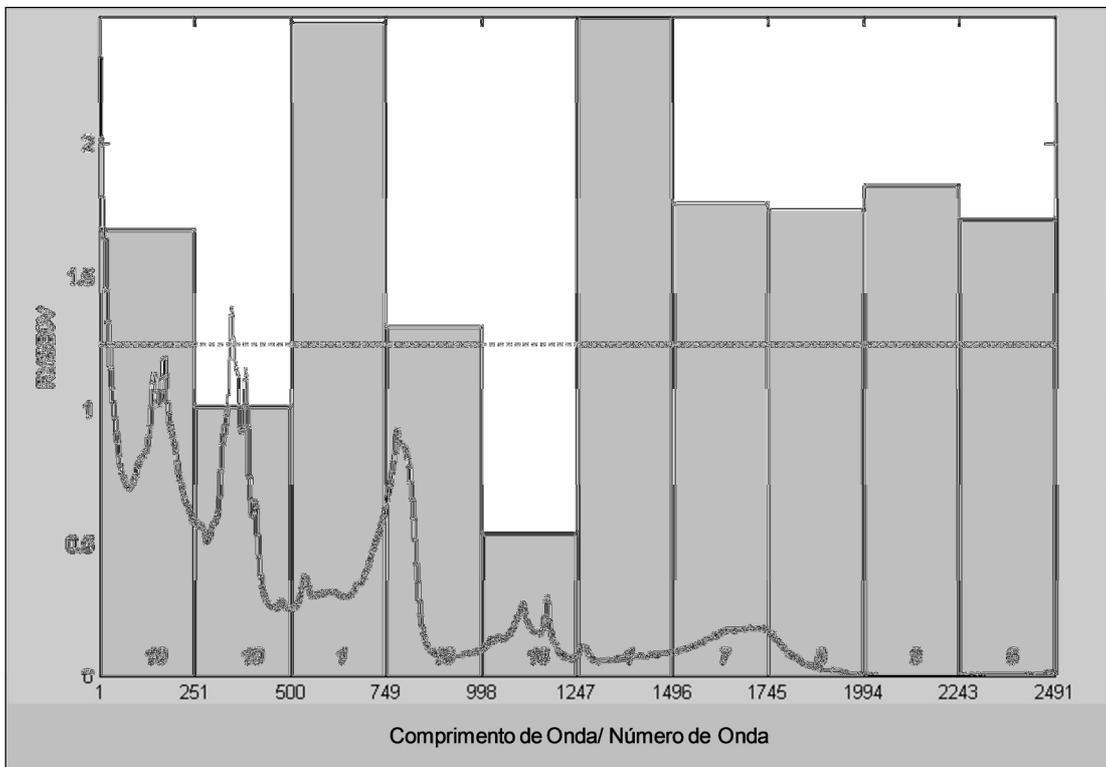


Figura 31. Gráfico do modelo PLS para a propriedade DCPD.

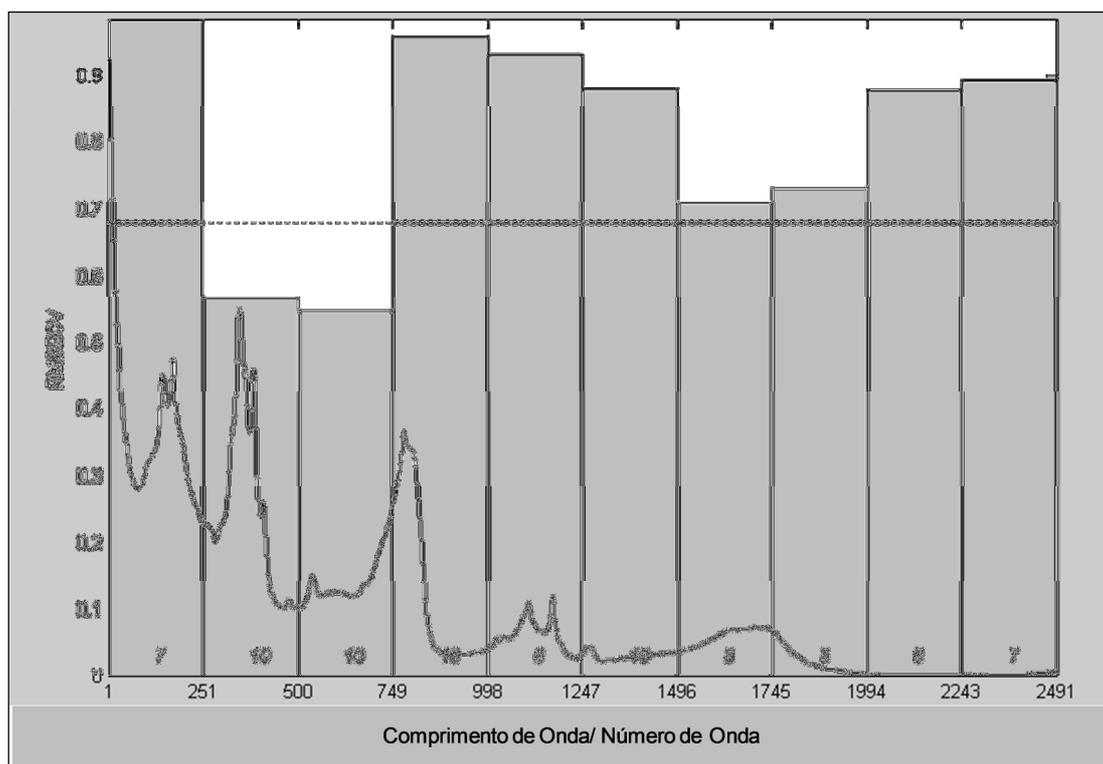


Figura 32. Gráfico do modelo iPLS para a propriedade outros.

4.3.5. Resultados do planejamento experimental e otimização dos modelos de calibração multivariada

Após definição do conjunto de calibração de cada componente, os experimentos do planejamento experimental foram executados, a partir dos níveis definidos na matriz Doehlert (Tabela 8), e os resultados são apresentados nas Tabelas de 12 a 20.

Tabela 12. Resultados do planejamento experimental da propriedade acetona (n=109).

Experimento	Faixa de comprimento de onda (cm ⁻¹)	Técnica de pré-processamento ^a	Seleção de variáveis ^b	Resultados da regressão e validação interna			Resultados da validação externa			
				PC	RMSECV (% m/m)	r ² _{regr}	RMSEP (% m/m)	r ² _{valid}	Slope Valid	Pvalue t pareado
1	3699 - 6102	PRE4	Com 1 Jack knife	10	2,01	0,8742	1,22	0,9379	0,922	0,265
2	3699 - 6102	PRE2	Sem Seleção	7	1,42	0,9345	1,12	0,9434	0,965	0,296
3	4661 - 6102	PRE4	Com 1 Jack knife	5	1,93	0,8796	1,29	0,9212	1,023	0,760
4	3699 - 6102	PRE6	Sem Seleção	7	1,14	0,9577	1,07	0,9493	0,946	0,459
5	3699 - 4661	PRE4	Com 1 Jack knife	6	1,78	0,8976	1,22	0,9294	0,985	0,997
6	4181 - 4661	PRE1	Com 1 Jack knife	4	1,58	0,9192	1,13	0,9463	0,931	0,238
7	4181 - 4661 e 5622 - 6102	PRE1	Com 1 Jack knife	5	1,37	0,9390	1,00	0,9591	0,950	0,040
8	4181 - 4661 e 5622 - 6102	PRE5	Com 2 Jack knife	5	1,79	0,8958	1,20	0,9348	1,058	0,918
9	4181 - 4661	PRE3	Sem Seleção	7	1,30	0,9449	0,96	0,9583	0,948	0,877
10	4181 - 4661	PRE5	Com 2 Jack knife	6	1,51	0,9264	1,11	0,9425	0,948	0,356
11	4181 - 4661 e 5622 - 6102	PRE3	Sem Seleção	6	1,47	0,9301	1,15	0,9414	0,991	0,054
12	4181 - 4661 e 5622 - 6102	PRE7	Com 1 Jack knife	5	1,21	0,9522	0,84	0,9752	0,918	0,174
13	4181 - 4661	PRE7	Com 1 Jack knife	6	1,28	0,9473	0,86	0,9751	0,915	0,085

^aDetalhes dos códigos de pré-processamento de dados na Tabela 9.

^bQuantidade de rodadas do algoritmo Jack Knife para seleção de variáveis.

Tabela 13. Resultados do planejamento experimental da propriedade acetonitrila (n=106).

Experimento	Faixa de comprimento de onda (cm ⁻¹)	Técnica de pré-processamento ^a	Seleção de variáveis ^b	Resultados da regressão e validação interna			Resultados da validação externa			
				PC	RMSECV (% m/m)	r ² _{regr}	RMSEP (% m/m)	r ² _{valid}	Slope Valid	p _{value} t pareado
1	4181 - 4661 e 5622 - 6102	PRE4	Com 1 Jack knife	5	4,88	0,9334	3,02	0,9404	1,072	0,017
2	4181 - 4661 e 5622 - 6102	PRE2	Sem Seleção	2	4,96	0,9313	2,61	0,9643	1,014	<0,0001
3	4661,2 - 5622	PRE4	Com 1 Jack knife	4	4,43	0,9450	2,68	0,9477	0,958	0,05
4	4181 - 4661 e 5622 - 6102	PRE6	Sem Seleção	2	3,67	0,9623	1,23	0,9893	0,982	0,529
5	3699 - 6102	PRE4	Com 1 Jack knife	3	5,21	0,9241	3,16	0,9352	1,047	0,334
6	4181 - 4661	PRE1	Com 1 Jack knife	2	3,66	0,9625	1,47	0,9851	1,000	0,193
7	5622 - 6102	PRE1	Com 1 Jack knife	2	3,89	0,9576	1,25	0,9913	0,981	0,016
8	5622 - 6102	PRE5	Com 2 Jack knife	2	6,12	0,8952	3,40	0,9491	1,085	<0,0001
9	4181 - 4661	PRE3	Sem Seleção	2	3,70	0,9617	1,49	0,9870	0,985	0,106
10	4181 - 4661	PRE5	Com 2 Jack knife	5	4,23	0,9498	2,41	0,9707	1,067	0,007
11	5622 - 6102	PRE3	Sem Seleção	2	3,91	0,9572	1,31	0,9913	0,981	0,016
12	5622 - 6102	PRE7	Com 1 Jack knife	2	3,54	0,9650	1,11	0,9914	1,011	0,456
13	4181 - 4661	PRE7	Com 1 Jack knife	1	3,48	0,9662	1,31	0,9897	0,999	0,016

^aDetalhes dos códigos de pré-processamento de dados na Tabela 9.

^bQuantidade de rodadas do algoritmo Jack Knife para seleção de variáveis.

Tabela 14. Resultados do planejamento experimental da propriedade água (n=105).

Experimento	Faixa de comprimento de onda (cm ⁻¹)	Técnica de pré-processamento ^a	Seleção de variáveis ^b	Resultados da regressão e validação interna			Resultados da validação externa			
				PC	RMSECV (% m/m)	r ² _{regr}	RMSEP (% m/m)	r ² _{valid}	Slope Valid	p _{value} t pareado
1	4661 - 6102	PRE4	Com 1 Jack knife	6	1,73	0,9763	1,30	0,9717	0,990	0,334
2	4661 - 6102	PRE2	Sem Seleção	4	2,25	0,9600	1,70	0,9522	0,967	0,394
3	6582 - 7543	PRE4	Com 1 Jack knife	5	2,89	0,9340	1,82	0,9486	0,935	0,389
4	4661 - 6102	PRE6	Sem Seleção	3	1,43	0,9839	1,10	0,9818	0,990	0,0003
5	3699 - 4661	PRE4	Com 1 Jack knife	4	2,49	0,9510	1,87	0,9417	0,973	0,367
6	3699 - 7543	PRE1	Com 1 Jack knife	2	1,48	0,9816	1,19	0,9776	1,019	0,053
7	3699 - 5622 e 6582 - 7543	PRE1	Com 1 Jack knife	2	1,56	0,9807	1,27	0,9742	1,014	0,066
8	3699 - 5622 e 6582 - 7543	PRE5	Com 2 Jack knife	2	2,44	0,9527	1,62	0,9569	0,970	0,338
9	3699 - 7543	PRE3	Sem Seleção	2	1,53	0,9816	1,19	0,9779	1,021	0,050
10	3699 - 7543	PRE5	Com 2 Jack knife	3	2,26	0,9596	1,75	0,9497	0,938	0,995
11	3699 - 5622 e 6582 - 7543	PRE3	Sem Seleção	2	1,57	0,9805	1,26	0,9746	1,017	0,074
12	3699 - 5622 e 6582 - 7543	PRE7	Com 1 Jack knife	5	1,88	0,9722	1,58	0,9592	1,036	0,055
13	3699 - 7543	PRE7	Com 1 Jack knife	4	1,92	0,9710	1,46	0,9642	1,016	0,527

^aDetalhes dos códigos de pré-processamento de dados na Tabela 9.

^bQuantidade de rodadas do algoritmo Jack Knife para seleção de variáveis.

Tabela 15. Resultados do planejamento experimental da propriedade n-propanol (n=108).

Experimento	Faixa de comprimento de onda (cm ⁻¹)	Técnica de pré-processamento ^a	Seleção de variáveis ^b	Resultados da regressão e validação interna			Resultados da validação externa			
				PC	RMSECV (% m/m)	r ² _{regr}	RMSEP (% m/m)	r ² _{valid}	Slope Valid	p _{value} t pareado
1	4181 - 4661 e 5622 - 6102	PRE4	Com 1 Jack knife	8	1,43	0,7453	1,11	0,7121	1,079	0,556
2	4181 - 4661 e 5622 - 6102	PRE2	Sem Seleção	9	0,92	0,8942	0,60	0,9441	1,079	0,012
3	3699 - 4661	PRE4	Com 1 Jack knife	7	1,40	0,7567	1,06	0,8204	0,873	0,006
4	4181 - 4661 e 5622 - 6102	PRE6	Sem Seleção	7	0,92	0,8954	0,51	0,9558	1,073	0,295
5	3699 - 6102	PRE4	Com 1 Jack knife	9	1,15	0,8354	0,94	0,8216	0,954	0,802
6	4181 - 4661	PRE1	Com 1 Jack knife	8	0,96	0,8862	0,64	0,9161	1,062	0,055
7	5622 - 6102	PRE1	Com 1 Jack knife	9	0,99	0,8794	0,45	0,9615	0,973	0,758
8	5622 - 6102	PRE5	Com 2 Jack knife	7	1,23	0,8118	0,84	0,8365	1,053	0,167
9	4181 - 4661	PRE3	Sem Seleção	10	0,87	0,9057	0,50	0,9437	1,019	0,805
10	4181 - 4661	PRE5	Com 2 Jack knife	8	1,00	0,8772	0,66	0,9243	1,022	0,187
11	5622 - 6102	PRE3	Sem Seleção	9	0,96	0,8847	0,49	0,9564	0,957	0,722
12	5622 - 6102	PRE7	Com 1 Jack knife	9	0,95	0,8888	0,79	0,8647	0,980	0,847
13	4181 - 4661	PRE7	Com 1 Jack knife	8	0,90	0,8998	0,64	0,9200	0,974	0,734

^aDetalhes dos códigos de pré-processamento de dados na Tabela 9.

^bQuantidade de rodadas do algoritmo Jack Knife para seleção de variáveis.

Tabela 16. Resultados do planejamento experimental da propriedade i-propanol (n=108).

Experimento	Faixa de comprimento de onda (cm ⁻¹)	Técnica de pré-processamento ^a	Seleção de variáveis ^b	Resultados da regressão e validação interna			Resultados da validação externa			
				PC	RMSECV (% m/m)	r ² _{regr}	RMSEP (% m/m)	r ² _{valid}	Slope Valid	p _{value} t pareado
1	3699 - 6102	PRE4	Com 1 Jack knife	6	1,47	0,6516	0,95	0,7510	1,197	0,270
2	3699 - 6102	PRE2	Sem Seleção	9	1,04	0,8253	0,79	0,8336	0,983	0,999
3	4181 - 4661	PRE4	Com 1 Jack knife	6	1,41	0,6785	0,90	0,7889	1,126	0,013
4	3699 - 6102	PRE6	Sem Seleção	9	1,19	0,7695	0,80	0,7953	0,960	0,763
5	4661 - 6102	PRE4	Com 1 Jack knife	6	1,76	0,5016	1,16	0,6437	1,221	0,031
6	3699 - 4661	PRE1	Com 1 Jack knife	5	1,41	0,6799	0,83	0,7799	1,028	0,372
7	3699 - 4181	PRE1	Com 1 Jack knife	6	1,41	0,6782	0,82	0,7881	1,090	0,810
8	3699 - 4181	PRE5	Com 2 Jack knife	7	1,27	0,7405	0,94	0,7890	1,089	0,278
9	3699 - 4661	PRE3	Sem Seleção	9	1,26	0,7415	0,93	0,7557	0,823	0,957
10	3699 - 4661	PRE5	Com 2 Jack knife	5	1,51	0,6299	1,08	0,6883	1,320	0,440
11	3699 - 4181	PRE3	Sem Seleção	8	1,34	0,7081	0,87	0,7649	0,904	0,736
12	3699 - 4181	PRE7	Com 1 Jack knife	5	1,24	0,7513	1,15	0,6633	0,885	0,524
13	3699 - 4661	PRE7	Com 1 Jack knife	6	1,22	0,7589	0,90	0,8584	1,255	0,003

^aDetalhes dos códigos de pré-processamento de dados na Tabela 9.

^bQuantidade de rodadas do algoritmo Jack Knife para seleção de variáveis.

Tabela 17. Resultados do planejamento experimental da propriedade t-butanol (n=105).

Experimento	Faixa de comprimento de onda (cm ⁻¹)	Técnica de pré-processamento ^a	Seleção de variáveis ^b	Resultados da regressão e validação interna			Resultados da validação externa			
				PC	RMSECV (% m/m)	r ² _{regr}	RMSEP (% m/m)	r ² _{valid}	Slope Valid	p _{value} t pareado
1	3699 - 4661 e 5622 - 6102	4	Com 1 Jack knife	7	1,54	0,9356	1,16	0,9433	1,043	0,556
2	3699 - 4661 e 5622 - 6102	2	Sem Seleção	7	1,30	0,9541	0,98	0,9561	1,003	0,012
3	5141 - 6102	4	Com 1 Jack knife	6	2,22	0,8657	1,36	0,9206	0,936	0,006
4	3699 - 4661 e 5622 - 6102	6	Sem Seleção	6	1,00	0,9728	0,80	0,9689	0,948	0,296
5	3699 - 6102	4	Com 1 Jack knife	10	1,07	0,9689	1,00	0,9710	0,980	0,802
6	5622 - 6102	1	Com 1 Jack knife	5	1,90	0,9009	1,22	0,9361	0,952	0,055
7	3699 - 4661	1	Com 1 Jack knife	5	1,21	0,9601	0,90	0,9648	0,963	0,758
8	3699 - 4661	5	Com 2 Jack knife	4	1,35	0,9503	0,89	0,9651	1,022	0,167
9	5622 - 6102	3	Sem Seleção	10	1,22	0,9593	1,06	0,9592	0,911	0,804
10	5622 - 6102	5	Com 2 Jack knife	6	2,28	0,8583	1,36	0,9198	1,008	0,187
11	3699 - 4661	3	Sem Seleção	6	1,16	0,9632	0,89	0,9663	0,952	0,722
12	3699 - 4661	7	Com 1 Jack knife	3	1,12	0,9655	0,72	0,9747	0,961	0,847
13	5622 - 6102	7	Com 1 Jack knife	5	1,50	0,9388	1,42	0,9223	0,897	0,737

^aDetalhes dos códigos de pré-processamento de dados na Tabela 9.

^bQuantidade de rodadas do algoritmo Jack Knife para seleção de variáveis.

Tabela 18. Resultados do planejamento experimental da propriedade álcool alílico (n=107).

Experimento	Faixa de comprimento de onda (cm ⁻¹)	Técnica de pré-processamento ^a	Seleção de variáveis ^b	Resultados da regressão e validação interna			Resultados da validação externa			
				PC	RMSECV (% m/m)	r ² _{regr}	RMSEP (% m/m)	r ² _{valid}	Slope Valid	p _{value} t pareado
1	3699 - 6102	PRE4	Com 1 Jack knife	7	4,46	0,8872	3,38	0,8576	1,047	0,598
2	3699 - 6102	PRE2	Sem Seleção	4	5,21	0,8458	2,82	0,9478	1,139	<0,0001
3	4181 - 4661	PRE4	Com 1 Jack knife	4	5,14	0,8503	3,60	0,8793	1,127	0,889
4	3699 - 6102	PRE6	Sem Seleção	4	4,65	0,8774	2,78	0,9603	1,069	<0,0001
5	4661 - 6102	PRE4	Com 1 Jack knife	6	5,22	0,8455	2,58	0,9436	1,075	0,007
6	3699 - 4661	PRE1	Com 1 Jack knife	4	4,37	0,8914	2,68	0,9385	1,068	0,001
7	3699 - 4181	PRE1	Com 1 Jack knife	4	4,87	0,8652	2,54	0,9556	1,078	0,000
8	3699 - 4181	PRE5	Com 2 Jack knife	4	5,61	0,8213	3,24	0,9151	1,042	<0,0001
9	3699 - 4661	PRE3	Sem Seleção	5	4,26	0,897	2,69	0,9431	1,045	<0,0001
10	3699 - 4661	PRE5	Com 2 Jack knife	5	4,24	0,8977	3,36	0,8573	1,027	0,905
11	3699 - 4181	PRE3	Sem Seleção	7	4,59	0,8804	2,52	0,9522	1,054	0,001
12	3699 - 4181	PRE7	Com 1 Jack knife	3	5,13	0,8504	2,93	0,937	1,078	<0,0001
13	3699 - 4661	PRE7	Com 1 Jack knife	3	4,87	0,8552	3,23	0,9019	1,054	<0,0001

^aDetalhes dos códigos de pré-processamento de dados na Tabela 9.

^bQuantidade de rodadas do algoritmo Jack Knife para seleção de variáveis.

Tabela 19. Resultados do planejamento experimental da propriedade DCPD (n=108).

Experimento	Faixa de comprimento de onda (cm ⁻¹)	Técnica de pré-processamento ^a	Seleção de variáveis ^b	Resultados da regressão e validação interna			Resultados da validação externa			
				PC	RMSECV (% m/m)	r ² _{regr}	RMSEP (% m/m)	r ² _{valid}	Slope Valid	p _{value} t pareado
1	1-500 e 749-1247	PRE4	Com 1 Jack knife	6	1,05	0,8347	0,77	0,8959	0,888	0,562
2	1-500 e 749-1247	PRE2	Sem Seleção	10	0,64	0,9374	0,87	0,8826	0,845	0,172
3	5141 - 6102	PRE4	Com 1 Jack knife	7	0,77	0,9106	0,50	0,9638	0,973	<0,0001
4	3699 - 4661 e 5141,8 - 6102	PRE6	Sem Seleção	9	0,77	0,9103	0,67	0,9137	1,066	0,943
5	4181 - 4661	PRE4	Com 1 Jack knife	10	0,73	0,9206	0,80	0,9010	0,855	0,177
6	5622 - 6102	PRE1	Com 1 Jack knife	6	0,56	0,9525	0,44	0,9467	0,977	0,069
7	4181 - 4661 e 5622 - 6102	PRE1	Com 1 Jack knife	10	0,60	0,9456	0,60	0,9286	1,002	0,317
8	4181 - 4661 e 5622 - 6102	PRE5	Com 2 Jack knife	7	0,82	0,8990	0,69	0,9098	0,919	0,821
9	5622 - 6102	PRE3	Sem Seleção	6	0,58	0,9488	0,46	0,9432	0,986	0,036
10	5622 - 6102	PRE5	Com 2 Jack knife	6	0,67	0,9323	0,45	0,9458	0,963	0,040
11	4181 - 4661 e 5622 - 6102	PRE3	Sem Seleção	10	0,59	0,9473	0,59	0,9206	1,009	0,600
12	4181 - 4661 e 5622 - 6102	PRE7	Com 1 Jack knife	8	0,55	0,9547	0,47	0,9565	1,032	0,948
13	5622 - 6102	PRE7	Com 1 Jack knife	6	0,66	0,9343	0,62	0,8831	0,979	0,017

^aDetalhes dos códigos de pré-processamento de dados na Tabela 9.

^bQuantidade de rodadas do algoritmo Jack Knife para seleção de variáveis.

Tabela 20: Resultados do planejamento experimental da propriedade outros (n=107).

Experimento	Faixa de comprimento de onda (cm ⁻¹)	Técnica de pré-processamento ^a	Seleção de variáveis ^b	Resultados da regressão e validação interna			Resultados da validação externa			
				PC	RMSECV (% m/m)	r ² _{regr}	RMSEP (% m/m)	r ² _{valid}	Slope Valid	p _{value} t pareado
1	3700 - 7543	PRE4	Com 1 Jack knife	10	0,34	0,9593	0,34	0,9583	1,0646	0,852
2	3700 - 7543	PRE2	Sem Seleção	10	0,39	0,9460	0,39	0,9395	1,0699	0,633
3	4661 - 6102	PRE4	Com 1 Jack knife	7	0,49	0,9142	0,49	0,9178	1,0667	0,430
4	3700 - 7543	PRE6	Sem Seleção	10	0,50	0,9095	0,50	0,9347	1,1433	0,304
5	3700 - 4661	PRE4	Com 1 Jack knife	10	0,43	0,9347	0,43	0,9259	1,0997	0,186
6	4181 - 4661	PRE1	Com 1 Jack knife	7	0,47	0,9213	0,47	0,9551	1,1097	0,294
7	4181 - 4661 e 6582 - 7543	PRE1	Com 1 Jack knife	7	0,49	0,9150	0,49	0,9439	1,1461	0,170
8	4181 - 4661 e 6582 - 7543	PRE5	Com 2 Jack knife	8	0,41	0,9395	0,41	0,9576	1,0891	0,463
9	4181 - 4661	PRE3	Sem Seleção	10	0,40	0,9420	0,40	0,9595	1,0773	0,758
10	4181 - 4661	PRE5	Com 2 Jack knife	10	0,39	0,9459	0,39	0,9635	1,0743	0,837
11	4181 - 4661 e 6582 - 7543	PRE3	Sem Seleção	10	0,46	0,9236	0,46	0,9488	1,1330	0,070
12	4181 - 4661 e 6582 - 7543	PRE7	Com 1 Jack knife	8	0,64	0,8546	0,64	0,9233	1,1708	0,368
13	4181 - 4661	PRE7	Com 1 Jack knife	10	0,58	0,8823	0,58	0,926	1,1558	0,635

^aDetalhes dos códigos de pré-processamento de dados na Tabela 9.

^bQuantidade de rodadas do algoritmo Jack Knife para seleção de variáveis.

As variáveis de resposta, PC, RMSECV, r^2_{regr} , RMSEP, r^2_{valid} , *Slope Valid* e p_{value} , obtidas a partir dos modelos definidos no planejamento experimental, foram analisadas e, em seguida, otimizadas simultaneamente, através da função de desejabilidade.

Uma vez que os critérios de aceitação para cada resposta foram especificados, foi possível encontrar um valor ótimo para cada variável de entrada. Os gráficos de desejabilidade para cada propriedade podem ser visualizados nas Figuras de 33 a 41 e mostram a influência das variáveis de entrada, comprimento de onda, pré-processamento e seleção de variáveis, nas variáveis de resposta. As linhas azuis horizontais, nos três gráficos da última linha, sinalizam as condições de máxima desejabilidade global (D), a qual pode ser obtida com as condições ótimas apresentadas na Tabela 21. Os pontos ótimos são representados pelas linhas vermelhas verticais, também nos três gráficos da última linha. Nos casos em que os valores sugeridos como ótimos foram diferentes dos níveis codificados no planejamento experimental, o nível mais próximo ao valor encontrado foi selecionado. A desejabilidade global para todas as propriedades também estão apresentadas na Tabela 21.

De um modo geral, os perfis de desejabilidade global em função do pré-processamento mostram que esse fator pode variar numa faixa razoável sem influenciar significativamente o valor de D, exceto para as propriedades acetona, água, n-propanol e álcool alílico, em que alterações nos valores do pré-processamento podem provocar uma queda no valor da desejabilidade global e, portanto, torna-se um fator muito importante na definição do ponto ótimo.

A seleção de variáveis pouco influencia as variáveis de resposta para a maioria das propriedades e, nestes casos, as inclinações observadas nos gráficos de desejabilidade individual são levemente acentuadas e o valor da desejabilidade global quase que se mantém constante. Entretanto, ainda que a influência da seleção de variáveis não seja tão significativa, observa-se que existem níveis dessa variável que não atingem a D máxima e, portanto, há um ponto ótimo dessa variável. Já os perfis de desejabilidade global para as propriedades acetona e álcool

alílico indicam que mudanças no nível do planejamento experimental para essa variável pode levar a uma diminuição da desejabilidade global.

Para as propriedades acetona, DCPD, n-propanol e álcool alílico, os gráficos de desejabilidade global indicam que a variável comprimento de onda é um fator significativo para determinação do ponto ótimo e que alterações no seu valor podem levar a um decréscimo no valor de D. Para a propriedade t-butanol, mudanças no valor do comprimento de onda não provocam efeito relevante no valor de D.

Em todos os casos, as faixas de comprimento de onda indicadas através da função de desejabilidade englobam as principais bandas de absorção dos componentes em estudo, conforme discutido no item 4.1.

É válido ressaltar que todas essas inclinações são muito instrutivas, porque fornecem uma idéia da influência de uma dada variável em uma dada resposta, a qual define a condição ótima para essa variável.

É possível observar também a influência de pelo menos uma das variáveis de entrada em alguma das respostas obtidas com a validação externa dos modelos, revelando a importância de avaliar tais parâmetros de desempenho ainda na etapa de construção dos modelos de calibração multivariada.

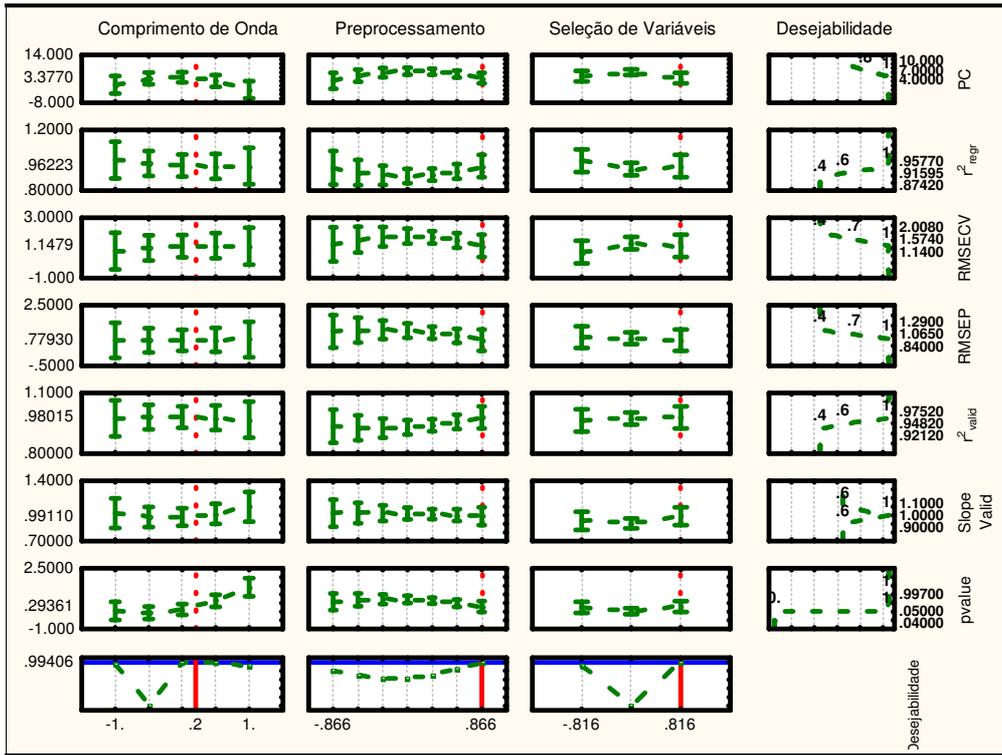


Figura 33. Gráficos de desejabilidade para a propriedade acetona.

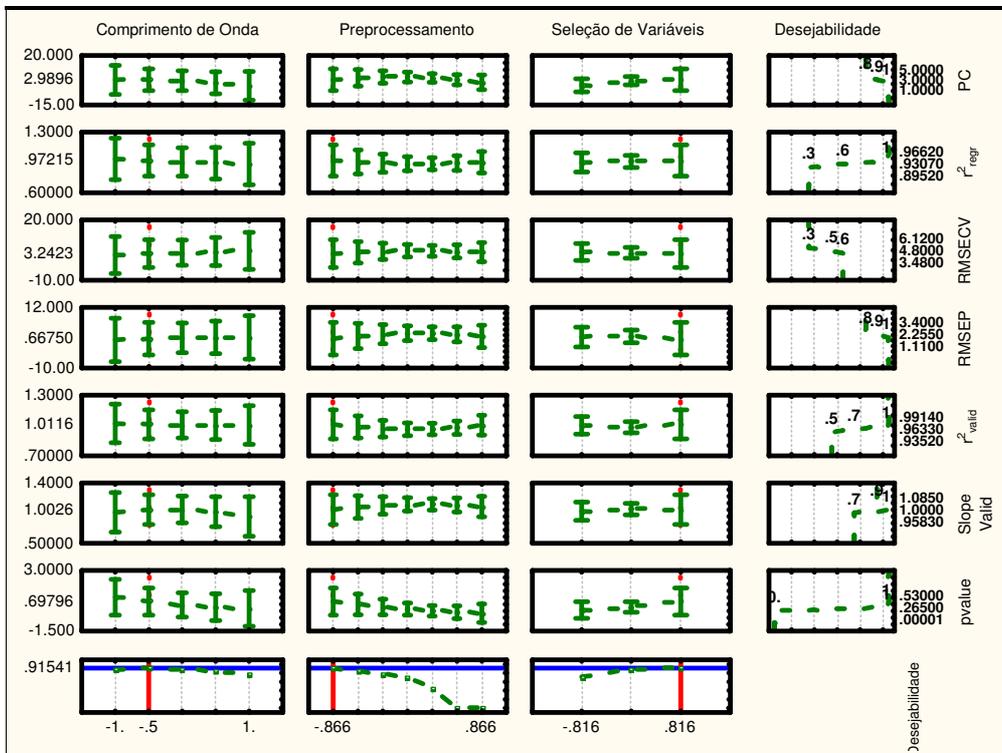


Figura 34. Gráficos de desejabilidade para a propriedade acetonitrila.

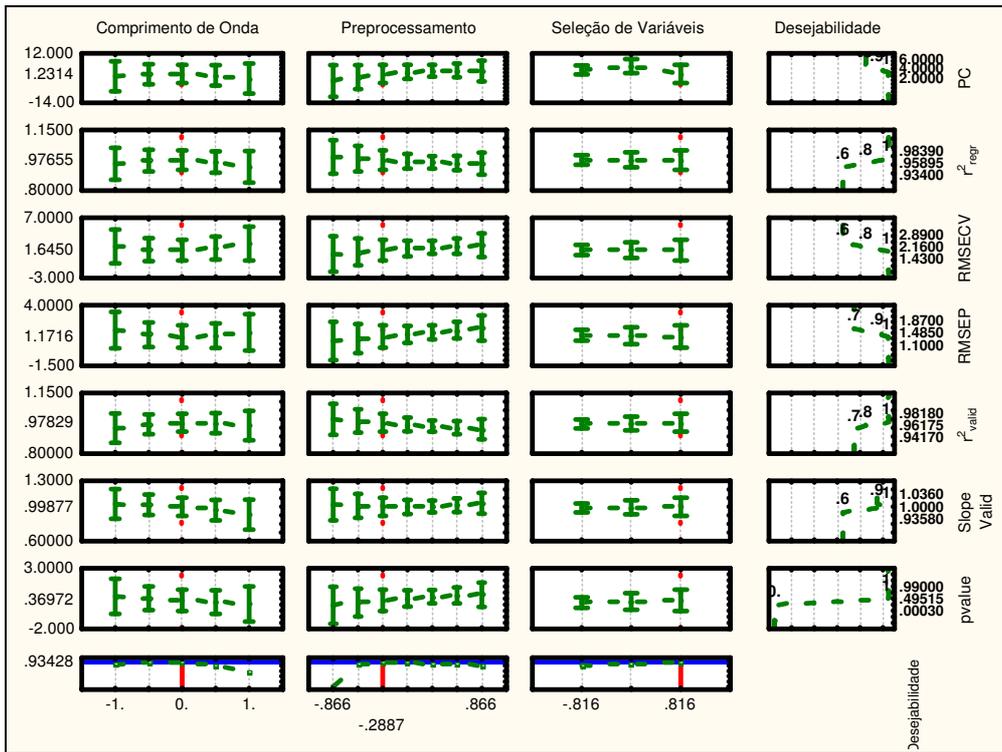


Figura 35. Gráficos de desejabilidade para a propriedade água.

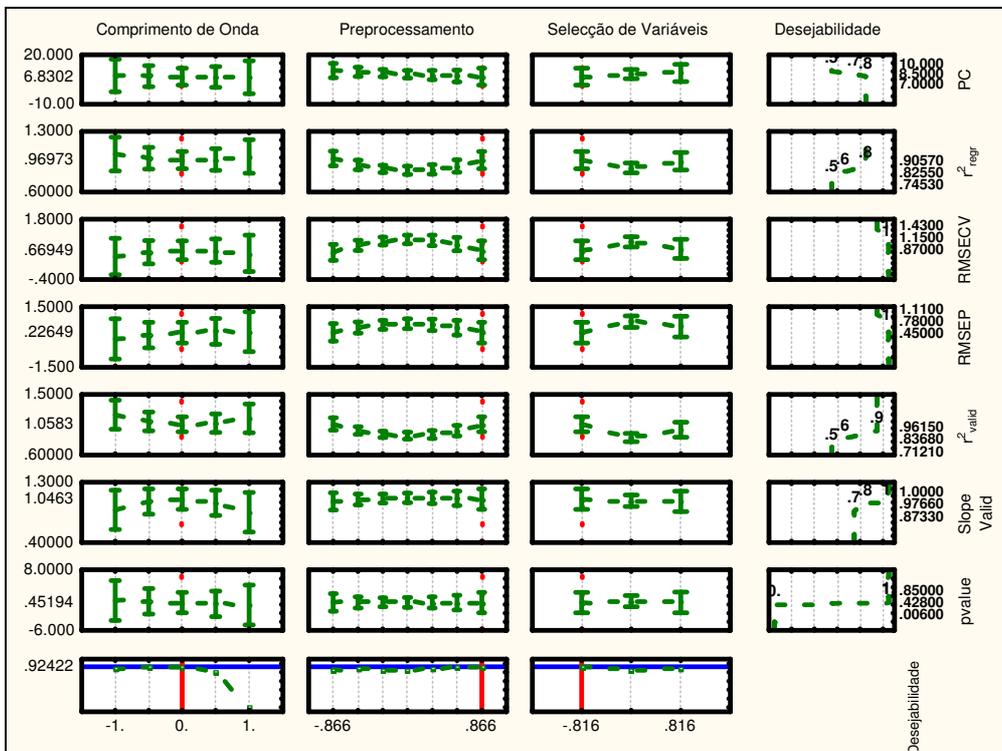


Figura 36. Gráficos de desejabilidade para a propriedade n-propanol.

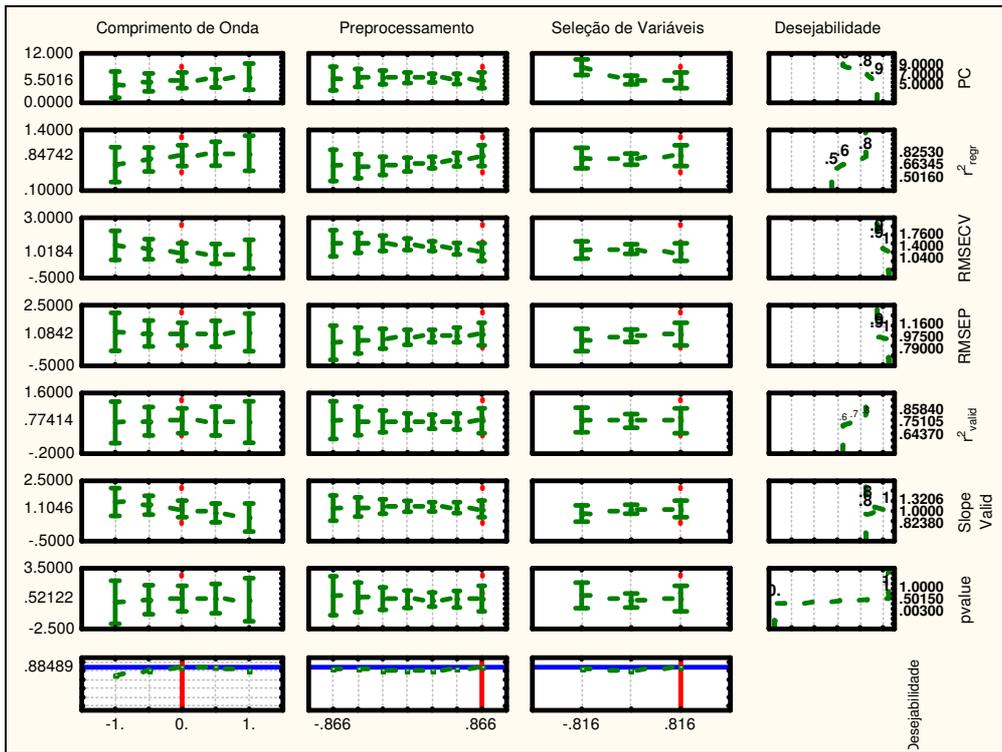


Figura 37. Gráficos de desejabilidade para a propriedade i-propanol.

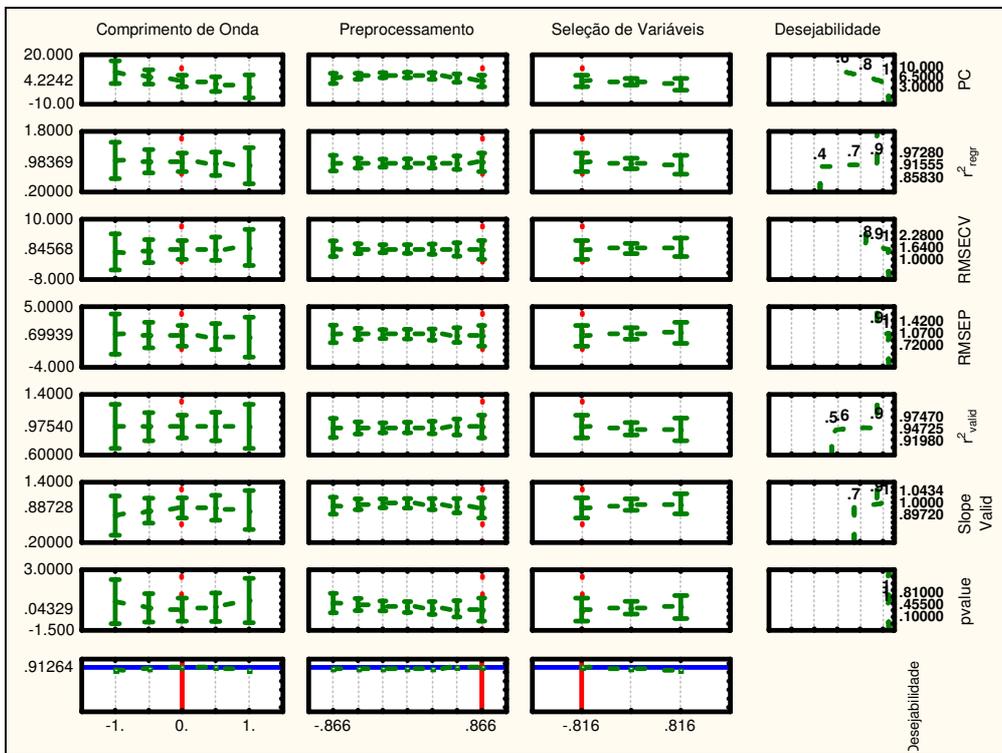


Figura 38. Gráficos de desejabilidade para a propriedade t-butanol.

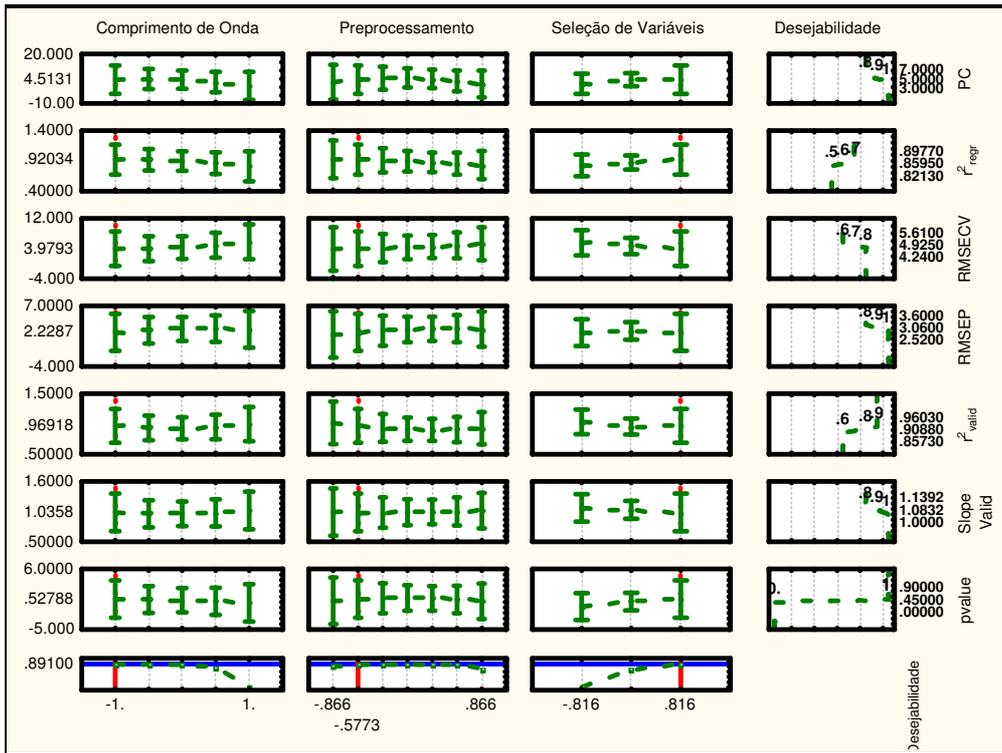


Figura 39. Gráficos de desejabilidade para a propriedade álcool alílico.

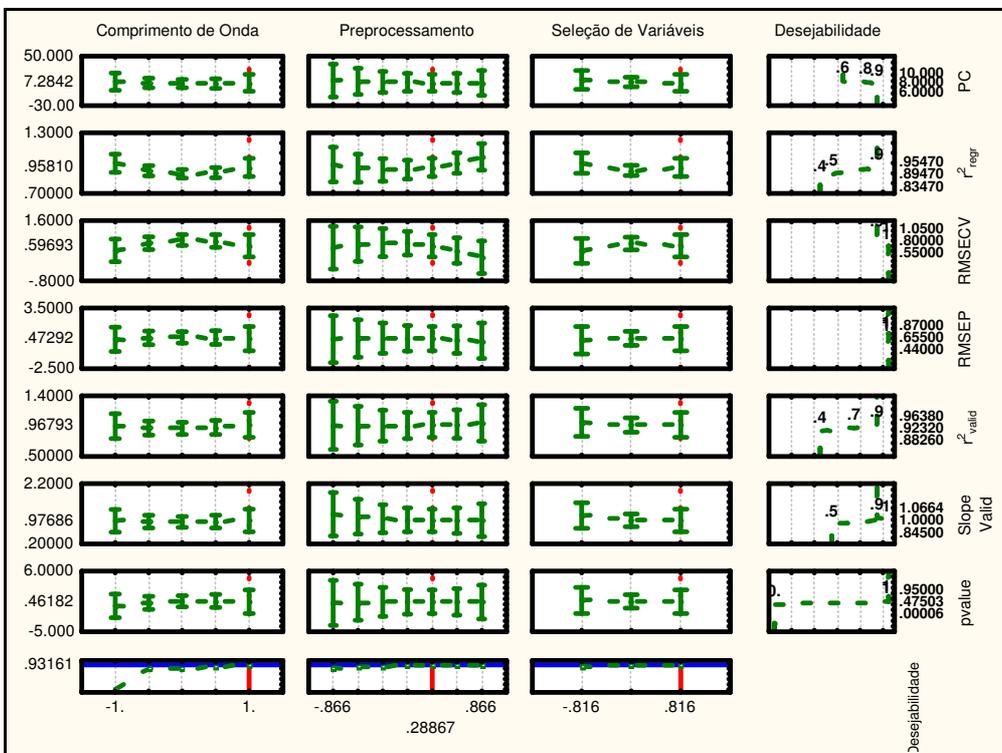


Figura 40. Gráficos de desejabilidade para a propriedade DCPD.

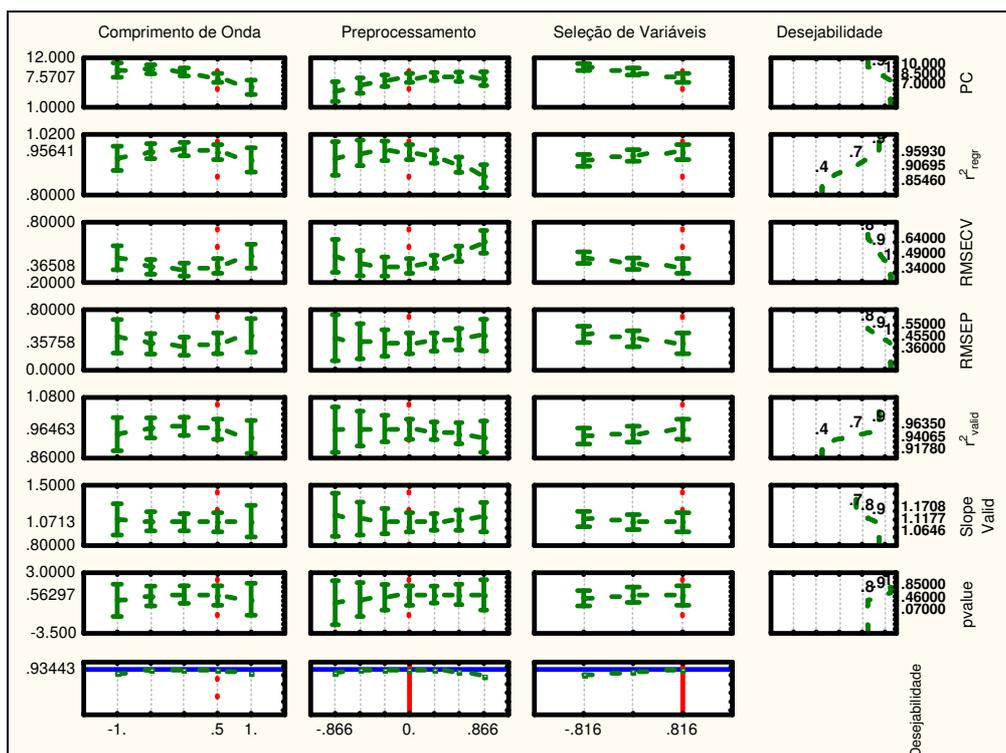


Figura 41. Gráficos de desejabilidade para a propriedade outros.

Tabela 21. Condições experimentais para obtenção dos modelos otimizados.

Componente	D ^a	Faixa de comprimento de onda (cm ⁻¹)	Técnica de pré-processamento ^b	Seleção de variáveis ^c
Acetona	0.94	3700 - 6102	PRE7	Com 2 Jack knife
Acetonitrila	0.91	4181- 4661	PRE1	Com 2 Jack knife
Água	0.93	4661- 6102	PRE3	Com 2 Jack knife
n-Propanol	0.92	4181- 4661e 5622 - 6102	PRE7	Sem Seleção
i-Propanol	0.88	3700 - 6102	PRE7	Com 2 Jack knife
t-Butanol	0.91	3700 – 4661 e 5622 - 6102	PRE7	Sem Seleção
Ácool Alílico	0.89	4661- 6102	PRE2	Com 2 Jack knife
DCPD	0.93	5141,8 - 6102	PRE5	Com 2 Jack knife
Outros	0.93	4181- 4661e 6582,6 - 7543	PRE4	Com 2 Jack knife

^aD é a desejabilidade global

^bDetalhes dos códigos de pré-processamento de dados na Tabela 9.

^cQuantidade de rodadas do algoritmo Jack Knife para seleção de variáveis.

4.3.6. Refinamento dos modelos

Estabelecidas as condições ideais de cada variável (Tabela 21), novos modelos de calibração foram obtidos nessas condições, os quais foram sucessivamente refinados, excluindo-se as amostras que apresentaram resíduos de concentrações elevados.

As figuras de mérito dos melhores modelos de calibração obtidos após o refinamento estão apresentados na Tabela 22.

Tabela 22. Figuras de mérito dos modelos de calibração.

Componente	n ^a	PC ^b	Faixa de calibração ^c (% m/m)	r ² _{regr}	RMSECV (% m/m)
Acetona	101	4	0,00 - 20,04	0,9836	0,70
Acetonitrila	98	2	8,68 - 89,97	0,9811	2,53
Água	93	1	0,00 - 45,78	0,9931	0,99
n-Propanol	85	10	0,00 - 9,52	0,9888	0,27
i-Propanol	89	7	0,00 - 10,00	0,9825	0,31
t-Butanol	99	5	0,00 - 19,23	0,9820	0,81
Ácool Alílico	91	6	0,00 - 52,05	0,9677	2,32
DCPD	97	6	0,00 - 9,44	0,9734	0,37
Outros	104	7	0,00 - 4,95	0,9239	0,44

^aNúmero de amostras do modelo de calibração.

^bNúmero de componentes principais do modelo PLS.

^cFaixa de aplicação de cada modelo, baseada no conjunto de calibração.

Pelo menos 85 amostras de calibração foram usadas em cada modelo, sendo necessário entre 1 e 10 fatores para assegurar o melhor desempenho. Para os componentes n-propanol e i-propanol um maior número de fatores foi requerido, provavelmente devido à alta complexidade da amostra e a presença de outros alcoóis na amostra com concentrações mais elevadas. Também foi necessário um

número elevado de fatores para modelagem da propriedade outros, uma vez que essa propriedade representa a soma de várias espécies químicas.

Os valores típicos do processo (Tabela 3) estiveram sempre dentro da faixa de aplicação de cada modelo, baseada nos dados do conjunto de calibração, de forma que o risco de ocorrência de *outliers* por extrapolações no uso dos modelos foi minimizado. Os resultados confirmam o bom desempenho dos modelos PLS em prever a composição química de um solvente complexo orgânico-aquoso, uma vez que os coeficientes de determinação de validação cruzada (r^2_{regr}) foram maiores que 0,92 para todos os modelos.

Vale ressaltar o excelente desempenho de componentes majoritários da amostra como água, acetona e acetonitrila, os quais apresentaram coeficientes de regressão da validação cruzada acima de 0,98. Além disso, menos de 4 fatores foram necessários para a obtenção de um modelo PLS satisfatório para essas predições, notadamente para a água. Tal componente, além de apresentar-se em alta concentração na amostra, possui bandas intensas e bem definidas na região NIR, de forma que sua determinação por essa técnica é bastante favorecida. O álcool alílico, outro componente presente em alta concentração, apresentou um bom desempenho, embora a presença de outros alcoóis na amostra influenciem nessa determinação, como pode-se observar pelo maior número de fatores requerido bem como pelo menor r^2_{regr} e maior RMSECV obtidos.

4.4. Validação estatística dos modelos

Depois de refinados, os modelos foram submetidos ao procedimento de validação estatística. Os valores dos erros médios quadráticos de previsão (RMSEP) foram comparados com os valores de reprodutibilidade intralaboratorial do método de referência. Os resultados estão apresentados na Tabela 23.

Observa-se que os valores obtidos para RMSEP apresentaram-se superiores aos valores de reprodutibilidade intralaboratorial, embora na mesma ordem de grandeza para a maioria dos componentes. Vale ressaltar que, utilizando técnicas e equipamentos distintos, o uso de valores de reprodutibilidade, não disponíveis para o método de referência, permitiria uma comparação mais assertiva com o RMSEP. Os componentes água, acetonitrila e álcool alílico apresentaram os maiores valores de RMSEP, visto que esses são os compostos presentes em alta concentração na amostra.

No entanto, no desenvolvimento de um método analítico deve-se buscar sua adequação ao propósito analítico. Dessa forma, os valores obtidos para o RMSEP estão dentro dos limites de aceitação definidos para essa aplicação industrial, aceitando-se uma incerteza de até 3%*m/m* nos níveis de concentração dos componentes principais, uma vez que trata-se do controle interno de uma corrente intermediária do processo.

É importante mencionar que a predição do componente t-butanol apresentou erro de predição significativamente maior do que a reprodutibilidade intralaboratorial do método de referência, possivelmente devido à baixa concentração deste componente na amostra, a qual é constituída também por diversos tipos de alcoóis com concentrações mais elevadas. Entretanto, nenhum dos valores observados para o p_{value} foi menor que o valor de p_{value} crítico (95% de confiança). Dessa forma, a hipótese nula de que não existe diferença significativa entre os valores preditos e medidos não deve ser rejeitada e os modelos foram aceitos. A boa correlação entre o método NIR e o método de referência pode ser visualizada através dos gráficos de correlação para os dados das amostras de validação externa através das Figuras de 42 a 49. Nesses gráficos, pode-se observar ainda uma maior densidade de dados nas regiões próximas aos valores típicos observados no processo.

Tabela 23. Resultados da validação estatística dos modelos PLS.

Componente	Método de Referência	NIR		
	Reprodutibilidade intralaboratorial ^a (% m/m)	Validação externa		
		N ^b	RMSEP (% m/m)	p _{value} t pareado ^c
Acetona	0,15	51	0,54	0,50
Acetonitrila	1,71	48	1,30	0,86
Água	1,2	48	1,05	0,58
n-Propanol	0,08	52	0,56	0,34
i-Propanol	0,07	50	0,43	0,39
t-Butanol	0,004	50	0,70	0,67
Ácool Alílico	0,41	45	1,66	0,36
DCPD	0,06	55	0,39	0,08
Outros	-	54	0,41	0,96

^aRefere-se aos valores de precisão intermediária ou reprodutibilidade intralaboratorial do método de referência calculados sobre os valores típicos mostrados na Tabela 2. A precisão intermediária refere-se aos resultados observados quando o fator analista varia, e os outros fatores como método e equipamento são mantidos constantes.

^bnúmero de amostras usadas na etapa de validação externa.

^cBaseado em um intervalo de confiança de 95%.

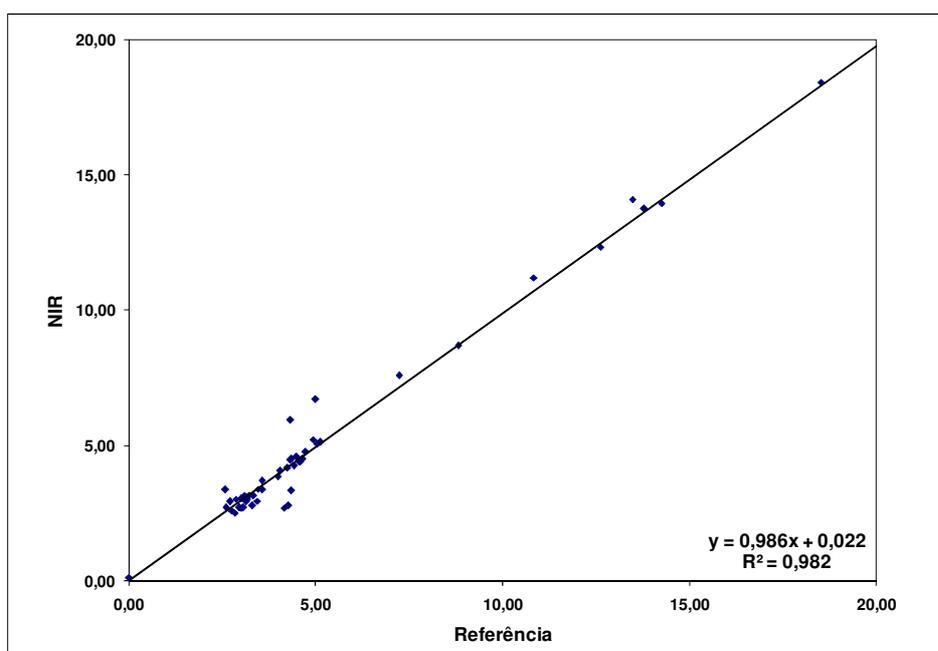


Figura 42. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade acetona.

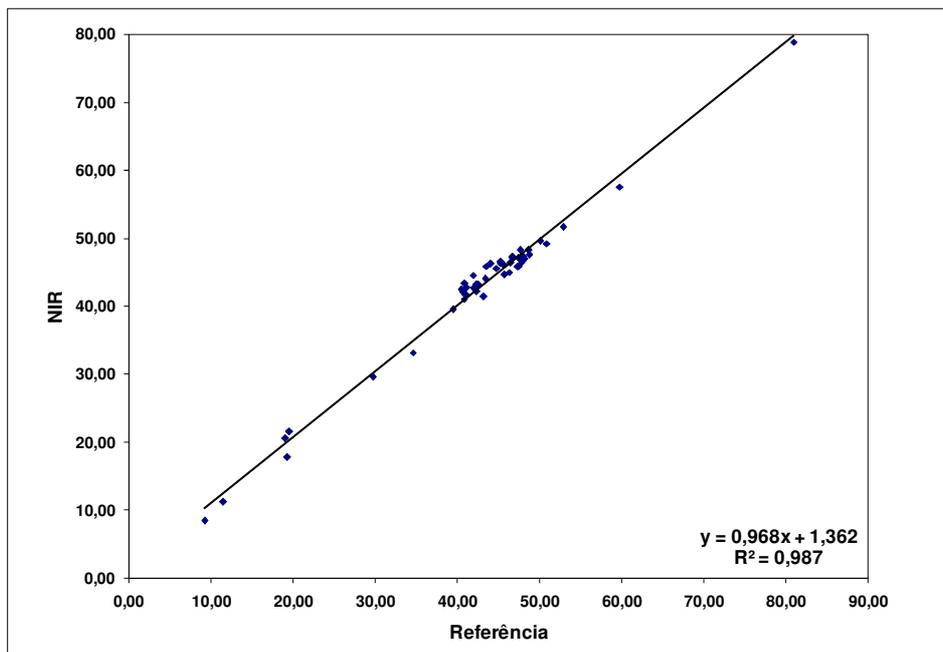


Figura 43. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade acetonitrila.

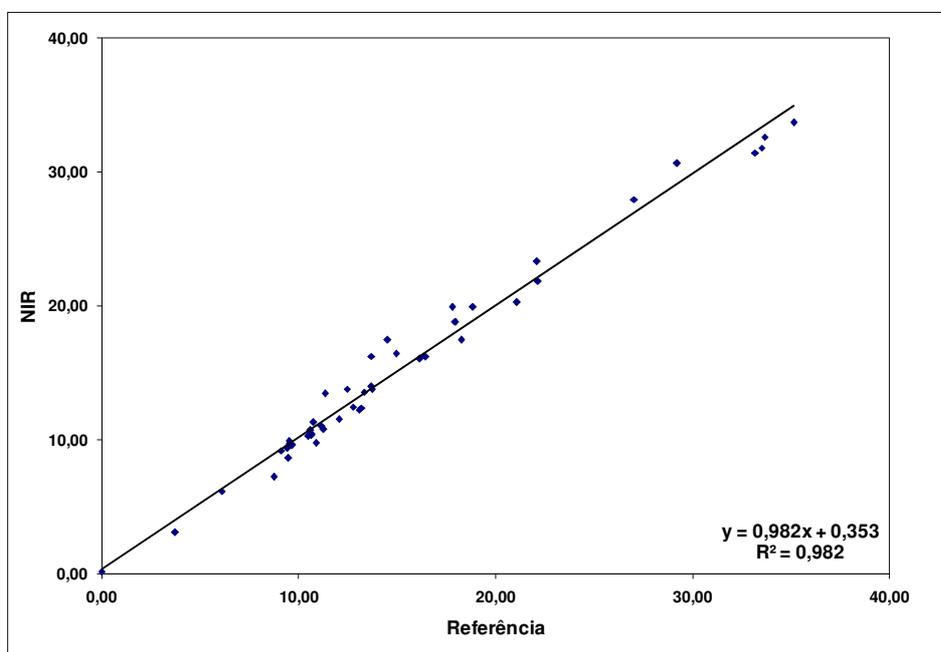


Figura 44. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade água.

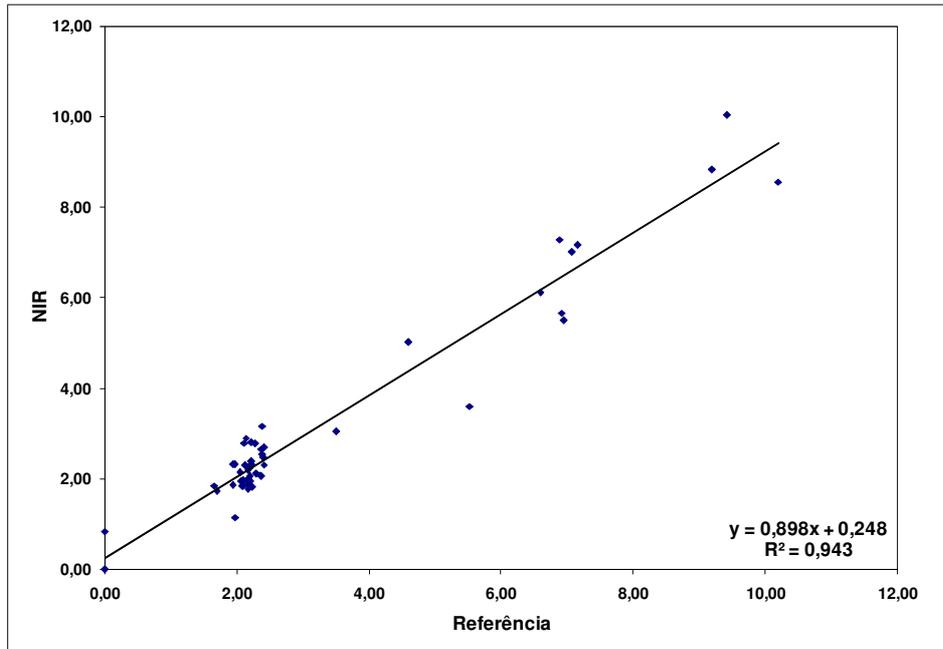


Figura 45. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade n-propanol.

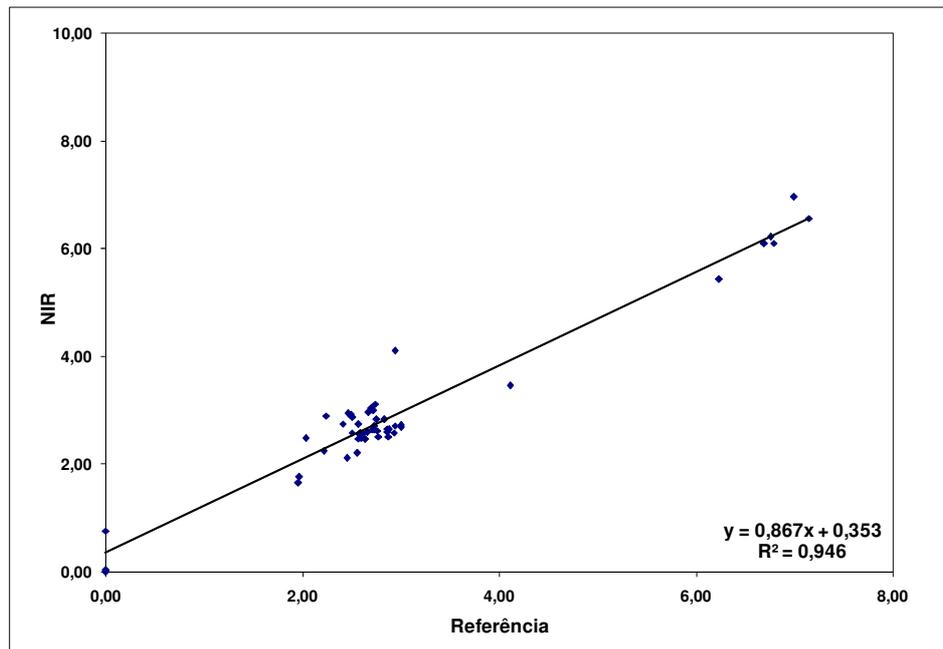


Figura 46. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade i-propanol.

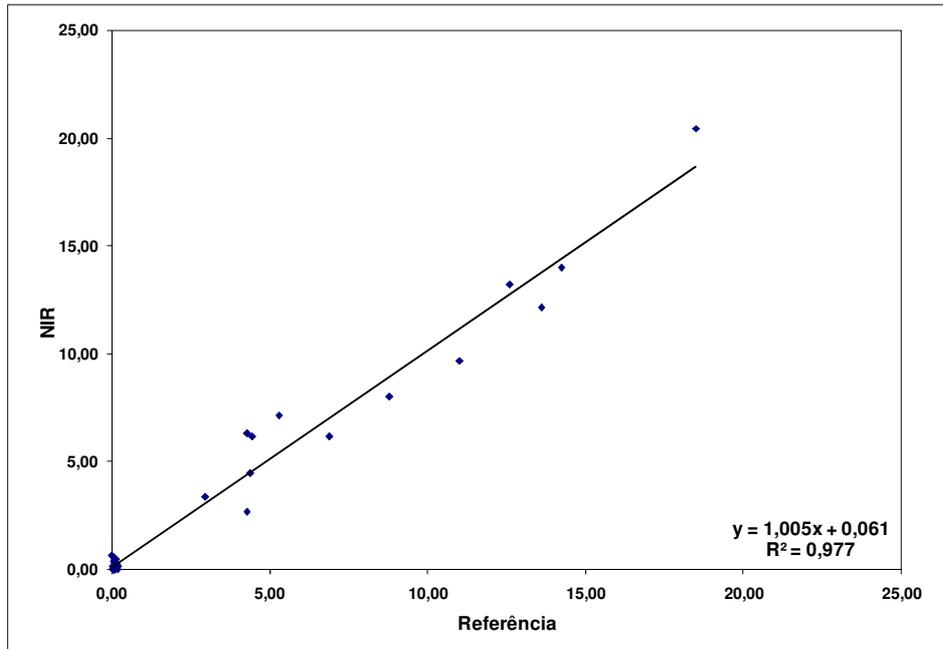


Figura 47. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade t-butanol.

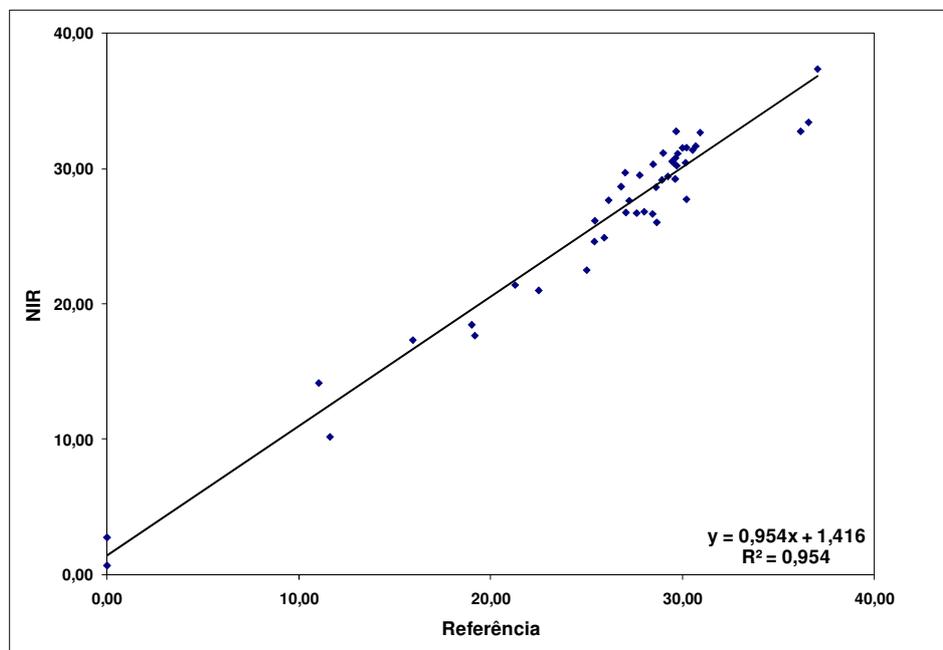


Figura 48. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade álcool alílico.

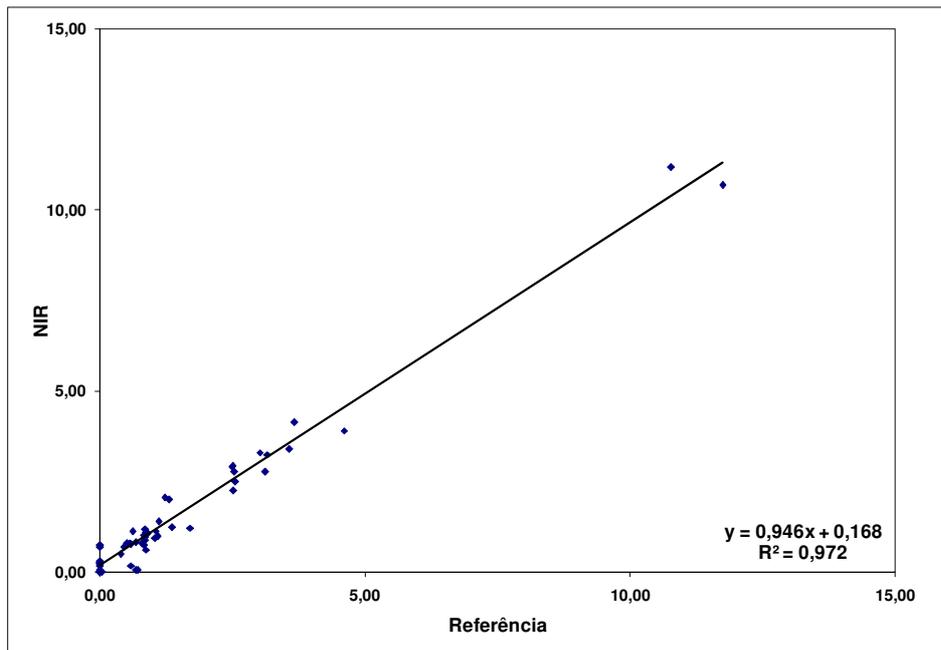


Figura 49. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade DCPD.

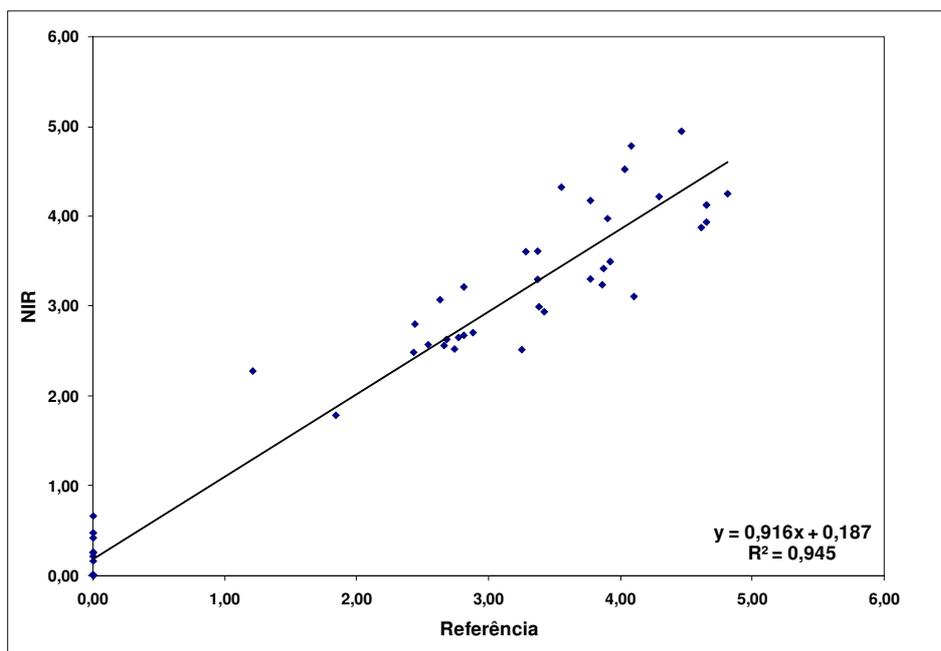


Figura 50. Gráfico de correlação dos valores de concentração preditos vs medidos para a propriedade Outros.

5. CONCLUSÕES

O desenvolvimento deste trabalho demonstrou a possibilidade de utilização da espectroscopia NIR como um método rápido e exato na predição da composição química de amostras complexas de um solvente orgânico-aquoso utilizado no processo de produção de isopreno, na planta petroquímica da Braskem.

Faz-se necessário enfatizar a grande vantagem em análises de rotina na obtenção da composição química completa a partir de um único espectro NIR. Além disso, a metodologia proposta provê respostas rápidas ao processo produtivo, garantindo a robustez necessária quanto à presença de particulados e compostos pesados. Por exemplo, a análise cromatográfica é realizada em aproximadamente 20 minutos, enquanto que menos de 5 minutos são requeridos para todo o procedimento de predição, incluindo limpeza da célula, acondicionamento da amostra, e aquisição dos espectros de ar (referência) e da amostra.

Uma contribuição importante da metodologia de construção dos modelos de calibração foi a utilização da matriz Doehlert no planejamento dos experimentos. A matriz aumentou a eficiência do processo de construção dos modelos PLS, pois permitiu um maior número de combinações das variáveis que interferem no desempenho dos modelos, além de ter possibilitado obter os modelos realizando um menor número de experimentos.

A função de deseabilidade, utilizada no processo de otimização dos modelos, possibilitou a obtenção das melhores condições para construção de novos modelos PLS. Os modelos construídos proporcionaram resultados satisfatórios graças à avaliação das multirespostas obtidas para as variáveis: comprimento de onda, pré-processamento e a técnica de seleção de variáveis Jack Knife.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Macho, S.; Larrechi, M. S.; Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry, **Trends in analytical chemistry**, v. 21, p. 799-806, 2002.
- [2] Chung, H.; Applications of Near-Infrared Spectroscopy in Refineries and Important Issues to Address, **Applied Spectroscopy Reviews**, v. 42, p. 251-285, 2007.
- [3] Honorato, F. A.; Barros Neto, B.; Pimentel, M. F.; Stragevitch, L.; Galvão, R. K. H.; Using principal component analysis to find the Best calibration settings for simultaneous spectroscopic determination of several gasoline properties, **Fuel**, v. 87, p. 3706-3709, 2008.
- [4] Kelly, J. J.; Callis, J. B.; Nondestructive Analytical Procedure for Simultaneous Estimation of the Major Classes of Hydrocarbon Constituents of Finished Gasolines, **Analytical Chemistry**, v. 62, p. 1444-1451, 1990.
- [5] Kelly, J. J.; Barlow, C. H.; Jinguji, T. M.; Callis, J. B.; Prediction of Gasoline Octane Numbers from Near-Infrared Spectral Features in the Range 660-1215 nm, **Analytical Chemistry**, v. 61, p. 313-320, 1989.
- [6] Côcco, L. C.; Yamamoto, C. I.; Von Meien, O. F.; Study of correlations for physicochemical properties of Brazilian gasoline, **Chemometrics and Intelligent Laboratory Systems**, v. 76, p. 55 – 63, 2005.
- [7] Bohács, G.; Ovádi, Z.; Prediction of gasoline properties with near infrared spectroscopy, **Journal of Near Infrared Spectroscopy**, v. 6, p. 341 – 348, 1998.
- [8] Choquette, S. J.; Chesler, S. N.; Duewer, D. L.; Identification and Quantitation of Oxygenates in Gasoline Ampules Using Fourier Transform Near-Infrared and Fourier Transform Raman Spectroscopy, **Analytical Chemistry**, v. 68, p. 3525 – 3533, 1996.
- [9] Parisi, A. F.; Nogueiras, L.; Prieto, H.; On-line determination of fuel quality parameters using near-infrared spectrometry with fibre optics and multivariate calibration, **Analytica Chimica Acta**, v. 238, p. 95-100, 1990.

- [10] Rebouças, M. V.; Santos, E. C.; Vieira, F. S. V.; Feasibility of quality process control of a naphtha fractioning unit based on near-infrared spectroscopic prediction of physical and chemical properties of medium naphtha streams, **Vibrational Spectroscopy**, v. 44, p. 187-191, 2007.
- [11] Lee, Y.; Chung, H.; Kim, N.; Spectral Range Optimization for the Near-Infrared Quantitative Analysis of Petrochemical and Petroleum Products: Naphtha and Gasoline, **Applied Spectroscopy**, v. 60, p. 892 – 897, 2006.
- [12] Falla, F. S.; Larini, C.; Le Roux, G. A. C.; Quina, F. H.; Moro, L. F. L.; Nascimento, C. A. O.; Characterization of crude petroleum by NIR, **Journal of Petroleum Science and Engineering**, v.51, p. 127-137, 2006.
- [13] Pasquini, C.; Bueno, A. F.; Characterization of petroleum using near-infrared spectroscopy: Quantitative modeling for the true boiling point curve and specific gravity, **Fuel**, v. 86, p. 1927–1934, 2007.
- [14] Rebouças, M. V.; Santos, J. B.; Domingos, D.; Massa, A. R. C. G.; Near-infrared spectroscopic prediction of chemical composition of a series of petrochemical process streams for aromatics production, **Vibrational Spectroscopy**, v. 52, p. 97-102, 2010.
- [15] Rebouças, M. V.; Barros Neto, B.; Near infrared spectroscopic prediction of physical properties of aromatics-rich hydrocarbon mixtures, **Journal of Near Infrared Spectroscopy**, v. 9, p. 263-273 , 2001.
- [16] Chung, H.; Lee, J-S.; Ku, M-S.; Feasibility of Simultaneous Measurement of Xylene Isomers and Other Hydrocarbons in p-Xylene Production Process Using Near-Infrared Spectroscopy, **Applied Spectroscopy**, v. 52, p. 885-889, 1998.
- [17] Fernandes, H. L.; Raimundo Jr., I. M.; Pasquini, C.; Rohwedder, J. J. R.; Simultaneous determination of methanol and ethanol in gasoline using NIR spectroscopy: Effect of gasoline composition, **Talanta**, v. 75, p. 804-810, 2008.
- [18] Mendes, L. S.; Oliveira, F. C. C.; Suarez, P. A. Z.; Rubim, J. C.; Determination of ethanol in fuel ethanol and beverages by Fourier transform (FT)-near infrared and FT-Raman spectrometries, **Analytica Chimica Acta**, v. 493, p. 219-231, 2003.

- [19] Garrigues, S.; Gallignani, M.; De la Guardia, M.; Flow-injection determination of water in organic solvents by near-infrared spectrometry, **Analytica Chimica Acta**, v. 281, p. 259-264, 1993.
- [20] Herschel, W.; Investigation of the Powers of the Prismatic Colours to Heat and Illuminate Objects, **Philosophical Transactions**, v. 90, p. 255-283, 1800.
- [21] Herschel, W.; Experiments on the Refrangibility of the Invisible Rays on the Sun, **Philosophical Transactions**, v. 90, p. 284-292, 1800.
- [22] Hart, J. R.; Norris, K. H.; Golumbic, C.; Determination of the moisture content of seed by near-infrared spectrophotometry of their methanol Extracts, **Cereal Chemistry**, v. 39, p. 94-98, 1962.
- [23] Pasquini, C.; Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Application, **Journal of Brazilian Chemical Society**, v. 14, p. 198-219, 2003.
- [24] Workman Jr, J.; Weyer, L.; Practical Guide to Interpretive Near-Infrared Spectroscopy, CRC Press, Boca Raton, 2008.
- [25] Workman Jr, J.; Handbook of organic compounds: Methods and Interpretation, v. 1, 1st ed., Academic Press, San Diego, 2001.
- [26] Bokobza, L.; Near infrared spectroscopy, **Journal of Near Infrared Spectroscopy**, v. 6, p. 3-17, 1998.
- [27] Wold, S.; Chemometrics; what do we mean with it, and what do we want from it?, **Chemometrics and Intelligent Laboratory Systems**, v. 30, p. 109-115, 1995.
- [28] Beebe, K. R.; Randy, J. P.; Seasholtz, M. B., Chemometrics: A practical guide, Wiley-Interscience, New York, 1998.
- [29] Zeaiter, M.; Roger, J.M.; Bellon-Maurel, V.; Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods, **Trends in Analytical Chemistry**, v. 24, p. 437-445, 2005.

- [30] Oliveira, F. C.; Souza, A. T. P. C.; Dias, J. A.; Dias, S. C. L.; Rubim, J. C.; A escolha da faixa espectral no uso combinado de métodos espectroscópicos e quimiométricos, **Química Nova**, v. 27, p. 218-225, 2004.
- [31] Leardi, R.; González, A. L.; Genetic Algorithms applied to feature selection in PLS regression: how and when to use them, **Chemometrics and Intelligent Laboratory Systems**, v. 41, p. 195-207, 1998.
- [32] Norgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L; Engelsen, B.; Interval partial least-square regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy, **Applied Spectroscopy**, v.54, p.413-4193, 2000.
- [33] Pontes, M. J. C.; Galvão, R. K. H.; Araújo, M. C. U.; Moreira, P. N. T.; Pessoa Neto, O. D.; José, G. E.; Saldanha, T. C. B.; The successive projections algorithm for spectral variable selection in classification problems, **Chemometrics and Intelligent Laboratory Systems**, v.78, p.11-18, 2005.
- [34] Araújo, M. C. U.; Saldanha, T. C. B.; Galvão, R. K. H.; Yoneyama, T.; Chame, H. C.; Visani, V.; The successive projections algorithm for variable selection in spectroscopy multicomponent analysis, **Chemometrics and Intelligent Laboratory Systems**, v.57 (2), p.65-73, 2001.
- [35] Brereton, R. G.; Introduction to multivariate calibration in analytical chemistry, **The Analyst**, v.125, p.2125-2154, 2000.
- [36] Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S., Volpe, P. L. O.; Quimiometria I: calibração multivariada, um tutorial, **Química Nova**, v. 22, p. 724-731, 1999.
- [37] Dardenne, P.; Sinnaeve, G.; Baeten, V.; Multivariate calibration and chemometrics for near infrared spectroscopy: which method?, **Journal of Near Infrared spectroscopy**, v. 8, p. 229-237, 2000.
- [38] Dantas Filho, H. A.; Galvão, R. K. H.; Araújo, M. C. U.; Silva, E. C.; Saldanha, T. C. B.; José, G. E.; Pasquini, C.; Raimundo Jr., I. M.; Rohwedder, J. J. R.; A Strategy for selecting calibration samples for multivariate modeling, **Chemometrics and Intelligent Laboratory Systems**, v.72, p.83-91, 2004.

- [39] Galvão, R. K. H.; Araújo, M. C. U.; José, G. E.; Pontes, M. J. C.; Silva, E. C.; Saldanha, T. C. B.; A method for calibration and validation subset partitioning, **Talanta**, v. 67, p. 736-738, 2005. [35] Costa Filho, P.A; Poppi, R.J.; Algoritmo genético em química, **Química Nova**, v. 22, p. 405-411, 1999.
- [40] Honorato, F. A.; Barros Neto, B.; Martins, M. N.; Galvão, R. K. H.; Pimentel, M. F.; Transferência de calibração em métodos multivariados, **Química Nova**, v. 30, p. 1301-1312, 2007.
- [41] Martens, H., Martens, M.; Modified Jack-Knife estimation of parameter uncertainty in bilinear modeling by partial least squares regression (PLSR), **Food Quality and Preference**, v. 11, p. 5 – 16, 2000.
- [42] Wold, S.; Sjostrom, M.; Eriksson, L.; PLS-regression: a basic tool of chemometrics, **Chemometrics and Intelligent Laboratory Systems**, v. 58, p. 109 – 130, 2001.
- [43] Wehrens, R.; Putter, H.; Buydens, L. M. C.; The bootstrap: a tutorial, **Chemometrics and Intelligent Laboratory Systems**, v. 54, p. 35 – 52, 2000.
- [44] Helland, I.S.; Some Theoretical aspects of partial least squares regression, **Chemometrics and Intelligent Laboratory Systems**, v. 58, p. 97-107, 2001.
- [45] Geladi, P.; Kowalski, B. R.; Partial Least Squares Regression: A Tutorial, **Analytical Chemistry Acta**, 1-17, 1986.
- [46] Larrechi, M.S.; Callao, M.P.; Strategy for introducing NIR spectroscopy and multivariate calibration techniques in industry, **Trends in Analytical Chemistry**, v. 22, p. 635-636, 2003.
- [47] Barros Neto, B.; Scarminio, I. S.; Bruns, R. E.; Como Fazer Experimentos, Editora da Unicamp, Campinas, 2001.
- [48] Teófilo, R. F.; Ferreira, M. M. C.; Quimiometria II: planilhas eletrônicas para cálculos de planejamentos experimentais, um tutorial, **Química Nova**, v. 29, p. 338-350, 2006.

- [49] Ferreira, S. L. C.; Santos, W. N.; Quintella, C. M.; Barros Neto, B. B.; Bosque-Sendra, J. M.; Doehlert matrix: a chemometric tool for analytical chemistry - a review, **Talanta**, v. 63, p. 1061-1067, 2004.
- [50] Bosque-Sendra, J. M.; Nechar, M.; Rodriguez, L. C.; Molina, M. F. M.; Optimization of Analytical Methods by Using Doehlert`s Designs, **Analytical Proceedings Including Analytical Communications**, v. 32, p. 375-377, 1995.
- [51] Rebouças, M. V.; Determinação de Arsênio em nafta e condensado de petróleo por espectrometria de absorção atômica com atomização eletrotérmica utilizando lantânio como modificador, Universidade Federal da Bahia, 2004.
- [52] Doehlert, D. H.; Uniform Shell Designs, **Applied Statistics**, v.19, p. 231-239, 1970.
- [53] Jun, G.; Yandong, S.; Jibin, D.; Bo, Z.; Jiangtao, X.; Determination of Olefin Hydrocarbons in Gasoline by Partial Least-Squares, **Intelligent Computation Technology and Automation (ICICTA)**, v. 3, p. 79 – 81, 2010.
- [54] Caneca, A. R.; Pimentel, M. F.; Galvão, R. K. H.; Matta, C. E.; Carvalho, F. R.; Raimundo Jr., I. M.; Pasquini, C.; Rohwedder, J. J. R.; **Talanta**, v. 70, p. 344 – 352, 2006.
- [55] Rebouças, M. V.; Santos, J. B.; Pimentel, M. F.; Teixeira, L. S. G.; A novel approach for development of a multivariate calibration model using a Doehlert experimental design: Application for prediction of key gasoline properties by Near-infrared Spectroscopy, **Chemometrics and Intelligent Systems**, v. 107, p. 185 – 193, 2011.
- [56] Libnau, F. O.; Kvalheim, O. M.; Christy, A. A.; Toft, J; Spectra of water in the near- and mid-infrared region, **Vibrational Spectroscopy**, v. 7, p. 243 – 254, 1994.
- [57] Maeda, H.; Wang, Y.; Ozaki, Y.; Suzuki, M.; Czarnecki, M. A.; Iwahashi, M.; A near-infrared study of hydrogen bonds in alcohols – comparison of chemometrics and spectroscopic analysis, **Chemometrics and Intelligent Systems**, v. 45, p. 121 – 130, 1999.
- [58] Czarnecki, M. A.; Wojtków, D.; Effect of varying water content on the structure of

butyl alcohol/water mixtures: FT-NIR two-dimensional correlation and chemometrics studies, **Journal of Molecular Structure**, v. 883 – 884, p. 203 – 208, 2008.