



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

**JAMILE BATISTA DOS SANTOS**

**CALIBRAÇÃO MULTIVARIADA NO INFRAVERMELHO PRÓXIMO  
PARA PREDIÇÃO DA COMPOSIÇÃO QUÍMICA DE CORRENTES  
PETROQUÍMICAS DO PROCESSO DE PRODUÇÃO DE  
AROMÁTICOS**

Salvador  
2011

**JAMILE BATISTA DOS SANTOS**

**CALIBRAÇÃO MULTIVARIADA NO INFRAVERMELHO PRÓXIMO  
PARA PREDIÇÃO DA COMPOSIÇÃO QUÍMICA DE CORRENTES  
PETROQUÍMICAS DO PROCESSO DE PRODUÇÃO DE  
AROMÁTICOS**

Tese apresentada como requisito para obtenção do título de mestre em Química pela Universidade Federal da Bahia.

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

Salvador  
2011

Aos meus pais Augusto (em nossos corações) e Rita por  
todo amor, compreensão e por ter sempre acreditado  
em mim e ser o meu maior exemplo de vida!  
Às minhas irmãs, Nayara e Caedma  
Ao meu filhinho, Augusto Neto  
Ao meu esposo, Aldeir  
Com amor e gratidão

Dedico

## **AGRADECIMENTOS**

- A Deus, por me permitir conquistar meus sonhos e renovar as minhas forças a cada manhã, por ter me dado uma família abençoada e amigos maravilhosos. Por tudo que tenho e que sou.
- Aos meus familiares pela paciência e compreensão. Em especial a minha mãe Rita, minhas irmãs Nayara e Caedma.
- Ao meu esposo Aldeir pelo apoio, compreensão, carinho e dedicação. A meu filhinho, por servir como fonte de inspiração e incentivo.
- Ao meu co-orientador Márcio Rebouças pela excelente orientação, apoio, incentivo, amizade, e pela oportunidade de realização deste trabalho.
- Ao meu orientador Leonardo Teixeira pela orientação, apoio e compreensão.
- Ao pessoal do laboratório de controle de qualidade da UNIB em especial a: Denise Brandão, Daniela Domingos, Leilacy, Ana Rosa, Ezequiel, Marcos, a todos que porventura não citei, mas foram igualmente importantes para esta jornada, agradeço pelo apoio, convivência e amizade.
- A Ludmila Ramos pelo incentivo e compreensão, ao pessoal do laboratório da Bahia Specialty Cellulose, em especial a Ehmilia e Adeilza pelo apoio e amizade.
- A Professora Fernanda Pimentel e seus alunos, pela acolhida durante a visita nos dias que passei com eles.
- A minha querida amiga Danúsia sempre fiel e companheira.
- A petroquímica Braskem S.A.
- Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento no início deste projeto.
- Ao Instituto Nacional de Ciências e Tecnologias Analíticas Avançadas-INCTAA.

- À todos os Professores, Funcionários e Colegas do Instituto de Química da UFBA, que contribuíram, direta ou indiretamente, para o sucesso deste trabalho.
- À coordenação de Pós-Graduação em Química e aos professores do Departamento de Química pela minha formação acadêmica.
- Finalmente, a todos aqueles que de alguma forma contribuíram para à realização deste gratificante trabalho.

## SUMÁRIO

<b>ÍNDICE DE FIGURAS.....</b>	<b>ix-xi</b>
<b>ÍNDICE DE TABELAS.....</b>	<b>xii-xiii</b>
<b>LISTA DE ABREVIATURAS E SIGLAS.....</b>	<b>xiv</b>
<b>RESUMO.....</b>	<b>xv</b>
<b>ABSTRACT.....</b>	<b>xvi</b>
<b>CAPÍTULO 1- INTRODUÇÃO.....</b>	<b>1-2</b>
<b>CAPÍTULO 2 – FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>3-34</b>
<b>2.1. A INDÚSTRIA PETROQUÍMICA E A PRODUÇÃO DE</b>	
<b>AROMÁTICOS.....</b>	<b>4-7</b>
2.1.1. Características e aplicações dos aromáticos.....	4-5
2.1.2. Processos Petroquímicos para Obtenção dos Xilenos.....	5-7
<b>2.2. ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO.....</b>	<b>8-13</b>
2.2.1 Histórico e Aplicações. ....	8-9
2.2.2. Princípios da espectroscopia no infravermelho próximo.....	9-12
2.2.3. Principais Bandas de absorção dos aromáticos na região NIR...13	
<b>2.3. MÉTODOS DE ANÁLISE MULTIVARIADA.....</b>	<b>14-17</b>
2.3.1. Quimiometria.....	14-17
2.3.2. Calibração Multivariada .....	14
2.3.2.1 Métodos de Regressão Multivariada.....	14-17
<b>2.3.3. TRATAMENTO DOS DADOS.....</b>	<b>16-23</b>
2.3.3.1 processamento dos dados espectrais.....	17
2.3.3.1.1. Aplicação de derivadas.....	18

2.3.3.1.2. Normalização.....	18
2.3.3.1.3. Centralização na Média.....	18
2.3.3.1.4. Escalonamento.....	18
<b>2.3.3.2 SELEÇÃO DE VARIÁVEIS E AMOSTRAS.....</b>	<b>19</b>
2.3.3.2.1 Algoritmo iPLS.....	19-21
2.3.3.2.2 Algoritmo Jack-Knife.....	21
2.3.3.2.3 Algoritmo de Kennard-Stone.....	22
2.3.3.2.4 Algoritmo SPXY.....	22-23
<b>2.3. PLANEJAMENTO EXPERIMENTAL.....</b>	<b>23-24</b>
2.3.4.1. Matriz Doehlert.....	25-28
2.3.4.2. Experimentos com Multi-respostas Usando uma Função de Desejabilidade.....	8-30
<b>2.4. APLICAÇÕES DA ESPECTROSCOPIA NIR NA INDÚSTRIA PETROQUÍMICA.....</b>	<b>31-33</b>
<b>2.5. OBJETIVOS.....</b>	<b>34</b>
2.5. 1. Objetivos Específicos.....	34
<b>CAPÍTULO 3 - PARTE EXPERIMENTAL.....</b>	<b>36-45</b>
3.1. Amostras .....	36-37
3.2. Reagentes.....	37
3.3. Instrumentação.....	37-38
3. 3.1. Obtenção dos espectros NIR.....	38
3.4. Programas Estatísticos.....	38
3.5. Construção dos modelos – Análise multivariada.....	39-40
3.6. Detecção de amostras anômalas ( <i>Outleir</i> ).....	41
3.7. Divisão do conjunto de amostra: seleção das amostras para calibração e validação.....	41

3.8. Seleção das variáveis espectrais.....41  
3.9. Planejamento Experimentos Empregando Matriz Doehlert.....42-45

**CAPÍTULO 4 - RESULTADOS E DISCUSSÕES.....49-55**

**4.1. ATRIBUIÇÃO DAS BANDAS DOS ESPECTROS NIR.....47**

**4.2. CONSTRUÇÃO DOS MODELOS.....48**

4.2.1. Detecção de amostras anômalas.....48

4.2.2. Seleção das variáveis espectrais.....49-55

**4.3. DESENVOLVIMENTO DOS MODELOS DE CALIBRAÇÃO.....56-71**

4.3.1. Otimização do modelo de calibração de Não Aromáticos.....56-58

4.3.2. Otimização dos modelos de calibração de Benzeno.....58-60

4.3.3. Otimização dos modelos de calibração de Tolueno.....61-62

4.3.4. Otimização dos modelos de calibração de Etil-benzeno.....62-64

4.3.5. Otimização dos modelos de calibração de o-xileno, m-xileno e p-xileno.....64-68

4.3.6. Otimização dos modelos de calibração de Aromáticos C8s+ e C9s+.....68-71

**4.4. OTIMIZAÇÃO DOS MODELOS.....71-73**

**4.5. VALIDAÇÃO EXTERNA DOS MODELOS.....73-79**

**CAPÍTULO 5 – CONSIDERAÇÕES FINAIS.....80-81**

**REFERÊNCIAS BIBLIOGRAFICA.....82-89**

## LISTA DE FIGURAS

Figura 1. Diagrama simplificado do processo petroquímico para a produção dos compostos aromáticos.....	6
Figura 2. Diagrama de energia potencial: (1) oscilador harmônico, (2) oscilador anarmônico.....	12
Figura 3. Gráfico do modelo iPLS do número de intervalos versus RMSECV.....	20
Figura 4. Gráficos do modelo iPLS: faixa escolhida para a construção do modelo PLS.....	21
Figura 5. Representação geométrica do planejamento tridimensional Doehlert para três variáveis.....	26
Figura 6: Três projeções planas possíveis para o planejamento Doehlert com três variáveis.....	27
Figura 7. Exemplo da aplicação da função de desejabilidade para cada variável.....	30
Figura 8: Instrumento (com detalhe da cela de transmitância).....	38
Figura 9. Principais etapas desenvolvidas para a obtenção do melhor modelo de calibração.....	40
Figura 10: Espectro NIR típico de cada amostra (A-E).....	47
Figura 11: Gráfico de escores PC1 versus PC2 para a Análise por Componentes Principais (PCA).....	48
Figura 12. Intervalos das variáveis selecionados pelo iPLS para Não Aromáticos. ....	50
Figura 13. Intervalos das variáveis selecionados pelo iPLS para Benzeno.....	50

Figura 14. Intervalos das variáveis selecionados pelo iPLS para Tolueno.....	51
Figura 15. Intervalos das variáveis selecionados pelo iPLS para Etilbenzeno.....	52
Figura 16. Intervalos das variáveis selecionados pelo iPLS para para-Xileno.....	53
Figura 17. Intervalos das variáveis selecionados pelo iPLS para meta-Xileno.....	53
Figura 18. Intervalos das variáveis selecionados pelo iPLS para orto-Xileno.....	54
Figura 19. Intervalos das variáveis selecionados pelo iPLS para Aromáticos C8s+.....	55
Figura 20. Intervalos das variáveis selecionados pelo iPLS para Aromáticos C9s+.....	55
Figura 21: Gráficos da função de desejabilidade para Não Aromáticos.....	57
Figura 22. Gráficos da função de desejabilidade para Benzeno.....	60
Figura 23. Gráficos da função de desejabilidade para Tolueno.....	62
Figura 24. Gráficos da função de desejabilidade para Etil-Benzeno.....	63
Figura 25. Gráficos da função de desejabilidade para-xileno.....	66
Figura 26. Gráficos da função de desejabilidade para meta-xileno.....	67
Figura 27. Gráficos da função de desejabilidade para orto-xileno.....	68

Figura 28. Gráficos da função de desejabilidade para Aromáticos C8s+.....	70
Figura 29. Gráficos da função de desejabilidade para Aromáticos C9s+.....	71
Figura 30. Gráfico de correlação dos valores concentração predito versus medido para Não Aromáticos.....	75
Figura 31. Gráfico de correlação dos valores concentração predito versus medido para Benzeno.....	75
Figura 32. Gráfico de correlação dos valores concentração predito versus medido para Tolueno.....	76
Figura 33. Gráfico de correlação dos valores concentração predito versus medido para Etil-Benzeno.....	76
Figura 34. Gráfico de correlação dos valores concentração predito versus medido para para-xileno.....	77
Figura 35. Gráfico de correlação dos valores concentração predito versus medido para meta-xileno.....	77
Figura 36. Gráfico de correlação dos valores concentração predito versus medido para orto-Benzeno.....	78
Figura 37. Gráfico de correlação dos valores concentração predito versus medido para Aromáticos C8s+.....	78
Figura 38. Gráfico de correlação dos valores concentração predito versus medido para Aromáticos C9s+.....	78

## LISTA DE TABELAS

Tabela 1. Regiões espectrais do infravermelho.....	10
Tabela 2. Valores codificados das coordenadas planas do planejamento Doehlert para três variáveis.....	28
Tabela 3. Composição típica das amostras do processo.....	36
Tabela 4. Planejamento Experimental da mistura.....	37
Tabela 5. Matriz Doehlert para as variáveis: comprimento de onda, pré-processamento e seleção de variáveis, com os valores reais e suas respectivas coordenadas.....	43
Tabela 6: Códigos referentes aos pré-processamentos utilizados.....	43
Tabela 7: Resultados para todas as variáveis de resposta usada para otimização do modelo de Não Aromáticos.....	56
Tabela 8: Resultados para todas as variáveis de resposta usada para otimização do modelo de Benzeno.....	59
Tabela 9: Resultados para todas as variáveis de resposta usada para otimização do modelo de Tolueno.....	61
Tabela 10: Resultados para todas as variáveis de resposta usada para otimização do modelo de Etil-benzeno.....	63
Tabela 11: Resultados para todas as variáveis de resposta usada para otimização do modelo de para-xileno.....	64
Tabela 12: Resultados para todas as variáveis de resposta usadas para otimização do modelo de meta-xileno.....	65
Tabela 13: Resultados para todas as variáveis de resposta usadas para otimização do modelo de orto-xileno.....	65
Tabela 14: Resultados para todas as variáveis de resposta usadas para otimização do modelo de aromático C8s+.....	69
Tabela 15: Resultados para todas as variáveis de resposta usadas para otimização do modelo de aromático C9s+.....	72

Tabela 16: Condições experimentais para obtenção dos modelos otimizados.....	72
Tabela 17: Figuras de mérito dos modelos de calibração.....	73
Tabela 18: Figuras de mérito dos modelos PLS.....	74

## LISTA DE ABREVIATURAS E SÍMBOLOS

BTEX – Mistura de Benzeno, tolueno, etil-benzeno e orto-, meta-, e para-xileno.

CV – Validação Cruzada (*Cross validation*).

iPLS – Mínimos Quadrados Parciais por Intervalo (*Interval Partial Least Squares*).

NIR – Infravermelho Próximo (*Near Infrared*).

NIRS – Infravermelho Próximo (*Near Infrared Spectroscopy*).

PC – Componente Principal (*Principal Component*), fator ou variável latente.

PCA – Análise de Componentes Principais (*Principal Components Analysis*).

PCR – Regressão por Componentes Principais (*Principal Component Regression*).

PLS – Regressão por Mínimos Quadrados Parciais (*Partial Least Squares*).

RMSEC – Raiz Quadrada do Erro Médio Quadrático da Calibração (*Root Mean Square Error of Calibration*).

RMSECV – Raiz quadrada do Erro Médio Quadrático de Validação Cruzada (*Root Mean Square Error of Cross Validation*).

RMSEP – Raiz Quadrada do Erro Médio Quadrático de Previsão (*Root Mean Squares Error of Prediction*).

## RESUMO

Título: Calibração Multivariada no Infravermelho Próximo para Predição da Composição Química de Correntes Petroquímicas do Processo de Produção de Aromáticos

Autora: Jamile Batista

Orientador: Prof. Dr. Leonardo Teixeira

Este trabalho teve como objetivo desenvolver um método analítico para predição da composição química de amostras geradas no processo de produção de aromáticos utilizando a Espectroscopia no Infravermelho Próximo (NIR) associado a técnicas de calibração multivariada. Os conjuntos de calibração e validação foram selecionados através do algoritmo de Kennard-Stone e a técnica estatística utilizada para a calibração multivariada foi a Regressão por Mínimos Quadrados Parciais (PLS). As regiões espectrais selecionadas na etapa de construção dos modelos foram obtidas através do algoritmo de seleção de variáveis por regressão de mínimos quadrados parciais por intervalo (iPLS). Para escolher as condições experimentais mais adequadas para a modelagem PLS foi realizado um planejamento experimental com matriz Doehlert usando três variáveis (pré-processamento, faixa de comprimento de onda e seleção de variáveis espectrais com o algoritmo Jack-knife). Foram desenvolvidos modelos de calibrações para a previsão da concentração de não aromáticos, benzeno, tolueno, etil-benzeno, para-xileno, meta-xileno, orto-xileno, aromáticos C8s+ e aromáticos C9s+ em amostras de correntes petroquímicas; e os RMSEPs encontrados foram 0,88; 0,38; 2,43; 1,19; 1,08; 1,13; 1,29; 3,87; 1,47% (m/m), respectivamente. O desempenho do melhor modelo de calibração de cada propriedade foi avaliado por meio de parâmetros da validação externa. Com os resultados obtidos, pôde-se demonstrar que os modelos construídos foram satisfatórios e os erros encontrados são aceitáveis para controle de processo na indústria.

Palavras-chave: Propriedades de aromáticos, Espectroscopia NIR, planejamento experimental.

## ABSTRACT

Title: Near Infrared Multivariate Calibration for Prediction of Chemical Composition of Petrochemical Streams in the Aromatics Production Process

Author: Jamile Batista

Advisor: Prof. Dr. Leonardo Teixeira

This work aims at developing an analytical method for predicting the chemical composition of samples generated in the production of aromatics using Near Infrared Spectroscopy (NIRS) combined with multivariate calibration techniques. The calibration and validation sets were selected by the Kennard-Stone algorithm. Partial Least Squares Regression (PLS) was the statistical technique used in the multivariate calibration. The spectral regions selected for the model development were obtained using the algorithm for spectral variable selection based on the partial least squares regression interval (iPLS). The most appropriate experimental conditions for PLS modeling were chosen using an experimental design based on a three-variable Doehlert matrix (pre-processing, wavelength range and selection of spectral variables with the Jack-knife algorithm). Calibration models were developed for predicting non-aromatic, benzene, toluene, ethyl benzene, p-xylene, m-xylene, o-xylene, aromatic C<sub>8</sub>s+ and aromatic C<sub>9</sub>s+. The performance of the best calibration model for each property was evaluated by external validation parameters. The PLS prediction of the properties presented RMSEP 0,88; 0,38; 2,43; 1,19; 1,08; 1,13; 1,29; 3,87; 1,47%w/w respectively. The PLS models showed satisfactory results and errors were found to be acceptable to the industry process control.

Key-words: Production Aromatics, NIR Spectroscopy, Experimental Design.

# **CAPÍTULO I**

## **INTRODUÇÃO**

## 1. Introdução

A indústria petroquímica representa um importante segmento produtivo da indústria química nacional. A petroquímica brasileira Braskem S.A é líder do mercado americano de resinas termoplásticas, sendo a primeira petroquímica integrada do país, isto é, que combina operações da primeira e da segunda geração da cadeia produtiva do plástico em uma única empresa.

Esse segmento industrial é o responsável pela produção dos insumos fundamentais utilizados nas indústrias de 2ª geração que produzem as resinas termoplásticas, empregando como matéria-prima principal a nafta, que é obtida através do fracionamento do petróleo. A nafta destaca-se pela diversidade de compostos que dela podem ser obtidos através de craqueamento e/ou reforma catalítica. Eteno, propeno, buteno-1, butadieno-1,3, isopreno, benzeno, tolueno, p-xileno são exemplos desses compostos.

A nafta é uma fração clara do petróleo enriquecida de compostos com 5 a 12 carbonos parafínicos, isoparafínicos e naftênicos, com menor concentração de compostos aromáticos e olefínicos. A nafta alimenta os fornos de pirólise em um processo de craqueamento para produção de eteno, ou é utilizada na reforma catalítica onde os compostos cíclicos (naftênicos) são convertidos em compostos aromáticos. Compostos aromáticos com elevada pureza, tais como benzeno, tolueno, o-xileno e p-xileno são obtidos através de processos de reação catalítica, extração e fracionamento<sup>[1]</sup>.

O processo de produção dos compostos aromáticos é geralmente controlado por analisadores em linha através da cromatografia gasosa (GC), com o objetivo de monitorar variáveis críticas do processo produtivo, garantir a qualidade dos produtos finais e correntes intermediárias, e também para avaliar as condições operacionais que indiquem a necessidade de regeneração do catalisador. Em alguns pontos do processo esse monitoramento ocorre também através da coleta de algumas amostras para obtenção da composição química

completa por GC em bancada. A análise de GC em bancada é realizada geralmente em cerca de 30 a 50 minutos. Desta forma, o desenvolvimento de métodos mais rápidos é importante para fornecer em menor tempo a resposta para o controle do processo.

A espectroscopia na região do infravermelho próximo (NIRS), associada à técnicas de regressão múltipla como PLS (regressão por mínimos quadrados parciais), tem se mostrado apropriada para aplicações petroquímicas, fornecendo a base para prever a concentração de produtos químicos<sup>[2-5]</sup>.

Esse trabalho teve como objetivo determinar a composição de correntes obtidas no processo de produção de aromáticos, utilizando calibração multivariada no infravermelho próximo. Para tal, otimizou-se o processo de construção dos modelos PLS através de planejamento experimental utilizando matriz Doehlert<sup>[6]</sup> com três variáveis. Os efeitos das variáveis nos modelos PLS foram avaliados através da análise de gráficos de desejabilidade, conforme metodologia de análise multi-resposta<sup>[7]</sup>. Dessa forma, pôde-se encontrar o ponto ótimo para cada variável estudada para construção dos modelos de calibração nessas condições.

# **CAPÍTULO II**

## **FUNDAMENTAÇÃO TEÓRICA**

## 2.1 A indústria petroquímica e a produção de Aromáticos

### 2.1.1 Características e Aplicações dos Aromáticos

As frações estudadas neste trabalho foram obtidas na Unidade de Produção de Aromáticos de uma indústria petroquímica, e, portanto, tais amostras são compostas quase totalmente de hidrocarbonetos aromáticos tais como benzeno, tolueno, etil-benzeno e xilenos em elevada concentração.

O benzeno ( $C_6H_6$ ), nas condições normais de temperatura e pressão, é um líquido incolor, possui odor característico, ponto de ebulição relativamente baixo ( $80,1\text{ }^\circ\text{C}$ ) e uma elevada pressão de vapor, o que provoca a sua rápida evaporação à temperatura ambiente. É altamente inflamável e pouco solúvel em água, mas miscível com a maioria dos solventes orgânicos. É usado, principalmente, para produzir outras substâncias químicas, sendo que seus derivados mais largamente produzidos incluem<sup>[8]</sup>: o estireno, o qual é usado para produzir polímeros e plásticos; fenol para resinas e adesivos (via cumeno); e cicloexano, o qual é usado na manufatura de nylon. Quantidades menores de benzeno são usadas para produzir alguns tipos de borrachas, lubrificantes, corantes, detergentes, fármacos, explosivos e pesticidas.

O tolueno é geralmente usado como um substituto para o benzeno, por ser menos tóxico<sup>[8]</sup>. Tolueno é o nome usual do metil-benzeno ( $C_7H_8$ ), solvente aromático de alta pureza. Uma das características principais deste produto é a sua rápida evaporação, combinando com um alto poder de solvência. É usado na fabricação de tintas e vernizes, sendo amplamente utilizado como solvente ou diluente para grande número de resinas. Também vem sendo utilizado nas formulações de adesivos à base de borracha, como solvente em tintas à base de alumínio para repintura de botijões de gás e como matéria-prima na fabricação de explosivos. Outras aplicações incluem a produção de secantes, aditivos, desinfetantes e ácido benzóico.

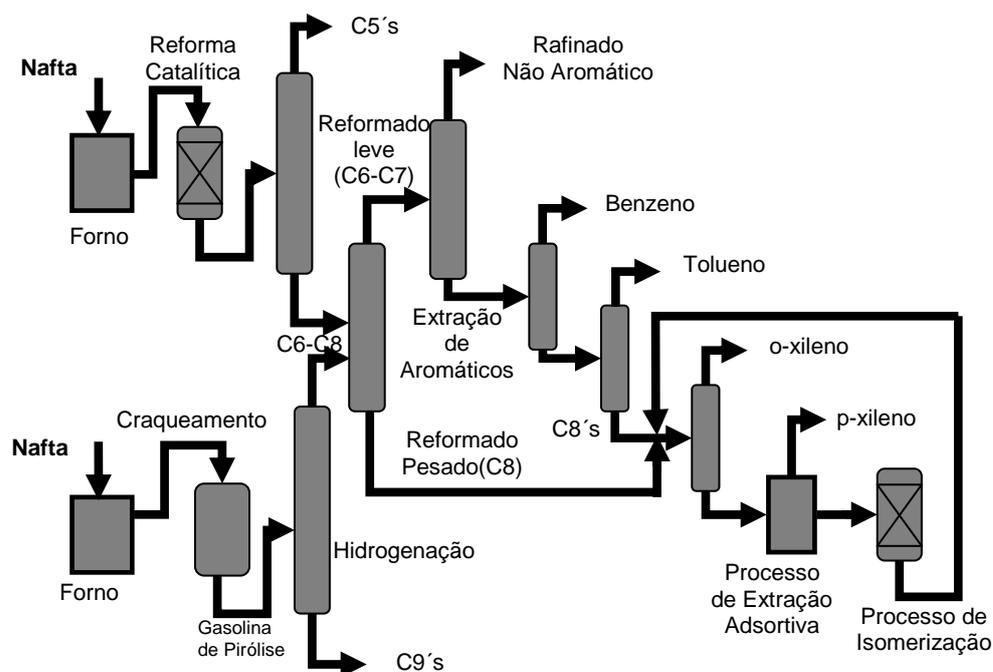
O produto xileno corresponde a uma mistura constituída quase totalmente de hidrocarbonetos aromáticos, com faixa de destilação compreendida entre 137 e 140,2 °C. É essencialmente uma mistura de três isômeros: para-xileno, orto-xileno, e meta-xileno, com pequena quantidade de etil-benzeno. Essa mistura de isômeros possui alto poder de solvência e baixo teor de voláteis. Sua utilização se dá nas indústrias de tintas, colas e adesivos, borracha, corantes, resinas, vernizes, metalurgia, defensores agrícola e eletroeletrônica.

### **2.1.2. Processos Petroquímicos para Obtenção de Aromáticos**

Na indústria petroquímica, a matéria-prima nafta pode alimentar unidades de reforma catalítica, onde compostos cíclicos são convertidos em compostos aromáticos. A nafta, pode também abastecer os fornos da unidade de craqueamento<sup>[8]</sup>, passando por processo no qual irá ocorrer quebra na sua cadeia carbônica, gerando alguns compostos que formam a corrente denominada gasolina de pirólise, que posteriormente será hidrogenada. As frações ricas em compostos com 6 a 8 átomos de carbono obtidas a partir da gasolina de pirólise hidrogenada ou da unidade de reforma catalítica, são processadas nas unidades de Extração de Aromáticos. Nesta unidade é obtida uma corrente composta de benzeno, tolueno e xilenos (BTX) que será submetida a processos de fracionamento, para obtenção no final, de compostos aromáticos com elevada pureza, tais como benzeno, tolueno, o-xileno e p-xileno. A Figura 1 mostra o diagrama simplificado do processo para obtenção de Aromáticos em uma indústria.

Unidades de fracionamento de aromáticos (UFA)<sup>[1]</sup> são comumente encontradas em centrais petroquímicas. A corrente de alimentação da UFA é composta por hidrocarbonetos de volatilidade relativa distinta (corte de hidrocarbonetos C6, C7 e C8) e, portanto, o princípio básico de separação nesta unidade é a destilação multicomponente. Os produtos da UFA são obtidos com alta pureza. O benzeno é obtido com pureza de

99,95 % (m/m), tolueno com pureza mínima de 99,7% (m/m) e uma mistura de xilenos, que produzirá orto-xileno com pureza mínima de 98,0% (m/m) na unidade de fracionamento de xilenos e para-xileno com no mínimo 99,8% (m/m) de pureza na unidade de adsorção de para-xileno.



**Figura 1.** Diagrama simplificado do processo petroquímico para a produção dos compostos aromáticos<sup>[1]</sup>.

Segundo as indicações mostradas na Figura 1, o processo de obtenção dos aromáticos puros é iniciado a partir da fração rica em compostos aromáticos C6-C8, obtida nas unidades de extração de aromáticos, a qual é submetida a processos de fracionamento que visa obter no final, compostos aromáticos, tais como benzeno, tolueno e xilenos. A próxima etapa é a separação dos isômeros de xilenos, principalmente o-xileno e p-xileno.

O processo de fracionamento de aromáticos é composto basicamente de quatro seções: tratamento da carga com argila; coluna de fracionamento para obtenção do benzeno; coluna de fracionamento para obtenção do tolueno; coluna de fracionamento para obtenção de xilenos.

A seção de tratamento da carga com argila ativada visa à remoção de componentes insaturados presentes na carga como olefinas e diolefinas que afetariam a pureza dos aromáticos produzidos na unidade. A corrente de BTX tratada pela torre de argila é enviada para as seções de fracionamento onde ocorre a separação dos componentes aromáticos. A coluna de benzeno produz benzeno produto na retirada lateral. A corrente de fundo da coluna de benzeno é composta de tolueno e xilenos, que são enviados para a coluna de tolueno, onde o tolueno é separado no topo e os xilenos no fundo da coluna. O tolueno separado nesta coluna é enviado para o tanque de produto final.

A corrente de fundo da coluna de tolueno, composta de xilenos e uma pequena quantidade de compostos pesados (C9s e C10s+) são enviadas para a coluna de xilenos. Os xilenos ricos em etil-benzeno (XMEB) são obtidos na corrente de topo e enviados para o tanque de produto final. A fração de xilenos com compostos pesados (orto-xileno e aromáticos C9s+) são separados no fundo e enviados para o fracionamento de Xilenos.

Neste trabalho as amostras utilizadas para construção dos modelos de calibração foram coletadas em pontos no processo de produção de aromáticos para serem analisadas, com objetivo de verificar a sua composição. A determinação da composição dessa corrente pode auxiliar na avaliação do rendimento do processo, no ajuste das condições operacionais, e na unificação do atendimento às especificações dos produtos finais.

## **2.2. Espectroscopia no Infravermelho**

### **2.2.1. Histórico e Aplicações**

Apesar da região do infravermelho próximo (NIR) ter sido a primeira região não-visível de radiação eletromagnética a ser descoberta, as baixas intensidades relativas de bandas de sobretom e combinação e a grande sobreposição entre bandas levaram ao pouco uso dessa região espectral para fins analíticos até a década de 70. Alguns trabalhos de

revisão envolvendo o estudo de espectros de líquidos orgânicos<sup>[9]</sup> foram citados durante esse período, mas ainda assim a espectroscopia no NIR era descrita como uma área de estudo negligenciada<sup>[10]</sup>. Mesmo no início da década de 80, um trabalho refere-se à espectroscopia no NIR como uma técnica analítica adormecida<sup>[11]</sup>.

Apesar dessa técnica ter sido descrita desde o início da década de 30, só durante a década de 60 ocorreu uma aplicação prática, com o grupo de pesquisadores do Professor Karl Norris<sup>[12]</sup>. O trabalho consistiu em determinar o teor de umidade em produtos agrícolas utilizando refletância difusa. O trabalho provocou uma revolução no uso da técnica, pois foi o primeiro a realizar uma determinação quantitativa, utilizando medidas em mais de um comprimento de onda (análise multivariada), contrapondo a utilização clássica de apenas um comprimento de onda definida pela Lei de Beer. Outro acontecimento que contribuiu significativamente com a espectroscopia NIR nas décadas de 80 e 90 foi o avanço da instrumentação relacionada à técnica e o desenvolvimento dos microcomputadores, contribuindo com o desenvolvimento da Quimiometria<sup>[13]</sup> e consequentemente com a análise multivariada<sup>[14-15]</sup>.

Atualmente, a espectroscopia no NIR tem sido aplicada como uma técnica analítica que fornece resultados eficientes para determinação de moléculas orgânicas e variáveis qualitativas. Esta técnica atingiu praticamente todas as áreas, devido ao nível de desenvolvimento tecnológico nos últimos anos, tendo inúmeras aplicações, tais como, nas indústrias alimentícia<sup>[16]</sup>, farmacêutica<sup>[17]</sup>, petroquímica<sup>[18-19]</sup>, têxtil<sup>[20]</sup> e na agricultura<sup>[21-22]</sup>.

Deve-se ressaltar, ainda, a série de vantagens que a espectroscopia NIR apresenta<sup>[15]</sup>:

- Possibilidade de determinações diretas (sem pré-tratamento da amostra);
- Rapidez para a obtenção dos resultados;
- Técnica não-destrutiva, permitindo o uso posterior da amostra;

- Técnica não invasiva, porém com alta penetração do feixe de radiação;
- Vasta aplicação para compostos orgânicos, considerando que pode ser empregada a qualquer molécula que possua ligações C-H, N-H, S-H ou O-H;
- Possibilidade de determinações simultâneas empregando ferramentas de calibração multivariada;
- A possibilidade do desenvolvimento de aplicações em linha (analisadores de processo). Apesar de todas essas vantagens, deve-se citar que a espectroscopia NIR é uma técnica pouco sensível e apresenta baixa seletividade, sendo necessário o uso de técnicas quimiométricas para extrair as informações relevantes.

### 2.2.2. Princípios

A região espectral que corresponde ao infravermelho compreende a faixa com comprimentos de onda entre 780 a 100000 nm. Essa região é dividida em três partes, que recebem denominações de acordo com a sua proximidade do espectro visível. Na Tabela 1 são apresentadas as características das três regiões do espectro infravermelho<sup>[23]</sup>.

**Tabela 1.** Regiões espectrais do infravermelho

Região	Intervalo de número de onda ( $\nu$ ), $\text{cm}^{-1}$	Região em comprimento de onda ( $\lambda$ ), nm
Próximo (NIR)	12800 - 4000	780 - 2500
Médio (MID)	4000 - 200	2500 - 5000
Distante(FAR)	200 - 10	5000 - 100000

A intensidade da radiação na região do infravermelho próximo, normalmente não é suficiente para proporcionar transições eletrônicas em

uma molécula, influenciando apenas no comportamento vibracional das ligações químicas. Para a molécula absorver radiação com comprimento de onda no infravermelho, a mesma precisa sofrer uma variação no momento de dipolo como consequência do movimento vibracional ou rotacional. Apenas nessas circunstâncias o campo elétrico alternado da radiação pode interagir com a molécula e causar variações na amplitude de um de seus movimentos. O momento dipolar é determinado pela magnitude da diferença de carga e a distância entre os dois centros de carga. Desta forma, uma vez que nenhuma variação efetiva ocorre no momento de dipolo durante a vibração ou rotação de uma molécula homonuclear, como O<sub>2</sub>, N<sub>2</sub> ou Cl<sub>2</sub>, essas substâncias não absorvem radiação na região do infravermelho<sup>[23-24]</sup>.

Para se compreender a interação da radiação NIR com uma molécula, é preciso estudar o comportamento vibracional das ligações químicas. As vibrações moleculares são divididas em duas categorias: estiramentos e deformações angulares. 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, aproximam-se as características da vibração de estiramento às de um modelo mecânico representado por duas massas ligadas por uma mola. Sob esta ótica, as frequências fundamentais de quaisquer dois átomos ligados podem ser calculadas assumindo que a energia segue o comportamento de um oscilador harmônico que obedece a Lei de Hooke<sup>[25]</sup>. Assim, as massas representam os átomos e a mola representa a ligação química entre eles. Uma perturbação de uma das massas ao longo do eixo da mola resulta em uma vibração denominada de movimento harmônico simples e a força restauradora é proporcional ao deslocamento (lei de Hooke) e tende a restaurar as massas para sua posição original. A frequência da vibração do modelo é dada por<sup>[25]</sup>:

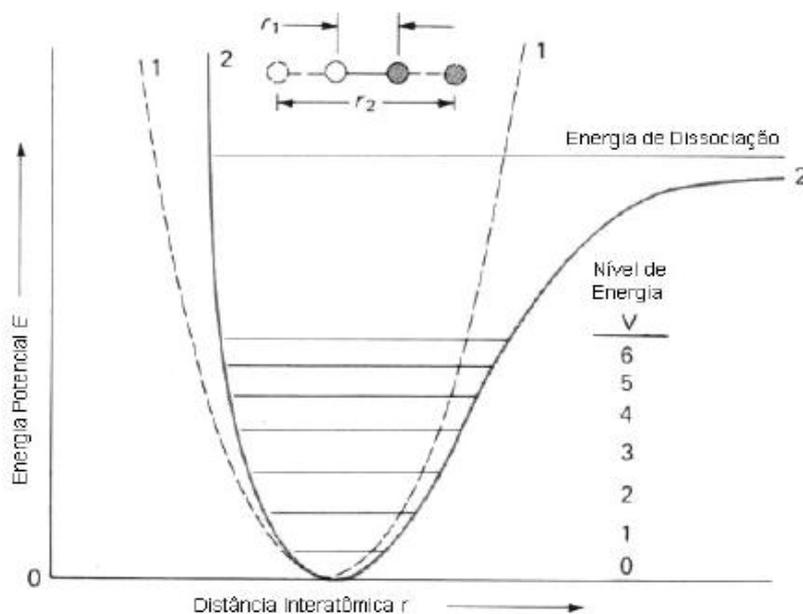
$$v = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (1)$$

sendo  $v$  a frequência da vibração,  $k$  a constante de força da ligação e  $\mu$  massa reduzida dos dois átomos de massas  $m_1$  e  $m_2$ , definida como:

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

Este modelo funciona bem para o cálculo de frequências fundamentais de moléculas diatômicas simples, apresentando resultados próximos aos valores médios encontrados para o estiramento de uma ligação entre dois átomos em uma molécula poliatômica<sup>[23,24]</sup>.

No oscilador harmônico as diferenças de energia entre dois níveis energéticos são sempre iguais e a única transição possível é entre níveis de energia adjacentes. No entanto, em um sistema molecular real é necessário levar em consideração: as forças de repulsão de Coulomb, quando ocorre aproximação entre os dois núcleos atômicos; e a distorção da ligação (até sua ruptura), se houver um aumento significativo na distância entre eles. Em ambos os casos ocorrerá uma distorção no modelo do oscilador harmônico originando um oscilador anarmônico<sup>[26]</sup>. Neste modelo, são permitidas transições diretas entre mais de um nível energético e existem diferenças entre os espaçamentos destes níveis energéticos. Na Figura 2 é mostrado o comportamento da energia potencial em um oscilador anarmônico.



**Figura 2:** Diagrama de energia potencial[15]. (1) oscilador harmônico, (2) oscilador anarmônico.

Evidências experimentais mostram que as moléculas se comportam como osciladores anarmônicos. O modelo anarmônico descreve três tipos de transições energéticas: as transições energéticas chamadas de fundamentais; as transições energéticas que se originam de um nível energético mais excitado, e recebem o nome de bandas quentes; e os sobretons que são as transições diretas entre mais de um nível energético.

Além dos sobretons previstos pelo modelo anarmônico, outras ocorrências espectrais podem ser observadas na região de infravermelho próximo, destacando-se as bandas de combinação, que são encontradas quando um fóton excita simultaneamente dois modos vibracionais<sup>[23]</sup>. A frequência da banda de combinação é aproximadamente a soma ou a diferença das duas frequências fundamentais. Esse fenômeno ocorre quando um quantum de energia é absorvido por duas ligações de uma vez.

Em resumo, pode-se afirmar que para uma molécula absorver radiação NIR é necessário que o deslocamento dos átomos em uma vibração

cause uma alteração no momento de dipolo da molécula ou do grupo de átomos em vibração criando assim um mecanismo de interação entre o modo vibracional e a radiação eletromagnética. Também é preciso que a molécula receba radiação em uma frequência capaz de fornecer exatamente a energia ( $h\nu$ ) necessária para que ocorra sobretons ou bandas de combinação de duas ou mais vibrações<sup>[15]</sup>.

Com relação às principais bandas de absorção dos compostos aromáticos, pode-se afirmar que a região de 6000-5500  $\text{cm}^{-1}$  normalmente corresponde à região de primeiro sobreton de vibração C-H de metil, e estiramento de aromáticos. A região de 5500-4500  $\text{cm}^{-1}$  corresponde a bandas de combinação de C-H de aromáticos. Bandas de 4500-4000  $\text{cm}^{-1}$  são resultantes da combinação de C-H de metil e C-H de anel aromático.

## **2.3. Métodos de Análise Multivariada**

### **2.3.1. Quimiometria**

É importante observar que como o espectro na região do NIR não apresenta bandas específicas, sua interpretação e uso como técnica qualitativa ou quantitativa torna-se difícil. No entanto, quando associada a métodos quimiométricos, tem alcançado bastante sucesso. A inserção de microcomputadores em laboratórios químicos, o desenvolvimento de técnicas instrumentais, a disponibilidade de softwares e o uso de ferramentas mais sofisticadas capazes de extrair informações de conjuntos de dados muito complexos contribuíram para o surgimento da área da química conhecida como Quimiometria. Pode-se definir a Quimiometria como a ciência que utiliza métodos matemáticos e estatísticos para extrair de dados químicos complexos de natureza multivariada o máximo de informações relevantes<sup>[27-28]</sup>.

Na Quimiometria existem várias técnicas extremamente úteis. Dentre elas, pode-se citar as técnicas de pré-processamento de dados<sup>[29-30]</sup>, planejamento e otimização experimental<sup>[31]</sup>, calibração multivariada<sup>[32-33]</sup>, seleção de variáveis<sup>[34-35]</sup>, e transferência de calibração<sup>[36-37]</sup>.

### 2.3.2. Calibração Multivariada

A calibração multivariada<sup>[38-39]</sup>, quando aplicada à espectroscopia NIR, consiste em estabelecer uma relação matemática entre os dados espectrais e o parâmetro físico ou químico de interesse previamente determinado por uma técnica independente. O procedimento de calibração é realizado em duas etapas<sup>[40]</sup>. A primeira consiste na construção de modelos de calibração, a partir da correlação entre a matriz de dados das variáveis medidas (matriz **X**) com a matriz de dados das propriedades de interesse (matriz **Y**) correspondente às concentrações dos analitos ou a outras propriedades das amostras. Na segunda etapa, o modelo é avaliado utilizando-se ferramentas estatísticas de validação. Por fim, o modelo validado é aplicado para prever propriedades de amostras desconhecidas, que possuam a mesma natureza das amostras do conjunto de calibração, a partir do seu sinal espectral<sup>[39]</sup>.

#### 2.3.2.1 Métodos de Regressão Multivariada

Uma diversidade de métodos de regressão vem sendo utilizada em química analítica para a construção de modelos de calibração multivariada. Dentre esses, os mais empregados são: regressão por componentes principais (PCR)<sup>[38]</sup>, regressão por mínimos quadrados parciais (PLS)<sup>[41-43]</sup> e regressão linear múltipla (MLR)<sup>[38,43]</sup>.

Na MLR a informação contida na matriz **X** é utilizada diretamente na modelagem do parâmetro de interesse o que pode resultar em problema de colinearidade<sup>[33]</sup>. Além disso, existem algumas restrições que muitas vezes limitam a sua aplicação, como por exemplo, o número de variáveis não pode ser maior do que o número de amostras de calibração. Portanto, é necessário selecionar um número de variáveis inferior ao número de amostras e que sejam relevantes para predição do valor de interesse. Todavia, esse problema pode ser resolvido empregando um processo de seleção de variáveis.

Os métodos PCR e PLS geralmente utilizam toda faixa espectral selecionada a partir da matriz  $X$  [38]. Esses métodos têm como fundamento encontrar poucas combinações lineares (componentes ou fatores), não correlacionadas, dos valores originais de  $X$ , para usar apenas essas combinações mais relevantes na regressão. A eliminação da correlação resolve o problema da colinearidade, e possibilita obter equações de regressão e predição mais estáveis. Geralmente, utiliza-se apenas a parte mais relevante da variação de  $X$  na regressão, através do uso de técnicas de seleção de variáveis.

A regressão por componentes principais e PLS se baseiam em uma análise de componentes principais (PCA), que consiste numa projeção da matriz de dados com objetivo de representar as variações presentes em muitas variáveis através de um número menor de fatores [38]. Assim, são obtidas novas coordenadas das amostras e um novo sistema de eixos (denominados fatores, componentes principais ou variáveis latentes), no qual a natureza multivariada dos dados pode ser visualizada em poucas dimensões. Geometricamente as componentes principais (PC) descrevem a variação ou a dispersão entre os pontos (amostras) utilizando o menor número de eixos que se alinham com os dados. As novas coordenadas das amostras no novo sistema de eixos das componentes principais são denominadas de escores (*scores*). Estes apontam as diferenças e similaridades entre as amostras. A construção de cada nova variável é o resultado da combinação linear das variáveis originais e os pesos (*loadings*) são os coeficientes dessa combinação. Eles indicam quanto cada variável original contribui para uma dada componente principal, enquanto os escores apontam diferenças e similaridades entre as amostras.

O número de PC's pode ser determinado utilizando-se um conjunto externo de amostra (conjunto de teste) ou por validação cruzada (*cross validation*). A validação cruzada consiste em, a partir de uma matriz contendo as amostras de calibração, aplicar o procedimento de modelagem em que se faz a retirada de uma ou mais amostras para a predição, e as amostras remanescentes são utilizadas para modelagem.

Repete-se este procedimento até que se tenha feito a predição com todas as amostras. A raiz quadrada do erro médio quadrático de validação cruzada (RMSECV) é calculada de acordo com a Equação (3). O melhor número de PC's será aquele capaz de produzir o menor (RMSECV)<sup>[43]</sup>:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_{ref} - y_{cal})^2}{n}} \quad (3)$$

em que,  $y_{ref,i}$  é o valor de referência da concentração ou da propriedade e  $y_{cal,i}$  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 concentração ou propriedades de amostras que não participaram da modelagem. A raiz quadrada do erro médio quadrático de previsão (RMSEP) é então calculada de forma semelhante à Equação (3). Como critério de aceitação do modelo, este valor normalmente é comparado com o valor da repetitividade ou da reprodutibilidade do método convencional<sup>[40]</sup>.

Para garantir a robustez dos modelos de calibração multivariada, as amostras que compõem o conjunto de calibração devem varrer a variabilidade natural da concentração ou propriedade de interesse e devem ser em número tal que possam representar a variabilidade e a complexidade da matriz.

Diante do exposto observa-se que a construção do modelo de calibração deve passar pelas seguintes etapas: seleção de amostras, determinação das propriedades para essas amostras, obtenção dos seus espectros, organização e análise exploratória dos dados, seleção de amostras de calibração e validação, desenvolvimento do modelo de calibração usando diferentes técnicas de calibração multivariada e

diferentes pré-processamentos, validação do modelo e aplicação do modelo para analisar amostras desconhecidas.

### **2.3.3. Tratamento de Dados**

#### **2.3.3.1 Pré-processamento dos Dados**

Pré-processamento é o nome dado para qualquer tratamento prévio dos dados antes da modelagem multivariada propriamente dita. Muitas vezes o pré-processamento é usado para remover ou reduzir informações irrelevantes à modelagem<sup>[44]</sup>

Existem diversos métodos de pré-processamento, dentre eles pode-se destacar: normalização, correção da linha de base, cálculo da derivada, correção multiplicativa de sinal, escalonamento e filtragem digital para remoção de ruído<sup>[38]</sup>.

Outro procedimento bastante empregado é a seleção de variáveis, o qual permite eliminar os termos que não são relevantes na modelagem, gerando um subconjunto com um menor número de variáveis, e que apresente maior sensibilidade para as propriedades de interesse<sup>[45]</sup>.

##### **2.3.3.1.1. Aplicação das Derivadas**

Um dos pré-processamentos mais usados é a derivada<sup>[38]</sup>. A primeira derivada remove deslocamentos constantes (*offsets*) da linha de base e a segunda derivada elimina uma variação linear da linha de base, normalmente devido a efeito de espalhamento. Como o cálculo das derivadas é feito utilizando-se diferenças entre valores de pontos adjacentes, a relação sinal-ruído torna-se pior. Por isso, antes da diferenciação é comum aplicar aos dados algum tipo de suavização. Vários algoritmos podem ser empregados para o cálculo de derivadas. A técnica de Savitzky-Golay<sup>[43-46]</sup> é usada para calcular derivadas espectrais pelo primeiro enquadramento em um polinomial de baixa ordem para pontos de uma janela selecionada pelo usuário. Em seguida, há um

deslocamento de um ponto nesta janela, ajustando-se outro polinômio (da mesma ordem), e assim sucessivamente até o final do espectro.

#### **2.3.3.1.2. Normalização**

A normalização dos dados consiste em dar mesmo peso a todas as absorvâncias, com objetivo de remover variações sistemáticas, ao dividir ou multiplicar cada variável de um vetor amostra por uma constante<sup>[44]</sup>.

#### **2.3.3.1.3. Centralização na Média**

Este pré-processamento é aplicado às colunas da matriz de dados<sup>[47]</sup>. Obtém-se o valor médio para cada variável que é subtraído de cada um dos valores da variável correspondente. Com isso, ocorre a mudança do sistema de coordenadas para o centro dos dados.

#### **2.3.3.1.4. Escalonamento**

O escalonamento consiste em dividir cada elemento do vetor coluna (variável) pelo desvio padrão dos elementos dessa coluna. O autoescalonamento é a denominação dada quando as técnicas de centralização na média e o escalonamento são aplicadas simultaneamente<sup>[38]</sup>.

### **2.3.3.2. Seleção de Variáveis e Amostras**

Os métodos de seleção de variáveis têm demonstrado que a escolha das regiões espectrais permite melhorar significativamente o desempenho das técnicas de calibração<sup>[45]</sup>. Esses métodos têm como função escolher regiões específicas do espectro (um comprimento de onda ou um conjunto de comprimentos de onda) em que a colinearidade não é significativa, gerando modelos mais estáveis, robustos e mais simples de interpretar para modelos de regressão MLR<sup>[45-48]</sup>.

Existem diferentes técnicas para seleção de variáveis que diferem com relação ao procedimento realizado para a seleção da região espectral. Dentre os métodos utilizados, pode-se destacar o método de mínimos quadrados parciais por intervalo<sup>[49-50]</sup> (iPLS, *Interval Partial Least Square*); o algoritmo das projeções sucessivas (SPA)<sup>[34,51]</sup>; o algoritmo Jack-Knife<sup>[43]</sup>; o algoritmo genético<sup>[52-53]</sup> (GA, *Genetic Algorithm*).

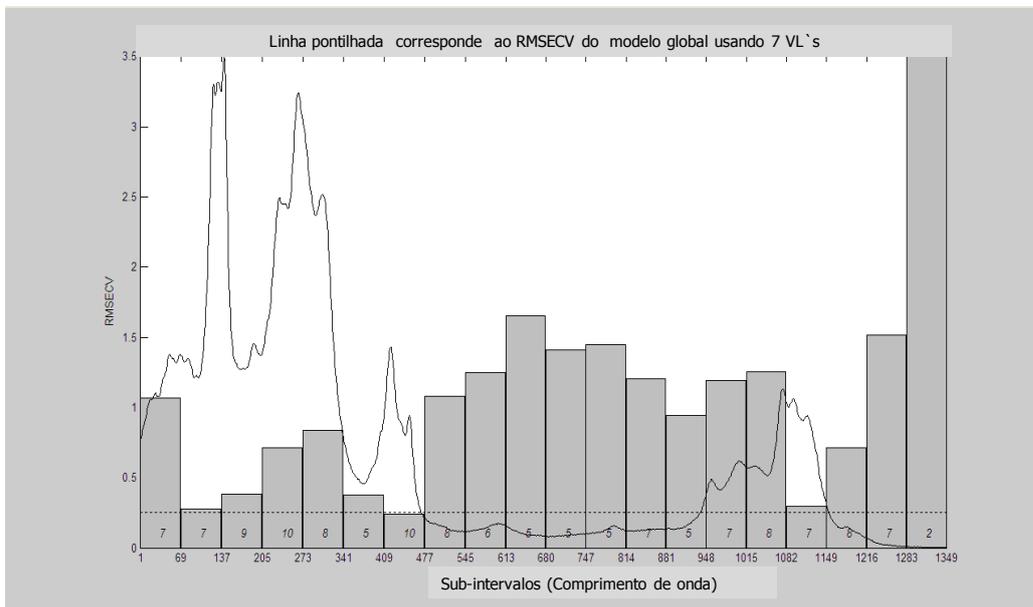
O algoritmo de Kennard-Stone<sup>[54]</sup> é usado para selecionar o conjunto de amostras de calibração e validação. Os algoritmos iPLS, Jack-Knife e Kennard-Stone, foram utilizados no desenvolvimento deste trabalho, e serão apresentados em mais detalhe nas seções seguintes.

#### **2.3.3.2.1. Algoritmo iPLS**

O método iPLS é uma extensão interativa desenvolvida para o PLS, em que desenvolve-se modelos locais PLS em subintervalos equidistantes de toda a região do espectro. Desta forma, é avaliada a relevância da informação nas diferentes subdivisões espectrais, de forma a remover as regiões espectrais cujas variáveis se apresentam como supostamente de menor relevância e/ou interferentes. A partir deste ponto, um novo modelo PLS é construído a partir das variáveis selecionadas<sup>[50]</sup>.

Os modelos locais PLS para cada região do espectro subdividido são comparados em um gráfico com o modelo PLS desenvolvido para todo o espectro (modelo global)<sup>[49-50]</sup>. A análise baseia-se no parâmetro de validação RMSECV (erro quadrático médio da validação cruzada), calculado para cada um dos modelos. O espectro é dividido em tantas partes quanto 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. Outro parâmetro utilizado é o coeficiente de correlação ( $r$ ) da reta de correlação entre os valores reais e previstos pelo modelo. Amostras e/ou medidas anômalas detectadas pelo PLS devem ser geralmente removidas antes da aplicação do iPLS. As regiões espectrais com valores de RMSECV menor que o modelo global

são selecionadas para a construção dos modelos PLS. Nas Figuras 03 e 04 são exemplificados os gráficos do iPLS.



**Figura 03.** Gráfico do modelo iPLS do número de intervalos versus RMSECV.

Na figura 03, é apresentada a divisão dos dados em 20 subgrupos. Para cada subgrupo é desenvolvido um modelo PLS, e os resultados de RMSECV são mostrados como barras na Figura 03. Os números no interior das faixas no gráfico são os números de variáveis latentes (VL's ou PC's) indicados pelo modelo iPLS 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, pode-se notar que o 2º, o 7º e o 17º intervalo apresentam valor de RMSECV de mesma grandeza do modelo global e por isso podem ser selecionados para o desenvolvimento do modelo de calibração. Na Figura 04, é apresentada a faixa das variáveis escolhidas (69-137). Assim, nesse exemplo o modelo PLS pode ser construído a partir de 68 variáveis do espectro original (variáveis 69-137), ao invés das 1349 variáveis iniciais.

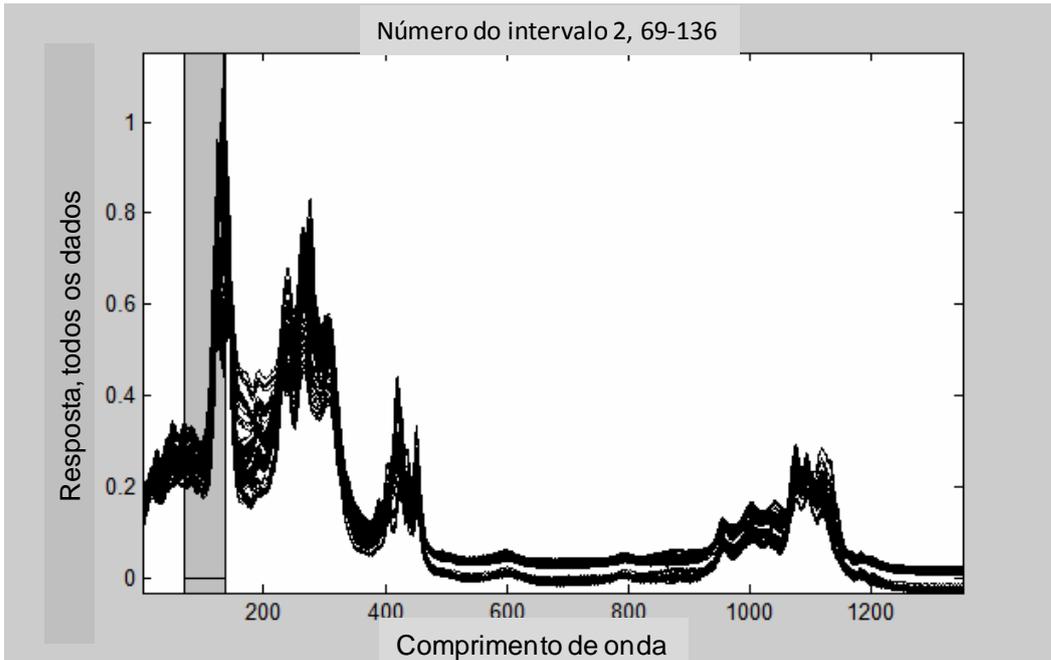


Figura 04 - Gráficos do modelo iPLS: faixa escolhida para a construção do modelo PLS.

#### 2.3.3.2.2. Algoritmo Jack-Knife

O algoritmo Jack-Knife pode ser utilizado para selecionar as variáveis mais informativas que serão utilizadas na construção dos modelos de regressão PLS. Nesse algoritmo<sup>[43]</sup>, 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.

#### 2.3.3.2.3 Algoritmo de Kennard-Stone

Um fator muito importante para o desenvolvimento dos modelos multivariados é a seleção de um subconjunto representativo de validação a partir de um conjunto grande de amostras. Existem vários métodos para realizar esta separação, e dentre esses pode-se citar o o algoritmo de Kennard-Stone<sup>[54,55]</sup>.

Neste trabalho, a separação dos conjuntos de dados entre os conjuntos de calibração e validação foi realizada através do algoritmo de Kennard-Stone (KS). Para assegurar a uniformidade de distribuição de cada subconjunto ao longo do espaço de dados  $X$  (respostas instrumentais), o algoritmo KS segue o procedimento por etapas de modo que uma nova seleção é feita nas regiões do espaço longe das amostras já selecionadas<sup>[56]</sup>.

Para este propósito, 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 4<sup>[56]</sup>:

$$d_x(a, b) = \sqrt{\sum_{j=1}^J [X_a(j) - X_b(j)]^2}; a, b \in [1, N] \quad (4)$$

Para dados espectrais,  $X_a(j)$  a  $X_b(j)$  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 começa tomando o par  $(a_1, a_2)$  de amostras nas quais a distância  $d_x(a_1, a_2)$  é maior, subsequentemente, o algoritmo seleciona a amostra que exibe a maior distância em relação a uma amostra já selecionada. O procedimento é repetido até que o número de amostras requerido seja encontrado.

#### 2.3.3.2.4. Algoritmo SPXY

O algoritmo SPXY é uma extensão do algoritmo KS que leva em consideração tanto as diferenças de  $X$  quanto as de  $Y$  no cálculo das distâncias inter-amostras<sup>[56]</sup>. Este algoritmo calcula cada distância  $d_x(a,b)$  (Equação 4) e cada distância  $d_y(a,b)$  (Equação 5), 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 (Equação 6).

$$d_y(a, b) = \sqrt{(y_a - y_b)^2} = |y_a - y_b|; a, b \in [1, N] \quad (5)$$

$$d_{xy}(a, b) = \frac{d_x(a, b)}{\max_{a, b \in [1, N]} d_x(a, b)} + \frac{d_y(a, b)}{\max_{a, b \in [1, N]} d_y(a, b)}; a, b \in [1, N] \quad (6)$$

### 2.3.4 Planejamento Experimental

No desenvolvimento de um método analítico deve ser realizado um estudo para otimização das diferentes variáveis que afetam o processo com objetivo de obter o melhor resultado analítico possível. A otimização pode ser feita monitorando-se a influência de uma das variáveis por vez sobre a resposta experimental, denominada otimização univariada. Sua principal desvantagem é o tempo gasto e a impossibilidade acerca das interações entre as variáveis que afetam a resposta, o que pode resultar em uma otimização ineficiente.

Os procedimentos para otimização multivariada, onde todas as variáveis importantes são avaliadas simultaneamente, baseiam-se em técnicas estatísticas de planejamento experimental e análise de dados. Tais procedimentos tem como vantagens a necessidade de um menor número de experimentos, maior rapidez e eficiência, além da possibilidade de avaliar os efeitos de interação entre as variáveis<sup>[7,31]</sup>.

A partir de um planejamento experimental é possível também identificar quais fatores tem influência ou não sobre a resposta. Os fatores são as variáveis independentes a serem controladas no processo, podendo ser quantitativas ou qualitativas, e a resposta é a variável dependente que, em uma análise química, corresponde ao valor medido nos experimentos.

Os planejamentos experimentais podem ser de primeira ordem, que permitem construir modelos lineares, ou de segunda ordem, que permitem ajustar modelos quadráticos. Nos planejamentos experimentais de primeira ordem, a função resposta e as variáveis seguem uma equação do tipo  $R = a + bX + cY + dXY$  em que “a” é o termo independente, “b” e “c” são os coeficientes lineares das variáveis X e Y,

respectivamente, e “d” é o coeficiente que corresponde a interação entre X e Y. Nos planejamentos experimentais de segunda ordem, a função resposta e as variáveis seguem a equação do tipo  $R = a + bX + cY + dX^2 + eY^2 + fXY$ , onde “a” é o termo independente, “b” e “c” são os coeficientes lineares de X e Y, “d” e “e” são os coeficientes quadráticos, e “f” é o coeficiente que representa a interação entre X e Y. A função quadrática é uma solução a modelagem de uma eventual curvatura presente no sistema. Em suma, os modelos lineares e quadráticos procuram descrever a resposta a ser obtida em função das variáveis estudadas.

O planejamento fatorial, completo ou fracionário, é um planejamento de primeira ordem muito utilizado. Tendo em vista sua simplicidade é indicado para fase inicial do desenvolvimento experimental quando há necessidade de se definir as variáveis independentes mais importantes e estudar os efeitos sobre a variável resposta escolhida<sup>[31]</sup>. Os planejamentos de segunda ordem são mais completos, e por isso são mais indicados para a fase de otimização multivariada.

Nos planejamentos experimentais onde as variáveis são exploradas em dois níveis é comum codificá-los usando os sinais (+) e (-). A atribuição destes sinais aos níveis superiores ou inferiores é feita de forma arbitrária e não interfere na realização dos experimentos ou interpretação dos resultados, além de permitir tratar o planejamento na forma de matrizes de planejamento.

#### **2.3.4.1 Matriz Doehlert**

O planejamento utilizando a matriz Doehlert é um sistema de otimização de experimentos<sup>[57-58]</sup> de segunda ordem que possui seus pontos distribuídos uniformemente por todo espaço experimental<sup>[31]</sup>. As características deste planejamento são:

- Os pontos experimentais são uniformemente distribuídos sobre uma estrutura romboidal; para duas variáveis um hexágono regular, para três variáveis um cuboctaedro. A uniformidade na distância entre os pontos na região de interesse é uma condição necessária para

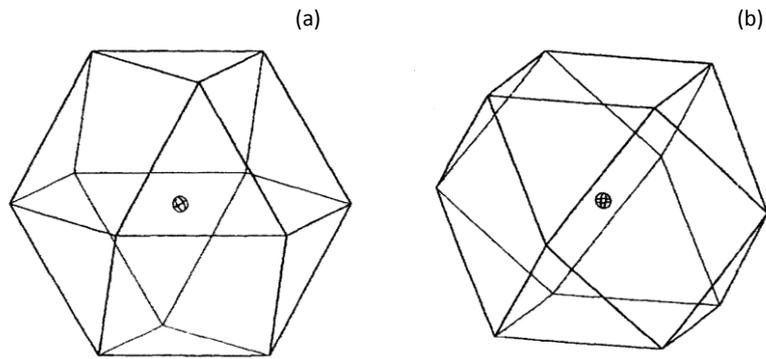
garantir uma boa interpolação, independente da função de ajuste utilizada.

- O número de experimentos necessários (N) corresponde a  $N = k^2 + k + C$ , onde k é o número de variáveis e C é o número de replicatas do ponto central.
- Cada variável é estudada com diferentes números de níveis. Por exemplo, para  $k=3$ , tem-se variáveis com sete, cinco e três níveis. Essa é uma característica importante, pois possibilita usar um número maior de níveis para uma variável mais crítica.
- A estrutura da matriz de planejamento permite a varredura de um amplo domínio experimental, através da seleção adequada dos níveis, com a vantagem em relação ao planejamento fatorial do tipo  $2^n$  de agregar informação de pontos intermediários entre o nível mínimo e máximo de cada variável.
- O deslocamento do planejamento para fora do domínio experimental inicial pode ser feito aproveitando-se pontos do planejamento inicial, dando a este planejamento grande mobilidade.

Os valores codificados das coordenadas de cada ponto experimental que compõem as matrizes Doehlert podem ser calculados a partir de relações trigonométricas e, portanto, sua obtenção não é tão imediata.

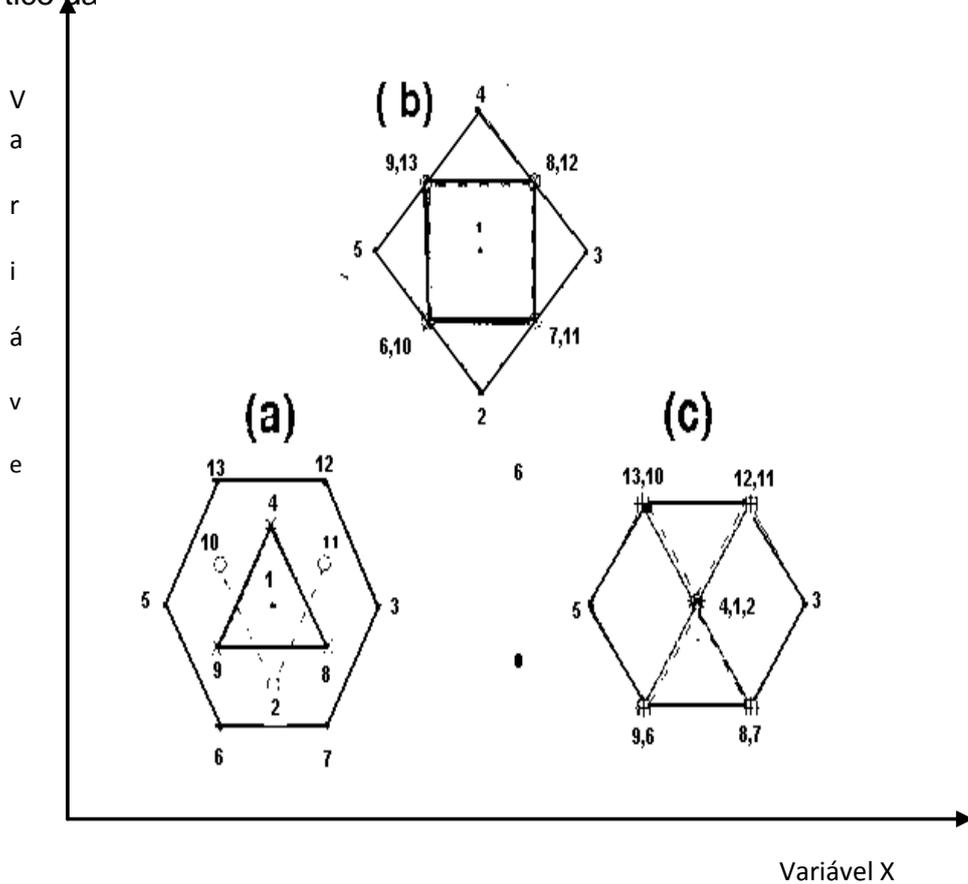
### **Planejamento Doehlert para três variáveis**

O planejamento Doehlert para três variáveis consiste de uma figura tridimensional: um cubo com oito vértices simetricamente truncados produzindo oito triângulos equiláteros cujas arestas são iguais àquelas das faces quadradas adjacentes (um cuboctaedro), representado na Figura 5.



**Figura 5.** Representação geométrica do planejamento tridimensional Doehlert para três variáveis: a) Cuboctaedro centrado sobre uma face triangular, b) Cuboctaedro centrado sobre uma face quadrada.

Três projeções planas são possíveis para esse planejamento (Figura 06 **Erro! Fonte de referência não encontrada.**). As projeções podem ser obtidas sobre uma face triangular, uma face quadrada ou ainda sobre um vértice da



**Figura 6:** 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.

As projeções planas do planejamento geram uma matriz teórica onde os valores experimentais encontram-se codificados pelas coordenadas de cada projeção. Os valores reais podem ser deduzidos através do conhecimento sobre o domínio experimental.

**Tabela 2.** Valores codificados das coordenadas planas do planejamento Doehlert para três variáveis. As colunas A, B e C correspondem às projeções da Figura 06 **Erro! Fonte de referência não encontrada.**

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)

#### 2.3.4.2 Experimentos com Multi-respostas usando uma Função de Desejabilidade

É relativamente simples encontrar as condições ótimas para uma única resposta usando-se planejamentos multivariados, porém isso não acontece quando se têm várias respostas. A solução para esse problema seria utilizar a técnica de otimização multivariada usando uma função multi-resposta com base no conceito de desejabilidade, através de transformações do tipo exponencial<sup>[31]</sup>. Essas transformações exigem a definição de um valor mínimo e um valor máximo que sejam aceitáveis para cada resposta e certos parâmetros específicos para que estas transformações expressem a desejabilidade do usuário. Sendo assim, para que uma função maximize uma resposta, deve-se assumir que existe um valor alvo para ela, acima do qual a condição estabelecida

esteja totalmente satisfeita e abaixo do qual o resultado não deva ser aceitável, considerando-se um limite de tolerância de aceitação.

Uma vez que a função de desejabilidade tenha sido especificada para todas as respostas, devemos combiná-las numa desejabilidade global (Equação 7), normalmente dada pela média geométrica das  $m$  desejabilidades individuais<sup>[31]</sup>:

$$D = \sqrt[m]{d_1 d_2 \dots d_m} \quad (7)$$

Com esse artifício, a otimização simultânea das várias respostas se reduz à maximização de um único valor, a desejabilidade global. 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.

Suponhamos que a resposta tenha um valor alvo  $A$ , situado em algum ponto dentro de uma faixa de aceitação, cujos limites inferior e superior é representado por  $LI$  e  $LS$ , respectivamente. Desta forma, a função de desejabilidade será dada por (equação 8, 9 e 10)<sup>[31]</sup>:

$$d = \left( \frac{y - LI}{A - LI} \right)^s, \quad \text{para } LI \leq y \leq A; \quad (8)$$

$$d = \left( \frac{y - LS}{A - LS} \right)^t, \quad \text{para } A \leq y \leq LS; \quad (9)$$

$$d = 0, \quad \text{para } y \text{ fora do intervalo } (LI, LS) \quad (10)$$

O valor de  $d$  está restrito ao intervalo  $[0,1]$ . 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.

A análise de cada gráfico (Figura 7) pode ser útil para fornecer uma informação visual sobre a influência e o nível de contribuição dos fatores para cada variável de resposta. As curvas mostram como as respostas variam com cada fator, mantidos fixos os níveis dos outros fatores nos valores especificados. Os seis gráficos da última coluna mostram os perfis das desejabilidade das seis respostas usadas nesse trabalho, definidos de acordo com os valores escolhidos para os limites e expoentes (códigos). As linhas tracejadas verticais sinalizam as condições de máxima desejabilidade global que nesse caso representa 1,0, como mostra a última linha de gráficos. Na Figura 7 é exemplificada a aplicação do método mostrando os resultados do algoritmo de otimização<sup>[31]</sup>.

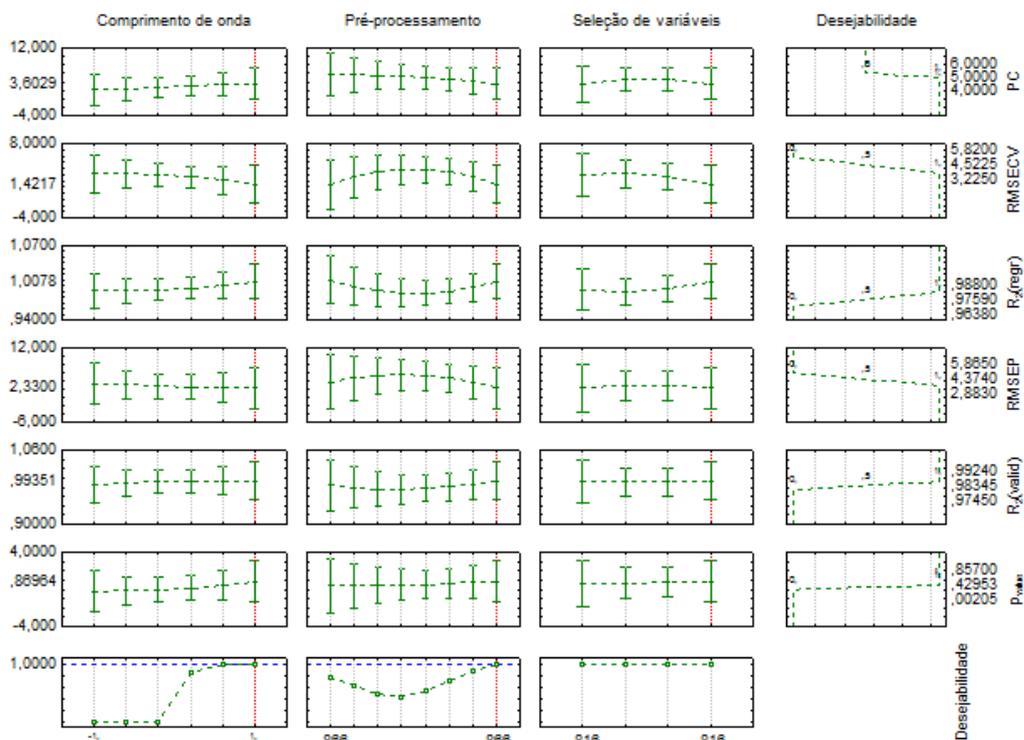


Figura 7: Exemplo da aplicação da função de desejabilidade para cada variável.

## 2.4. Aplicações da Espectroscopia NIR na indústria petroquímica

A espectroscopia NIR tem sido utilizada para diversos fins desde a análise de matérias-primas, de correntes intermediárias até produtos derivados do refino de petróleo (nafta<sup>[3]</sup>, querosene<sup>[59]</sup>, diesel<sup>[60]</sup> e gasolina<sup>[5]</sup>).

Muitos trabalhos já foram desenvolvidos sobre a aplicação da espectroscopia NIR para análise de gasolina. Este combustível, provavelmente, foi um dos primeiros derivados do petróleo a ser analisado através dessa técnica. Nestes trabalhos foram previstas propriedades como octanagem, teor de aromáticos, teor de saturados, densidade e teor de etanol, para aplicações tanto em laboratório como em processo<sup>[61-63]</sup>.

Outro grande potencial da espectroscopia NIR em refinarias e petroquímicas está no monitoramento em tempo real das diferentes propriedades de combustíveis. Larrechi *et al.* (2003) discutiram algumas estratégias<sup>[40]</sup> para avaliar a possibilidade de desenvolvimento de métodos PLS para monitorar propriedades de interesse industrial. Nesse trabalho os autores discutiram algumas questões, exemplificando com modelos PLS, construídos para determinação de hidrocarbonetos e densidade em amostras de nafta, e determinação de etileno e viscosidade em copolímeros de polipropileno. As principais questões abordadas foram: fontes de erros que afetam os resultados, relação entre propriedade de interesse e espectro NIR, amostragem e estratégias para garantir a validade do modelo ao longo do tempo.

Rebouças e Barros Neto (2001) desenvolveram um método de calibração multivariada no infravermelho próximo para predição<sup>[63]</sup> de propriedades físicas (densidade relativa e parâmetros de destilação) de uma mistura de hidrocarbonetos aromáticos. Os espectros NIR foram obtidos na região de 1670-1800 nm e 2100-2500 nm. Os modelos foram desenvolvidos usando a técnica de regressão PLS. Os erros de previsão obtidos foram próximos aos valores de repetitividade do método de

referência. Neste trabalho os autores discutiram a viabilidade de diferenciação dos espectros de aromáticos presentes na mistura, através do estudo da bandas de absorção dos compostos aromáticos puros e concluíram que as técnicas quimiométricas são capazes de diferenciar esses espectros por causa dos grupos substituintes no anel aromático.

Hoeil Chung *et al.* (1998) construíram modelos PLS para determinar o-xileno, m-xileno, p-xileno, etilbenzeno, tolueno, em amostra do processo de produção de p-xileno a partir de espectros NIR obtidos na faixa de 1100 a 2500 nm<sup>[64]</sup>. A concentração de cada componente da mistura foi estimada a partir dos espectros NIR, obtendo uma excelente correlação quando comparado com os resultados obtidos por cromatografia gasosa. Os resultados demonstraram a possibilidade de usar a espectroscopia NIR para monitorar os componentes presentes no processo de produção de p-xileno para controle do processo.

Honorato *et al.* (2008) realizaram um estudo<sup>[65]</sup> para identificar as condições adequadas para construir um modelo de calibração multivariada, para determinar algumas propriedades da gasolina. Para isso, avaliaram a influência de seleção de comprimento de onda, pré-processamento e técnica de regressão. Para avaliar a seleção do comprimento de onda utilizaram o algoritmo genético e algoritmo das projeções sucessiva; para o pré-processamentos usaram a primeira derivada Norris, suavização de Savitzky Golay, e a combinação da primeira derivada Norris com a suavização de Savitzky Golay, os métodos de regressão avaliados foram PLS, PCR e MLR. Obtiveram as condições adequadas para construção dos modelos de calibração através da análise das componentes principais (PCA) dos valores do erro de previsão (RMSEP) obtidos. Os melhores resultados encontrados, foram utilizando: o método de regressão PLS, e o algoritmo genético para seleção de comprimento de onda. Nesse estudo os pré-processamentos utilizados não interferiram nos resultados de previsão da gasolina.

Fernandes *et al.*(2008) desenvolveram modelos PLS usando diferentes faixas espectrais para determinar metanol e etanol em gasolina<sup>[66]</sup>. Os

espectros NIR foram coletados na faixa 714 a 2500 nm. Os autores realizaram também um estudo do efeito da variação da composição dos solventes da gasolina para predição do metanol e etanol, e verificaram que alguns solventes, como tolueno e o-xileno, interferem na previsão de conteúdo de etanol, enquanto isooctano, 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 e demonstrando que é possível determinar simultaneamente metanol e etanol em gasolina.

Lee *et al.* (2001) demonstraram a importância de selecionar a faixa espectral correspondente a uma determinada propriedade, pois observaram que as informações espectrais relacionadas a uma determinada propriedade, geralmente corresponde a uma faixa específica do espectro NIR<sup>[67]</sup>. Foi relatada a dificuldade para localizar essas regiões do espectro em amostras petroquímicas devido à complexidade de sua composição. Desta forma, para otimização da faixa espectral utilizou-se o MW-PLS, obtendo modelos de calibração com menor número de fatores.

Percebe-se que a espectroscopia no infravermelho próximo, aliada às técnicas multivariadas de análise, vem sendo utilizada em substituição à metodologias convencionais, trazendo diversas vantagens como: menor consumo de reagentes tóxico, com menor exposição do analista; mínimo de preparo e manipulação da amostra; e menor tempo de análise, reduzindo custos.

O presente trabalho constitui-se de fundamental importância, tanto para indústria como fins acadêmicos. Para indústria Petroquímica no que diz respeito à obtenção de uma técnica capaz de controlar o processo e monitor os produtos especificados. Para fins acadêmicos, em relação a sua contribuição com abordagem de otimização do processo de construção dos modelos.

## **2.5. OBJETIVOS**

Este trabalho teve como objetivo estudar a aplicação da espectroscopia na região do infravermelho próximo, associada à técnicas de calibração multivariada, para a previsão da composição de correntes do processo petroquímico de produção de aromáticos.

## **2.6. OBJETIVOS ESPECÍFICOS**

- Construção dos modelos para determinar a composição de correntes do processo petroquímico de produção de aromáticos.
- Otimizar o processo de construção dos modelos PLS através de planejamento experimental;
- Definir as variáveis que serão estudadas e avaliar os efeitos dessas variáveis através da análise de gráficos de desejabilidade;
- Obtenção do ponto ótimo para cada variável estudada.

# **CAPÍTULO III**

## **PARTE EXPERIMENTAL**

### 3. Parte Experimental

#### 3.1. Amostras

Para o presente estudo, 225 amostras foram coletadas entre maio e outubro de 2008 e de janeiro a março de 2010, em seis diferentes pontos de amostragem codificados como A, B, C, D, E e F, ao longo do processo petroquímico de produção de aromáticos mostrado na Figura 1. A composição química típica das amostras do processo é mostrada na Tabela 3.

**Tabela 3.** Composição típica das amostras do processo.

Componentes	Concentração (m/m)					
	A	B	C	D	E	F
Total de Não Aromáticos	2,22	47,65	10,56	10,15	3,76	5,89
Benzeno	0,08	9,52	0,09	0,14	2,41	2,82
Tolueno	0,27	40,43	0,87	1,08	36,68	41,78
Etil Benzeno	10,13	0,7	22,37	13,94	0,34	0,20
p-Xileno	12,63	0,49	7,27	17,63	55,44	47,71
o-Xileno	17,48	0,1	9,21	16,2	0,06	0,04
m-Xileno	29,01	1,05	49,53	39,23	1,19	1,48
Cumeno	0,01	0,01	0,01	0,01	0,01	0,01
Total de Aromáticos C8s	69,25	2,36	88,38	87,00	57,14	49,43
Total de Aromáticos C9s+	27,98	0,07	0,10	0,99	0,07	0,07

Adicionalmente, foram preparadas em laboratório 9 misturas de acordo com as proporções definidas na Tabela 4. Os componentes (x e y) correspondem às frações de topo e fundo da coluna de fracionamento de xilenos. Essas misturas visam simular situações de descontrole operacional, com o objetivo de atribuir maior robustez e variabilidade ao conjunto de calibração.

As amostras obtidas foram armazenadas em geladeira a 5 °C com finalidade de reduzir as perdas de componentes voláteis, o que alteraria a sua composição e, portanto, seus espectros. Antes da análise, as amostras foram retiradas da geladeira e analisadas somente após atingir o equilíbrio com a temperatura ambiente (20 ± 2 °C).

**Tabela 4.** Composição das misturas preparadas no laboratório.

Misturas	x (% v/v)	Y (% v/v)
1	0,5	0,5
2	0,6	0,4
3	0,4	0,6
4	0,7	0,3
5	0,3	0,7
6	0,8	0,2
7	0,2	0,8
8	0,9	0,1
9	0,1	0,9

### 3.2. REAGENTES

#### Reagentes

-Clorofórmio p.a. para limpeza da célula do FT-NIR (Merck, com 99% de pureza).

-Tolueno grau espectroscópico como padrão (Merck, com 99,9% de pureza).

### 3.3 INSTRUMENTAÇÃO

As determinações da composição química, através do método de referência ASTM D-2427, foram feitas em cromatógrafo Agilent 6890N com injetor automático 7683B, com detector de ionização de chama (CG-FID) e coluna HP-INOWAX (60 m x 0,32 mm de diâmetro interno x 0,50  $\mu$ m).

Para a obtenção dos espectros de infravermelho próximo foi utilizado espectrômetro NIR com Transformada de Fourier (FT-NIR) Bomem modelo MB-160 (Bomem/Hartmann & Braun, Canada), com interface com microcomputador, equipado com fonte de tungstênio e detector de InAs. Foi utilizada a cela de amostra de CaF<sub>2</sub>, modelo S1-4, com caminho ótico de 0,5 mm. Na Figura 08 são apresentados o espectrômetro e célula utilizados.



**Figura 08:** Instrumento empregado para as medidas na região NIR (com detalhe da célula de medida).

### 3.3.1 OBTENÇÃO DOS ESPECTROS NIR

Os espectros de absorvância foram obtidos no intervalo de 1000-2500 nm ( $10.000$  a  $4.000$   $\text{cm}^{-1}$ ) através da média de 16 varreduras, com resolução de  $4$   $\text{cm}^{-1}$ , utilizando como espectro de referência o ar ambiente. A temperatura do laboratório durante a coleta dos espectros foi mantida na faixa de  $20 \pm 2$  °C. O tempo para obtenção de um espectro foi de cerca de 60 segundos. A exatidão da medida espectral, em termos de comprimento de onda, foi avaliada utilizando tolueno como padrão para verificação do pico em  $4667,24$   $\text{cm}^{-1}$ .

### 3.4. Programas estatísticos

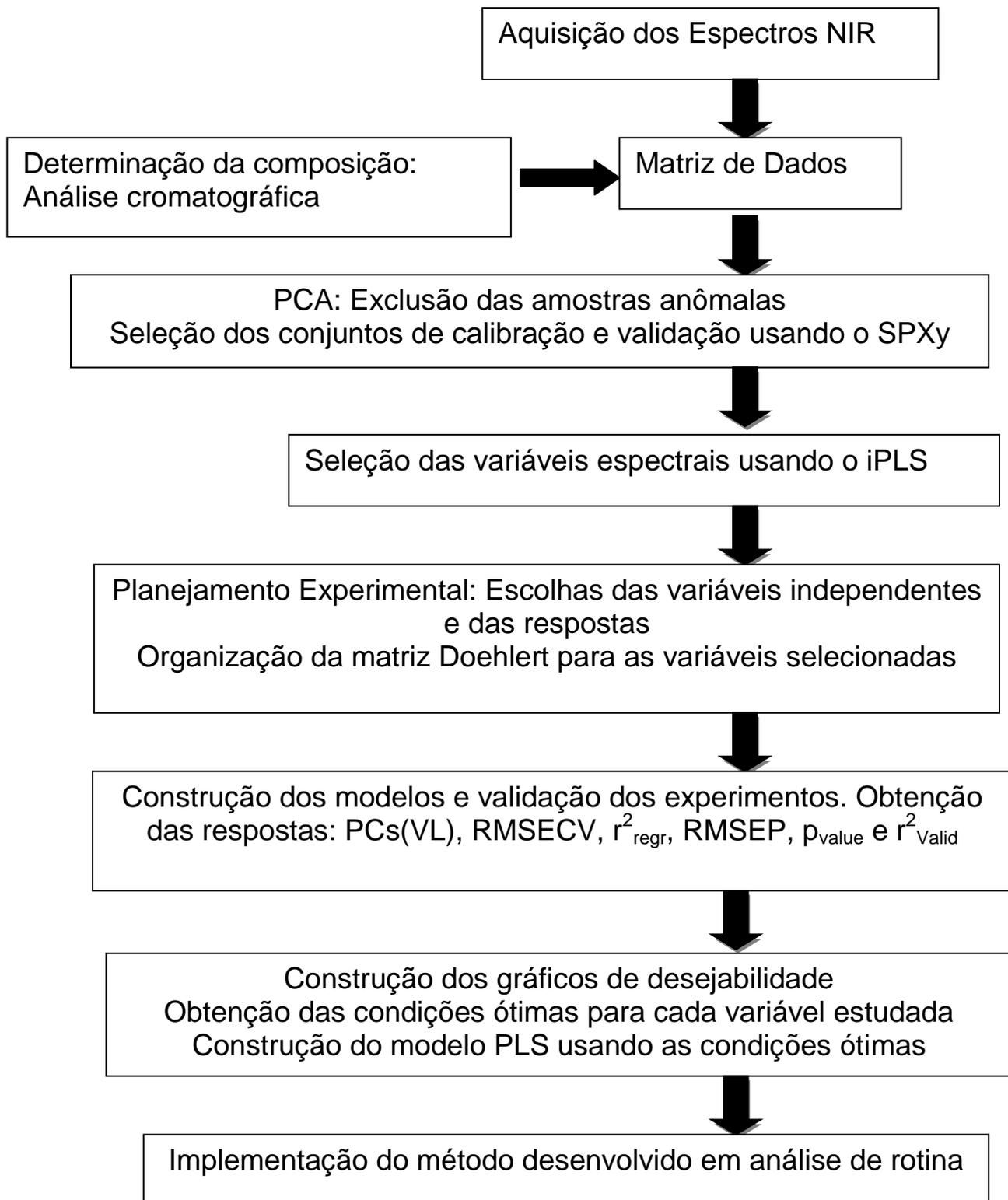
O programa UNSCRAMBLER® versão 9.7 (CAMO Process A.S) foi usado para fazer a análise exploratória dos dados e a regressão por mínimos quadrados parciais.

O programa MATLAB®, versão 6.5, foi utilizado para executar o algoritmo SPXY<sup>[56]</sup> para a seleção de amostras do conjunto de calibração e validação. Para seleção de variáveis espectrais foi utilizado o pacote iPLS (PLS-Toolbox 2.1) baseado no método PLS1.

O programa STATISTICA<sup>[68]</sup>, versão 6.0, foi usado para tratar os resultados do planejamento de experimentos da matriz Doehlert e os gráficos de desejabilidade.

### **3.5. Construção dos modelos – Análise Multivariada**

Antes da construção dos modelos, foi realizada a análise exploratória dos dados para detecção de amostras anômalas e a seleção das amostras que para compor os conjuntos de calibração e validação. Posteriormente, as faixas espectrais relevantes para os constituintes em estudo foram selecionadas. Para otimização do processo de calibração multivariada foi realizado um planejamento experimental utilizando a matriz Doehlert. As variáveis selecionadas 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, e esses modelos foram posteriormente aperfeiçoados excluindo algumas amostras com alto resíduo de concentração e espectral. Os modelos construídos usando as condições ótimas para cada variável foram implementados em rotina sendo necessário um monitoramento contínuo, a fim verificar a necessidade de atualização dos modelos de calibração. Na Figura 9 está ilustrado o fluxograma das etapas utilizadas para o desenvolvimento dos modelos de calibração.



**Figura 9:** Principais etapas desenvolvidas para a obtenção do melhor modelo de calibração.

### **3.6. Detecção de amostras anômalas (*outliers*)**

Uma das exigências para a construção do modelo de calibração multivariada é que seu conjunto de dados não possua amostras anômalas (amostras que possuem um comportamento extremo, não característico do conjunto de dados). Desta forma, a ferramenta utilizada para identificação foi o *Hotteling test*, a Análise de Componentes Principais (PCA) e a análise do gráfico de resíduos de *student vs. Leverage*.

No presente trabalho, a identificação dos *outliers* nos conjuntos de calibração e validação seguiram também as recomendações da norma E1655-05 da ASTM (American Society for Testing and Materials)<sup>[69]</sup>. Amostras fora da elipse de *Hotteling*, com resíduos altos e/ou com grande distância de Mahalanobis<sup>[70]</sup> (alto *leverage*) foram descartadas.

### **3.7. Seleção do Conjunto de Amostras**

Usando o algoritmo SPXY<sup>[56]</sup>, o conjunto de dados foi fracionado, sendo uma parte das amostras separada para calibração e outra para validação, de acordo com as recomendações da norma ASTM E1655-05.

### **3.8. Seleção de variáveis espectrais**

O algoritmo iPLS foi empregado através do software MatLab 6.5 usando o pacote PLS-Toolbox 2.1, para seleção das variáveis espectrais. Esse algoritmo seleciona a(s) faixa(s) espectral(is) mais relevante(s). Para isso, o espectro derivativo foi dividido em 20 intervalos equidistantes e modelos PLS foram construídos nesses intervalos. O melhor intervalo, ou seja, a melhor seleção de variáveis, foi aquele que apresentou o menor ou igual valor de RMSECV em relação ao modelo global usando o espectro inteiro (todas as variáveis).

### **3.9. Planejamento de Experimento Empregando Matriz Doehlert**

Para otimizar<sup>[31]</sup> o processo de construção dos modelos de calibração foi realizado um planejamento experimental baseado na matriz Doehlert<sup>[31, 58]</sup> para definição das melhores condições de modelagem PLS para cada propriedade.

As variáveis estudadas foram: pré-processamento (de acordo com item 2.3.3.1), faixa comprimento de onda (de acordo com item 2.3.3.2.4) e seleção de variáveis espectrais através do algoritmo Jack-Knife (de acordo com item 2.3.3.2.2). As variáveis de entrada mais importantes estudadas foram a faixa de comprimento de onda e a técnica de pré-processamento de dados. Desta forma, estas variáveis foram definidas na matriz Doehlert com o número máximo de níveis: cinco e sete níveis, respectivamente. A terceira variável escolhida foi a aplicação da técnica de seleção das variáveis espectrais com o algoritmo Jack-knife, usando três níveis.

Para o planejamento experimental Doehlert, com três variáveis independentes, foram necessários 13 experimentos. Na Tabela 5, é mostrada a matriz Doehlert aplicada com os valores reais das variáveis independentes e entre parênteses suas respectivas coordenadas codificadas.

**Tabela 5:** Matriz Doehlert para as variáveis comprimento de onda, pré-processamento e seleção de variáveis, com os valores reais e suas respectivas coordenadas (codificadas).

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

<sup>a</sup> Detalhes dos códigos de pré-processamento de dados na Tabela 6.

Para a variável faixa de comprimentos de onda, os níveis foram definidos levando em consideração as principais regiões de absorção do espectro das amostras estudadas. Assim no nível (-0,5), utilizou-se a faixa espectral completa; no nível (1), dados obtidos através de referências da literatura<sup>[63]</sup>; e nos demais níveis, as faixas foram selecionadas usando o algoritmo iPLS (de acordo com o item 3.8).

As várias combinações das técnicas de pré-processamento usadas foram: só autoescalamento; normalização padrão(SNV); Suavização Savitzky-Golay (filtro = 3); normalização (área); normalização + correção da linha de base; primeira derivada Savitzky-Golay (número de pontos da janela = 21); segunda derivada Savitzky-Golay (número de pontos da janela = 21). Em todos os experimentos além do pré-processamentos citados acima foi realizado também o autoescalamento. Detalhes são apresentadas na Tabela 6.

A técnica de seleção das variáveis Jack Knife foi utilizada para selecionar as variáveis espectrais mais relevantes para construção dos modelos PLS. Os três níveis usados corresponderam a: não utilizou da

técnica; utilização da técnica por uma vez; e por duas vezes sucessivas. A ordem da realização dos experimentos foi aleatória.

A partir dos dados obtidos nos experimentos propostos na matriz Doehlert, foram construídos os gráficos de desejabilidade com objetivo de determinar os pontos ótimos de cada uma das variáveis.

**Tabela 6:** Códigos referentes aos pré-processamentos utilizados.

Código	Pré-processamento <sup>a</sup>
PRE 1	Nenhum
PRE 2	SNV
PRE 3	Suavização Savitsky-Golay (filtro = 3)
PRE 4	Normalização (área)
PRE 5	Normalização + correção da linha de base
PRE 6	Primeira derivada Savitsky-Golay (número de pontos da janela = 21)
PRE 7	Segunda derivada Savitsky-Golay (número de pontos da janela = 21)

<sup>a</sup> Em todos os experimentos foi aplicado o autoescalamento.

As variáveis de resposta do planejamento experimental foram estabelecidas com base nas duas etapas de validação do modelo (interna e externa). Os parâmetros usados para monitorar o desempenho do modelo na etapa de validação interna foram: número de componentes principais (PCs) ou variáveis latentes (VL), RMSECV e  $r^2_{\text{regr}}$  (coeficiente de determinação da etapa de validação cruzada). Os parâmetros usados para avaliação do desempenho da validação externa foram: RMSEP e  $r^2_{\text{valid}}$  (coeficiente de determinação da etapa de validação externa) e o valor de p ( $p_{\text{value}}$ ) para o teste t pareado.

Para cada propriedade modelada, as variáveis de resposta foram avaliadas simultaneamente através da função de desejabilidade<sup>[31]</sup>, usando o software Statística versão 6.0 (Stasoft Inc., Tulsa, EUA)<sup>[68]</sup>. Desta forma, pode-se encontrar um valor ótimo dentro de uma faixa de aceitação para cada propriedade, de acordo com os valores escolhidos para os limites.

Para as variáveis PC, RMSECV e RMSEP buscou-se um valor mínimo, dentre os valores obtidos, para os quais definiu-se a desejabilidade 1. Já para as variáveis  $r^2(\text{regr})$  e  $r^2(\text{valid})$  buscou-se um valor máximo dentre os

valores obtidos, aos quais foram fixados desejabilidade 1. Para valor de  $p$ (teste t pareado) buscou-se valores maiores que 0,05, e assim foi fixada desejabilidade 1 para os valores maiores que 0,05.

A partir da condição ideal para cada variável de entrada, foram construídos modelos, que foram posteriormente refinados, através da remoção das amostras com maior resíduo de concentração e resíduo espectral. Esse procedimento foi realizado buscando obter um melhor desempenho do modelo.

Depois do refinamento final, os modelos foram submetidos a procedimento de validação estatística<sup>[71]</sup>. Um conjunto de dados foi usado para validação externa para avaliar a precisão dos modelos. Depois da avaliação do primeiro conjunto de validação externa, os melhores modelos foram usados para analisar amostras de rotina e submetidos à segunda etapa final de validação externa.

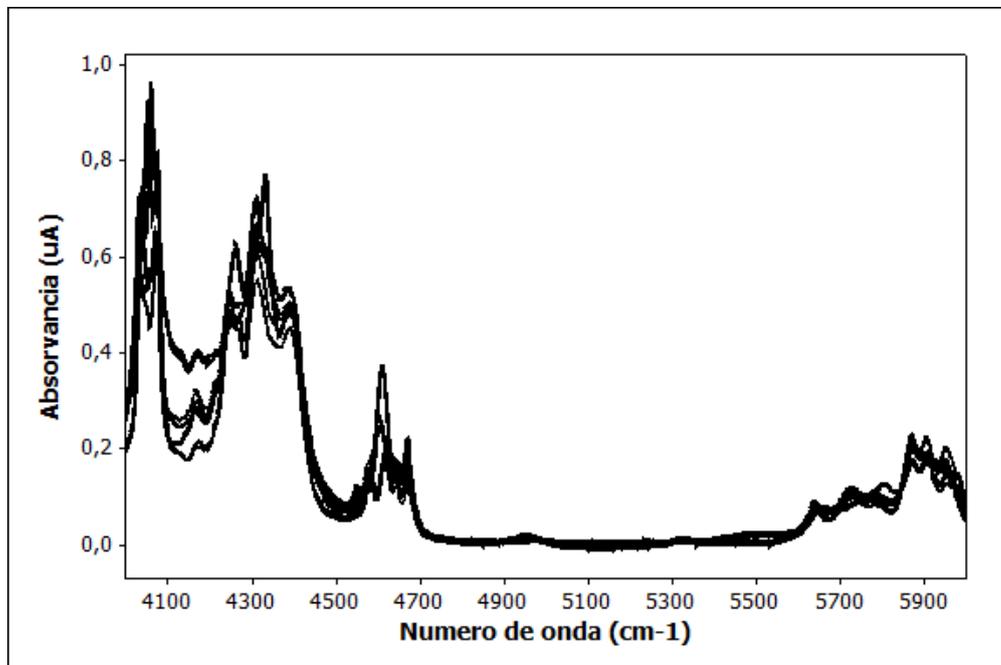
O desempenho do modelo foi avaliado através de comparação entre os valores preditos e medidos usando o teste t pareado<sup>[60]</sup>. O erro de predição (RMSEP) foi comparado com a repetitividade do método de referência. A precisão do modelo foi avaliada a partir dos valores de repetitividade obtidos através da análise dez replicatas da mesma amostra, em comparação com os valores da reprodutibilidade do método de referência.

# **CAPÍTULO IV**

## **RESULTADOS E DISCUSSÃO**

#### 4.1. Atribuição das Bandas do Espectro NIR

Os espectros NIR das amostras de calibração foram registrados na região 10.000-4000  $\text{cm}^{-1}$ .



**Figura 10:** Espectros NIR típicos de cada amostras codificadas (A-E).

Analisando os espectros, pode-se perceber que as regiões espectrais compreendidas abaixo de 4000  $\text{cm}^{-1}$  e acima de 6000  $\text{cm}^{-1}$  não fornecem informações relevantes. Dessa forma, as regiões onde verificou-se pouca informação espectral e grande ruído foram previamente excluídas e não foram utilizadas para construção dos modelos. Utilizando para construção dos modelos apenas a região de 6000- 4000  $\text{cm}^{-1}$ , conforme Figura 10.

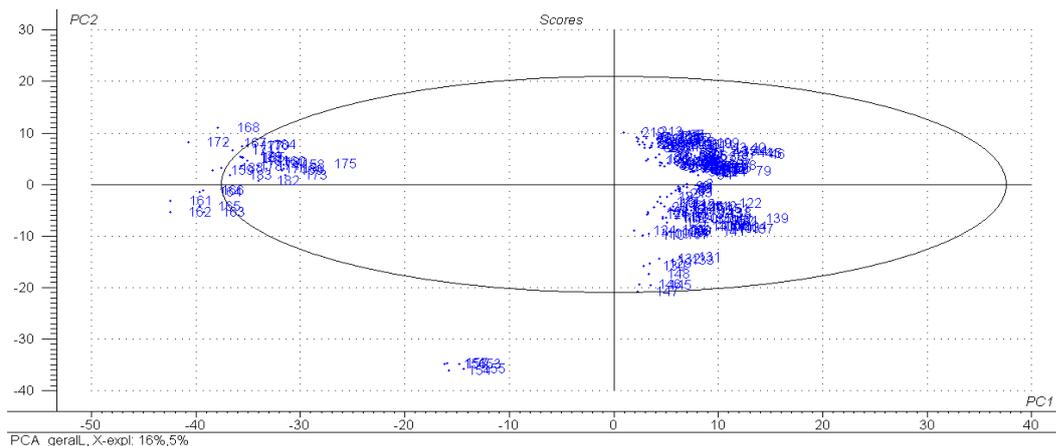
Ainda em relação aos espectros, pode-se afirmar que a região<sup>[72]</sup> de 6000-5500  $\text{cm}^{-1}$  normalmente corresponde à região de primeiro sobretom de vibração C-H de metil, e estiramento de aromáticos. A região de 5500-4500  $\text{cm}^{-1}$  corresponde a bandas de combinação de C-H de aromáticos. Bandas de 4500-4000  $\text{cm}^{-1}$  são resultantes da combinação de C-H de metil e C-H de anel aromático.

## 4.2. 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 de correntes petroquímicas.

### 4.2.1. Detecção de amostras anômalas e seleção das variáveis espectrais.

Inicialmente a matriz de dados X (espectros) e y(propriedades) foi previamente tratada no UNSCRAMBLER®, para detecção de amostras anômalas, através de aplicação da PCA e da elipse – Hotelling test, (Figura 11). Assim o conjunto de dados constituído de 225 amostras foi avaliado, e foram excluídas 16 amostras.



**Figura 11:** Gráfico de escores PC1 versus PC2 para detecção de amostras anômalas.

Após a exclusão das amostras anômalas, presentes no conjunto de dados foi realizada a seleção das amostras para calibração e validação usando o algoritmo SPXY<sup>[56]</sup>. Foram selecionadas 141 amostras para calibração e 68 amostras para a validação. Para comprovar se as amostras selecionadas para validação eram representativas dentro da faixa de trabalho, os gráficos dos escores de PC1 versus PC2 foram avaliados.

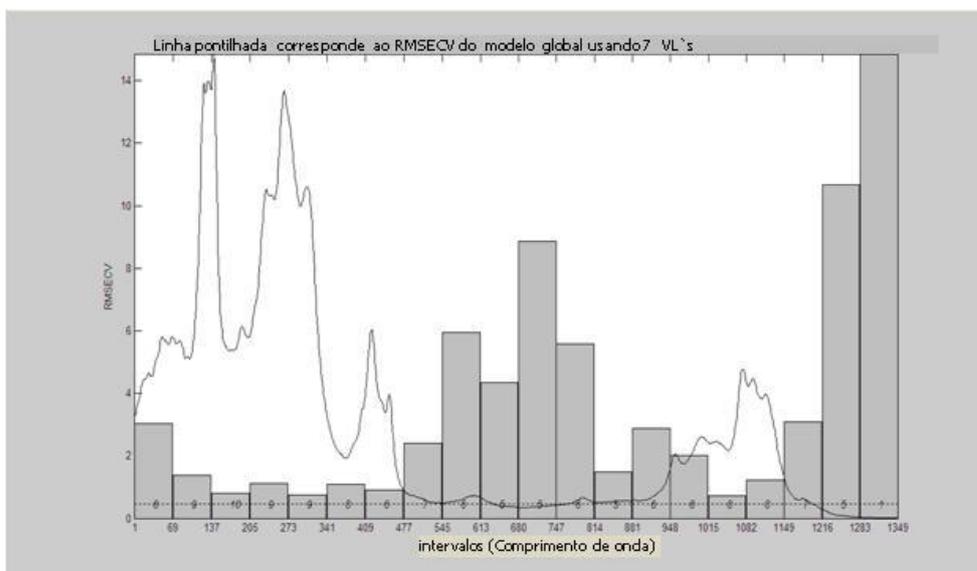
#### 4.2.2 Seleção da região espectral de trabalho

As regiões espectrais selecionadas para construção dos modelos foram obtidas de acordo com o algoritmo iPLS, conforme descrito na Seção 2.3. 3.2.1. O critério adotado pelo iPLS para a escolha da melhor região foi baseado no menor RMSECV<sup>[43]</sup>.

Nas Figuras de 12 a 20 são mostrados os resultados obtidos pela aplicação do iPLS para os oito parâmetros modelados (concentração de não-aromáticos, benzeno, tolueno, etil-benzeno, para-xileno, meta-xileno, orto-xileno, aromático C8s+ e aromático C9s+). Em cada uma dessas figuras é também mostrado um espectro bruto para uma melhor interpretação dos gráficos. As regiões selecionadas pelo iPLS apresentaram o valor do RMSECV igual ou um pouco maior ao valor obtido para o modelo global usando toda faixa (linha tracejada).

A faixa indicada pelo iPLS foi utilizada para construção dos modelos PLS. De acordo com a matriz Doehlert (Tabela 5), os modelos foram obtidos usando a faixa espectral completa, as três regiões que apresentaram menores valores de RMSECV pelo iPLS e uma faixa indicada por referência da literatura<sup>[63]</sup>. É importante ressaltar que como a região selecionada pelo iPLS corresponde a uma banda muito estreita, em alguns casos foram utilizadas as combinações das melhores regiões que apresentaram menores valores de RMSECV.

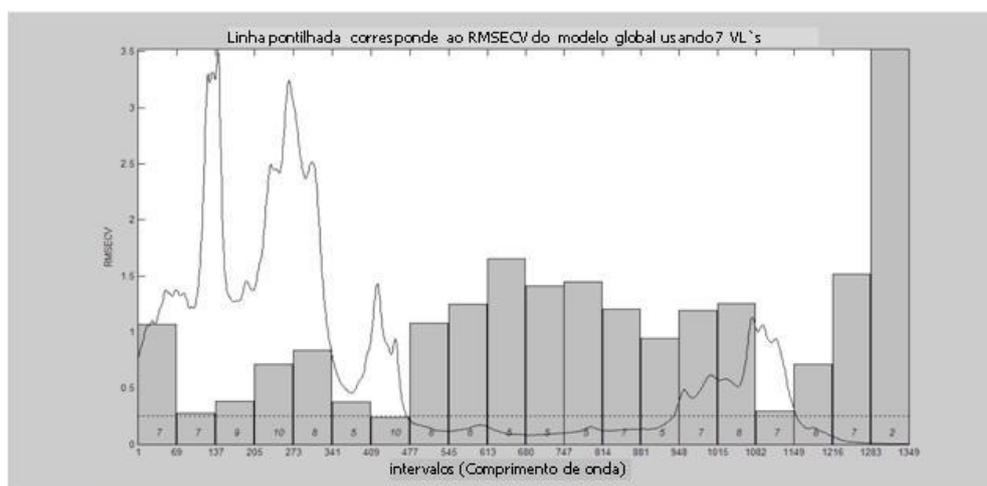
Para o modelo PLS de Não Aromáticos a região indicada pelo iPLS (Figura 12) foi de 273-341, correspondente a 4500 a 4300 ( $\text{cm}^{-1}$ ).



**Figura 12.** Intervalos das variáveis selecionados pelo iPLS para Não Aromáticos.

Essa região apresentou o valor do RMSECV maior que o valor obtido para o modelo global (linha tracejada) usando toda faixa. Desta forma, as faixas utilizadas para encontrar o ponto ótimo na matriz Doehlert foram: faixa completa de 1-1349 ( $6400-3800\text{ cm}^{-1}$ ), 1-477 ( $4700-3800\text{-cm}^{-1}$ ); 948-1149 ( $6000-5600\text{-cm}^{-1}$ ); a combinação das faixas 1-477 e 948-1149; e a região indicada por referências na literatura<sup>[63]</sup> ( $4690-4000\text{ cm}^{-1}$ ).

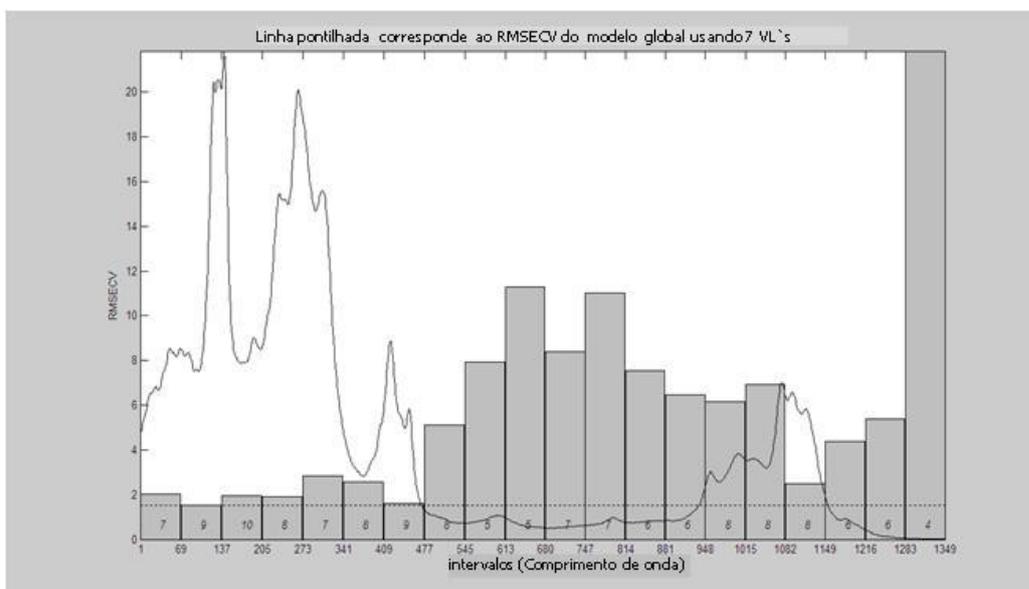
Para o Benzeno (Figura 13), a sub-região de menor valor de RMSECV indicada pelo iPLS foi a região compreendida de 409-477, correspondente a ( $4700-4500\text{ cm}^{-1}$ ).



**Figura 13.** Intervalos das variáveis selecionados pelo iPLS para Benzeno.

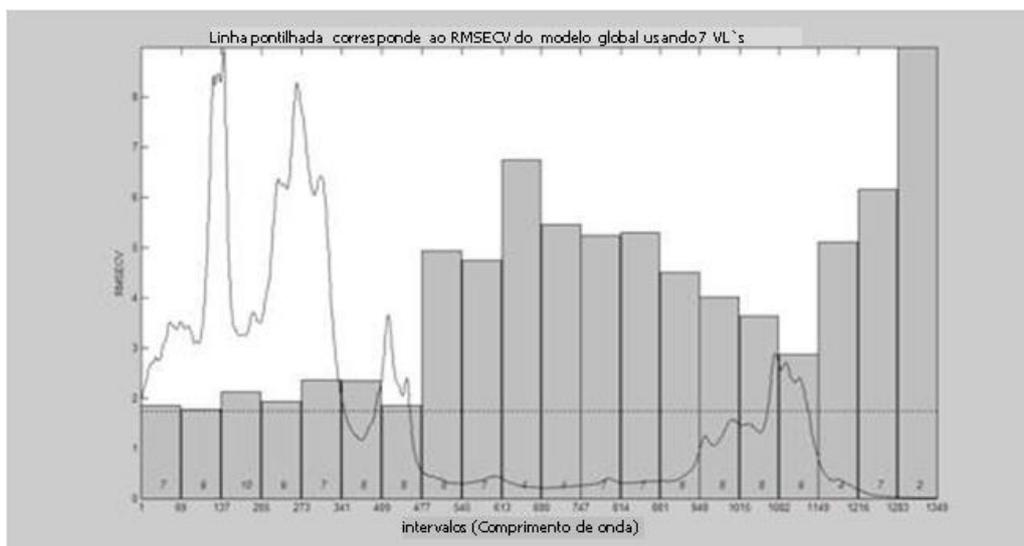
Nesse caso, a região selecionada obteve valor de RMSECV igual ao obtido para o modelo global. As faixas utilizadas para encontrar o ponto ótimo foram: faixa completa de 1-1349 ( $6400\text{-}3800\text{ cm}^{-1}$ ); 69-205( $4190\text{-}3930\text{- cm}^{-1}$ ) e 341-477 ( $4700\text{-}4455\text{- cm}^{-1}$ ); 1082-1149 ( $6000\text{-}5884\text{- cm}^{-1}$ ); a combinação das faixas 69-205, 341-477 e 1082-1149; e a região indicada por referências da literatura<sup>[63]</sup> 69-477 ( $4800\text{-}3900\text{ cm}^{-1}$ ) e 881-1216 ( $6200\text{-}5500\text{ cm}^{-1}$ ).

A sub-região de menor valor de RMSECV selecionada pelo iPLS para o Tolueno (Figura 14), foi na faixa de 69-137 ( $4100\text{-}3900\text{ cm}^{-1}$ ). Essa região apresentou o valor do RMSECV igual ao o valor obtido para o modelo global usando toda faixa. As faixas utilizadas para encontrar o ponto ótimo foram: faixa completa de 1-1349 ( $6400\text{-}3800\text{ cm}^{-1}$ ); 1-477 ( $4700\text{-}3800\text{- cm}^{-1}$ ); 1082-1149 ( $6000\text{-}5884\text{- cm}^{-1}$ ); a combinação das faixas 1-477 e 1082-1149; e a região indicada por referências da literatura<sup>[63]</sup> 69-477 ( $4800\text{-}3900\text{ cm}^{-1}$ ) e 881-1216( $6200\text{-}5500\text{ cm}^{-1}$ ).



**Figura 14.** Intervalos das variáveis selecionados pelo iPLS para Tolueno.

A região indicada pelo iPLS (Figura 15) para o modelo PLS de Etilbenzeno foi compreendida na faixa de 69-137, correspondente a ( $4100\text{-}3900\text{ cm}^{-1}$ ).

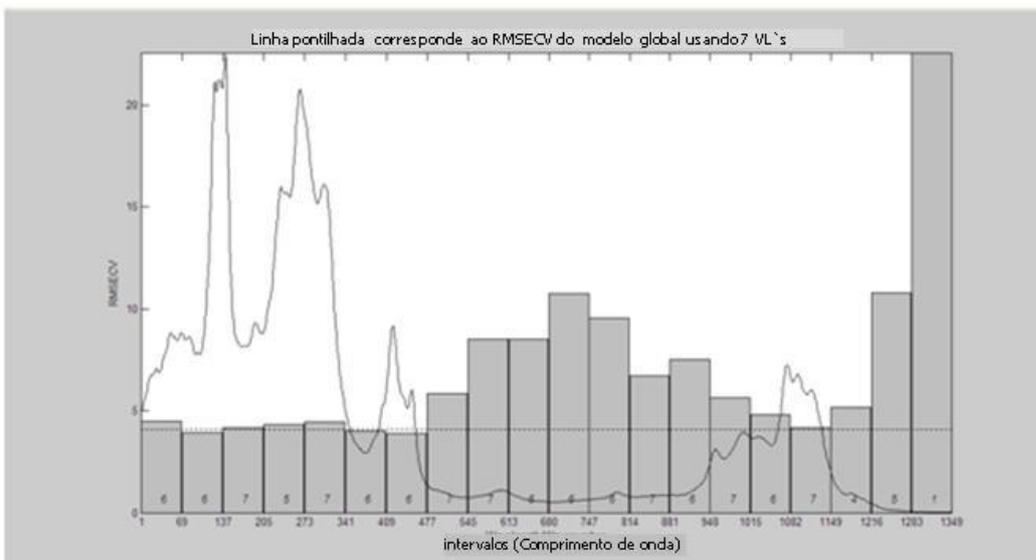


**Figura 15.** Intervalos das variáveis selecionados pelo iPLS para Etil-benzeno.

Essa região apresentou o valor do RMSECV igual ao valor obtido para o modelo global usando toda faixa. Nesse caso, as faixas utilizadas para encontrar o ponto ótimo foram: faixa completa de 1-1349 ( $6400\text{-}3800\text{ cm}^{-1}$ ), 1-137 ( $3800\text{-}4000\text{ cm}^{-1}$ ) e 205-273 ( $4300\text{-}4190\text{ cm}^{-1}$ ); 409-477 ( $4700\text{-}4500\text{ cm}^{-1}$ ); a combinação das faixas 1-137, 205-273 e 409-477; e a região indicada por referências na literatura<sup>[63]</sup> ( $4690\text{-}4000\text{ cm}^{-1}$ ).

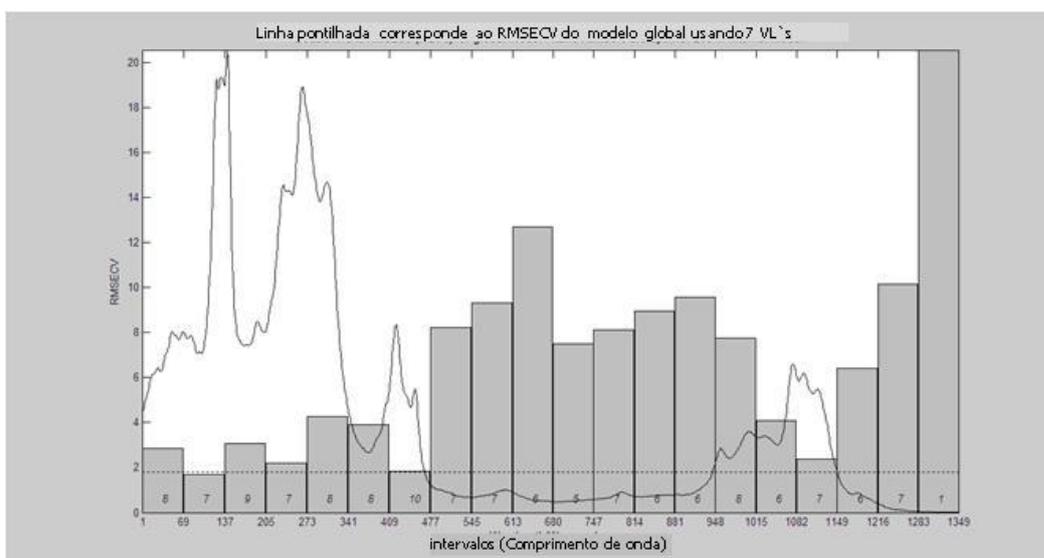
Para o para-xileno e meta-xileno (Figuras 16 e 17) a sub-região de menor valor de RMSECV selecionada pelo iPLS foi na faixa de 69-137 ( $4100\text{-}3900\text{ cm}^{-1}$ ). Essa região apresentou o valor do RMSECV igual ao valor obtido para o modelo global usando toda faixa.

As faixas utilizadas para otimização da construção dos modelos de para-xilenos foram: faixa completa de 1-1349 ( $6400\text{-}3800\text{ cm}^{-1}$ ); 69-137 ( $4100\text{-}3900\text{ cm}^{-1}$ ) e 341-477 ( $4700\text{-}4500\text{ cm}^{-1}$ ); 1082-1149 ( $6000\text{-}5884\text{ cm}^{-1}$ ); a combinação das faixas 1-477 e 1082-1149; e a região indicada por referências da literatura<sup>[63]</sup> 69-477 ( $4800\text{-}3900\text{ cm}^{-1}$ ) e 881-1216 ( $6200\text{-}5500\text{ cm}^{-1}$ ).



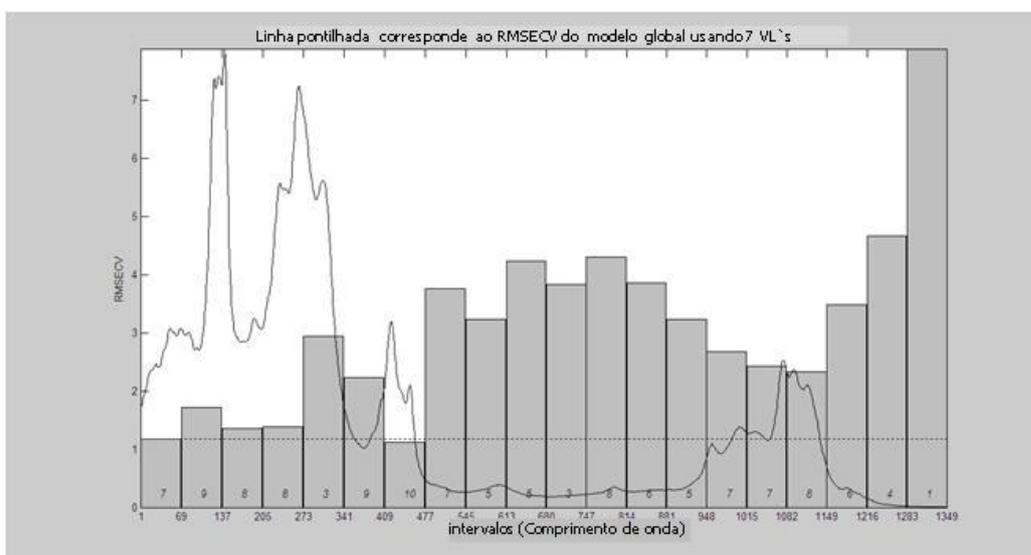
**Figura 16.** Intervalos das variáveis selecionados pelo iPLS para para-Xileno.

Para o meta-xileno as faixas utilizadas para otimização da construção dos modelos PLS foram: faixa completa de 1-1349 ( $6400-3800\text{ cm}^{-1}$ ); 69-137 ( $4100-3900\text{ cm}^{-1}$ ) e 409-477 ( $4700-4500\text{ cm}^{-1}$ ); 1-477 ( $4700-3800\text{ cm}^{-1}$ ); a combinação da faixas 1-477 a 1082-1149( $6000-5884\text{ cm}^{-1}$ ); e a região indicada por referências da literatura<sup>[63]</sup> 69-477 ( $4800-3900\text{ cm}^{-1}$ ) e 881-1216( $6200-5500\text{ cm}^{-1}$ ) .



**Figura 17.** Intervalos das variáveis selecionados pelo iPLS para meta-Xileno.

A sub-região de menor valor de RMSECV selecionada pelo iPLS para o orto-xileno (Figura 18) correspondeu a faixa de 409-477(4700-4500  $\text{cm}^{-1}$ ), Essa região apresentou o valor do RMSECV menor que o valor obtido para o modelo global.

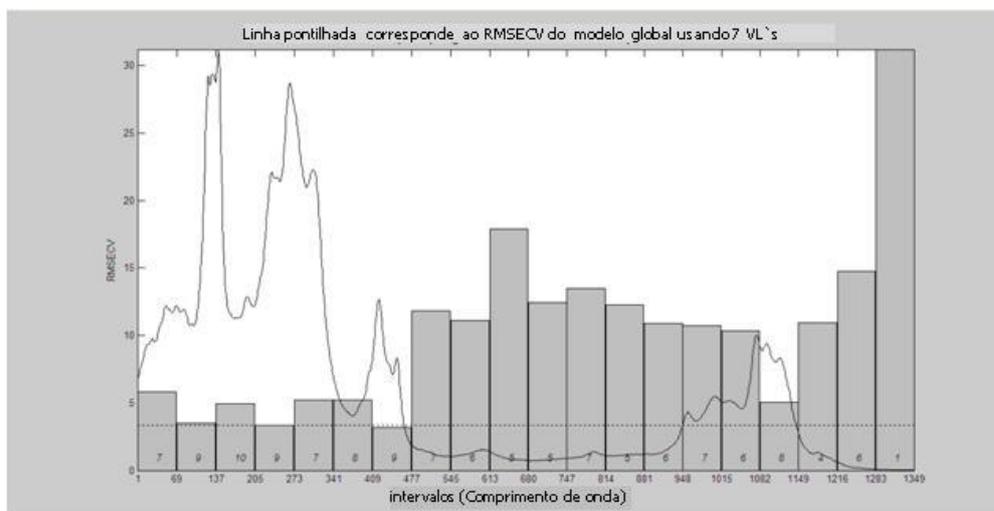


**Figura 18.** Intervalos das variáveis selecionados pelo iPLS para orto-Xileno.

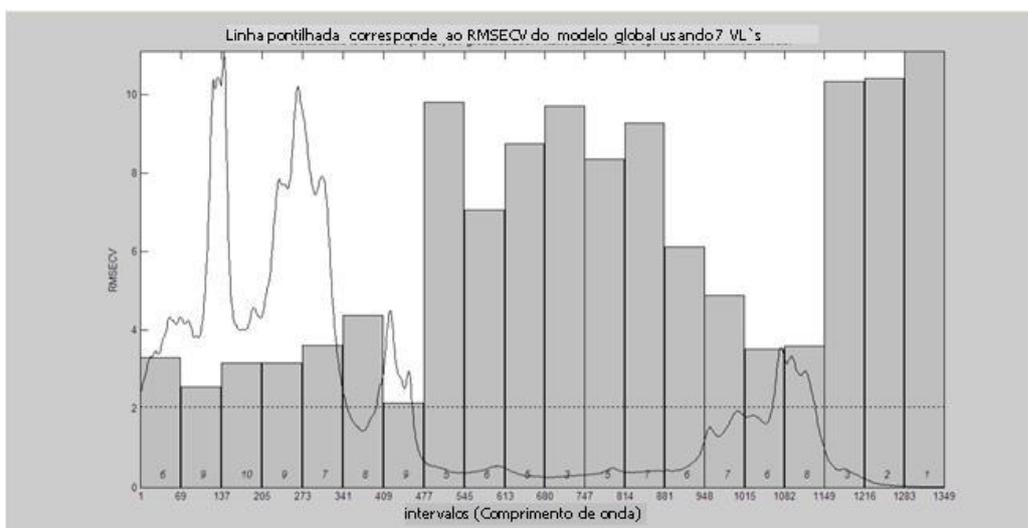
As faixas utilizadas para encontrar o ponto ótimo foram: faixa completa de 1-1349 ( $6400\text{-}3800\text{ cm}^{-1}$ ); 1-273 ( $4300\text{-}3800\text{ cm}^{-1}$ ); 409-477 ( $4700\text{-}4500\text{ cm}^{-1}$ ); a combinação das bandas 1-273 e 409-477; e a região indicada por referências da literatura<sup>[63]</sup> 69-477 ( $4800\text{-}3900\text{ cm}^{-1}$ ) e 881-1216 ( $6200\text{-}5500\text{ cm}^{-1}$ ).

Para os Aromáticos C8s e C9s+ (Figuras 19 e 20) a sub-região de menor valor de RMSECV selecionada pelo iPLS foi de 409-477 ( $4700\text{-}4500\text{ cm}^{-1}$ ). Essa região apresentou o valor do RMSECV igual ao o valor obtido para o modelo global (linha tracejada) usando toda faixa. Neste caso, as faixas utilizadas para otimização da construção dos modelos foram: faixa completa de 1-1349 ( $6000\text{-}4000\text{ cm}^{-1}$ ); 69-137 ( $4100\text{-}3900\text{ cm}^{-1}$ ); 1-477 ( $4700\text{-}3800\text{ cm}^{-1}$ ); 409-477 ( $4700\text{-}4500\text{ cm}^{-1}$ ); e a região

indicada por referências da literatura<sup>[63]</sup> 69-477 (4800-3900  $\text{cm}^{-1}$ ) e 881-1216 (6200-5500  $\text{cm}^{-1}$ ).



**Figura 19.** Intervalos das variáveis selecionados pelo iPLS para Aromáticos C8s.



**Figura 20.** Intervalos das variáveis selecionados pelo iPLS para Aromáticos C9s+.

É importante observar que existe uma faixa com baixo valor de RMSECV comum a praticamente todos os constituintes acima, exceto para Não aromáticos, essa faixa corresponde ao intervalo 409-477 (4700-4500  $\text{cm}^{-1}$ ), de acordo com as referências<sup>[63,72]</sup> essa região corresponde a bandas de absorção de C-H de aromáticos, e portanto a inclusão dessa região no modelo PLS, é justificada para análise desses constituintes.

### 4.3. Desenvolvimento dos modelos de calibração

O processo de construção dos modelos ocorreu conforme sequência descrita na Figura 9. Os experimentos foram executados, a partir dos valores definidos na matriz Doehlert para cada variável estudada, conforme Tabela 5.

#### 4.3.1. Otimização do modelo de calibração de Não Aromáticos

Os resultados do planejamento experimental de Não Aromáticos para todas as variáveis de resposta estão apresentados na Tabela 7.

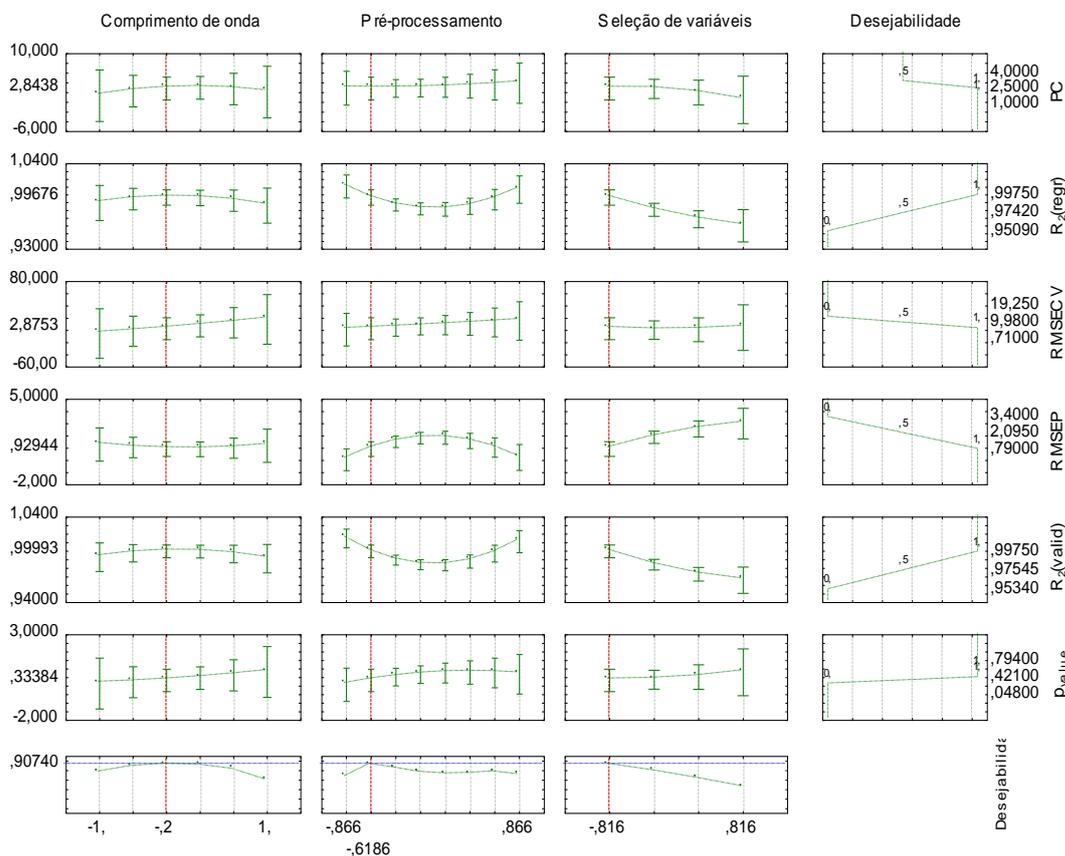
**Tabela 7:** Resultados para as variáveis de resposta usadas para otimização do modelo para Não Aromáticos.

Experimento	Resultados da regressão e validação interna			Resultados de validação externa		
	PC	RMSECV (% , m/m)	$r^2_{regr}$	RMSEP (% , m/m)	$r^2_{valid}$	pvalue teste $t_{pareado}$
1	3	2,78	0,9632	2,98	0,9648	0,57
2	4	0,71	0,9975	0,79	0,9975	0,55
3	2	3,21	0,9509	3,4	0,9534	0,59
4	4	11,95	0,9932	1,06	0,9956	0,79
5	2	2,94	0,959	3,14	0,9604	0,79
6	2	1,78	0,9849	1,55	0,991	0,05
7	2	1,87	0,9833	1,66	0,9890	0,74
8	1	2,93	0,9592	3,21	0,9601	0,53
9	2	2,07	0,9796	1,99	0,9843	0,39
10	2	3,02	0,9568	3,23	0,9585	0,40
11	2	19,25	0,9824	1,75	0,9877	0,75
12	4	1,24	0,9927	1,14	0,9950	0,10
13	4	12,10	0,9930	1,11	0,9953	0,09

Esses resultados foram analisados, a fim de otimizar simultaneamente as várias respostas, usando a função de multi-resposta da função de desejabilidade (Figura 21). Os três gráficos da última linha mostram os perfis das desejabilidades global, sendo que o primeiro gráfico da última linha representa a desejabilidade global em função do comprimento de onda. Analisando esse gráfico, pode-se perceber que este fator pode

variar numa faixa razoável sem influenciar significativamente o valor de desejabilidade. Por outro lado, percebe-se que no terceiro gráfico da última linha qualquer alteração no valor da técnica seleção de variáveis Jack Knife provocará alteração na desejabilidade.

No entanto, observando as desejabilidades individuais para cada resposta, verificou-se que o pré-processamento é também uma variável importante para determinação do ponto ótimo para a modelagem de Não Aromáticos, pois apresentou inclinações mais pronunciadas. Todas essas inclinações são instrutivas, porque fornecem uma ideia da influência de uma dada variável em uma dada resposta, que define a condição ótima para essa variável.



**Figura 21.** Gráficos da função de desejabilidade para Não Aromáticos.

A desejabilidade global, representada pela linha tracejada horizontal azul nos três gráficos da última linha, nesse caso chegou a 0,90. Desta forma, os ótimos sugeridos (intersecção de cada curva com a linha

tracejada da desejabilidade total) podem ser utilizados para otimizar a análise multivariada, visto que uma alta desejabilidade foi alcançada (máximo 1).

Sempre que os valores sugeridos como ótimos foram diferentes dos níveis codificados no planejamento experimental, o nível mais próximo da desejabilidade encontrado foi selecionado. Assim, a otimização da calibração de Não Aromáticos foi obtida para os níveis de comprimento de onda, pré-processamento e seleção de variáveis, -0,2; -0,61 e -0,816, respectivamente. Dessa forma, as melhores condições de comprimento de onda indicadas para modelagem de Não Aromáticos foi a faixa de (6400-3800  $\text{cm}^{-1}$ ), correspondente a todo espectro NIR(-0,5). Esse resultado já era esperado, visto que definiu-se para função de desejabilidade a seleção de regiões com menores valores de RMSECV. Além disso, foi verificado, através do iPLS, que o modelo global construído com a faixa inteira apresentava menores valores que os modelos construídos usando as faixas selecionadas pelo iPLS.

Em relação ao pré-processamento, foi indicado centralização na média com normalização (-0,577) (de acordo com a Tabela 6) e para técnica de seleção de variáveis Jack-Knife: sem seleção (-0,816).

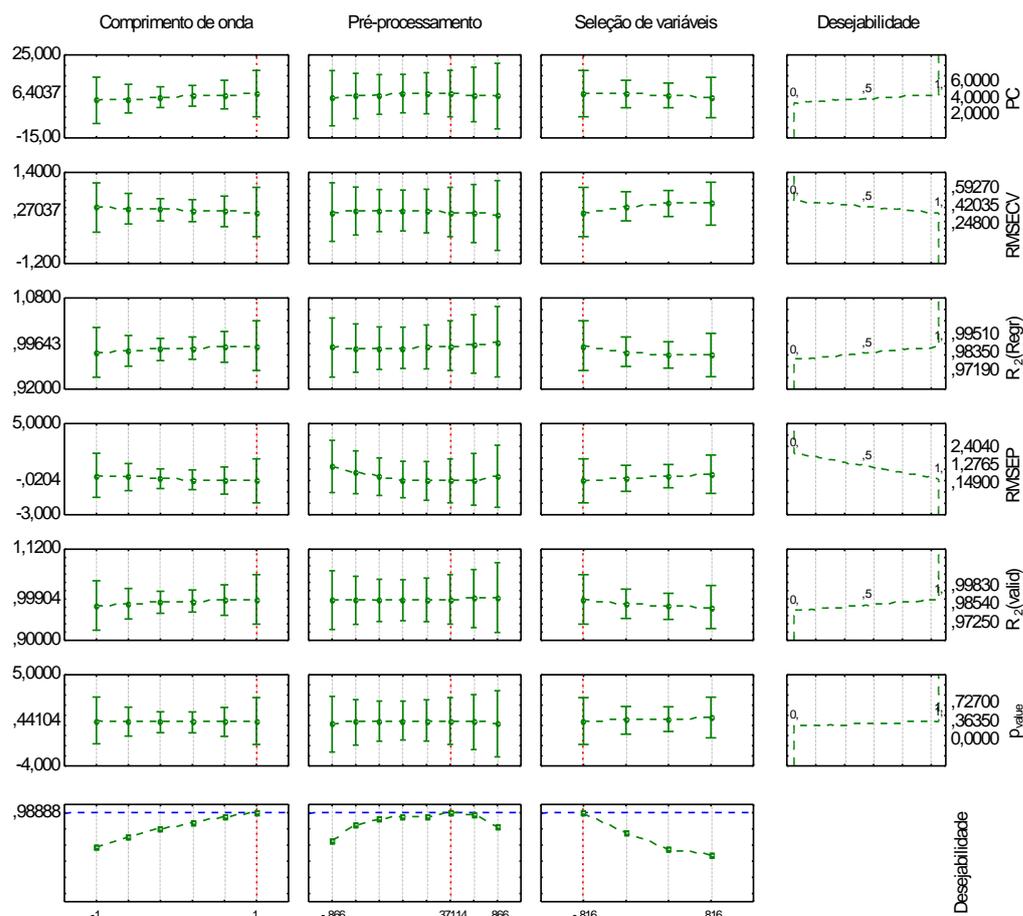
#### **4.3.2. Otimização do modelo de calibração de Benzeno**

Os resultados das variáveis de respostas dos experimentos propostos pela matriz Doehlert para otimização do modelo para o benzeno estão apresentados na Tabela 8.

**Tabela 8:** Resultados para as variáveis de resposta usadas para otimização do modelo para Benzeno.

Experimento	Resultados da regressão e validação interna			Resultados de validação externa		
	PC	RMSECV (% m/m)	$r^2_{\text{regr}}$	RMSEP (% m/m)	$r^2_{\text{valid}}$	p <sub>value</sub> teste t <sub>pareado</sub>
1	6	0,493	0,9805	0,532	0,9820	0,41
2	5	0,248	0,9951	0,149	0,9983	0,74
3	6	0,493	0,9805	0,567	0,9827	0,72
4	5	0,253	0,9949	0,162	0,9981	0,35
5	6	0,469	0,9824	0,536	0,9814	0,61
6	5	0,395	0,9875	2,093	0,9806	0
7	6	0,250	0,9950	2,404	0,9960	0
8	6	0,487	0,9810	0,526	0,9825	0,55
9	2	0,484	0,9812	0,547	0,9780	0,22
10	3	0,542	0,9765	0,588	0,9743	0,60
11	6	0,397	0,9874	0,395	0,9899	<0,0001
12	3	0,448	0,9839	0,441	0,9857	0,15
13	2	0,593	0,9719	0,638	0,9725	0,04

Os resultados obtidos (Figura 22) foram analisados, através da função de desejabilidade, para encontrar um valor ótimo dentro de uma faixa de aceitação para cada resposta.



**Figura 22.** Gráficos da função de desejabilidade para Benzeno.

A desejabilidade global, obtida para benzeno, chegou a 0,988, e as linhas tracejadas vermelhas indicam as condições ótimas sugeridas. Para otimização da calibração de benzeno foi indicado como ótimo para comprimento de onda, pré-processamento e técnica de seleção de variáveis: 1,0; 0,371 e -0,816, respectivamente. Assim, a melhor faixa de comprimento de onda indicada para construção do modelo de benzeno foi o nível (1,0) (vide Tabela 4), que corresponde à faixa de comprimento de onda de  $6200-5500 \text{ cm}^{-1}$  e  $4800-3900 \text{ cm}^{-1}$ . Essa faixa foi escolhida com auxílio da literatura<sup>[63]</sup>.

O pré-processamento e a técnica de seleção indicados foram: centralização na média com normalização e correção da linha de base (0,289) (Tabela 5), e sem usar a técnica de seleção Jack knife (-0,816).

### 4.3.3. Otimização modelo de calibração do Tolueno

Os resultados obtidos relativos as variáveis de resposta do planejamento experimental para o Tolueno estão apresentados na Tabela 9.

**Tabela 9:** Resultados para as variáveis de resposta usadas para otimização do modelo de Tolueno.

Experimento	Resultados da regressão e validação interna			Resultados de validação externa		
	PC	RMSECV		RMSEP		teste t <sub>pareado</sub>
		(%, m/m)	r <sup>2</sup> <sub>regr</sub>	(%, m/m)	r <sup>2</sup> <sub>valid</sub>	
1	3	5,36	0,9386	7,39	0,9117	0,72
2	3	3,98	0,9661	5,47	0,9484	0,32
3	5	2,05	0,9910	1,44	0,9963	0,16
4	2	4,24	0,9616	5,25	0,9519	0,43
5	6	3,35	0,9760	3,49	0,9846	0,16
6	6	1,88	0,9924	1,80	0,9944	0,06
7	6	1,89	0,9924	1,84	0,9942	0,10
8	6	3,35	0,9761	3,75	0,9799	0,26
9	6	2,09	0,9907	1,90	0,9934	0,02
10	4	2,72	0,9842	3,50	0,9800	0,06
11	6	2,10	0,9905	1,90	0,9934	0,03
12	4	2,43	0,9873	2,80	0,9872	0,46
13	4	2,44	0,9873	2,79	0,9873	0,47

Os resultados obtidos foram analisados, através da função de desejabilidade, para encontrar um valor ótimo dentro de uma faixa de aceitação para cada resposta, conforme ilustrado na Figura 23.

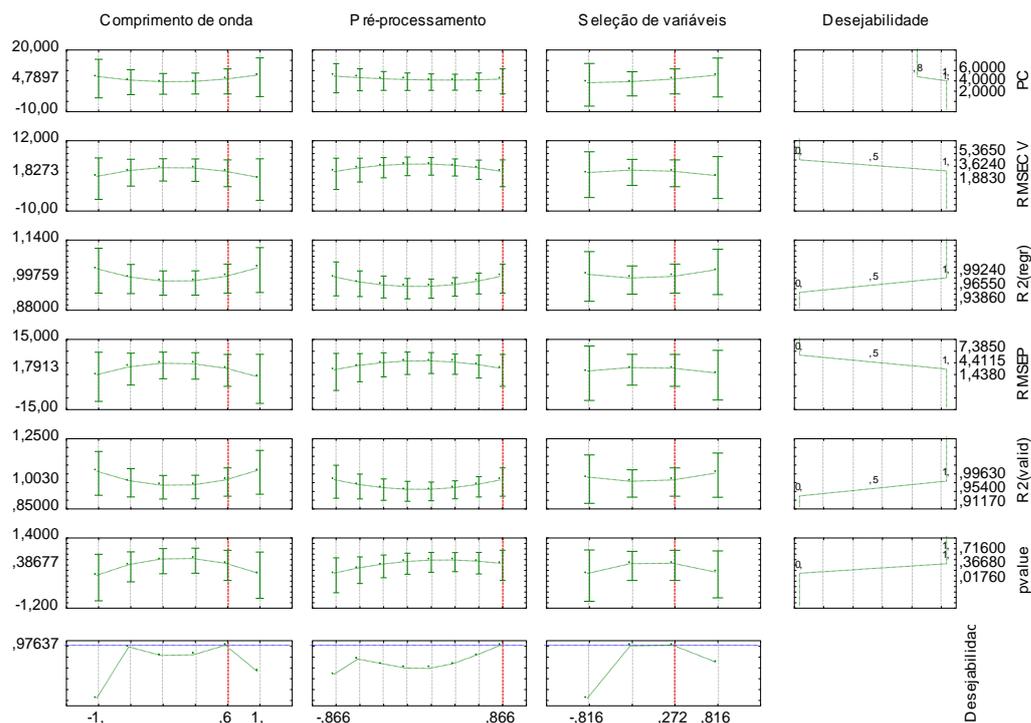


Figura 23. Gráficos da função de desejabilidade para Tolueno.

A desejabilidade global obtida foi de 0,976. Na otimização da calibração de tolueno obteve-se 0,6; 0,866 e 0,272 para comprimento de onda, pré-processamento e seleção de variáveis, respectivamente. Assim as melhores condições para modelagem foram: comprimento de onda  $6000-5800\text{ cm}^{-1}$  e  $4700-3800\text{ cm}^{-1}$  (essa faixa foi selecionada através do iPLS), pré-processamento aplicando segunda derivada Savitsky-Golay com filtro 21 (0,866), e seleção de variáveis com 1 rodada do algoritmo Jack-knife (0), de acordo com a Tabela 5.

#### 4.3.4. Otimização modelo de calibração para Etil-benzeno

Os resultados das variáveis de reposta para otimização do modelo para etil-benzeno estão apresentados na Tabela 10.

**Tabela 10:** Resultados para as variáveis de resposta usadas para otimização do modelo de Etil-benzeno.

Experimento	Resultados da regressão e validação interna			Resultados de validação externa		
	PC	RMSECV (% m/m)	$r^2_{reg}$	RMSEP (% m/m)	$r^2_{val}$	$p_{teste}$ $t_{pareado}$
1	6	2,06	0,9455	2,08	0,9556	0,71
2	6	1,66	0,9647	1,08	0,9892	0,10
3	6	2,21	0,9371	2,24	0,9568	0,63
4	7	1,73	0,9617	0,95	0,9916	0,35
5	6	2,05	0,9457	2,06	0,9632	0,33
6	6	2,05	0,9459	1,69	0,9730	0,34
7	6	1,78	0,9591	1,24	0,9851	0,88
8	5	1,96	0,9504	1,51	0,9782	0,46
9	6	2,05	0,9461	1,66	0,9732	0,35
10	5	2,02	0,9477	1,49	0,9771	0,35
11	7	1,70	0,9629	1,09	0,9851	0,04
12	11	1,64	0,9655	1,07	0,9897	0,48
13	6	1,77	0,9598	1,09	0,9872	0,31

Os resultados obtidos foram analisados, através da função de desejabilidade para se encontrar um valor ótimo dentro de uma faixa de aceitação para cada resposta, conforme ilustrado na Figura 24.

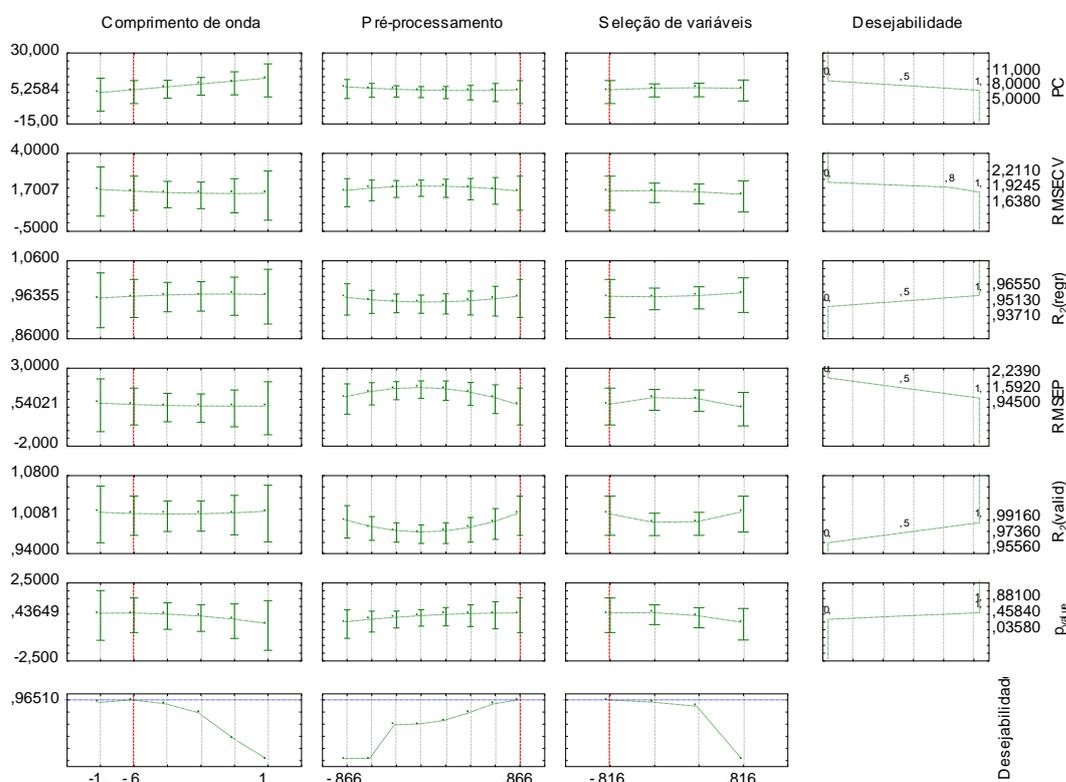


Figura 24. Gráficos da função de desejabilidade para Etil-Benzeno.

A desejabilidade global obtida foi de 0,965. Para otimização da calibração de Etil-benzeno foi utilizado comprimento de onda, pré-processamento e seleção de variáveis correspondente a -0,6; 0,866 e -0,816, respectivamente. Assim, as melhores condições para modelagem foram usando comprimento de onda 6400-3800  $\text{cm}^{-1}$  faixa completa (-0,5), pré-processamento com segunda derivada Savitsky-Golay com filtro 21 (0,866), e seleção de variáveis: sem seleção (-0,816).

#### 4.3.5. Otimização modelo de calibração para p-xileno, m-xileno e o-xileno

Os resultados das variáveis de resposta para otimização do modelo de para-xileno, meta-xileno e orto-xileno estão apresentados nas Tabelas 11, 12 e 13, respectivamente.

**Tabela 11:** Resultados para as variáveis de resposta usadas para otimização do modelo de para-xileno.

Experimento	Resultados da regressão e validação interna			Resultados de validação externa		
	PC	RMSECV (%, m/m)	$r^2_{reg}$	RMSEP	$r^2_{valid}$	$p_{value}$
				(%, m/m)		$t_{pareado}$
1	6	4,80-	0,9520	3,41	0,9797	<0,0001
2	4	4,53	0,9573	3,79	0,9702	0,14
3	5	4,93	0,9493	3,20	0,9848	<0,0001
4	2	4,78	0,9523	4,62	0,9530	3,62
5	4	5,29	0,9417	3,26	0,9789	0,07
6	5	3,95	0,9676	1,68	0,9947	<0,0001
7	5	3,95	0,9675	1,66	0,9949	<0,0001
8	4	4,00	0,9666	2,12	0,9923	<0,0001
9	5	3,96	0,9673	1,63	0,9945	<0,0001
10	4	3,99	0,9667	2,25	0,9923	<0,0001
11	5	3,97	0,9672	1,65	0,9948	<0,0001
12	3	4,36	0,9604	2,81	0,9814	0,34
13	3	4,33	0,9609	2,72	0,9820	0,32

**Tabela 12:** Resultados para as variáveis de resposta usadas para otimização do modelo de meta-xileno.

Experimento	Resultados da regressão e validação interna			Resultados de validação externa		
	PC	RMSECV (%, m/m)	$r^2_{\text{regr}}$	RMSEP (%, m/m)	$r^2_{\text{valid}}$	Pvalue teste
						$t_{\text{pareado}}$
1	6	3,20	0,9734	3,28	0,9808	0,507
2	5	1,68	0,9926	1,12	0,9969	0,07
3	5	3,73	0,9638	3,88	0,9652	0,21
4	4	1,87	0,9909	1,29	0,9959	0,004
5	5	3,70	0,9645	3,95	0,9646	0,21
6	5	2,56	0,9830	2,19	0,9884	<0,0001
7	5	2,59	0,9826	2,23	0,9877	0,01
8	4	2,68	0,9813	2,81	0,9842	<0,0001
9	6	2,03	0,9892	1,71	0,9928	0,3
10	4	2,63	0,9820	2,72	0,9853	<0,0001
11	6	2,08	0,9887	1,76	0,9924	0,3
12	4	2,53	0,9834	2,35	0,9869	0,005
13	4	2,53	0,9836	2,32	0,9872	0,005

**Tabela 13:** Resultados para as variáveis de resposta usadas para otimização do modelo de orto-xileno.

Experimento	Resultados da regressão e validação interna			Resultados de validação externa		
	PC	RMSECV (%, m/m)	$r^2_{\text{regr}}$	RMSEP (%, m/m)	$r^2_{\text{valid}}$	Pvalue teste
						$t_{\text{pareado}}$
1	6	1,47	0,9616	1,32	0,9745	0,946
2	5	1,27	0,9714	2,16	0,9902	0,02
3	6	1,50	0,9603	1,82	0,9611	0,35
4	6	1,22	0,9735	0,84	0,9901	0,30
5	6	1,52	0,9591	1,88	0,9584	0,36
6	6	1,27	0,9713	1,38	0,9681	0,36
7	6	1,36	0,9672	1,37	0,9725	0,87
8	5	1,44	0,9633	1,09	0,9660	0,86
9	7	1,17	0,9756	1,12	0,9795	0,08
10	5	1,34	0,9683	1,40	0,9674	0,37
11	7	1,30	0,9700	1,20	0,9741	0,96
12	5	1,54	0,9578	1,40	0,9653	0,09
13	6	1,22	0,9735	0,87	0,9894	0,026

Os resultados obtidos foram analisados, através da função de desejabilidade, para encontrar um valor ótimo dentro de uma faixa de aceitação para cada resposta, conforme ilustrado nas Figuras 25, 26 e 27.

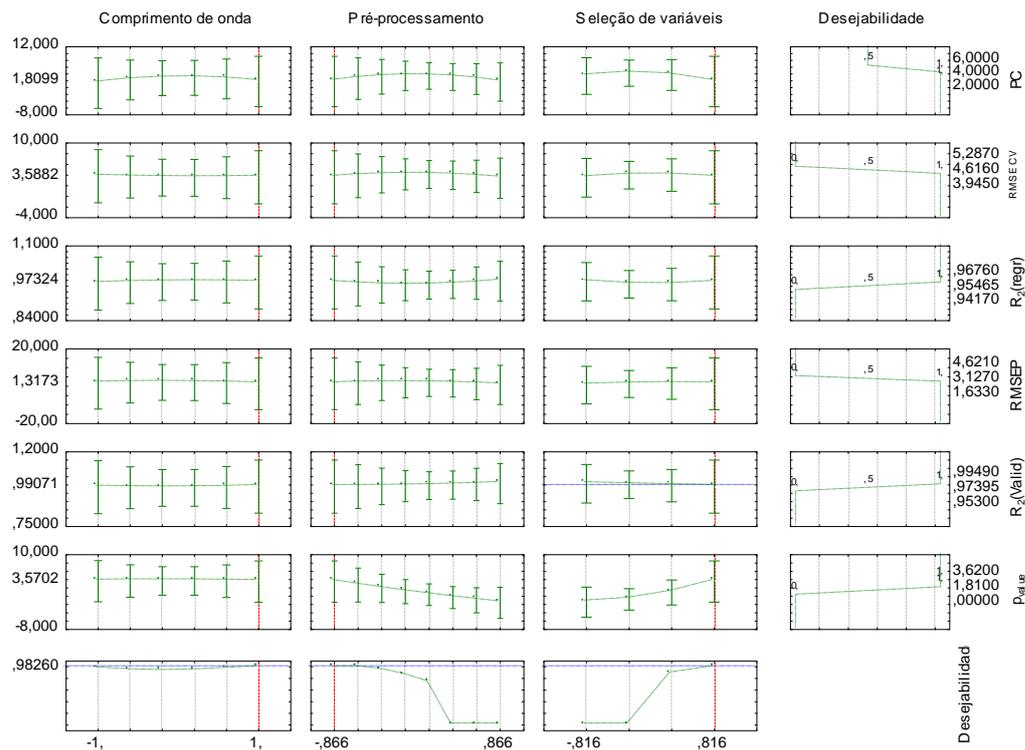


Figura 25. Gráficos da função de desejabilidade do para-xileno.

A desejabilidade global obtida para o para-xileno foi de 0,982, indicando para obtenção da curva de calibração as seguintes condições para comprimento de onda, pré-processamento e seleção de variáveis: 1,0; -0,866 e 0,816, respectivamente. Assim, as melhores condições para modelagem foram: comprimento de onda na faixa de 6200-5500  $\text{cm}^{-1}$  e 4800-4000  $\text{cm}^{-1}$ , conforme referência literatura<sup>[63]</sup> (1,0); pré-processamento: somente autoescalamento (-0,866); e seleção de variáveis com duas rodadas do algoritmo Jack-knife (0,816), de acordo com a Tabela 5.

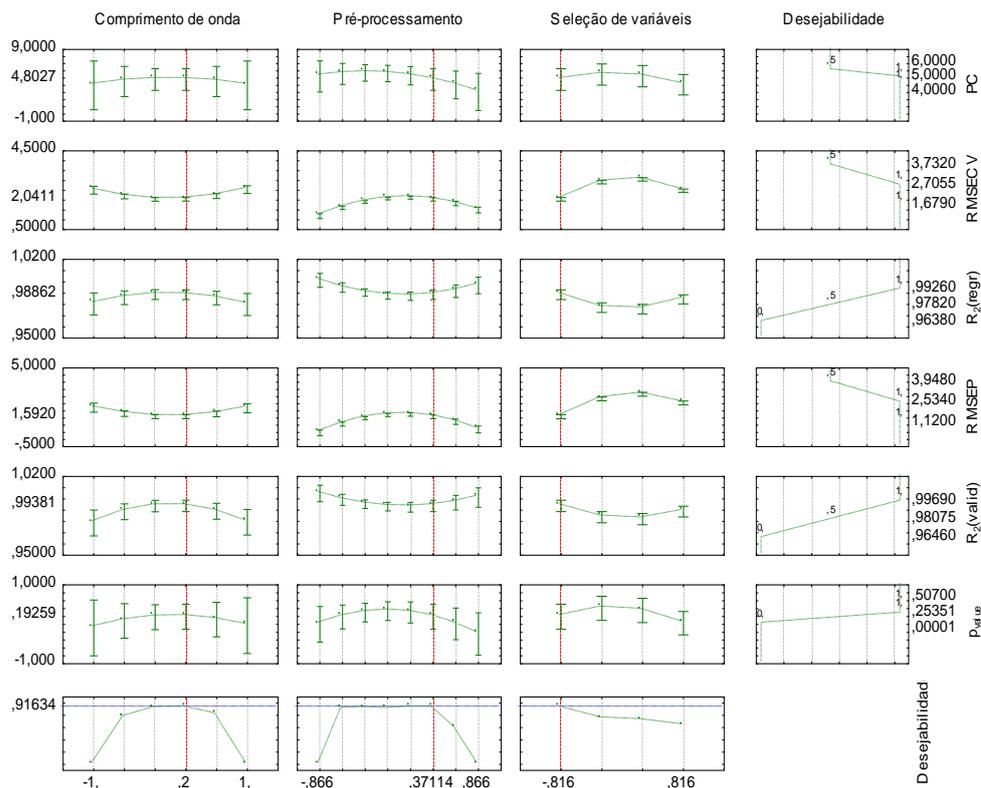


Figura 26. Gráficos da função de desejabilidade para meta-xileno.

Para o meta-xileno, a desejabilidade global obtida foi de 0,916. As condições indicadas para comprimento de onda, pré-processamento e seleção de variáveis na otimização da curva foram: 0,2; 0,37 e -0,816, respectivamente. Assim, as melhores condições para construção dos modelos PLS foram: comprimento de onda na faixa de  $(4700-4500 \text{ cm}^{-1})$  e  $(4000-3930 \text{ cm}^{-1})$ , obtida através do iPLs (0); pré-processamento com centralização na média, normalização e correção da linha de base (0,289); e sem utilizar a técnica de seleção de variáveis (-0,816), conforme a Tabela 5.

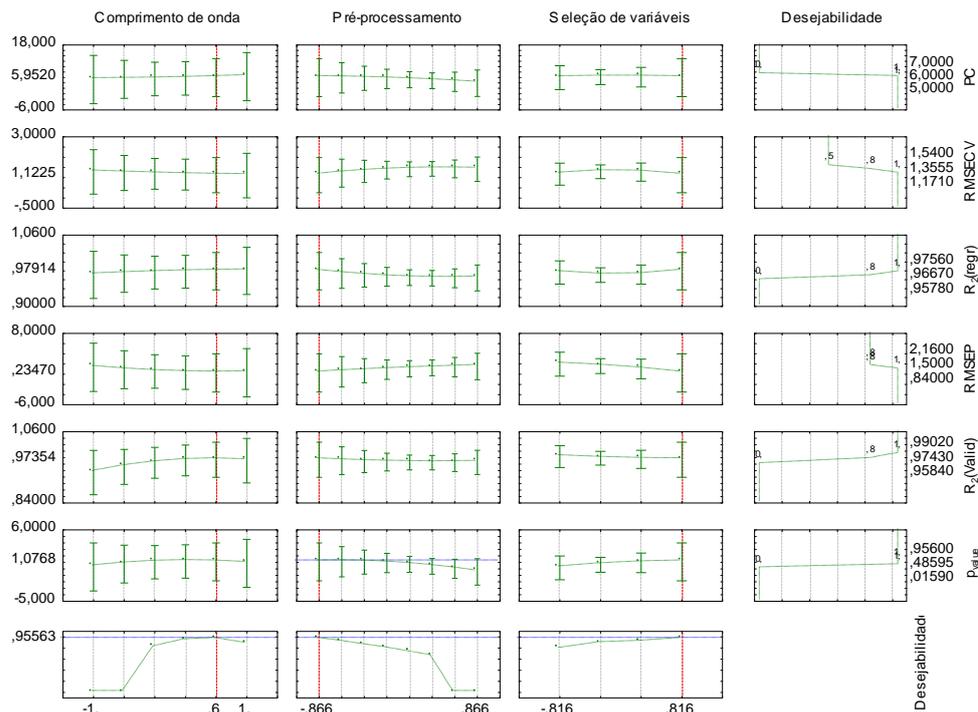


Figura 27. Gráficos da função de desejabilidade para orto-xileno.

Para otimização da calibração de orto-xileno, a desejabilidade global obtida foi de 0,955, e foi obtida usando comprimento de onda, pré-processamento e seleção de variáveis com valores codificados de 0,6; -0,866 e 0,816, respectivamente. Assim, as melhores condições para modelagem foram usando comprimento de onda na faixa  $4700\text{-}3800\text{ cm}^{-1}$  e  $5800\text{-}6000\text{ cm}^{-1}$ , selecionadas pelo iPLS (0,5); pré-processamento somente com autoescalonamento (-0,866); e seleção de variáveis com duas rodadas da técnica Jack Knife (0,816).

#### 4.3.6. Otimização do modelo de calibração de aromáticos C8s+ e aromáticos C9s+

Os resultados das variáveis de resposta para otimização dos modelos para Aromáticos C8s e C9s+ estão apresentados nas Tabelas 14 e 15.

**Tabela 14:** Resultados para as variáveis de resposta usadas para otimização do modelo de Aromáticos C8s+.

Experimento	Resultados da regressão e validação interna			Resultados de validação externa		
	PC	RMSECV (%, m/m)	$r^2_{regr}$	RMSEP (%, m/m)	$r^2_{valid}$	$P_{value}^{d_{teste}}$ $t_{pareado}$
1	6	5,82	0,9638	5,75	0,9765	0,10
2	5	3,86	0,9841	3,09	0,9915	0,08
3	6	5,31	0,9699	5,70	0,9753	0,44
4	5	3,61	0,9861	2,88	0,9924	0,14
5	6	5,32	0,9697	5,87	0,9752	0,3
6	6	3,35	0,9880	3,98	0,9856	0,5
7	6	3,46	0,9871	4,22	0,9832	0,3
8	4	4,60	0,9774	5,08	0,9789	0,03
9	6	3,76	0,9848	4,60	0,9797	0,86
10	4	4,82	0,9752	5,28	0,9765	0,002
11	4	5,58	0,9667	5,40	0,9745	0,17
12	5	3,23	0,9880	2,98	0,9918	0,46
13	5	3,24	0,9880	2,99	0,9916	0,10

**Tabela 15:** Resultados para as variáveis de resposta usadas para otimização do modelo de Aromáticos C9s+.

Experimento	Resultados da regressão e validação interna			Resultados de validação externa		
	PC	RMSECV (%, m/m)	$r^2_{regr}$	RMSEP (%, m/m)	$r^2_{valid}$	$P_{value}^{d_{teste}}$ $t_{pareado}$
1	5	3,09	0,9194	2,64	0,9517	0,15
2	5	2,60	0,9429	1,79	0,9754	0,10
3	6	2,56	0,9446	0,31	0,942	0,94
4	6	2,71	0,9382	1,76	0,9799	<0,0001
5	6	2,71	0,9379	2,26	0,9670	0,63
6	7	1,97	0,9674	1,34	0,9834	0,44
7	7	1,96	0,9678	1,22	0,9905	<0,0001
8	6	1,94	0,9682	1,21	0,9867	0,14
9	7	2,02	0,9656	1,43	0,9810	0,65
10	6	2,04	0,9649	1,45	0,9828	0,03
11	7	1,99	0,9660	1,33	0,9835	0,43
12	5	2,21	0,9590	0,61	0,9756	0,61
13	5	2,21	0,9589	1,62	0,9754	0,68

Os resultados obtidos foram analisados, através da função de desejabilidade, para encontrar um valor ótimo dentro de uma faixa de aceitação para cada resposta, conforme ilustrado na Figura 28 e 29.

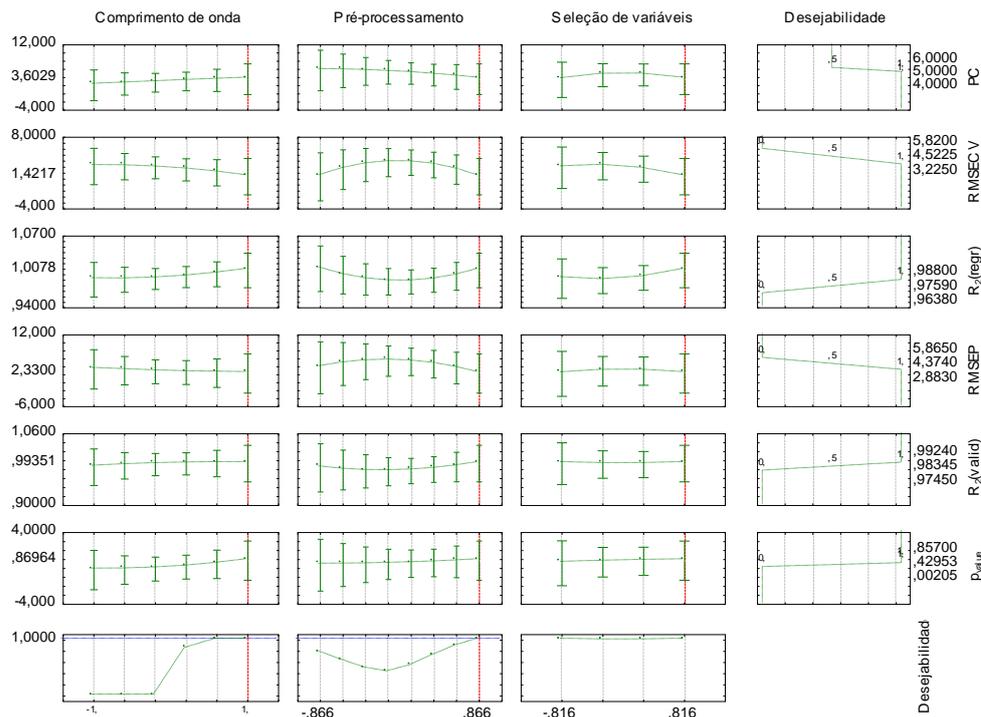


Figura 28. Gráficos da função de desejabilidade para Aromáticos C8s+.

As linhas tracejadas verticais vermelhas indicam as condições ótimas sugeridas para Aromático C8s+, enquanto que a desejabilidade global obtida foi de 1,0. A otimização da calibração foi obtida para comprimento de onda, pré-processamento e seleção de variáveis com valores codificados de 1,0; 0,866 e -0,816, respectivamente. Assim, as melhores condições para modelagem foram: comprimento de onda na faixa de ( $6200-5500\text{ cm}^{-1}$ ) e ( $4800-4000\text{ cm}^{-1}$ ), conforme referência<sup>[63]</sup> (1,0); pré-processamento com segunda derivada Savitsky-Golay e filtro 21 (0,866); e seleção de variáveis: sem seleção (0,816).

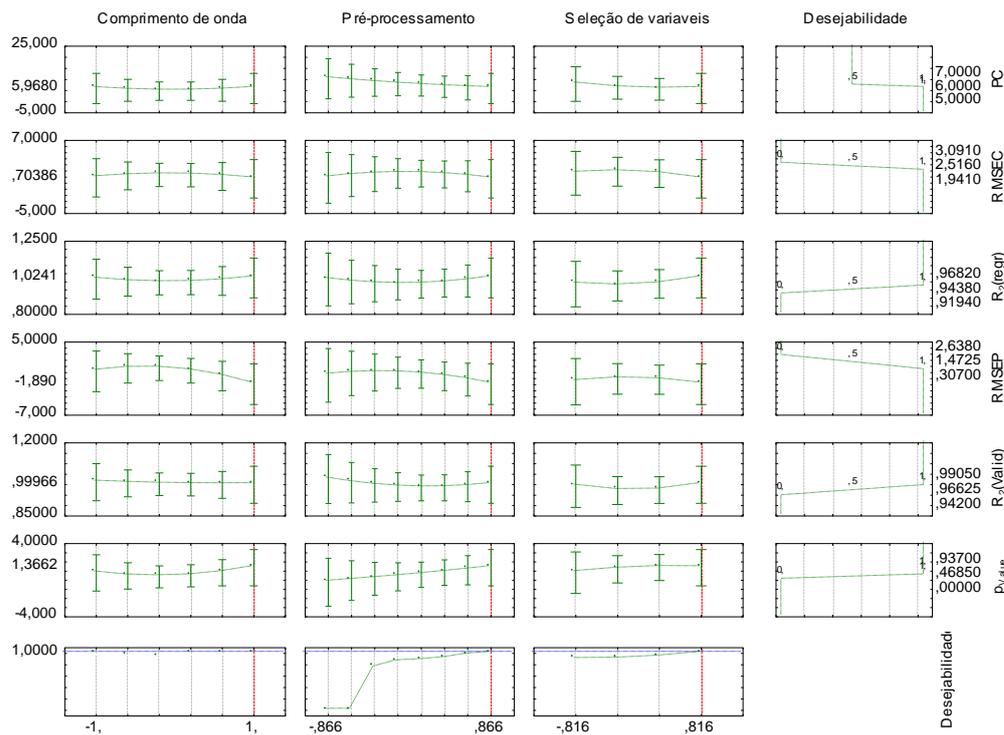


Figura 29. Gráficos da função de desejabilidade para Aromáticos C9s+.

A otimização da calibração de aromáticos C9s+ foi obtida usando comprimento de onda, pré-processamento e seleção de variáveis com valores codificados de 1,0; 0,866 e 0,816, respectivamente. Obteve-se, portanto, as mesmas condições que os constituintes Aromáticos C8s+.

#### 4.4. Otimização dos Modelos

Após a obtenção dos pontos ótimos para cada propriedade, foram construídos os modelos de calibração PLS, de acordo com as condições resumidas na Tabela 16, indicadas pela função de desejabilidade para obtenção dos modelos otimizados.

**Tabela 16:** Condições experimentais para obtenção dos modelos otimizados.

Componente <sup>a</sup>	Faixa de comprimento de onda (cm <sup>-1</sup> )	Código Pre <sup>b</sup>	Técnica de Seleção de variáveis <sup>c</sup>
Não Aromáticos	6400-3800	PRE2	Sem Seleção
Benzeno	6200-5500 e 4800-3900	PRE5	Sem Seleção
Tolueno	6000-5800 e 4700-3800	PRE7	1 Jack Knife
Etil-benzeno	6400-3800	PRE7	Sem Seleção
para-xileno	6200-5500 e 4800-4000	PRE1	2 Jack Knife
meta-xileno	4700-4500 e 4000-3930	PRE5	Sem Seleção
orto-xileno	6000-5800 e 4700-3800	PRE1	2 Jack Knife
Aromáticos C8s+	6200-5500 e 4800-4000	PRE7	Sem Seleção
Aromáticos C9s+	6200-5500 e 4800-4000	PRE7	Sem Seleção

<sup>a</sup>Refere-se à concentração (% m/m) dos componentes.

<sup>b</sup>Código do pré-processamento aplicado a cada modelo, de acordo com a Tabela 5.

<sup>c</sup>Quantidade de rodadas do algoritmo Jack-Knife para seleção de variáveis.

Os modelos obtidos de acordo com as condições acima foram refinados, excluindo as amostras que apresentaram resíduos espectrais e de concentrações altos.

Os resultados obtidos para os modelos são apresentados na Tabela 17. Pelo menos 90 amostras de calibração foram usadas em cada modelo. Para assegurar o desempenho, foram necessários entre 3 e 6 fatores. O número de fatores dos modelos PLS e os valores encontrados para RMSECV foram similares aos trabalhos disponíveis na literatura<sup>[64]</sup> para determinação de xilenos e outros hidrocarbonetos (fatores de 5 a 8 e RMSECV de 0,20 a 6,0% m/m). Os coeficientes de determinação de validação cruzada  $r^2_{\text{reg}}$  foram maiores que 0,98 para todos os modelos.

**Tabela 17:** Figuras de mérito dos modelos PLS de calibração.

Componente	n <sup>a</sup>	PC <sup>b</sup>	Faixa de calibração <sup>c</sup> (%, m/m)	Valores típicos (%, m/m)	r <sub>2</sub> (regr) <sup>d</sup>	RMSECV <sup>e</sup>
Não Aromáticos	104	3	0,59-52,45	12,71	0,9984	0,63
Benzeno	99	4	0,01-11,05	2,39	0,9860	0,37
Tolueno	121	3	0,10-76,64	17,64	0,9937	2,02
Etil-benzeno)	99	5	0,02-39,26	9,07	0,9896	0,72
para-xileno	105	5	0,25-84,22	21,15	0,9984	0,85
meta-xileno	128	4	0,01-55,76	23,48	0,9966	1,14
orto-xileno	111	6	0,01-19,18	8,29	0,9926	0,66
Aromáticos						
C8s+	107	6	1,17-92,10	61,99	0,9931	2,48
Aromáticos						
C9s+	90	5	0,01-32,67	5,28	0,9903	1,11

<sup>a</sup> n é o número de amostras do modelo de calibração.

<sup>b</sup> PC é o número de fatores do modelo PLS.

<sup>c</sup> Faixa de calibração é a faixa de aplicação de cada modelo, baseada no conjunto de calibração.

<sup>d</sup> Representa a correlação entre os valores predito e medidos para o conjunto de validação interna (validação cruzada), pode ser chamado de coeficiente de determinação de  $r^2_{\text{regr}}$ .

<sup>e</sup> Raiz Quadrado do Erro Médio Quadrático de Validação Cruzada.

Deve-se ressaltar que o comportamento dos modelos sempre vai depender da matriz e dos níveis de concentração dos componentes. Dessa forma, o conhecimento da composição química dos componentes deve também ser usado para avaliar o desempenho dos modelos. Avaliando os modelos de calibração, percebe-se que o benzeno apresentou o menor coeficiente de determinação, e isso está relacionado com a complexidade de se determinar compostos em baixas concentrações em misturas complexas.

#### 4.5. Validação Externa dos Modelo

Depois do refinamento final, os modelos foram submetidos a procedimento de validação estatística. Os valores do erro de predição RMSEP apresentados na Tabela 18, foram comparados com os valores de repetitividade dos métodos de referência, sendo encontrado valores aceitáveis, podendo, portanto, ser utilizados para análise de amostras em rotina.

**Tabela 18:** Figuras de mérito dos modelos PLS.

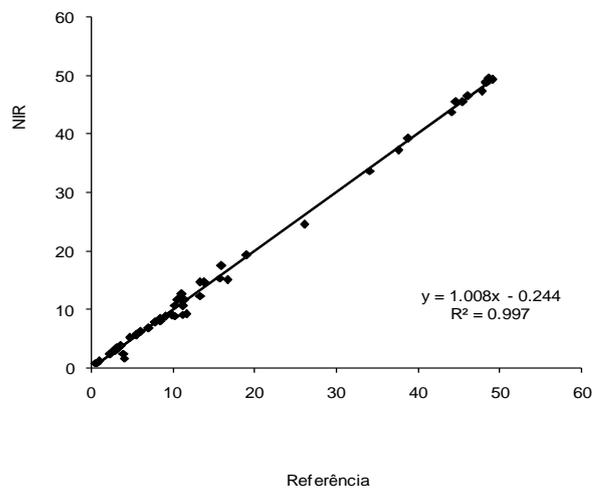
Componente	Referência		NIR	
	Método de referência	Precisão Intermediária <sup>a</sup>	Validação Externa	
			N <sup>b</sup>	p <sub>value</sub> (teste t pareado) <sup>c</sup>
Não Aromáticos (%m/m)	0,73	55	0,88	0,42
Benzeno (%m/m)	0,42	56	0,38	0,79
Tolueno (%m/m)	2,36	63	2,43	0,09
Etil-benzeno (%m/m)	0,28	55	1,19	0,18
para-xileno(%m/m)	0,38	60	1,08	0,26
orto-xileno (%m/m)	0,35	60	1,13	0,24
meta-xileno (%m/m)	0,49	58	1,29	0,09
Aromáticos C8s+ (%m/m)	0,61	62	3,87	0,31
Aromáticos C9s+ (%m/m)	0,68	64	1,47	0,06

<sup>a</sup> Representa a repetibilidade do método de referência.

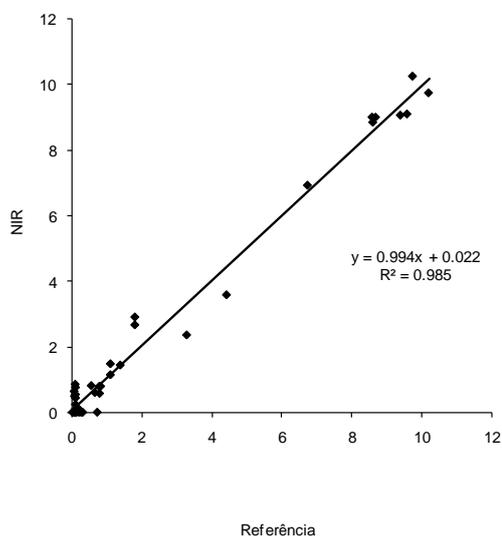
<sup>b</sup> N é o número de amostras usadas na etapa de validação externa

<sup>c</sup> Valores obtidos através teste t pareado para a comparação dos resultados preditos NIR com o método de referência para cada modelo.

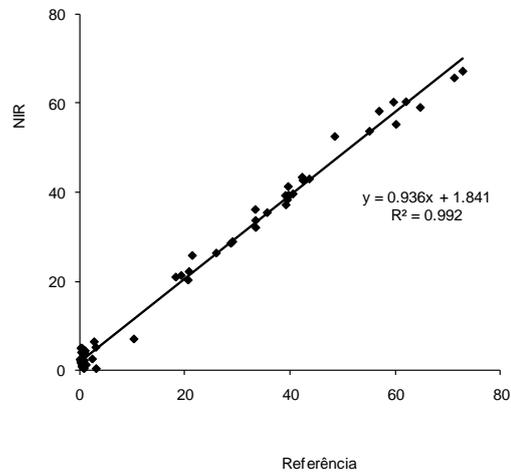
O desempenho dos modelos foi avaliado através da comparação entre os valores preditos e medidos usando o teste t pareado, e nenhum dos valores observados para p foi menor que o valor de p crítico (95% de confiança). Dessa forma, a hipótese nula de que não existe diferença significativa entre os valores preditos e medidos não deve ser rejeitada. Os modelos foram aceitos, encontrando excelentes correlações entre o método NIR e o método de referência. Nas Figuras de 30 a 37, são apresentados os gráficos de correlação para os dados das amostras de validação externa.



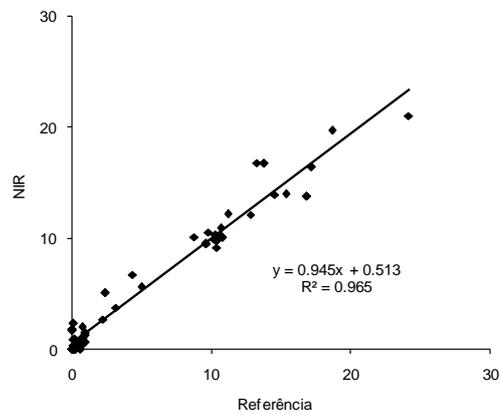
**Figura 30.** Gráfico de correlação dos valores concentração predito versus medido para Não Aromáticos.



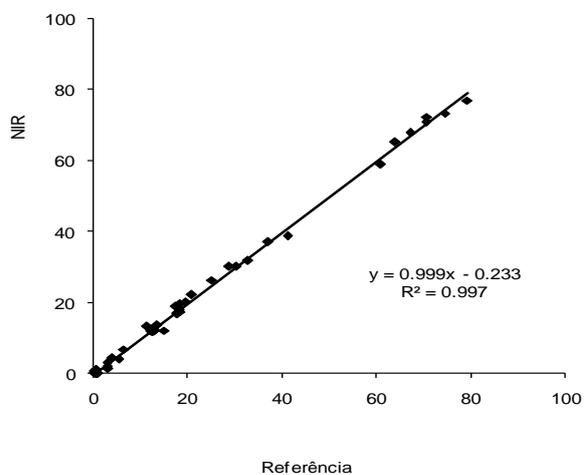
**Figura 31.** Gráfico de correlação dos valores concentração predito versus medido para Benzeno.



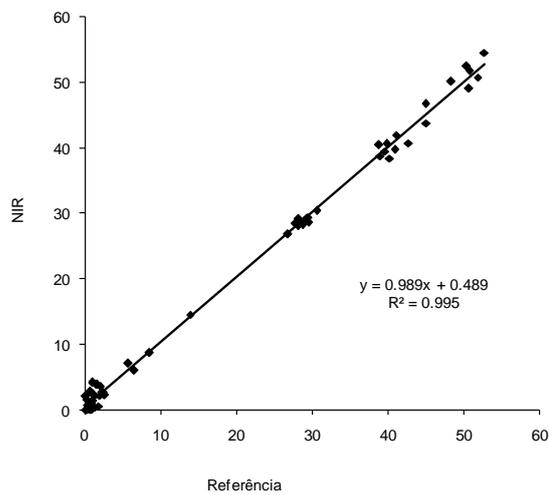
**Figura 32.** Gráfico de correlação dos valores concentração predito versus medido para Tolueno.



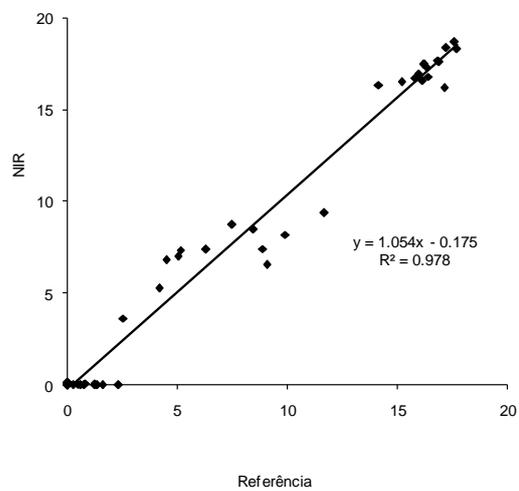
**Figura 33.** Gráfico de correlação dos valores concentração predito versus medido para Etil-Benzeno.



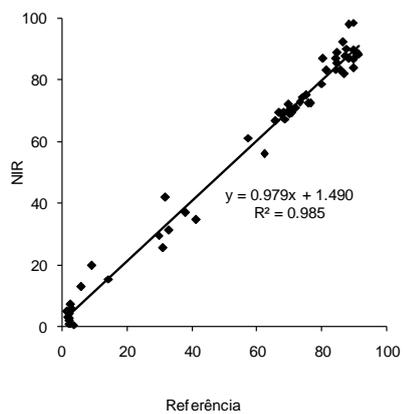
**Figura 34.** Gráfico de correlação dos valores concentração predito versus medido de para-xileno.



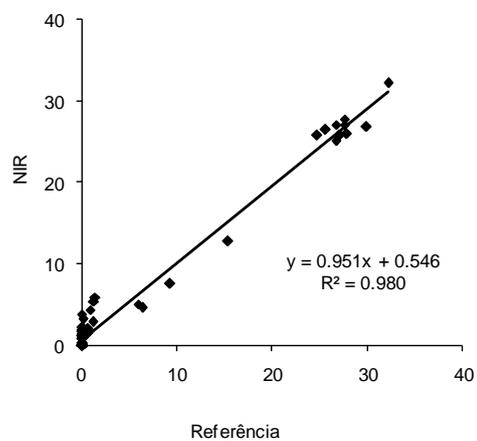
**Figura 35.** Gráfico de correlação dos valores concentração predito versus medido para meta-xileno.



**Figura 36.** Gráfico de correlação dos valores concentração predito versus medido para orto-xileno.



**Figura 37.** Gráfico de correlação dos valores concentração predito versus medido para Aromáticos C8s+.



**Figura 38.** Gráfico de correlação dos valores concentração predito versus medido para Aromáticos C9s+.

## 5. Considerações Finais

A espectroscopia NIR, aliada às técnicas de calibração multivariada PLS, possibilitou o desenvolvimento de metodologias de análise para prever a composição de diferentes amostras no processo de produção de aromáticos na planta petroquímica da Braskem. Os modelos desenvolvidos podem ser utilizados de forma satisfatória para monitorar os parâmetros químicos necessários para controle de qualidade do processo, desta forma, apresentando-se como uma alternativa vantajosa em relação ao método tradicional de análise. A substituição do método de cromatografia gasosa pelo NIR tem diversas vantagens tais como: a possibilidade de analisar amostra de seis diferentes pontos usando o mesmo modelo; rapidez da análise, pois a análise cromatográfica leva cerca de 30 minutos enquanto que no NIR é cerca de 5 minutos; e a possibilidade de obter a composição completa da amostra através de um único espectro.

Uma contribuição importante da metodologia de otimização desenvolvida, foi que através da matriz Doehlert, aumentou a eficiência do processo de construção dos modelos PLS, pois essa matriz permitiu um maior número de combinações das variáveis que interferem no desempenho do modelo, além de ter possibilitado obter os modelos realizando um menor número de experimento.

Neste trabalho, foi possível obter as melhores condições para construção dos modelos PLS, através da técnica de planejamento experimental matriz Doehlert e da função de desejabilidade. 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.

Desta forma, a metodologia NIR-PLS desenvolvida e validada, apresentou um bom desempenho para estimar a concentração de correntes petroquímicas, podendo ser implementada como metodologia para o controle de qualidade da produção.

## REFERÊNCIAS BIBLIOGRAFICA

- [1] LANGERHORST, M.; “Estudo do Comportamento Dinâmico de uma Coluna Industrial de Fracionamento de Tolueno com Integração Energética”; Tese de mestrado, Escola Politécnica- UFBA(2001).
- [2] 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.
- [3] REBOUCAS, 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.
- [4] ALEMEA, H. G.; COSTAB, L. M. ; BARBEIRA, P. J. S.; Determination of ethanol and specific gravity in gasoline by distillation curves and multivariate analysis; **Talanta**, v. 78, p. 1422–1428, 2009.
- [5] Bao, X.; Dai. L; Partial least squares with outlier detection in spectral analysis: A tool to predict gasoline properties; **Fuel**, v. 88, p. 1216–1222, 2009.
- [6] DOEHLERT, D.H.; Uniform shell designs, **Applied Statistics**, v.19, p.231–239, 1970.
- [7] DERRINGER, G.; SUICH, R.; Simultaneous optimization of several response variables; **Journal of Quality Technology**, v. 12, p. 214-219, 1980.
- [8] [www.higieneocupacional.com.br/download/tolueno-innova.pdf](http://www.higieneocupacional.com.br/download/tolueno-innova.pdf)
- [9] ELLIS, J. W.; Trans. Faraday Soc. 1928, 25, 888.
- [10] WHEELER, O. H.; Near infrared spectra: A neglected field of spectral study; **Chemical Education**, v. 37, p. 234, 1960.

- [11] WETZEL, D. L.; Near-infrared reflectance analysis - Sleeper among spectroscopic techniques; **Analytical Chemistry**, v. 55, p. 1165A, 1983.
- [12] HART, J.R.; NORRIS, K.H.; GOLUBIC, C.; **Cereal Chem**, v.39, p. 94, 1961.
- [13] KOWALSKI, B. R.; Chemometrics: Theory and Application, American Chemical society, 1980.
- [14] VANDEGINSTE, B. G. M.; MASSART, D. L.; BUYDENS, L. M. C.; JING, S.; LEWI, P. J.; SMEYERS-VERBEKE, J.; Handbook of Chemometrics and Qualimetrics: Part B, Amsterdam: Elsevier, 1998.
- [15] PASQUIN, C.; Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications; **Journal of the Brazilian Chemical Society**, v.14, p.198, 2003.
- [16] P. WILLIAMS, K. NORRIS; Near-Infrared Technology in the Agricultural and Food Industries, second ed., American Association of Cereal Chemists, St. Paul, MN, USA, 2001.
- [17] SCAFI, S. H. F., PASQUINI, C., Identification of counterfeit drugs using near infrared spectroscopy; **Analyst**, v.126, p.2218, 2001.
- [18] 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.12, 2002.
- [19] PASQUIN, 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.
- [20] CLEVE, E.; BACH, E.; SCHOLLMEYE, E.; Using chemometric methods and NIR spectrophotometry in the textile industry; **Analytica Chimica Acta**, v. 420, p. 163–167, 2000.
- [21] NISHIYAMA, P.; ALVAREZ, M.; VIEIRA, L.G.E.; **Biology and Technology**, v.34, p. 361, 1991.

- [22] NISHIYAMA, P.; ALVAREZ, M.; VIEIRA, L.G.E.; J. SCI.; **Agriculture and Food**, v. 59, p. 277, 1992.
- [23] SKOOG, D.A.; HOLLER, F.J.; NIENAM, T.A; Princípios de análise Instrumental. 5a ed., Bookman, Porto Alegre, 2002.
- [24] COATES, J.; A review of current new technology: Used in instrumentation for industrial vibrational spectroscopy, **Spectroscopy**, v.14, p. 21-34,1999.
- [25] BURNS, D. A.; CIURCZAK, E. W. Handbook of Near-Infrared Analysis. New York: Marcel Dekker, 2001.
- [26] SCHRADER, B.; Infrared and Raman Spectroscopy: methods and Applications, Weinheim: VCH, 1995.
- [27] BLANCO, M.; VILLARROYA, I.; NIR spectroscopy: a rapid-response analytical tool, **Trends in analytical chemistry**, vol. 21, p.243, 2002.
- [28] BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E.; 25 anos de Quimiometria no Brasil, **Química Nova**, v.29, p.1401-1406, 2006.
- [29] GALVÃO, R. K. H.; ARAUJO, M. C. U.; SALDANHA, T. C. B.; VISANI, V.; PIMENTEL, M. F.; Estudo comparativo sobre filtragem de sinais instrumentais usando transformada de Fourier e de Wavelet; **Química Nova**, v.24 (6), p.874-884, 2001.
- [30] GALVÃO, R. K. H.; JOSÉ, G. E.; DANTAS FILHO, H. A.; ARAÚJO, M. C. U.; SILVA, E. C.; PAIVA, H. M.; SALDANHA, T. C. B.; SOUZA, E. S. O. N.; Optimal wavelet filter construction using X and Ydata, **Chemometrics and Intelligent Laboratory Systems**, v.70, p.1-10, 2004.
- [31] BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E; Planejamento e Otimização de Experimentos, 2nd Ed,. Editora da Unicamp,Campinas, Brasil, 1996
- [32] SMALL, G.W.; Chemometrics and near-infrared spectroscopy: Avoiding the pitfalls, **Trends in Analytical Chemistry**, v. 25, p.1062-1066, 2006.

- [33] BRERETON, R. G.; Introduction to multivariate calibration in analytical chemistry, **Analyst**, v.125, p.2125-2154, 2000.
- [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] DANTAS FILHO, H. A.; GALVÃO, R. K. H.; ARAUJO, 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.
- [36] FEUDALE, R. N.; WOODY, N. A.; TAN, H.; MYLES, A. J; BROWN , S. D.; FERRÉ, J.; Transfer of multivariate calibration models: A review, **Chemometrics and Intelligent Laboratory Systems**, v.64, p.181-192, 2002.
- [37] HONORATO, F. A.; GALVÃO, R. K. H.; PIMENTEL, M. F.; BARROS NETO, B.; ARAÚJO, M. C. U.; CARVALHO, F. R.; Robust modeling for multivariate calibration transfer by the successive projections algorithm, **Chemometrics and Intelligent Laboratory Systems**, v.76, p.65-72, 2005.
- [38] BEEBE, K. R.; RANDY, J. P.; SEASHOLTZ, M. B., *Chemometrics: A practical guide*, Wiley-Interscience, New York, 1998.
- [39] 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.
- [40] 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.

- [41] HELLAND, I.S.; Some Theoretical aspects of partial least squares regression, **Chemometrics and Intelligent Laboratory Systems**, v. 58, p. 97-107, 2001.
- [42] GELADI, P.; KOWALSKI, B. R.; Partial Least Squares Regression: A Tutorial, **Analytical Chemistry Acta**, 1-17, 1986.
- [43] NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T.; A user-friendly guide to Multivariate Calibration and Classification, Chichester- UK, NIR Publications, 2002.
- [44] 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.
- [45] 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.
- [46] CERQUEIRA, E. O.; POPPI, R. J.; KUBOTA, L. T.; MELO, C.; Utilização de filtro de transformada de Fourier para minimização de ruídos em sinais analíticos, **Química Nova**, 23: 690, 2000.
- [47] OTTO, M. Chemometrics – Statistics and Computer application in Analytical Chemistry, Weinheim: Wiley, 1999.
- [48] CENTNER, V.; MASSART, D.; Elimination of uninformative variables for multivariate calibration, **Analytical Chemistry**, v.68, p. 3851-3858, 1996.
- [49] 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.
- [50] BORIN, A.; POPPI, R. J.; Application of mid spectroscopy and iPLS for the quantification of contaminants in lubricating oil, **Vibrational Spectroscopy**, v.37, p.27-32, 2005.

- [51] PONTES, M. J. C.; GALVÃO, R. K. H.; ARAUJO, 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.
- [52] LEARDI, R.; Genetic algorithms in chemometrics and chemistry: a review, **Jornal of Chemometrics**, v.15, p. 559, 2001.
- [53] COSTA FILHO, P.A; POPPI, R.J.; Aplicação de algoritmos genéticos na seleção de variáveis em espectroscopia no infravermelho médio, determinação simultânea de glicose, maltose e frutose, **Química Nova**, v. 25, p. 46-52, 2002.
- [54] KENNARD, R. W.; STONE, L. A.; Computer aided desing of experiments, **Technometrics**, v. 11, p. 137-148,1969.
- [55] FILHO, H. A.D.; GALVÃO, R. K. H.; ARAÚJO, M.C .U.; DA SILVA, E. C.; SALDANHA, T. C. B.; JOSÉ, G. E.; PASQUINE, C.; (...); ROHWEDDER, J. J. R; A strategy for selecting calibration samples for multivariate modelling, **Quimiometria e Laboratório de Sistemas Inteligentes**, v.73, p. 83, 2004.
- [56] GALVÃO, R. K. H.; ARAUJO, M. C. U.; JOSÉ, G. E.; PONTES, M. J. C.; SILVA, E. C.; SALDANHA, T. C. B.; A method for calibration and validation subsetpartitioning, **Talanta**, v. 67, p. 736-738, 2005.
- [57] GÁZQUEZ, D.; SÁNCHEZ-VIÑAS, M.; BAGUR, M.G.;GARCIA,G.; Application of Doehlert designs in the optimization of experimental variables associated with the simultaneous solvent extraction of several metals, **Journal of Analytical Atomic Spectrometry**, v.13, p. 105-108, 1998.
- [58] FERREIRA, S.L.C.; SANTOS, W.N.L.; QUINTELLA, C.M.; NETO, B.B.; BOSQUE-SENDRA, J.M.; Doehlert matrix: a chemometric tool for analytical chemistry - a review, **Talanta**, v. 63, p. 1061-1067, 2004.
- [59] GOMES-CARRACEDO, M.P.; ANDRADE, J.M.; CALVIÑO, M.; FERNADEZ, E.; PRADA, D.; MUNIATEQUI, S.; Multivariate prediction of

eight Kerosene properties employing vapour-phase mid-infrared spectroscopy, **Fuel**, v. 82, p. 1211-1218, 2003.

[60] CANECA, A.R.; PIMENTEL, M.F.; GALVÃO, R.K.H.; MATTA, C.E.; CARVALHO, F.R.; PASQUINE, C.; ROHWEDDER, J.J.R.; Assessment of infrared spectroscopy and multivariate techniques for monitoring the service condition of diesel-engine lubricating oils, **Talanta**, v.70, p. 344–352, 2006.

[61] 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.

[62] ESPINOSA, A.; SANCHEZ, M.; OSTA, S.; BONIFACE, C.; On-line NIR Analysis and Advanced Control Improve Gasoline Blending, **Oil & Gas Journal**, v. 92, p. 42, 49-56, 1994.

[63] REBOUCAS, M. V.; BARROS NETO, B.; Near infrared spectroscopic prediction of physical properties of aromatics-rich hydrocarbon mixtures, **J. Near Infrared Spectroscopy**, v. 9, p. 263-273, 2001.

[64] CHUNG, H.; LEE, J-K.; KU, M-S.; Feasiability of simultaneous measurement of xylene Isomers and other hydrocarbons in p-xylene production processes using near infrared spectroscopy, **Applied Spectroscopy**, v 52, p. 885-889, 1998.

[65] HONORATO, F. A.; NETO, B.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.

[66] FERNANDES, L.H.; RAIMUNDO JR, I.M.; PASQUINE, 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.

[67] 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.

[68] Statistica for Windows, StatSoft, Inc., 2300 East 14th Street, Tulsa, OK 741014, USA, 2001.

[69] Annual Book of ASTM Standards. Standards practices for infrared multivariate quantitative analysis, E1655, vol 03.06. ASTM International, WestConshohocken, Pennsylvania, USA, 2005.

[70] MAESSCHALCK, R. DE; JOUAN-RIMBAUD, D.; MASSART, D. L.; The mahalanobis distance, **Chemometrics and Intelligent Laboratory Systems**, v.50, p.1-18, 2000.

[71] MILLER, J.C.; MILLER, J.N.; Statistics for Analytical Chemistry, third ed. Ellis Horwood PTR Prentice Hall, London, 1993.

[72] WORKMAN, J. J.; Interpretive spectroscopy for near infrared, **Applied Spectroscopy**, v.31, p.251, 1996.