

**UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA  
DEPARTAMENTO DE QUÍMICA FUNDAMENTAL**

**PREVISÃO DE PROPRIEDADES DAS GASOLINAS DO  
NORDESTE EMPREGANDO ESPECTROSCOPIA NIR/MIR  
E TRANSFERÊNCIA DE CALIBRAÇÃO**

**TESE DE DOUTORADO**

**Fernanda Araújo Honorato**

**U F P E**

**Orientadores: Profa. Dra. Maria Fernanda Pimentel**

**Prof. Dr. Benício de Barros Neto**

**Tese de Doutorado**



**PREVISÃO DE PROPRIEDADES DAS GASOLINAS DO  
NORDESTE EMPREGANDO ESPECTROSCOPIA NIR/MIR  
E TRANSFERÊNCIA DE CALIBRAÇÃO**

**LAC-UFPE**  
Laboratório  
de Combustíveis

*Submetido ao Programa de Pós Graduação em Química  
do Departamento de Química Fundamental da UFPE  
como parte dos requisitos para obtenção do título  
de Doutor em Química.*

**Fernanda Araújo Honorato**

**Recife, fevereiro de 2006**

**Honorato, Fernanda Araújo**

**Previsão de propriedades das gasolinas do Nordeste empregando espectroscopia NIR/MIR e transferência de calibração / Fernanda Araújo Honorato. – Recife : O Autor, 2006.**

**x, 91 folhas ; il., fig., tab.**

**Tese (doutorado) – Universidade Federal de Pernambuco. CCEN. Química, 2006.**

**Inclui bibliografia e anexos.**

**1. Química analítica – Quimiometria. 2. Propriedade das gasolinas – Espectroscopia NIR/MIR. 3. Análise de componentes principais – Técnica quimiométrica. 4. Transferência de modelos de calibração – Seleção de variáveis robustas. I. Título.**

**543.421  
543.57**

**CDU (2.ed.)  
CDD (22.ed.)**

**UFPE  
BC2006-185**

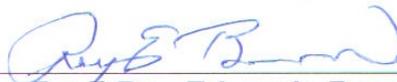
Tese submetida ao Corpo Docente do Programa de Pós-Graduação em Química do Departamento de Química Fundamental da Universidade Federal de Pernambuco como parte dos requisitos necessários para a obtenção do Grau de Doutor em Química.

Aprovada:



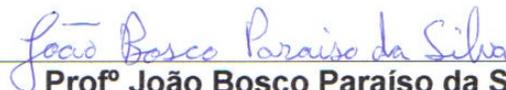
---

**Profª Maria Fernanda Pimentel**  
Departamento de Engenharia Química da UFPE  
(Orientadora)



---

**Profº Roy Edwards Bruns**  
Instituto de Química da UNICAMP



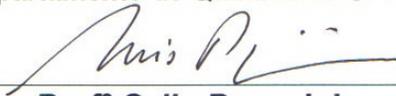
---

**Profº João Bosco Paraíso da Silva**  
Departamento de Química Fundamental da UFPE



---

**Profº Wallace Duarte Fragoso**  
Departamento de Química da UFPB



---

**Profº Celio Pasquini**  
Instituto de Química da UNICAMP

***“Previsão de Propriedades das Gasolinas do Nordeste  
Empregando Espectroscopia NIR/MIR e Transferência de  
Calibração”***

por

**Fernanda Araújo Honorato**

Departamento de Química Fundamental  
Centro de Ciências Exatas e da Natureza  
Universidade Federal de Pernambuco  
Recife-PE-Brasil  
23 de Fevereiro de 2006

*Aos meus pais, Geraldo e Leticia*  
*Às minhas irmãs, Jaqueline e Bianca*  
*À minha filhinha, Mariana*  
*Ao meu esposo, Ricardo*  
*Com amor e gratidão*

*Dedico.*

## **AGRADECIMENTOS**

---

- \* À professora e amiga Fernanda Pimentel, pela excelente orientação, apoio, amizade, incentivo e compreensão.
- \* Ao professor Benício de Barros Neto, pela co-orientação, ensinamentos na área de quimiometria e valiosas discussões e sugestões.
- \* Ao professor Roberto Kawakami da Divisão de Engenharia Eletrônica – ITA, pelas contribuições imprescindíveis na parte de transferência de calibração.
- \* Ao professor e amigo Mário César Ugolino, pelas contribuições e por permitir o uso do espectrômetro NIR para fins de transferência.
- \* Aos responsáveis pelo Laboratório de Combustíveis, professores Luís e Florival, pela acolhida e prestação de serviços incondicionais.
- \* Aos laboratórios participantes da Rede de Combustíveis e Lubrificantes (RECOL NO/NE), pela coleta das amostras e análises pelos métodos convencionais.
- \* Aos professores Celio Pasquini, Ivo Raimundo e Jarbas, que através das reuniões do projeto PROCAD/CAPES, proporcionaram importantes discussões científicas.
- \* Aos amigos Cláudio, Carla, Suzana, Flávio e José Edson e aos demais colaboradores do LAC (Fernandinho, André, Daniel, Janete, Márcia, Isaías, Jackson, Jâmison, Rogério, Clodomir, Emerson, Romário e Arnóbio), pelo agradável convívio no ambiente de trabalho e auxílio nas atividades de pesquisa.
- \* Aos professores do Departamento de Química Fundamental – UFPE, pelos importantes ensinamentos.
- \* Ao professor Ricardo Longo, pela solicitude como coordenador da pós-graduação.
- \* Ao CNPq, pela concessão da bolsa.
- \* À Ricardo, meu esposo e cúmplice, pelo apoio, compreensão, carinho e dedicação incondicionais, e pelas discussões e sugestões sobre o projeto de pesquisa.
- \* À minha filhinha Mariana, por servir de fonte de inspiração e incentivo no árduo caminho da ciência.

ÍNDICE DE FIGURAS.....	iii
ÍNDICE DE TABELAS.....	vi
LISTA DE ABREVIATURAS E SIGLAS.....	vii
RESUMO.....	ix
ABSTRACT.....	x
INTRODUÇÃO.....	1
PARTE 1 – PREVISÃO DE PROPRIEDADES DAS GASOLINAS DO NORDESTE DO BRASIL.....	3
1.1 - GASOLINAS: TIPOS E PROPRIEDADES.....	3
1.2 - ESPECTROSCOPIA NO INFRAVERMELHO.....	7
1.3 - TÉCNICAS QUIMIOMÉTRICAS.....	10
1.3.1 - Pré-processamento dos Dados.....	11
1.3.2 - Seleção de Variáveis.....	12
1.3.3 - Calibração Multivariada.....	16
1.4 - REVISÃO DA LITERATURA.....	19
1.5 – OBJETIVO.....	26
1.6 - EXPERIMENTAL.....	27
1.6.1 - Conjunto de Dados.....	27
1.7 - RESULTADOS E DISCUSSÃO.....	31
1.7.1 - Região NIR.....	31
1.7.2 - Região MIR.....	34
PARTE 2 - TRANSFERÊNCIA DE MODELOS DE CALIBRAÇÃO.....	44
2.1 - REVISÃO DA LITERATURA.....	44
2.1.1 - Adaptação do Modelo de Calibração e das Respostas Espectrais - Métodos de Padronização.....	45
2.1.1.1 - Padronização das Respostas Espectrais.....	45

2.1.1.1.1 - Padronização Direta (DS).....	47
2.1.1.1.2 - Padronização Direta em Etapas (PDS).....	48
2.1.1.1.3 - Padronização Univariada.....	49
2.1.1.2 - Padronização dos Valores Previstos.....	50
2.1.2 - Aumento da Robustez do Modelo.....	51
2.1.2.1 - Modelos Globais.....	51
2.1.2.2 - Pré-processamentos.....	52
2.1.2.3 - Seleção de Variáveis.....	55
2.1.3 - Seleção das Amostras de Transferência.....	55
2.2 - OBJETIVO.....	57
2.3 - FUNDAMENTAÇÃO TEÓRICA.....	57
2.3.1 - Notação.....	57
2.3.2 - Critério de Robustez.....	57
2.3.3 - Seleção das Amostras de Transferência.....	58
2.4 - EXPERIMENTAL.....	60
2.4.1 - Conjuntos de Dados.....	60
2.4.2 - Pré-processamento dos Dados.....	61
2.4.3 - Modelagem e Validação.....	61
2.5 - RESULTADOS E DISCUSSÃO.....	62
<b>CONCLUSÕES E PROPOSTAS FUTURAS.....</b>	<b>70</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>71</b>
<b>ANEXOS.....</b>	<b>82</b>
Tabela A - Valores absolutos dos RMSEP's obtidos para cada propriedade.....	82
Tabela B - Valores relativos dos RMSEP's obtidos para cada propriedade.....	84
Figura A.1 - Gráfico dos valores previstos x observados das amostras de previsão (a) Densidade, (b) PIE, (c) T 10%, (d) T 50%, (e) T 90%, (f) PFE, (g) MON, (h) RON, (i) Parafinas, (j) Isoparafinas, (l) Aromáticos, (m) Naftênicos, (n) Olefinas.....	86

## ÍNDICE DE FIGURAS

---

Figura 1.1 - Espectrômetro FT-IR Spectrum GX da Perkin Elmer.....	28
Figura 1.2 - Acessórios utilizados para aquisição dos espectros nas três faixas espectrais: (a) Célula de 1 mm, (b) Célula de 10 mm e (c) Sonda de ATR.....	28
Figura 1.3 - Sistema de fluxo para aquisição dos espectros com as células (a) compartimento para fixação das celas, (b) bomba peristáltica.....	29
Figura 1.4 - Perfil do espectro para a amostra AL 330 na região NIR com caminho óptico de 10 mm.....	31
Figura 1.5 - Perfil do espectro para a amostra AL 330 na região NIR com caminho óptico de 1 mm.....	32
Figura 1.6 - Gráfico dos escores de todas as amostras nas duas primeiras componentes para a região NIR com caminho óptico de 10 mm. PC1 e PC2 descrevem, respectivamente, 75% e 22% da variância nos dados espectrais.....	33
Figura 1.7 - Gráfico dos escores de todas as amostras nas duas primeiras componentes, para a região NIR com caminho óptico de 1 mm. PC1 e PC2 descrevem, respectivamente, 68% e 17% da variância nos dados espectrais.....	33
Figura 1.8 - Perfil do espectro para a amostra AL 330 na região MIR.....	34
Figura 1.9 - Gráfico dos escores de todas as amostras de gasolinas do NE nas duas primeiras componentes, na região MIR. PC1 e PC2 descrevem, respectivamente, 75% e 10% da variância nos dados espectrais.....	35
Figura 1.10 - Gráfico dos escores de todas as amostras de gasolinas do NE nas duas primeiras componentes da PCA baseada nos valores das treze propriedades de interesse. PC1 e PC2 descrevem, respectivamente, 75% e 8% da variância nos dados espectrais.....	35
Figura 1.11 - Gráfico dos pesos nas duas primeiras componentes da PCA baseada nos valores das treze propriedades de interesse.....	36
Figura 1.12 - Gráfico dos pesos obtido para os valores relativos de RMSEP das treze propriedades. PC1 e PC2 descrevem, respectivamente, 47% e 40% da variância nos dados.....	38

Figura 1.13 -Gráfico dos escores em PC1 e PC2, identificados de acordo com as faixas espectrais e os algoritmos de calibração. PC1 e PC2 descrevem, respectivamente, 47% e 40% da variância nos dados. NIRA equivale a NIR com caminho óptico de 1 mm e NIRB, a NIR com caminho óptico de 10 mm.....	38
Figura 1.14 - Gráfico dos escores em PC1 e PC2 dos valores relativos de RMSEP identificados de acordo com os pré-processamentos e os algoritmos de seleção de variáveis. PC1 e PC2 descrevem, respectivamente, 47% e 40% da variância nos dados.....	39
Figura 1.15 - Gráfico dos pesos obtido para os valores relativos de RMSEP das propriedades, excluindo Parafinas e Naftênicos. PC1 e PC2 descrevem, respectivamente, 55% e 17% da variância nos dados.....	40
Figura 1.16 - Gráfico dos escores em PC1 e PC2 dos valores relativos de RMSEP, excluindo Parafinas e Naftênicos, identificadas de acordo com as faixas espectrais e os algoritmos de calibração. PC1 e PC2 descrevem, respectivamente, 55% e 17% da variância nos dados.....	41
Figura 1.17 - Gráfico dos escores em PC1 e PC2 dos valores relativos de RMSEP excluindo Parafinas e Naftênicos identificados de acordo com os pré-processamentos e os algoritmos de seleção de variáveis. PC1 e PC2 descrevem, respectivamente, 55% e 17% da variância nos dados.....	41
Figura 2.1 - Espectro original das amostras de gasolinas obtidos nos equipamentos primário (a) e secundário (b).....	62
Figura 2.2 - Espectro original das amostras de milho obtidos nos equipamentos primário (a) e secundário (b).....	63
Figura 2.3 - Espectros derivados das amostras de gasolinas depois da pré-seleção das regiões espectrais: (a) primário, (b) secundário e (c) diferenças entre os espectros medidos no primário e no secundário.....	64
Figura 2.4 - Espectros derivados das amostras de milho: (a) primário, (b) secundário e (c) diferenças entre os espectros medidos no principal e no secundário.....	64

- Figura 2.5 - Comparação entre os resultados do APSV-MLR (O) e PDS-PLS ( ) para T 90% nos dados de gasolina utilizando-se dois métodos para seleção de amostras: (a) KS e (b) APSA. Os símbolos pretos indicam que os valores correspondentes são significativamente menores (ao nível de 95% de confiança) que os valores alternativos..... 66
- Figura 2.6 - Comparação entre os resultados do APSV-MLR (O) e PDS-PLS ( ) para teor de umidade nos dados de milho utilizando-se dois métodos para seleção de amostras: (a) KS e (b) APSA. Os símbolos pretos indicam que os valores correspondentes são significativamente menores (ao nível de 95% de confiança) que os valores alternativos..... 67
- Figura 2.7 - Valor do erro (E) em função do número de variáveis selecionadas pelo APSV considerando nove amostras de transferência. (a) Dados de gasolina com amostras de transferência selecionadas pelo APSA. (b) Dados de milho com amostras de transferência selecionadas pelo KS. As setas indicam o ponto de mínimo utilizado para escolher o número de variáveis..... 68

## **ÍNDICE DE TABELAS**

---

Tabela 1.1 – Combinações de técnicas de calibração e pré-processamentos utilizados no desenvolvimento dos modelos.....	30
Tabela 1.2 – Valores de RMSEP para cada propriedade considerando a condição de modelagem mais favorável.....	43
Tabela 2.1 – Valores dos RMSEP's obtidos pelos métodos PLS e APSV-MLR.....	65

## ***LISTA DE ABREVIATURAS E SÍMBOLOS***

---

AEAC:	Álcool Etílico Anidro Combustível
AG:	Algoritmo Genético
ANN:	Artificial Neural Networks
ANP:	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
AOTF:	AcoustoOptic Tunable Filters
APS:	Algoritmo das Projeções Sucessivas
APSA:	Algoritmo das Projeções Sucessivas para seleção de amostras
APSV:	Algoritmo das Projeções Sucessivas para seleção de variáveis
ASTM:	American Society for Testing and Materials
ATR:	Attenuated Total Reflectance
BTEX:	Benzeno, tolueno, etil benzeno, e orto-, meta- e paraxilenos
DS:	Direct Standardization
ETBE:	Etil terc-butil éter
ETOH:	Etanol
FIR:	Finite Impulse Response Filters
FT-IR:	Fourier Transform Infrared
IAD:	Índice Antidetonante
KS:	Algoritmo Kennard-Stone
MIR:	Medium Infrared
MLR:	Multiple Linear Regression
MON:	Motor Octane Number
MSC:	Multiplicative Signal Correction
MTBE:	Metil terc-butil éter
NBR:	Norma Brasileira de Regulamentação
NIR:	Near Infrared
OSC:	Orthogonal Signal Correction
PC:	Principal Component
PCA:	Principal Components Analysis

PCR:	Principal Components Regression
PDS:	Piecewise Direct Standardization
PFE:	Ponto final de evaporados
PIANO:	Parafinas, Isoparafinas, Aromáticos, Naftênicos e Olefinas
PIE:	Ponto inicial de evaporados
PLS:	Partial Least Squares
PONA:	Parafinas, Olefinas, Naftênicos e Aromáticos
RMSEP:	Root Mean Square Error of Prediction
RMSEP <sup>P</sup> :	RMSEP para o conjunto de predição do equipamento primário utilizando o modelo calculado nesse equipamento
RMSEP <sup>S</sup> <sub>p</sub> :	RMSEP para o conjunto de predição do equipamento secundário utilizando o modelo calculado no equipamento primário
RMSEP <sup>S</sup> <sub>PT</sub> :	RMSEP para o conjunto de predição do equipamento secundário utilizando o modelo calculado no primário, com técnicas de transferência
RMSEP <sup>S</sup> <sub>s</sub> :	RMSEP para o conjunto de predição do equipamento secundário utilizando o modelo calculado nesse equipamento
RMSET:	RMSEP para o conjunto de teste
RMSEV:	RMSEP para o conjunto de predição
RON:	Research Octane Number
SG:	Savitzky-Golay Smoothing
T 10%:	Temperatura para 10% de destilados
T 50%:	Temperatura para 50% de destilados
T 90%:	Temperatura para 90% de destilados
TAME:	Terc-amil metil éter
TOP:	Transfer by Orthogonal Projections

Este trabalho tem duas partes, ambas envolvendo espectroscopia no infravermelho e quimiometria. Na primeira parte, foram obtidos modelos de calibração multivariada baseados em dados espectrais nas regiões NIR e MIR para prever as principais propriedades de gasolinas comercializadas na região Nordeste. Foram coletadas 160 amostras de gasolinas e os modelos de calibração foram construídos considerando-se dados espectrais da região NIR (em dois caminhos ópticos diferentes - 1 e 10 mm) e MIR, dois algoritmos de calibração (mínimos quadráticos parciais - PLS e regressão linear múltipla - MLR), e diferentes pré-processamentos (derivada, alisamento e seleção de variáveis com o Algoritmo Genético, AG, ou o Algoritmo de Projeções Sucessivas, APS). Analisando-se os erros médios quadráticos relativos de previsão (RMSEPR) para os vários modelos, observou-se que todas as propriedades envolvidas podem ser preditas de forma satisfatória a partir do espectro NIR na faixa 1600-2500 nm (caminho óptico de 1 mm), com calibração por MLR e seleção de variáveis pelo algoritmo genético, com qualquer dos pré-processamentos utilizados. A outra parte trata do problema de transferência de calibração. Propôs-se uma nova estratégia para a construção de modelos de calibração robustos em relação a diferenças entre dois equipamentos. O APS foi utilizado para selecionar variáveis de forma a minimizar o erro de previsão para o conjunto de teste do equipamento primário, mas também para um pequeno conjunto de amostras medidas no equipamento secundário (amostras de transferência). Dois conjuntos de dados foram empregados: espectros MIR de gasolinas C, para previsão da propriedade T90% (temperatura para 90% de amostra destilada); e espectros NIR de amostras de milho para previsão do teor de umidade. Os modelos MLR robustos assim obtidos foram comparados a modelos PLS, utilizando-se padronização direta em etapas (PDS) para corrigir os espectros do equipamento secundário. Os erros de predição no equipamento secundário para os modelos MLR robustos foram comparáveis aos dos modelos PLS-PDS e levemente inferiores aos erros do modelo APSV-MLR.

Palavras chave: *Propriedades das gasolinas, Espectroscopia NIR/MIR, Transferência de Calibração.*

This work consists of two parts, both involving infrared spectroscopy and chemometrics. In the first part, multivariate calibration models based on NIR and MIR spectral data are developed to predict several properties of gasoline samples sold in stations of Northeastern Brazil. 160 gasoline samples were collected, and several calibration models based on data from these samples were built. These models differ according (a) to the spectral region – NIR (with two optical path lengths, 1 or 10 mm) or MIR; (b) to the calibration model (multiple or partial least squares regression); and (c) to the preprocessing method (derivation, smoothing and variable selection either with a genetic algorithm or the successive projections algorithm, SPA). From an analysis of the mean square errors of prediction corresponding to these models, it is concluded that all gasoline properties included in this study can be satisfactorily predicted from NIR spectral data in the 1600-2500 nm range (with 1-mm path length) by a multiple linear regression (MLR) model with variable selection by the genetic algorithm, using any of the preprocessing techniques. The second part deals with calibration transfer. A new transfer strategy is proposed to build calibration models that are robust with respect to differences between two spectrometers. The SPA algorithm was used to select variables that minimize the prediction errors not only for the validation set used with the master equipment, but also those of a small sample set whose data are recorded on the slave equipment (the transfer samples). Two data sets were used to test the proposed methodology: MIR spectra of type C gasolines with the 90% distillate temperature as the response, and NIR spectra of maize kernel samples, for humidity prediction. The robust MLR models thus constructed were compared with partial least squares (PLS) models on the same data, using piecewise direct standardization (PDS) to correct the spectra recorded on the slave equipment. The prediction errors of the MLR models for the slave equipment are comparable to those of the PLS-SDS models, and slightly smaller than to the errors produced by the MLR-SPA model.

**Keywords:** *Gasoline properties, NIR/MIR Spectrometry, Calibration Transfer.*

Na produção e comercialização, o controle de qualidade das gasolinas é determinado a partir de um conjunto de ensaios físicos e químicos que demandam elevado tempo de análise, requerem equipamentos especiais, grandes volumes de amostra e pessoal treinado para a sua execução. Sendo assim, a idéia de desenvolver novos sistemas que sejam capazes de determinar de forma rápida e simultânea os vários parâmetros de interesse, e que permita algum tipo de classificação para comprovar a qualidade e autenticidade do produto, torna-se atrativa e vantajosa.

A espectroscopia na região do infravermelho próximo (NIR) e médio (MIR) associada à quimiometria tem se mostrado eficiente na determinação de propriedades das gasolinas, de forma rápida, não destrutiva, de baixo custo, e com excelente confiabilidade.

Uma vez que um modelo tenha sido desenvolvido para uma dada aplicação analítica, seria desejável que o mesmo pudesse ser aplicado em outras condições ambientais e instrumentais, bem como em equipamentos diferentes daquele utilizado na etapa de modelagem. Desse modo, poderia ser evitada a duplicação de custo e esforços envolvidos no processo de calibração multivariada. A área conhecida como “Transferência de Calibração” tem por objetivo o desenvolvimento de técnicas de tratamento de dados e modelagem para abordar problemas como os exemplificados acima.

A primeira parte deste trabalho consiste em uma análise de diferentes faixas espectrais (NIR e MIR), pré-processamentos (seleção de variáveis, alisamento e derivada) e técnicas de calibração (PLS e MLR) na construção de modelos de calibração multivariados, com o propósito de selecionar uma estratégia analítica que seja satisfatória para determinar algumas das principais propriedades das gasolinas comercializadas na região Nordeste do Brasil. As propriedades consideradas foram: pontos da curva de destilação, índice de octanagem (MON e RON), densidade e composição (Parafinas, Isoparafinas, Aromáticos, Naftênicos e Olefinas).

A segunda parte propõe uma estratégia para construção de modelos MLR robustos às diferenças entre dois equipamentos, utilizando o Algoritmo das Projeções Sucessivas - APS para seleção de variáveis. Este algoritmo baseia-se na minimização do valor de um erro conjunto que considera não só o conjunto de amostras de teste do equipamento principal, mas também um pequeno conjunto de amostras medidas no equipamento secundário (amostras de transferência). Duas técnicas diferentes são utilizadas para selecionar as amostras de transferência: o algoritmo clássico de Kennard - Stone e a versão modificada do APS para seleção de amostras (APSA). A estratégia foi aplicada a dois conjuntos de dados obtidos por espectroscopia no infravermelho em dois equipamentos diferentes: espectros de gasolinas para prever a temperatura para 90% de amostra destilada (T 90%), e espectros de milho para determinar o teor de umidade.

## **PARTE 1 – PREVISÃO DE PROPRIEDADES DAS GASOLINAS DO NORDESTE DO BRASIL**

### **1.1 - GASOLINAS: TIPOS E PROPRIEDADES**

A gasolina é um combustível derivado do petróleo, constituído basicamente por hidrocarbonetos e, em menor quantidade, por produtos oxigenados, além de compostos de enxofre, nitrogênio e metálicos, todos em baixas concentrações. É um produto volátil, inflamável, tóxico, límpido e isento de material em suspensão, possuindo odor forte e característico.

Inicialmente a gasolina era obtida apenas pela destilação do petróleo, sendo simplesmente a fração mais volátil. Atualmente, vários processos de refino são utilizados na produção da gasolina.

Em geral, a primeira etapa é uma simples separação física (destilação) da qual se aproveita a nafta denominada nafta de destilação. A separação produz várias outras frações, uma das quais, denominada gasóleo, passa por um processo de craqueamento (quebra de moléculas maiores em menores que pertencem à fração da gasolina). Deste processo se obtém a nafta de craqueamento que, adicionada à nafta de destilação, produz a gasolina. Mais tarde, na tentativa de aumentar o rendimento desta extração, surgiram outras técnicas, como polimerização (conversão de olefinas gasosas em moléculas maiores dentro da faixa da gasolina), alquilação (processo que combina uma olefina com moléculas de isobutano, por exemplo) e isomerização. Estes são os procedimentos utilizados pelas refinarias.

Nas indústrias petroquímicas obtém-se a gasolina misturando-se correntes intermediárias do processo, que são dosadas até atingirem as especificações exigidas pela agência reguladora.

A composição da gasolina depende do tipo de processo utilizado para a sua produção. No Brasil, a gasolina é composta basicamente de nafta craqueada e nafta destilada, além de Álcool Etílico Anidro Combustível (AEAC), que é adicionado pelas companhias distribuidoras com teor especificado pela Agência Nacional do

Petróleo, Gás Natural e Biocombustíveis - ANP ( $25 \pm 1$  v/v% - Portaria No. 554 de 27/05/2003).

Quatro tipos de gasolina para consumo em veículos terrestres e aquáticos são definidos e especificados pela ANP:

- **Gasolina Comum "A"**: é a gasolina produzida pelas refinarias de petróleo e entregue diretamente às companhias distribuidoras, isenta de componentes oxigenados.
- **Gasolina Comum "C"**: é a gasolina mais simples encontrada no mercado. Trata-se da gasolina comum "A" recebida pelas distribuidoras, proveniente das refinarias, à qual é acrescentado AEAC. Pode ou não receber algum tipo de corante ou aditivo detergente dispersante, cuja finalidade é minimizar a formação de depósitos no carburador e nos bicos injetores, bem como no coletor e nas hastes das válvulas de admissão, além de serem anticorrosivos. A gasolina aditivada recebe um corante que lhe confere uma cor distinta daquela apresentada pela gasolina comum para evitar possíveis fraudes e adulteração do produto.
- **Gasolinas Premium "A" e "C"**: desenvolvidas pela PETROBRÁS com o objetivo principal de atender aos veículos nacionais e importados que apresentem altas taxas de compressão e alto desempenho. Assim como a gasolina comum C recebe 25% de AEAC nas distribuidoras e pode ou não ser aditivada.

A PETROBRÁS desenvolveu também a Gasolina Podium, recomendada para veículos de altíssimo desempenho, com a mesma tecnologia empregada para produzir a gasolina utilizada na Fórmula 1. Esta gasolina apresenta o maior índice de octanagem de todas as gasolinas, baixo teor de enxofre e reduz o acúmulo de resíduos nos motores, o que possibilita maiores intervalos entre manutenções, apresentando assim o padrão ambiental exigido pela Europa e Estados Unidos.

A gasolina comercial que é destinada a automóveis, motocicletas, pequenas embarcações e aviões de pequeno porte deve possuir as seguintes características:

- 
- Proporcionar queima limpa e sem detonação;
  - Não ser corrosiva aos materiais que compõem a máquina;
  - Ter pouca tendência à produção de resíduos e gomas;
  - Proporcionar partidas fáceis.

Para que estas características sejam atendidas, a gasolina deve apresentar parâmetros de qualidade de acordo com os valores especificados pela ANP. Os principais parâmetros de qualidade descritos nas tabelas de especificações dos Regulamentos Técnicos ANP No. 05/2001 são divididos em análises regulares e análises complementares. Dentro das análises regulares tem-se cor, aspecto, densidade, teor de AEAC, destilação (temperaturas para 10%, 50%, 90% de evaporados, pontos final (PFE) e inicial (PIE) de ebulição e resíduo), MON, IAD, teor de benzeno e teor de hidrocarbonetos olefínicos e aromáticos. As análises complementares são: pressão de vapor, goma atual lavada, período de indução, corrosividade ao cobre, teor de enxofre e de chumbo. Os parâmetros que serão avaliados neste trabalho são descritos a seguir.

#### □ **Densidade – ASTM D 4052 <sup>(1)</sup> /NBR 14065 <sup>(2)</sup>**

Propriedade relacionada com o teor energético total contido em uma determinada massa ou volume do combustível a uma dada temperatura, e que varia com a natureza do petróleo e com o processo que lhe deu origem. Um pequeno volume (aproximadamente 0,7 mL) de amostra líquida é introduzido em um tubo de amostra oscilante do aparelho e a mudança na frequência de oscilação, causada pela mudança na massa do tubo, é usada em conjunto com o dado de calibração na temperatura de trabalho para determinar a densidade da amostra. É útil para transformar volume em massa, ou vice-versa, para cálculos de quantidades na comercialização, bem como para detecção de contaminantes que apresentem densidade diferente. Esta análise leva em torno de 10 min/amostra.

### □ **Destilação – ASTM D 86** <sup>(3)</sup> / **NBR 9619** <sup>(4)</sup>

Determina as características de volatilidade do combustível, verificando se as frações leves e pesadas estão dentro das faixas de valores determinados pela norma. As características de destilação de hidrocarbonetos têm um importante efeito na segurança e desempenho, especialmente no caso de combustíveis. A volatilidade é o principal determinante da tendência de uma mistura de hidrocarbonetos produzir vapores potencialmente explosivos. Afeta o desempenho do motor devido a sua influência sobre o grau de evaporação do combustível nos coletores de admissão e nos cilindros antes e durante a combustão, além de influenciar a partida a frio, proporcionando uma boa mistura do ar com o combustível. Fornece informações sobre a composição, propriedades e o comportamento do combustível durante a estocagem e uso.

A temperatura para 10% de destilados (T 10%) indica a volatilidade relativa para a partida do motor. Quanto menor este valor, menor será a temperatura do motor necessária para a partida. A temperatura para 50% de destilados (T 50%) indica o desempenho de aceleração e aquecimento do motor sob condições de partida a frio. Quanto menor esta temperatura, melhor o desempenho do combustível. As temperaturas para 90% de evaporados (T 90%) e ponto final de ebulição (PFE) indicam a quantidade de componentes com altos pontos de ebulição presentes na gasolina. Altas temperaturas para estes parâmetros (T 90% e PFE) podem contribuir para melhorar a economia do combustível e a resistência à detonação. Entretanto, valores excessivamente altos podem provocar formação de depósitos na câmara de combustão e depósitos de goma no sistema de admissão do combustível. Esta análise requer 100 ml de amostra e dura cerca de 40 min.

### □ **Octanagem (MON e RON) – ASTM D 2699** <sup>(5)</sup> / **2700** <sup>(6)</sup>

A octanagem é um parâmetro que dá uma indicação da qualidade antidetonante da gasolina, estando relacionada com a presença de hidrocarbonetos ramificados e aromáticos. O número de octanas representa o percentual de isoctanos ( $C_8H_{18}$ ) e de

heptanos ( $C_7H_{16}$ ) contidos no combustível. Os combustíveis são normalmente identificados por dois tipos de octanagem: MON (*Motor Octane Number*) que se refere ao uso do motor em altas rotações e RON (*Research Octane Number*), para o uso do motor em baixas rotações. O Índice Antidetonante (IAD) é a média aritmética dos valores das octanagens determinadas pelos métodos MON e RON. O teste é feito em motores especiais (motores CRF – *Cooperative Fuel Research*) montados em base estacionária. O tempo de análise é de 180 min para cada ensaio, que consome entre 800 e 1000 ml de combustível.

#### □ PIANO – ASTM D 1319 <sup>(7)</sup>/NBR 14932 <sup>(8)</sup>

Descreve a composição da gasolina em termos dos seguintes grupos de hidrocarbonetos: parafinas, isoparafinas, aromáticos, naftênicos e olefinas. A concentração e a estrutura das moléculas de hidrocarbonetos que constituem a gasolina são de grande importância para a sua qualidade, uma vez que influenciam significativamente suas propriedades. Os aromáticos, apesar de promover alta octanagem, são bastante tóxicos. As olefinas também apresentam vantagens energéticas, mas participam da formação de ozônio. Nesta análise, que é feita por cromatografia, é consumido 1,0 ml de amostra, e o tempo de análise fica em torno de 150 min/amostra.

## 1.2 - ESPECTROSCOPIA NO INFRAVERMELHO

Espectroscopias vibracionais como Infravermelho Próximo (NIR), Infravermelho Médio (MIR) e Raman, utilizando instrumentos dispersivos ou interferométricos, têm sido extensivamente utilizadas nas duas últimas décadas em diferentes tipos de aplicações analíticas.

A Espectroscopia no Infravermelho Próximo – NIR <sup>(9-11)</sup> é um tipo de espectroscopia vibracional com energias de transição na região delimitada pelos comprimentos de onda 780 e 2500 nm ( $12800$  e  $4000\text{ cm}^{-1}$ ) do espectro eletromagnético.

As ocorrências espectrais nesta região correspondem a sobretons e combinações harmônicas das vibrações fundamentais de ligações C-H, O-H e N-H provocados pela grande anarmonicidade dessas vibrações. Outros efeitos, tais como acoplamentos ou ressonâncias entre diferentes vibrações do mesmo grupo funcional, podem ocorrer.

A ressonância de Fermi que acontece entre uma absorção fundamental e um sobreton quando a diferença de energia entre eles é muito baixa, e a ressonância de Darling-Dennison, resultante da interação entre dois sobretons de nível alto de uma molécula e uma banda de combinação, por exemplo, também podem ser observados, contribuindo para a complexidade do espectro NIR. A ocorrência desses tipos de ressonâncias possibilitará o aparecimento de duas bandas, ao invés de uma, na região de combinação (1600-2500 nm).

Essas ocorrências espectrais são de fraca intensidade, cerca de 100 a 1000 vezes menos intensas que as observadas no infravermelho médio (MIR) e distante (FIR). Entretanto, a diminuição de sensibilidade devida às fracas transições não constitui grande desvantagem, pois já se encontram disponíveis, para uso nesta região, fontes de radiação intensas e detectores de alta eficiência que contribuem para o aumento da relação sinal/ruído. Por outro lado, as baixas intensidades permitem o uso da técnica em amostras com alto teor de água e amostras sólidas.

Uma das vantagens do NIR com relação ao MIR é a utilização de instrumentação de menor custo e robusta, geralmente sem partes móveis e composta de fontes simples, como lâmpadas de tungstênio de alta intensidade. Um espectrômetro NIR pode ser construído com os componentes ópticos empregados em instrumentos que operam na região do Ultravioleta/Visível (UV-Vis). Os detectores mais utilizados são baseados em materiais fotocondutores de PbS e InGaAs, sendo este último dotado de alto poder de detecção e velocidade de resposta elevada. De acordo com a tecnologia utilizada para seleção de comprimentos de onda, os instrumentos NIR podem ser classificados em instrumentos à base de filtros (interferência – Fabri-Perrot e filtro óptico acústico

sintonizável – AOTF), portáteis que empregam LEDs como fontes de luz, dispersivos (rede plana ou côncava de feixe único, feixe duplo, multicanal e multiplexados) e inteferométricos (Transformada de Fourier).

O espectro na região NIR emprega radiação de menor comprimento de onda e, portanto de maior energia, o que pode favorecer uma maior penetração da radiação nas amostras. A informação (espectro) pode ser obtida através de medidas de transmitância/ absorbância (amostras transparentes), transflectância, refletância difusa (amostras sólidas), interactância e transmitância por dispersão média (amostras sólidas densas).

A região NIR foi pouco utilizada até 1950. A principal razão para isto foi atribuída à complexidade dos espectros, que apresentam bandas largas resultantes da sobreposição de picos individuais, o que dificulta sua interpretação por métodos univariados. Sendo assim, somente com a ajuda de métodos multivariados seria possível obter informações quali/quantitativa desses espectros. Com o desenvolvimento das técnicas matemáticas <sup>(12,13)</sup> (algoritmos para pré-processamento dos dados e regressão), a disponibilidade de softwares e o desenvolvimento de novas tecnologias instrumentais, a tecnologia NIR consagrou-se como uma ferramenta para análises rápidas e precisas, aplicada a uma variedade de produtos.

Um grande número de metodologias analíticas desenvolvidas com base em espectros NIR demonstra a capacidade da técnica <sup>(14-20)</sup>. A área agrícola foi a primeira a utilizá-la, sendo um de seus usos mais recentes o apoio à agricultura de precisão <sup>(15)</sup>, onde dados coletados em tempo real são correlacionados com a composição do solo. As indústrias farmacêuticas <sup>(19,20)</sup> utilizam a espectroscopia NIR na identificação de matéria-prima de partida e no monitoramento individual de medicamentos, técnica que já é reconhecida em algumas farmacopéias <sup>(21)</sup>. As indústrias petroquímicas fazem uso do NIR para caracterização do petróleo <sup>(22,23)</sup>, bem como no acompanhamento do processo de produção e controle de qualidade de combustíveis <sup>(24-31)</sup>. Resultados promissores têm sido obtidos também em análises bioquímicas não invasivas, em tomografia óptica clínica e *in-vivo* e em imagem NIR <sup>(32,33)</sup>.

O Infravermelho Médio (MIR) compreende a faixa do espectro eletromagnético que vai de 2500 a 25000 nm ( $4000$  a  $400\text{ cm}^{-1}$ ). As bandas mais intensas são devidas a transições fundamentais. O espectro na região MIR apresenta picos nítidos e estreitos, relacionados essencialmente a vibrações moleculares fundamentais. Esses picos apresentam absorbâncias excessivamente altas, o que pode ser evitado utilizando-se sistemas que possam atenuar estes sinais, por exemplo ATR (*Attenuated Total Reflectance*). Nas indústrias petroquímicas, o MIR encontra aplicação em controle de processos <sup>(30,31,34)</sup>.

### 1.3 – TÉCNICAS QUIMIOMÉTRICAS

A Espectroscopia NIR/MIR não teria alcançado tamanho sucesso, particularmente para determinações quantitativas, sem o uso de métodos quimiométricos. Estes métodos se baseiam em técnicas matemáticas e estatísticas para extrair informações relevantes de dados analíticos, neste caso de dados espectrais na região infravermelha. A Quimiometria auxilia a espectroscopia IR em várias situações, como pré-processamento dos dados espectrais para corrigir, por exemplo, espalhamento de luz, deslocamentos da linha de base, efeitos aditivos ou multiplicativos provenientes da forma de apresentação da amostra; seleção de variáveis que, além de permitir o uso de algoritmos que tornam os modelos de calibração mais simples e de fácil interpretação, encontram comprimentos de onda que podem ser utilizados na construção de instrumentos dedicados; identificação e classificação de amostras devido às bandas de absorção características desta região (C-H, O-H e N-H); calibração multivariada quantitativa, com ferramentas que em geral supõem a existência de uma relação linear entre os dados espectrais e a concentração ou valores de propriedades a serem determinados.

### 1.3.1 - Pré-processamento dos Dados

Pré-processamento é o nome genérico para qualquer manipulação matemática prévia dos dados antes da modelagem multivariada propriamente dita. Muitas vezes o pré-processamento visa remover ou reduzir fontes de variação irrelevantes para o objetivo da modelagem. O pré-processamento pode ser aplicado às amostras ou às variáveis.

Para pré-processar as amostras <sup>(12)</sup> pode-se utilizar a normalização ou atribuição de pesos, que removem variações sistemáticas ao dividir ou multiplicar cada variável de um vetor amostra por uma constante; alisamento, que reduz a quantidade de variação aleatória (ruídos de alta frequência) a fim de aumentar a relação sinal-ruído; e correções de linha de base que reduzem a variação sistemática (ruídos de baixa frequência) capaz de provocar mudanças no sinal de interesse e dominar a análise.

Um dos pré-processamentos mais utilizado é a derivada <sup>(12,13)</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 devida a efeitos 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 isto, antes da diferenciação é comum aplicar-se aos dados algum tipo de suavização. O algoritmo mais utilizado para este fim é o de Savitzky-Golay – SG <sup>(35)</sup>.

Outro pré-processamento muito utilizado é a correção de sinal multiplicativa<sup>(13)</sup> (MSC - *Multiplicative Signal Correction*), onde se faz uma regressão de cada espectro contra um espectro de referência, que normalmente é o espectro médio do conjunto de calibração. Assim os espectros são corrigidos subtraindo-se dos valores originais o coeficiente linear da reta de regressão e em seguida dividindo-se pela inclinação dessa mesma regressão. Para uma dada calibração, a MSC produz dois efeitos importantes: simplifica o modelo de calibração, por que reduz o número de componentes, e melhora a linearidade. Entretanto, por assumir que a diferença entre linhas de base e inclinação é constante em toda a região de

comprimento de onda, esta técnica não traz muitos benefícios para resultados de transferências de modelos de calibração.

As variáveis podem ser pré-processadas através da operação de centrar na média, que corresponde a subtrair o valor de cada elemento do vetor coluna (variável) pelo valor médio dos elementos dessa coluna. A atribuição de pesos (autoescalamento) onde os valores originais em cada coluna são subtraídos das respectivas médias e divididos pelo desvio padrão.

### 1.3.2 - Seleção de Variáveis

Os métodos de seleção de variáveis buscam encontrar variáveis contendo informações relevantes para o problema de interesse a fim de basear a modelagem apenas nestas variáveis. Isto permite que os modelos obtidos sejam mais simples e mais fáceis de interpretar.

Para um modelo de calibração, existem diferentes técnicas para seleção de variáveis <sup>(12)</sup>, dentre elas: seleção do melhor subconjunto baseada na comparação dos erros de predição de todos os modelos possíveis; seleção progressiva (*forward selection*) que seleciona as melhores variáveis aplicando testes F aos sucessivos modelos obtidos; eliminação progressiva (*backward elimination*) que começa com todas as variáveis e usa o mesmo tipo de teste para descartar as menos importantes; procedimentos passo a passo (*stepwise procedures*), que é equivalente a seleção progressiva diferenciando-se apenas pela eliminação prévia de variáveis que não contribuem para o modelo. Estes procedimentos têm como desvantagem a grande probabilidade de que a seleção se prenda a um determinado local ignorando assim a possibilidade de selecionar o melhor conjunto de variáveis.

O Algoritmo Genético <sup>(36)</sup> (AG) seleciona um conjunto de variáveis de forma mais aleatória e menos propensa a soluções locais. Este algoritmo é baseado na teoria da evolução de Darwin, selecionando modelos construídos com combinações de variáveis que fornecem os melhores valores de predição. Cada amostra (indivíduo) é representada por um “cromossomo” cujos genes são

estruturas binárias que codificam cada variável (“1” se a variável foi selecionada e “0” se foi descartada). Desta forma, cada amostra será representada por uma seqüência diferente de genes (diferentes posições de “uns” e “zeros”). Uma população é formada por um determinado número de seqüências diferentes, que são avaliadas a partir da capacidade preditiva do modelo. As melhores seqüências (aquelas cujos modelos forneceram os menores valores de erro para o conjunto de teste) sobrevivem, sofrem mutação e recombina-se de acordo com alguns critérios pré-definidos para formar os descendentes. A mutação é utilizada para incluir variáveis que não estão presentes nos modelos da população inicial (pais). A principal dificuldade é encontrar os valores adequados para os fatores de influência (tamanho da população, grau de mutação, regra de combinação, regra de seleção e número de gerações) que afetam o resultado final.

O algoritmo *jack-knife* <sup>(37)</sup> também pode ser usado para selecionar as variáveis que serão mantidas nos modelos de regressão PLS. Os erros padrão das estimativas dos coeficientes de regressão obtidos por PLS são avaliados de forma empírica a partir de uma série de subconjuntos de dados. A partir daí, pode-se calcular intervalos de confiança para os coeficientes de regressão e avaliar a significância de cada variável utilizada no modelo. O mesmo procedimento pode ser utilizado para regressão em componentes principais (PCR) ou qualquer outro método de regressão.

Finalmente, pode-se selecionar variáveis por meio do recém-proposto Algoritmo das Projeções Sucessivas - APS <sup>(38-43)</sup> (*Successive Projections Algorithm*), que utiliza operações simples para remover colinearidade do conjunto de dados de forma a melhorar o condicionamento numérico e reduzir a propagação de ruídos em um modelo MLR.

Dois aspectos motivam o uso do APS para seleção de variáveis. A redução da colinearidade do conjunto de dados, o que permite a calibração de modelos MLR que são consideravelmente mais simples que modelos PLS/PCR, o que é importante para transferência de calibração, o fato de o APS conduzir a modelos

mais generalizáveis <sup>(39,40)</sup> quando comparado a algoritmos estocásticos de busca como o AG, por exemplo.

O APS pertence a uma classe de métodos de seleção <sup>(44)</sup> que começa por uma determinada variável e segue progressivamente adicionando outras variáveis de acordo com algum critério de seleção, até que uma condição de parada seja satisfeita. Em qualquer estratégia de seleção três aspectos, que serão discutidos a seguir, devem ser considerados: a escolha da variável inicial, o critério de seleção para as variáveis restantes, e a condição de parada (número  $N$  de variáveis a serem selecionadas).

O APS trabalha com base em conjuntos de calibração e teste ambos contendo dados de respostas instrumentais ( $\mathbf{X}$ ) e valores dos parâmetros medidos por um método de referência ( $\mathbf{y}$ ). A principal operação do APS consiste em manipulações algébricas executadas na matriz  $\mathbf{X}_{\text{CAL}}$  ( $K_C \times J$ ), onde linhas e colunas correspondem a  $K_C$  amostras de calibração e  $J$  variáveis espectrais, respectivamente. A partir de uma coluna  $\mathbf{x}_0$ , correspondente à variável inicial da seleção, o APS determina qual das colunas restantes tem a maior projeção no subespaço  $S_0$  ortogonal a  $\mathbf{x}_0$ . Esta coluna, chamada de  $\mathbf{x}_1$ , pode ser considerada como a que contém a maior quantidade de informação não incluída em  $\mathbf{x}_0$ . Na próxima iteração, o APS restringe a análise ao subespaço  $S_0$ , tomando  $\mathbf{x}_1$  como a nova coluna de referência, e procede com os passos descritos. Dessa forma, o critério de seleção do APS favorece a minimização da colinearidade entre as variáveis. Além disso, permite que no máximo  $K_C$  variáveis possam ser selecionadas. Isto porque depois de cada operação de projeção, a dimensão do espaço coluna de  $\mathbf{X}_{\text{CAL}}$  é reduzida de um fator, ou seja, um grau de liberdade é removido. Assim, depois de  $K_C$  operações de projeção todos os vetores coluna de  $\mathbf{X}_{\text{CAL}}$  terão sido projetados sobre o espaço original, isto é,  $\mathbf{X}_{\text{CAL}}$  se tornará uma matriz nula.

Para determinar a melhor variável inicial (coluna de  $\mathbf{X}_{\text{CAL}}$ ) e o número ótimo de variáveis  $N$ , o APS segue o seguinte procedimento: se  $N$  é fixo,  $J$  subconjuntos de  $N$  variáveis podem ser selecionados utilizando cada uma das  $J$  variáveis

disponíveis como ponto de partida. Para cada um desses subconjuntos de variáveis, um modelo MLR é construído e o erro quadrático médio de predição no conjunto de teste ( $RMSEV$ ) é calculado de acordo com a equação:

$$RMSEV = \sqrt{\frac{1}{Kv} \sum_{k=1}^{Kv} (y_v^k - \hat{y}_v^k)^2} \quad (1.1)$$

Onde,  $y_v^k$  e  $\hat{y}_v^k$  são respectivamente os valores de referência e previsto para o parâmetro de interesse na  $K$ -ésima amostra de teste, e  $Kv$  é o número de amostras de teste. O menor  $RMSEV$  assim obtido é denotado por  $RMSEV^*(N)$ , onde o asterisco é usado para indicar o melhor resultado para os subconjuntos das  $N$  variáveis. Repetindo-se este procedimento para  $N = 1, 2, \dots, K_C$ , o  $N$  ótimo pode ser obtido do mínimo da curva do  $RMSEV^*(N)$ . Para economizar tempo computacional, o procedimento pode ser interrompido antes de  $N$  alcançar  $K_C$ , se o analista julgar que o ponto mínimo da curva do  $RMSEV^*(N)$  já está bem estabelecido.

É importante notar que o procedimento APS completo equivale a testar  $J \times K_C$  subconjuntos de variáveis (ou menos, se o procedimento for interrompido em  $N$  menor que  $K_C$ ), o que é exaustivo e demanda tempo computacional. Para comparação, se for feita uma busca exaustiva, o número de combinações de variáveis que serão testadas de forma a cobrir todos os subconjuntos de 1 até  $K_C$  variáveis será dado pela soma de binômios  $(J : 1) + (J : 2) + \dots + (J : K_C)$ . Por exemplo, se  $J = 100$  e  $K_C = 10$ , o APS testará apenas 1000 subconjuntos de variáveis das  $1,9 \times 10^{13}$  combinações possíveis. Um estudo simulado <sup>(49)</sup> revelou que o APS economiza tempo computacional. Neste estudo, o APS inicialmente conduziu a  $RMSEV$  maiores que a busca exaustiva. Entretanto, quando os modelos resultantes foram testados em um conjunto de dados não utilizado no procedimento de seleção, o APS forneceu erros de predição menores. Desta maneira, o melhor poder de generalização do APS pode ser atribuído ao fato de

que o APS considera não apenas a minimização dos erros de predição, mas também a eliminação explícita de colinearidade entre as variáveis. Esta última condição introduz restrições no processo de busca, o que torna o resultado da seleção menos dependente das particularidades do conjunto de teste.

### 1.3.3 - Calibração Multivariada

A Calibração Multivariada <sup>(12,13,45)</sup> é provavelmente a área da Quimiometria que tem atraído mais interesse em aplicações da espectroscopia na região do infravermelho. O processo geral de calibração <sup>(45)</sup> consiste em duas etapas, uma que estabelece o modelo de calibração, obtido relacionando-se a matriz de dados das variáveis medidas (Matriz **X**) com a matriz de dados das propriedades de interesse (Matriz **Y**), e outra que utiliza o modelo estabelecido para estimar propriedades desconhecidas de amostras a partir do seu sinal analítico. O propósito da calibração é estabelecer uma relação matemática entre os dados espectrais e o parâmetro físico ou químico de interesse previamente determinado por uma técnica independente. Uma vez estabelecido um modelo matemático satisfatório, a análise de amostras pode ser feita com base no seu espectro.

Existem muitas ferramentas quimiométricas para fazer uso das informações espectroscópicas para fins de calibração, sendo as mais comuns, a Regressão Linear Múltipla (MLR) <sup>(12,13)</sup>, a Regressão em Componentes Principais (PCR) <sup>(12,13)</sup> e a Regressão por Mínimos Quadrados Parciais (PLS) <sup>(12, 13, 46-48)</sup>.

Na MLR a informação contida na matriz **X** é utilizada diretamente na modelagem do parâmetro de interesse, o que pode levar a problemas de colinearidade <sup>(49)</sup>. Além disto, para que o problema tenha solução matemática, o número de amostras deve exceder o número de variáveis e, portanto, é necessário selecionar um número de variáveis inferior ao número de amostras e que, é claro, sejam relevantes para a predição do valor de interesse.

Os métodos PCR e PLS utilizam em princípio toda a informação espectral e também requerem os valores do parâmetro de interesse (**Y**) no conjunto de

calibração. Esses métodos têm como fundamento encontrar poucas combinações lineares (componentes ou fatores) dos valores originais de  $\mathbf{X}$ , não correlacionadas, e usar apenas estas combinações lineares na regressão. A eliminação da correlação resolve o problema da colinearidade, o que permite obter equações de regressão e predição mais estáveis. Além disto, em PLS, normalmente a parte mais relevante da variação de  $\mathbf{X}$  é usada na regressão.

Os dois métodos são consideravelmente mais eficientes para lidar com ruídos experimentais, e mais robustos, pois seus parâmetros praticamente não se alteram com a inclusão de novas amostras semelhantes no conjunto de calibração<sup>(12)</sup>. Ambos se baseiam em uma Análise de Componentes Principais (PCA)<sup>(12,13)</sup>. Esta consiste numa projeção da matriz de dados com o objetivo de representar as variações presentes em muitas variáveis através de um número menor de fatores. Para tanto, 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, é construído para representar as amostras. 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. Estes apontam as diferenças e similaridades entre as amostras. Cada componente principal é construída pela combinação linear das variáveis originais. Os coeficientes da combinação linear, que indicam quanto cada variável original contribui para uma determinada componente, são denominados *loadings* (pesos). Geometricamente, os pesos são os co-senos dos ângulos entre os eixos das variáveis e a componente principal.

O método PCR é simplesmente a regressão das respostas  $\mathbf{Y}$  nos escores obtidos na PCA realizada na matriz  $\mathbf{X}$ . Como os valores de  $\mathbf{Y}$  não influenciam o cálculo das PC's, o método PCR em geral requer mais componentes que o PLS, principalmente se a informação relevante corresponder a uma variação relativamente pequena em  $\mathbf{X}$ . O PLS contorna este problema incluindo a informação de  $\mathbf{Y}$  na obtenção das

componentes principais, que agora passam a descrever as direções que maximizam a covariância entre os blocos  $\mathbf{X}$  e  $\mathbf{Y}$ . Deixam portanto de ser componentes principais no sentido estrito, embora o termo ainda seja usado por força de costume.

Experiências com dados de NIR/MIR têm confirmado que a regressão PLS pode dar bons resultados de predição com menos componentes que a PCR, porque o número de componentes necessário para interpretar a informação em  $\mathbf{X}$  que é relacionada à de  $\mathbf{Y}$  é menor para PLS que para PCR. Usando o número ótimo de componentes em cada caso, os dois métodos freqüentemente fornecem resultados de predição comparáveis. Na prática, existem casos onde a regressão PLS é melhor, e outros em que PCR é melhor, mas a diferença normalmente não é significativa.

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 complexidade da matriz. Além disso, deve-se considerar a robustez do modelo com respeito às mudanças na instrumentação, condições laboratoriais, ou composição da amostra.

O número de PC's pode ser determinado utilizando-se um conjunto externo de amostras (amostras de teste) ou por validação cruzada (*cross-validation*). Na validação cruzada, uma ou mais amostras são removidas do conjunto de calibração e um modelo é construído com os dados restantes. A amostra ou segmento de amostras que foi removido tem seus valores previstos pelo modelo e um erro de calibração é calculado. Este processo é repetido para cada amostra ou segmento de amostras até que cada uma delas tenha sido excluída em algum momento, e o erro médio quadrático de validação cruzada é calculado (RMSECV). O melhor número de PC's é o que produz o menor valor de RMSECV.

Quando se tem um grande número de amostras é preferível utilizar uma parte dessas amostras para formar um conjunto de teste porque assim pode-se obter uma estimativa mais realista do erro que o modelo irá cometer na predição de novas amostras. Neste caso, o desempenho do modelo é medido pelo erro médio

quadrático de predição para o conjunto de teste (RMSEPT), calculado pela **Equação 1.2**:

$$RMSEPT = \sqrt{\frac{\sum_{i=1}^{N_p} (\hat{y}_i - y_i)^2}{N_p}} \quad (1.2)$$

onde,  $\hat{y}_i$  e  $y_i$  são respectivamente os valores previsto e de referência da concentração ou propriedade do conjunto de teste, e  $N_p$  o número de amostras.

Depois de construído, o modelo deve passar por um processo de validação antes da sua aplicação. Na validação, o modelo é testado prevendo concentrações ou propriedades de amostras que não participaram em nenhum momento da modelagem, e o erro médio quadrático de predição (RMSEP), semelhante ao apresentado na **Equação 1.2**, é calculado para essas amostras. Como critério de aceitação do modelo, este valor normalmente é comparado com o valor da reprodutibilidade do método convencional.

Do exposto, observa-se que a construção, implementação e manutenção de um modelo de calibração multivariada devem passar pelas seguintes etapas <sup>(50)</sup>: seleção das amostras de calibração, determinação das propriedades para essas amostras, obtenção dos seus espectros, cálculo do modelo de calibração usando diferentes técnicas de calibração multivariada e diferentes pré-processamentos, validação do modelo, aplicação do modelo para analisar amostras desconhecidas e transferência ou atualização do modelo de calibração.

## 1.4 – REVISÃO DA LITERATURA

Diferentes técnicas analíticas têm sido utilizadas para desenvolver modelos capazes de correlacionar a composição química com as propriedades físico-químicas das gasolinas.

Van Leeuwen *et al* <sup>(51)</sup> utilizaram técnicas de regressão não-linear e não-paramétricas (redes neurais artificiais – ANNs e *projection pursuit* – PPR) para modelar a relação não-linear entre os grupos PIANO (parafinas, isoparafinas, aromáticos, naftênicos e olefinas), obtidos por cromatografia gasosa, e o número de octanas, encontrando para RON valores de predição próximos à variância inerente ao teste padrão.

Côcco *et al* <sup>(52)</sup> também conseguiram bons resultados ao correlacionar as composições químicas das gasolinas brasileiras, obtidas por cromatografia gasosa, com algumas das propriedades mais importantes (densidade, curva de destilação e pressão de vapor Reid) utilizando redes neurais artificiais. Ultravioleta (UV) <sup>(53)</sup>, Espectrometria de Massa <sup>(54)</sup> (EM), Espectroscopia de Ressonância Magnética Nuclear <sup>(55)</sup> (RMN) e Espectroscopia Raman <sup>(56,57)</sup> também foram utilizadas para obter correlações entre as diferentes propriedades das gasolinas e suas composições.

A espectroscopia na região do infravermelho, especificamente na região do infravermelho próximo <sup>(58,22-27)</sup> (NIR) e médio <sup>(28-31)</sup> (MIR), tem se mostrado eficiente na indústria química e petroquímica, não só para determinar as diversas propriedades de misturas complexas como também para classificar produtos de petróleo <sup>(23)</sup>. As características que tornam esta técnica apropriada para essas aplicações incluem rapidez no registro dos dados (um minuto ou menos por amostra), baixo custo, dispensa do preparo prévio da amostra, além de se tratar de um método não destrutivo, de excelente confiabilidade, de aplicação quase universal (qualquer molécula contendo ligações C-H, N-H ou O-H) e da possibilidade de uso on-line.

Sendo a espectroscopia NIR/MIR uma técnica secundária, em que o resultado é obtido através de medidas indiretas da propriedade, é necessário fazer uso de algoritmos que relacionem o espectro da amostra com as propriedades medidas pelos métodos tradicionais padrão. Desta forma, é necessário construir um banco de dados espectrais obtidos de um conjunto de amostras previamente analisadas pelos métodos tradicionais. A partir deste banco de dados e com o uso de técnicas

quimiométricas são então construídos modelos matemáticos que serão utilizados para a predição das propriedades de interesse.

Muitos pesquisadores têm utilizado com sucesso a espectroscopia NIR/MIR para prever as diferentes propriedades das gasolinas. Kelly *et al*<sup>(58)</sup> utilizaram a região de 660-1215 nm com fibra óptica, denominada região do infravermelho próximo de comprimento de onda curto (SW-NIR), e calibração multivariada (MLR e PLS, com os dados pré-processados por primeira e segunda derivadas, alisamento e correção de linha de base) para prever diferentes parâmetros de qualidade (MON, RON, pressão de vapor Reid, densidade, teor de chumbo, bromo e enxofre, e conteúdo de aromáticos, olefinas e saturados). Os valores de desvio padrão obtidos foram próximos dos valores de referência.

Parisi *et al*<sup>(22)</sup> também utilizaram a região NIR de comprimento de onda curto (600-1100 nm) com fibra óptica para monitoramento remoto de algumas propriedades da gasolina (MON, RON e PONA – parafinas, olefinas, naftênicos e aromáticos) e do índice de cetano no óleo diesel. Durante o processo de medida os pesquisadores se depararam com a presença de bolhas na cela de absorção, que resultou em espalhamento de luz. Este inconveniente foi resolvido por meio de uma correção da linha de base com valores de absorbância a 1100 nm. Para MON e RON os erros absolutos obtidos são comparáveis aos desvios determinados pelos métodos padrão, para PONA os erros relativos foram sempre menores que 4%, e para o índice de cetano em óleo diesel o desvio padrão médio obtido foi comparável à tolerância do método padrão.

Choquette *et al*<sup>(24)</sup> utilizaram FT- NIR e FT- Raman com o método de regressão multivariada PLS para identificar e quantificar o conteúdo de oxigenados em ampolas de gasolinas contendo um ou vários materiais oxigenados (MTBE – *metil terc-butil éter*, ETBE – *etil terc-butil éter*, TAME – *terc-amil metil éter* ou ETOH - *etano*). Os espectros NIR foram obtidos na região de 830-1800 nm e, os espectros Raman cobriram a faixa de 500-3200  $\text{cm}^{-1}$ . O estudo foi dividido em três etapas. No primeiro estudo foi desenvolvido um modelo PLS para cada oxigenado e cada técnica espectroscópica. O

segundo estudo examinou a facilidade de identificar e quantificar oxigenados individualmente com um modelo global e, o terceiro examinou modelos para determinar mistura de oxigenados. Concluiu-se que para quantificar espécies oxigenadas nas ampolas a fim de rejeitá-las caso não atendessem às especificações certificadas, tanto FT-NIR quanto FT-Raman são eficientes. Entretanto, para identificar cada espécie oxigenada presente na mistura, a espectroscopia Raman mostrou-se mais satisfatória. Portanto, um sistema híbrido Raman/NIR pode ser utilizado com sucesso na quantificação destes tipos de amostras.

Guchardi *et al*<sup>(25)</sup> desenvolveram um método para determinar os aditivos etanol e MTBE em gasolinas. Os dados foram coletados em um espectrofotômetro NIR que utiliza como monocromador um filtro acústico óptico sintonizável (AOTF) operando na faixa de 1500-2400 nm. Utilizou-se no cálculo dos modelos a regressão linear múltipla (MLR) com seleção de variáveis pelo Algoritmo Genético (AG). O algoritmo selecionou sete comprimentos de onda o que levou à redução do número de variáveis (277 para 7) e conseqüentemente, à diminuição do tempo de aquisição dos dados (2,5 min para 4 s), além de permitir uma modelagem mais simples, porém tão eficiente quanto um modelo baseado em todas as variáveis. Comparando os resultados obtidos por MLR, em termos de erro médio quadrático de predição (RMSEP), com o método de regressão por mínimos quadrados parciais (PLS) a metodologia proposta forneceu resultados significativamente melhores para determinação de MTBE e comparáveis para etanol.

Bohács *et al*<sup>(26)</sup> utilizaram a espectroscopia NIR também com AOTF como monocromador, e fibras ópticas de alta qualidade juntamente com calibração multivariada (PLS) para determinar 12 diferentes propriedades químicas e físicas das gasolinas (MON, RON, benzeno, MTBE, teor de enxofre, quatro pontos da curva de destilação – pontos inicial e final, 70, 100 e 180 °C, pressão de vapor Reid e densidade). Os espectros foram obtidos na faixa 900-1700 nm com uma média de cem varreduras. Para MON, RON, benzeno e MTBE os erros padrão de predição

foram equivalentes às reprodutibilidades dos testes padrão; para as outras propriedades, foram maiores que os valores de referência.

Boschetti e Olivieri <sup>(27)</sup> optaram por trabalhar com regiões espectrais previamente selecionadas por uma versão modificada do algoritmo genético que atribui pesos a cada faixa de acordo com sua capacidade preditiva para um determinado parâmetro. Os espectros NIR foram obtidos na faixa 1000-2495 nm, com resolução de 12 cm<sup>-1</sup> e média de três varreduras, para prever as propriedades número de octanas, teor de benzeno, tolueno e aromáticos totais em gasolinas. Os modelos foram calculados utilizando-se PLS cujos erros relativos de predição foram comparáveis aos resultados do teste do motor para número de octanas e cromatografia para as demais propriedades.

Um procedimento para determinação direta e simultânea de benzeno, tolueno e MTBE em amostras de gasolina foi proposto por López-Anreus *et al* <sup>(28)</sup>. O método baseia-se na injeção de 1 µl de gasolina em um reator de vidro Pyrex aquecido eletricamente, no qual a amostra é volatilizada a 90 °C. O vapor assim gerado é transportado por meio de um fluxo de nitrogênio para uma cela de gás estreita e o espectro na região do infravermelho médio - MIR é registrado entre 1600 e 500 cm<sup>-1</sup> em função do tempo, com resolução de 4 cm<sup>-1</sup>. Os registros das injeções em fluxo obtidos para bandas nas regiões 671-675 cm<sup>-1</sup>, 727-731 cm<sup>-1</sup> e 1210-1214 cm<sup>-1</sup> foram utilizados para a determinação de benzeno, tolueno e MTBE, respectivamente, com diferentes critérios para correção de linha de base. Os resultados obtidos para as amostras comerciais foram concordantes com os resultados obtidos por cromatografia gasosa e outros procedimentos FT-IR.

Fodor *et al* <sup>(29)</sup> também exploraram a região MIR (4000-650 cm<sup>-1</sup>) para estabelecer um protocolo computacional e experimental que calcula modelos PLS a fim de prever simultaneamente diferentes propriedades das gasolinas (destilação, RON, MON, IAD ((MON + RON)/2), aromáticos, olefinas, parafinas, teor de benzeno, etanol, MTBE e oxigênio total) em até 1 min com menos de 2 ml de amostra. Os espectros foram obtidos com sonda de ATR e média de 32 varreduras.

A precisão do procedimento foi comparável à reprodutibilidade das análises padrão de laboratório.

Os espectros na região MIR apresentam picos com valores de absorbância relativamente altos que podem gerar não-linearidades. Desta forma, Andrade *et al*<sup>(30)</sup> utilizaram além de PLS outros algoritmos (PLS-polinomial, regressão local e redes neurais) capazes de modelar tais influências a fim de desenvolver modelos com capacidade preditiva aceitável para uso no monitoramento da qualidade industrial. Os espectros foram obtidos na região MIR (1600-600  $\text{cm}^{-1}$ ), com resolução de 4  $\text{cm}^{-1}$  e média de seis varreduras para prever a propriedade MON de gasolinas sem procedência e cataliticamente reformadas. Os valores obtidos foram concordantes com outros resultados da literatura obtidos por cromatografia, ressonância magnética nuclear, NIR e outras técnicas analíticas.

Pasadakis *et al*<sup>(31)</sup> empregaram a região MIR para desenvolver modelos de calibração utilizando redes neurais artificiais para prever pontos da curva de destilação e as propriedades frias, ponto de gota (*pour point* - PP) e ponto de névoa (*cloud point* - CP), do óleo diesel combustível. Os espectros foram registrados em toda a faixa MIR (4000-400  $\text{cm}^{-1}$ ) com resolução de 2  $\text{cm}^{-1}$  e média de 20 varreduras, numa célula de transmissão com caminho óptico de 10 m. Foram desenvolvidos dois grupos de modelos de predição. O primeiro tinha 21 modelos ANN, com absorbâncias na faixa 1700-600  $\text{cm}^{-1}$ , para prever a curva de destilação a cada fração de 5% evaporado. O segundo grupo, composto por dois modelos ANN, utilizou a 1ª derivada do sinal para prever as propriedades frias das mesmas amostras. Os erros padrão obtidos pelos modelos ANN para os conjuntos de treinamento e teste foram comparáveis à repetibilidade dos métodos padrão.

Felício *et al*<sup>(34)</sup> avaliaram as vantagens de combinar as espectroscopias NIR e MIR para a predição de RON e benzeno em gasolinas, e ponto de fulgor em óleo diesel. Os modelos foram calculados com diferentes algoritmos PLS (PLS simples, multiblocos, e serial), a título de comparação. Para a região NIR os espectros foram obtidos na faixa 1060-2000 nm utilizando-se uma sonda de transfectância com

caminho óptico de 6 mm, com resolução de  $16\text{ cm}^{-1}$  e média de 32 varreduras. Na região MIR, os espectros cobriram a faixa  $4000\text{-}600\text{ cm}^{-1}$ , e utilizaram sonda de ATR com resolução de  $8\text{ cm}^{-1}$  e 40 varreduras. Os modelos obtidos pelos diferentes algoritmos PLS foram avaliados pelos valores da variância explicada na validação cruzada, RMSEP e intervalos de 95% de confiança. O algoritmo PLS serial forneceu os melhores resultados em todos os casos, seguido de PLS simples com MIR ou NIR. Entretanto, para os parâmetros estudados, o melhor modelo de calibração foi considerando o PLS simples, pois os resultados foram igualmente exatos e obtidos em menos tempo.

Como exposto, a espectroscopia NIR/MIR, empregada juntamente com técnicas quimiométricas, permite monitorar em tempo real diferentes propriedades dos combustíveis, em especial as gasolinas. Entretanto, algumas considerações básicas devem ser feitas a fim de possibilitar a combinação NIR/MIR com calibração multivariada. Macho *et al*<sup>(59)</sup> discutem algumas considerações que ajudam a avaliar a possibilidade de aplicação dessa combinação para monitorar propriedades de interesse industrial. Questões como relação entre a propriedade de interesse e o espectro NIR, amostragem, fontes de erro que afetarão os resultados, validade do modelo ao longo do tempo, etc., são discutidas no contexto da determinação de hidrocarbonetos e densidade em amostras de nafta, e a determinação de etileno e viscosidade em copolímeros de polipropileno.

Cooper *et al*<sup>(60)</sup> compararam as três técnicas de análise vibracionais mais comuns (NIR, MIR e Raman) para a determinação dos componentes benzeno, tolueno, etil benzeno, e orto-, meta- e paraxilenos (BTEX) em gasolinas comerciais. Para cada um dos seis componentes BTEX foi construído um modelo PLS em cada técnica espectroscópica. A validação *leave-one-out* dos conjuntos de treinamento resultou em menores valores de erros padrão para Raman e MIR quando comparados ao NIR, mas em geral, os menores erros padrão foram obtidos utilizando-se a técnica MIR.

Santos Jr. *et al*<sup>(61)</sup> também fizeram um estudo comparativo entre as três técnicas espectroscópicas (NIR, IR e Raman) com transformada de Fourier, utilizando PLS e

redes neurais para predição das propriedades do diesel (índice de cetano, densidade, viscosidade, temperaturas de destilação a 50 e 85% de evaporado, e o teor de enxofre total). Os espectros IR foram obtidos utilizando-se duas células de ATR, uma convencional (ATR1) e uma de imersão (ATR2), tendo a célula ATR1 apresentado melhores resultados. Para densidade, viscosidade e temperatura para 50% de evaporados, os modelos PLS FTNIR e FTIR-ATR1 forneceram valores razoáveis de RMSEP. O modelo PLS/FT-Raman só se mostrou aceitável para a temperatura com 50% de evaporados. Nenhuma das técnicas conseguiu gerar modelos de calibração PLS aceitáveis para determinar o teor de enxofre. Com as redes neurais, os modelos baseados nos espectros FT-Raman apresentaram os melhores desempenhos para todas as propriedades. Para o teor de enxofre, tanto os modelos ANN/FT-Raman quanto ANN/FTIR-ATR1 apresentaram bons desempenhos de predição.

## 1.5 – OBJETIVO

De acordo com o exposto, para determinação das propriedades dos combustíveis, em especial das gasolinas, tem sido utilizado diferentes técnicas de calibração, com diferentes pré-processamentos em várias regiões espectrais do infravermelho. Entretanto, a literatura não reporta nenhuma condição de combinação destes parâmetros que seja capaz de determinar satisfatoriamente as diferentes propriedades das gasolinas. Sendo assim, a primeira etapa desse trabalho considera diferentes combinações de faixa espectral (regiões NIR e MIR), pré-processamentos (derivada, alisamento e seleção de variáveis) e técnicas de calibração (PLS, PCR e MLR) para obter uma condição de modelagem capaz de prever de forma satisfatória, diferentes propriedades das gasolinas comercializadas na região Nordeste.

## 1.6 - EXPERIMENTAL

### 1.6.1 - Conjunto de Dados

Para a construção dos modelos de calibração utilizou-se um conjunto de dados constituído de 160 amostras de Gasolina C comum comercializadas na região Nordeste (BA, SE, AL, PE, PB, RN, CE e PI). O conjunto foi dividido em 100 amostras de calibração, 30 de teste e 30 de predição.

Obtidas as amostras, procederam-se às seguintes etapas:

**1. Determinação dos parâmetros de qualidade (Matriz Y).** As análises foram feitas de acordo com os Regulamentos Técnicos da ANP N<sup>o</sup> 05/01, que determina os métodos padrão (ASTM) de análise. Como já descrito, as propriedades consideradas neste trabalho foram:

- Densidade;
- Destilação (PIE, T10%, T50%, T90% e PFE);
- Octanagem (em termos de MON e RON);
- Composição (PIANO);

**2. Obtenção dos espectros (Matriz X).** Os espectros foram registrados nas regiões NIR, em dois caminhos ópticos diferentes (10 mm: 780 – 1600 nm e 1 mm: 1600 – 2500 nm), e MIR (com sonda de ATR, 2500 – 15400 nm), utilizando-se um Espectrômetro FT-IR Spectrum GX da Perkin Elmer (**Figura 1.1**). Os acessórios utilizados (células de 1 e 10 mm, e sonda de ATR marca Axiom Analytical Incorporated, modelo DPR-210XY) podem ser visualizados na **Figura 1.2**. Os espectros foram obtidos a temperatura média de 22 °C, com resolução espectral de 8 cm<sup>-1</sup> e média de 16 varreduras.

As medidas com as células foram feitas em sistema de fluxo para evitar que a amostra evaporasse durante a medição, além de tornar o processo mais prático e rápido. As amostras eram bombeadas diretamente dos frascos de vidro, com o auxílio de uma bomba peristáltica, até preencher totalmente o volume das células (260 µl para a cela de 1 mm e 2600 µl para a cela de 10 mm). Nesse instante o

bombeamento cessava e a medida era realizada. Em seguida, a amostra era descartada e a célula lavada com a nova amostra a ser medida. A **Figura 1.3** apresenta a bomba e o compartimento onde as células eram fixadas para aquisição do sinal.



Figura 1.1 – Espectrômetro FT-IR Spectrum GX da Perkin Elmer.

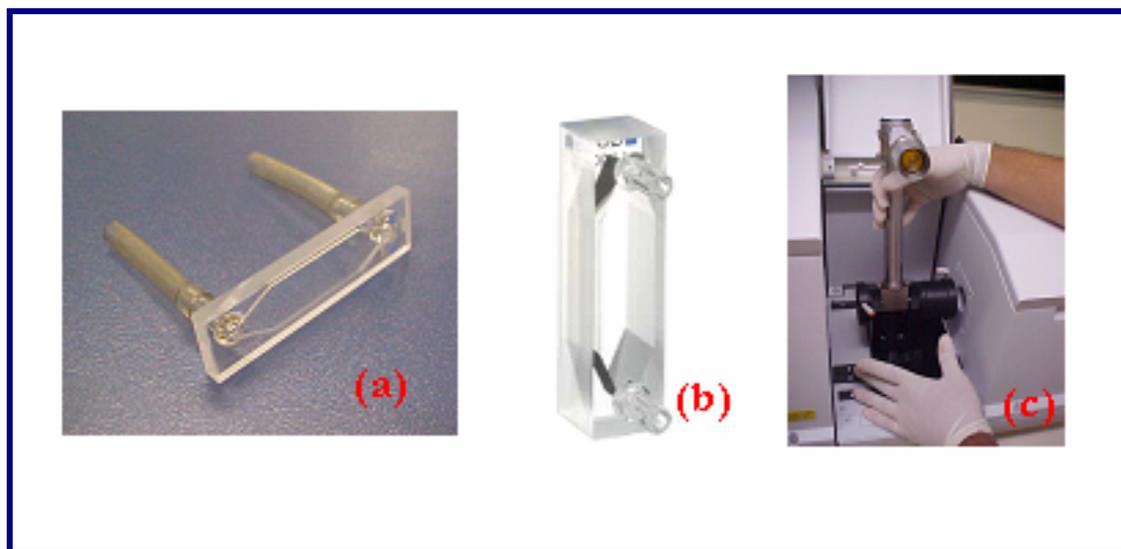


Figura 1.2 – Acessórios utilizados para aquisição dos espectros nas três faixas espectrais: (a) Célula de 1 mm, (b) Célula de 10 mm e (c) Sonda de ATR.

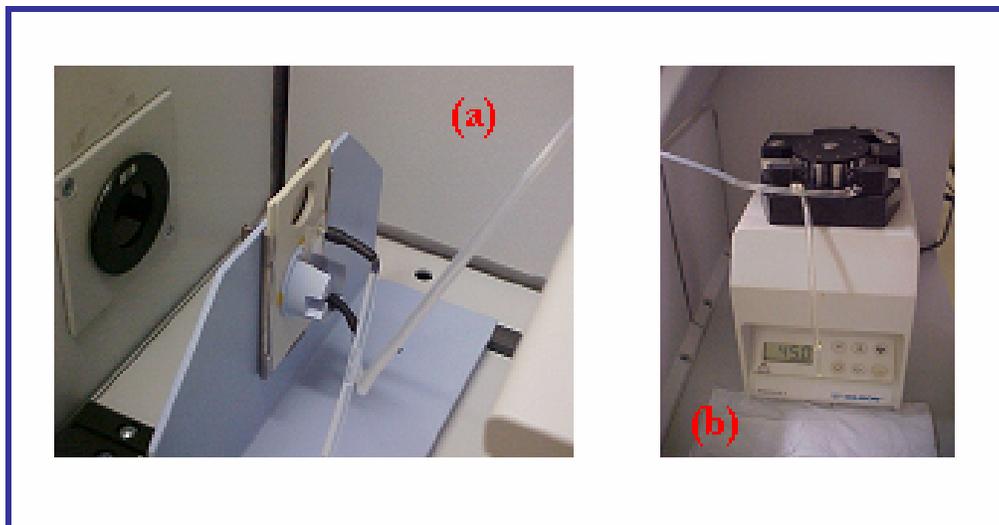


Figura 1.3 – Sistema de fluxo para aquisição dos espectros com as células (a) compartimento para fixação das células, (b) bomba peristáltica.

**3. Construção dos modelos de calibração.** Foram utilizadas 100 amostras na calibração, com as seguintes procedências: AL (6), BA (26), CE (26), PE (16) e RN (26). Os modelos foram obtidos variando-se a técnica de calibração e o tipo de pré-processamento, nas combinações indicadas na **Tabela 1.1**.

Todos os cálculos referentes aos pré-processamentos e às regressões foram feitos utilizando-se o software Unscrambler <sup>(62)</sup>. As seleções de variáveis pelos algoritmos AG e APS foram feitas com o software Matlab <sup>(63)</sup>, através de programas desenvolvidos pelo Profo. Dr. Roberto Kawakami da Divisão de Engenharia Eletrônica - ITA. O alisamento por filtros SG foi feito utilizando janela de 11 pontos ajustada com polinômio de segunda ordem. A seleção de variáveis com o AG considerou uma população de 100 indivíduos, 75 gerações, probabilidade de *cross-over* 60% e de mutação 10% e, assim como para o APS, número mínimo de variáveis a serem selecionadas 5, e máximo de 30.

**Tabela 1.1 – Combinações de técnicas de calibração e pré-processamentos utilizados no desenvolvimento dos modelos.**

PLS e PCR	MLR
Sem pré-processamento (SPP)	AG sem pré-processamento
1ª. Derivada de Norris (DER)	AG com 1ª. Derivada de Norris
Alisamento por Savitzky-Golay (SG)	AG com alisamento SG
1ª. Derivada e Alisamento (DER SG)	AG com derivada e alisamento
Seleção via AG spp (AG SPP)	APS sem pré-processamento (APS SPP)
Seleção via AG der (AG DER)	APS com 1ª. Deriv. de Norris (APS DER)
Seleção via AG sg (AG SG)	APS com alisamento sg (APS SG)
Seleção via AG der sg (AG DER SG)	APS com deriv. e alisam. (APS DER SG)

**4. Validação dos modelos.** Os conjuntos de teste e de predição, cada um deles composto por 30 amostras dos diferentes estados, foram escolhidos observando-se o gráfico dos escores do PLS e levando-se em conta também a faixa de variação de **Y**, a fim de assegurar a representatividade. O critério de aceitação destes modelos foi o valor do RMSEP ser menor que o valor da reprodutibilidade do método ASTM.

**5. Análise de Componentes Principais (PCA).** A combinação da faixa espectral com o pré-processamento e a técnica de regressão resultou em quarenta e oito modelos diferentes. Os valores de RMSEP produzidos por esses modelos foram, submetidos a uma análise de componentes principais, com o objetivo de identificar uma combinação de faixa espectral, técnica de regressão e pré-processamento satisfatória para determinar o conjunto de propriedades de interesse.

## 1.7 - RESULTADOS E DISCUSSÃO

### 1.7.1 - Região NIR

Os espectros na região NIR foram obtidos com dois caminhos ópticos diferentes, 10 mm e 1 mm. Os perfis característicos desses espectros para gasolina C (amostra AL 330) nessas regiões estão apresentados nas **Figuras 1.4 e 1.5**.

Para o caminho óptico de 10 mm, a região informativa do espectro está situada de 780 a 1600 nm. Nela aparecem o terceiro (800 a 1100 nm) e segundo (1100 a 1300 nm) sobretons, além de bandas de sobretons de combinações das vibrações fundamentais (1300 a 1420 nm) de estiramentos C-H, e primeiro sobreton das vibrações de estiramentos O-H (1420 a 1600 nm). Observa-se que a partir de 1650 nm ocorre a saturação do sinal.

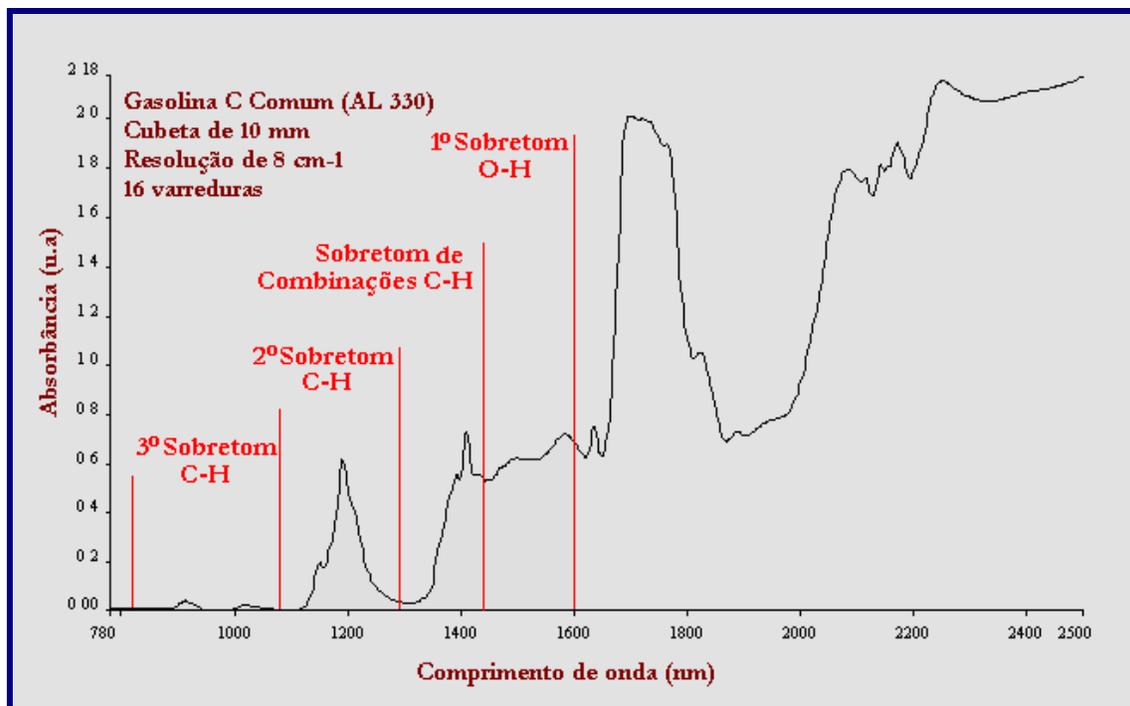


Figura 1.4 – Perfil do espectro para a amostra AL 330 na região NIR com caminho óptico de 10 mm.

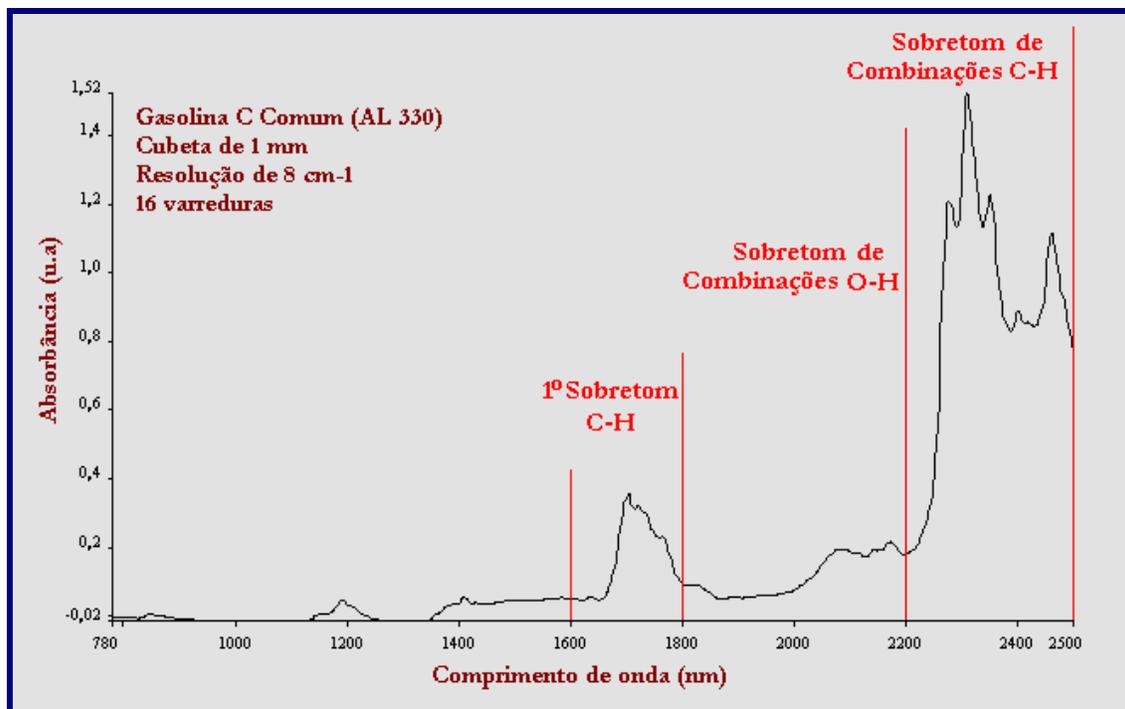


Figura 1.5 – Perfil do espectro para a amostra AL 330 na região NIR com caminho óptico de 1 mm.

A região informativa selecionada do espectro obtido para o caminho óptico de 1 mm compreende a faixa de 1600 a 2500 nm. Nesta faixa observa-se o primeiro sobretom (1600 a 1800 nm) e sobretoms das bandas de combinações das vibrações fundamentais de estiramentos C-H (2200 a 2500 nm) e de O-H (1800 a 2200 nm).

Uma vez obtidos os dados espectrais, efetuou-se uma análise de componentes principais com os dados espectrais de todas as 160 amostras (centrados na média), para um primeiro exame dos padrões de distribuição das amostras. Para o caminho óptico de 10 mm a matriz de dados tem 436 variáveis (780-1650 nm), e para o caminho óptico de 1 mm tem 450 variáveis (1600-2500 nm).

Para o caminho óptico de 10 mm, o gráfico dos escores em PC1 x PC2 é mostrado na **Figura 1.6**. Pode-se observar, ao longo de PC1, não só uma separação parcial entre os conjuntos de amostras provenientes dos vários estados, como também uma estratificação dentro de um mesmo estado. As amostras de Pernambuco, por exemplo, formam em PC1 três grupos bem separados: um com

escores negativos, outro com escores positivos, e um terceiro situado perto da origem.

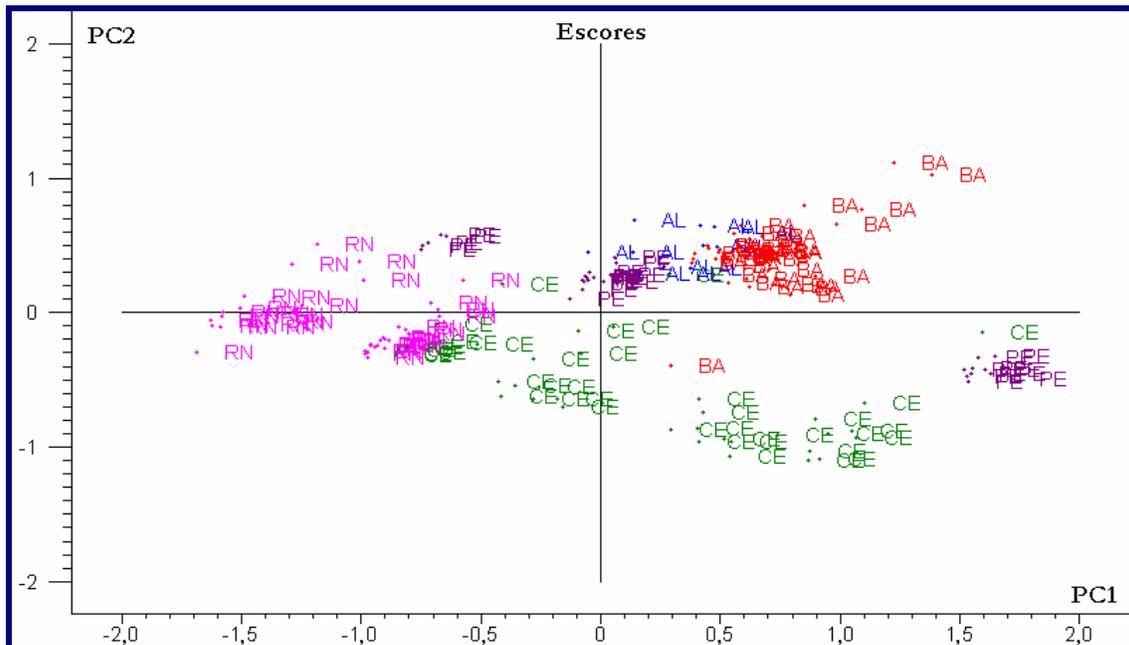


Figura 1.6 – Gráfico dos escores de todas as amostras nas duas primeiras componentes, para a região NIR com caminho óptico de 10 mm. PC1 e PC2 descrevem, respectivamente, 75% e 22% da variância nos dados espectrais.

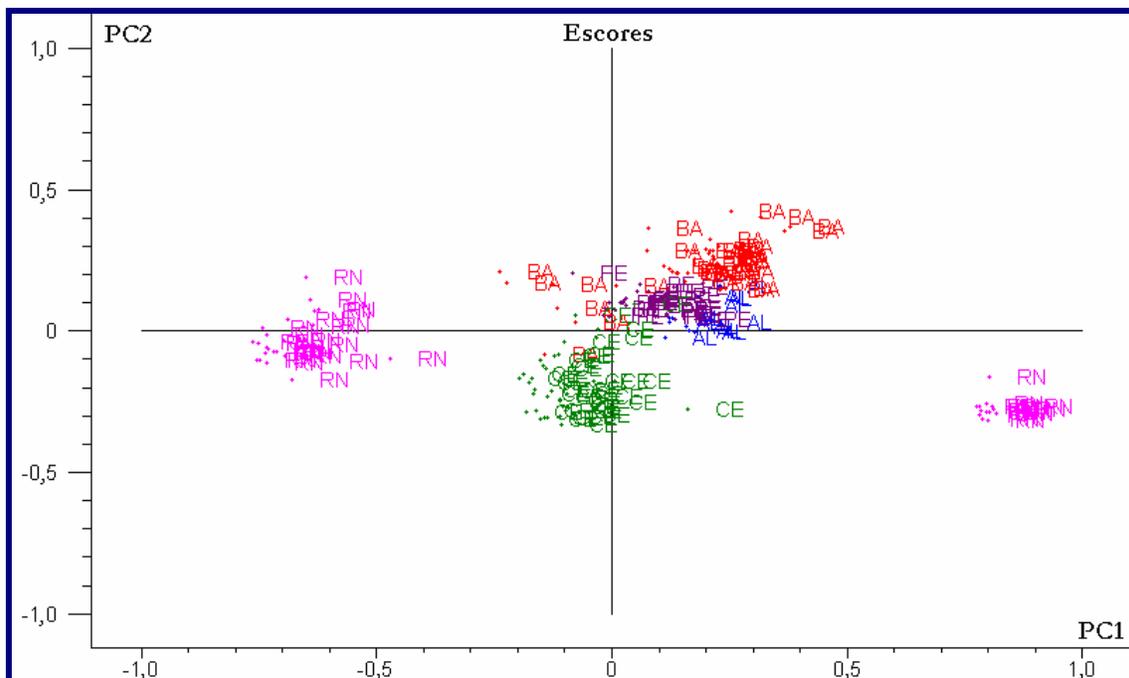


Figura 1.7 – Gráfico dos escores de todas as amostras nas duas primeiras componentes, para a região NIR com caminho óptico de 1 mm. PC1 e PC2 descrevem, respectivamente, 68% e 17% da variância nos dados espectrais.

Para o caminho óptico de 1 mm (**Figura 1.7**), o gráfico dos escores em PC1 é dominado essencialmente por um contraste entre dois grupos de amostras do Rio Grande do Norte, o que força as demais amostras a se concentrar em torno da origem. Tentou-se correlacionar os padrões observados nas duas análises NIR com as treze propriedades do bloco **Y**, mas não foi possível chegar a nenhuma explicação satisfatória para a ocorrência dos agrupamentos observados.

### 1.7.2 - Região MIR

Os espectros foram obtidos na faixa de 2500 - 15380 nm, e seu perfil típico está ilustrado na **Figura 1.8**. Pode-se observar pelo número de picos bem definidos que se trata de uma região bastante informativa (vibrações fundamentais). Observa-se bandas de vibrações das deformações axiais de ligações C-H (2800-3800 nm), vibrações de estiramento de C-H aromático (6500-7500 nm), vibrações de estiramentos das ligações O-H (8800-10000nm) e vibrações das deformações de C-H alifáticos (11000-12000 nm) e olefínicos (12100-15400 nm).

Para esta região, a matriz de dados tem 3222 variáveis. O gráfico dos escores para as duas primeiras componentes da PCA desses dados (**Figura 1.9**) é o que melhor discrimina os estados de origem (ao longo de PC1).

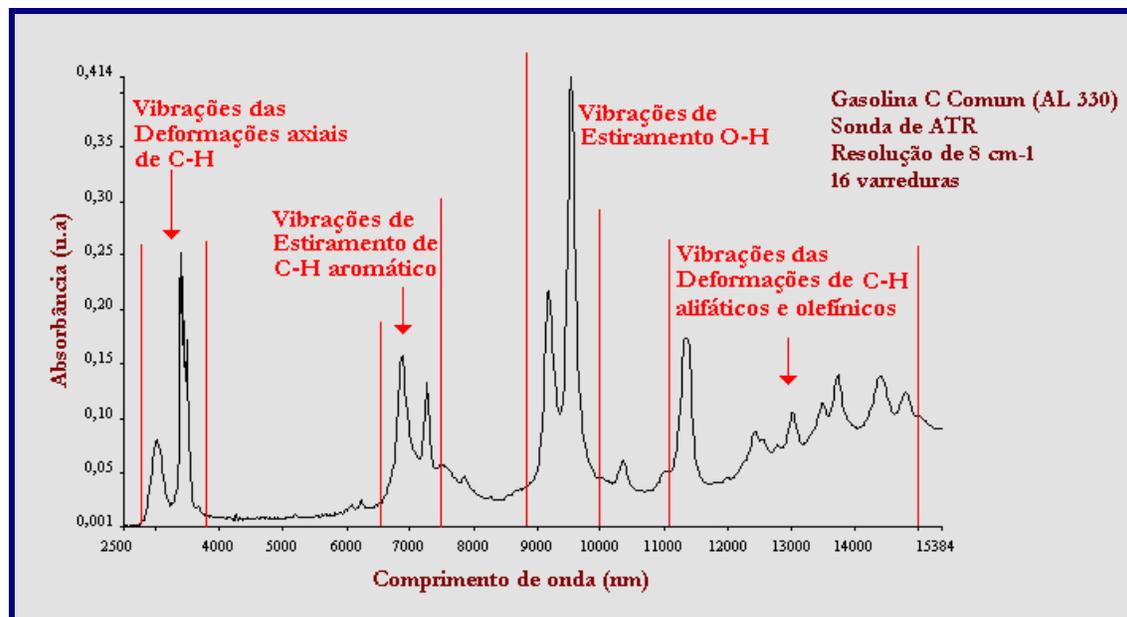


Figura 1.8 – Perfil do espectro para a amostra AL 330 na região MIR.

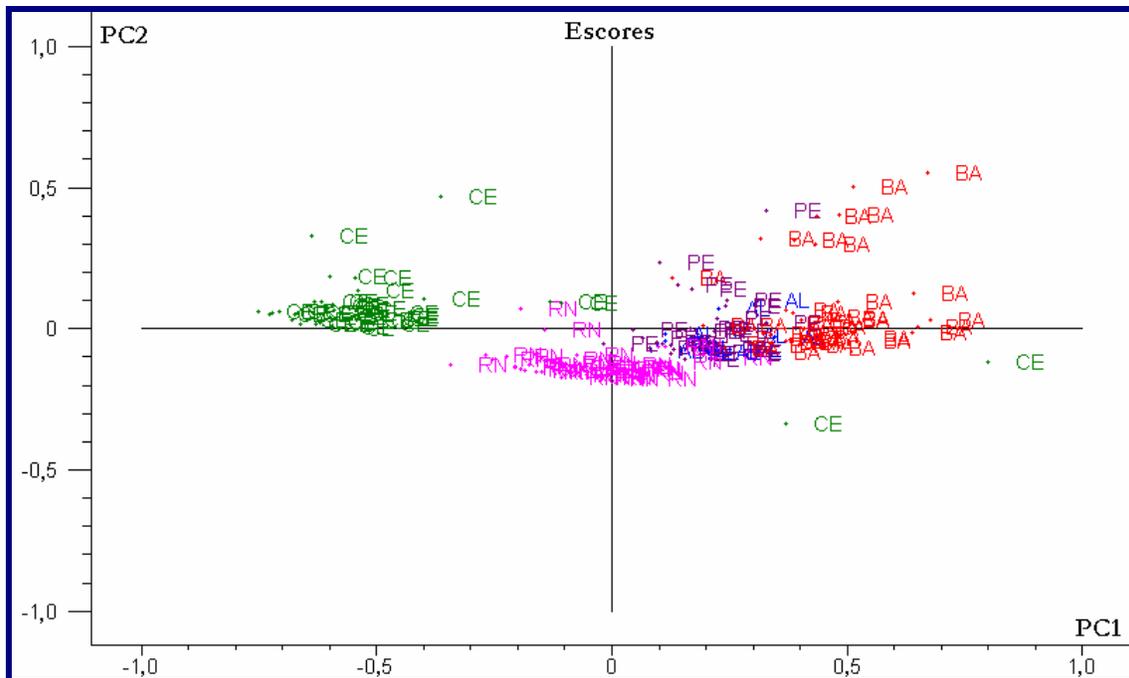


Figura 1.9 – Gráfico dos escores de todas as amostras de gasolinas do NE nas duas primeiras componentes, na região MIR. PC1 e PC2 descrevem, respectivamente, 75% e 10% da variância nos dados espectrais.

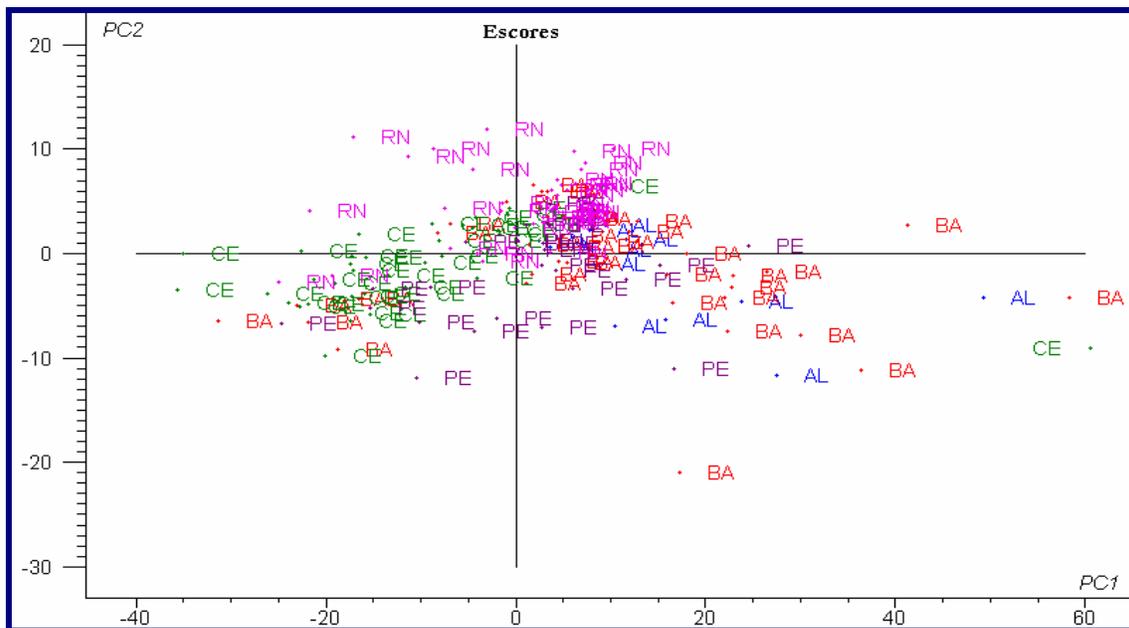
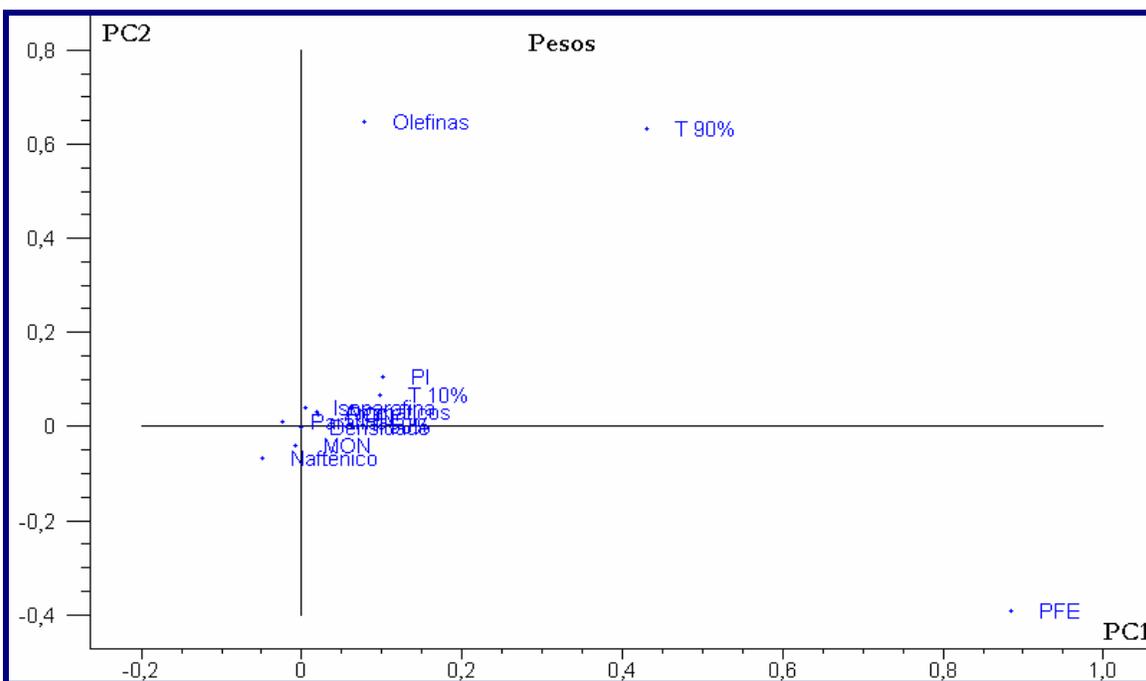


Figura 1.10 – Gráfico dos escores de todas as amostras de gasolinas do NE nas duas primeiras componentes da PCA baseada nos valores das treze propriedades de interesse. PC1 e PC2 descrevem, respectivamente, 75% e 8% da variância nos dados espectrais.

A **Figura 1.10** mostra os escores em PC1 e PC2 para a PCA de todas as amostras, porém realizada no bloco **Y**, e com os dados autoescalados. Embora este gráfico ordene (ao longo de PC1) as amostras com uma ordem semelhante à observada na **Figura 1.9**, a superposição é muito mais pronunciada.

O gráfico dos pesos na **Figura 1.11** mostra que a primeira componente principal explica 75% da informação total dos dados e é composta principalmente pela propriedade PFE. A PC2, representando 8% da variância total, é constituída por Olefinas e T90% (com pesos positivos) e em menor proporção por PFE (com pesos negativos).



**Figura 1.11** – Gráfico dos pesos nas duas primeiras componentes da PCA baseada nos valores das treze propriedades de interesse.

O cálculo dos modelos de regressão nas várias condições citadas (**Tabela 1.1**) forneceu para cada técnica de calibração (PLS, PCR e MLR) vinte e quatro valores de RMSEP (8 valores para cada faixa espectral – 1 mm, 10 mm e MIR), totalizando setenta e dois (72) valores de RMSEP para cada propriedade envolvida (13, no total). Os valores obtidos por PCR não foram significativamente diferentes do PLS, e por isto não serão considerados.

O desempenho de cada modelo variou com a propriedade estimada, o pré-processamento e a técnica de calibração. Para avaliar se existe uma combinação de faixa espectral, técnica de calibração e pré-processamento que possa produzir um modelo de calibração capaz de determinar satisfatoriamente estas treze propriedades foi efetuada uma análise de componentes principais dos valores relativos dos RMSEP's. As **Tabelas A1** e **A2** no Anexo correspondem aos valores de RMSEP's absolutos e relativos, respectivamente.

Para realizar esta PCA, considerou-se as propriedades (13) como sendo as variáveis e as diferentes combinações de faixa espectral, técnica de calibração e pré-processamento (48) como sendo os objetos. Os dados foram apenas centrados nas respectivas médias.

A **Figura 1.12** apresenta o gráfico dos pesos para as duas primeiras componentes, que juntas descreveram 87% da variância dos dados. Pode-se perceber claramente que PC1 e PC2 são determinadas pelas variâncias correspondentes às propriedades identificadas como Parafinas e Naftênicos, respectivamente. Como os pesos dessas duas variáveis nas componentes que elas dominam são positivos, pode-se concluir que as condições mais favoráveis para determinação das propriedades devem aparecer, no gráfico dos escores, no quadrante inferior esquerdo.

As **Figuras 1.13** e **1.14** apresentam os gráficos dos escores no plano PC1 x PC2, com os pontos identificados de acordo com as combinações utilizadas nos cálculos dos modelos. Os resultados mais favoráveis usam a região NIR A (NIR com caminho óptico de 1 mm), não importando muito se a regressão é feita por MLR ou PLS, com seleção de variáveis pelo algoritmo genético. O tipo de pré-processamento também não parece fazer muita diferença. A preferência pela região NIR A não é de surpreender, já que ela apresenta as bandas de combinação dos estiramentos C-H.

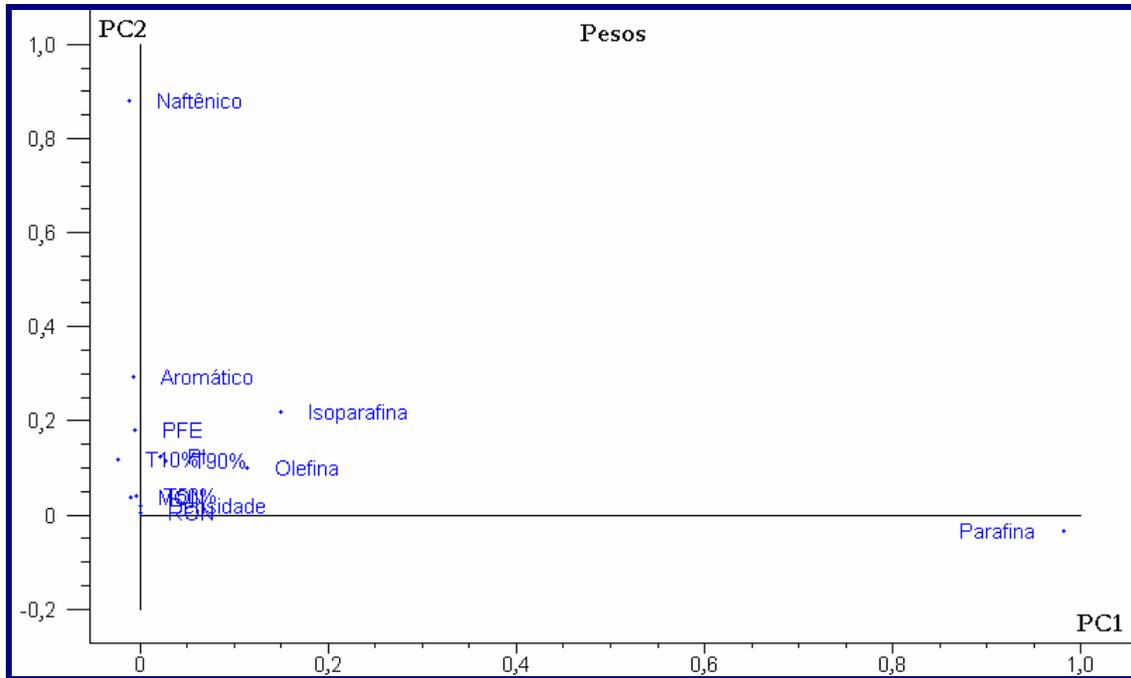


Figura 1.12 – Gráfico dos pesos obtido para os valores relativos de RMSEP das treze propriedades. PC1 e PC2 descrevem, respectivamente, 47% e 40% da variância nos dados.

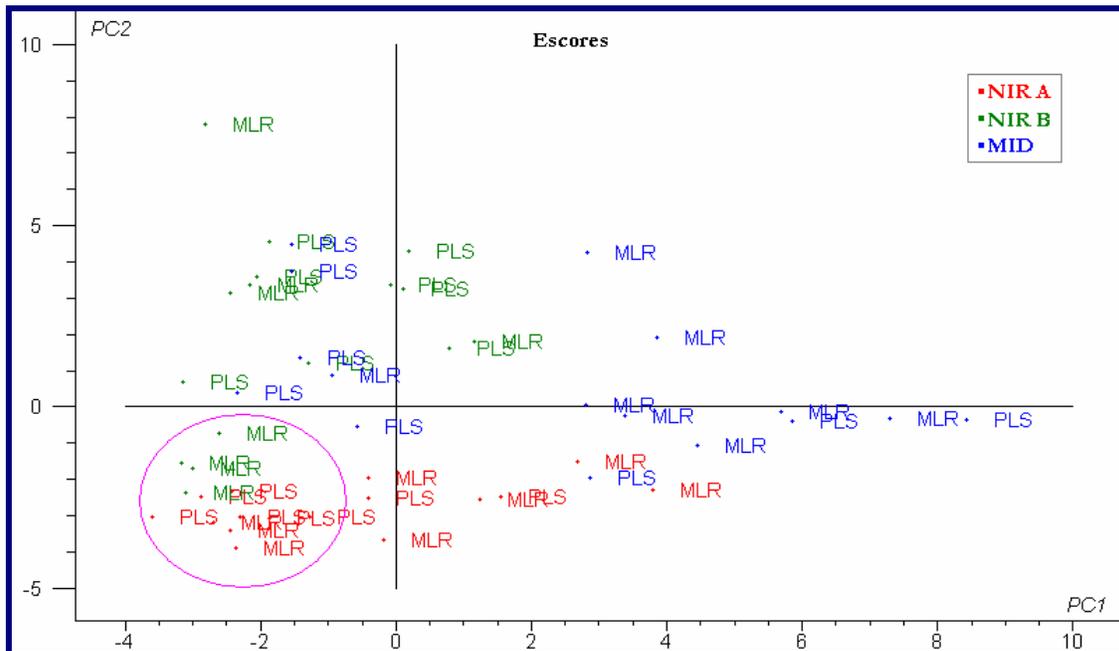


Figura 1.13 – Gráfico dos escores em PC1 e PC2, identificados de acordo com as faixas espectrais e os algoritmos de calibração. PC1 e PC2 descrevem, respectivamente, 47% e 40% da variância nos dados. NIR A equivale a NIR com caminho óptico de 1 mm e NIRB, a NIR com caminho óptico de 10 mm.

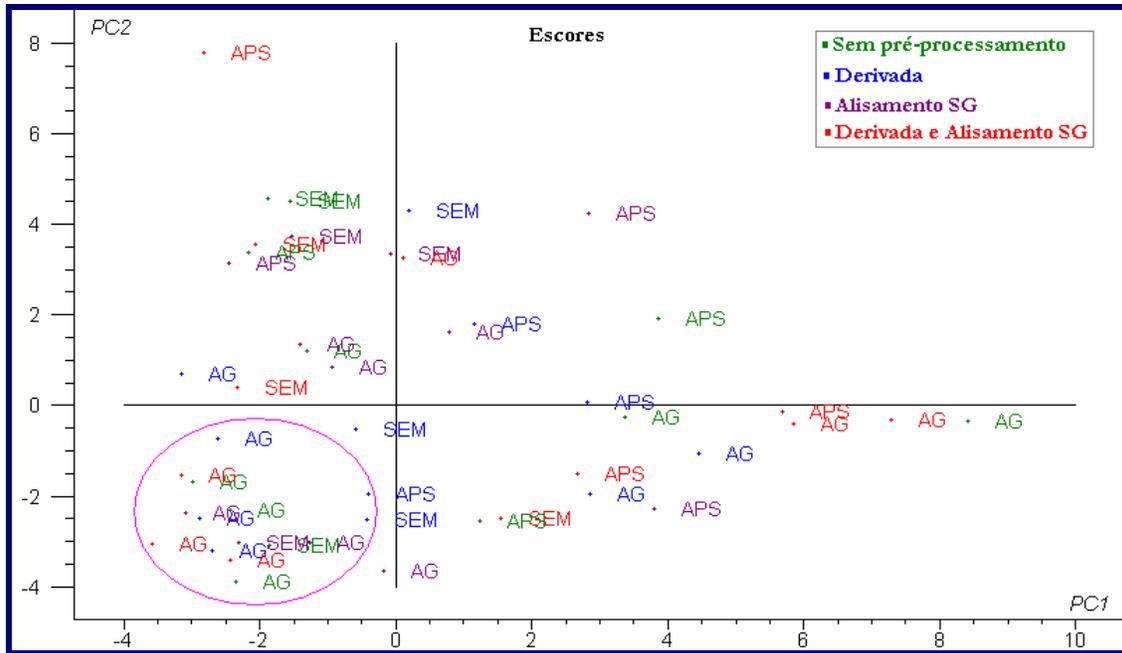


Figura 1.14 – Gráfico dos escores em PC1 e PC2 dos valores relativos de RMSEP identificados de acordo com os pré-processamentos e os algoritmos de seleção de variáveis. PC1 e PC2 descrevem, respectivamente, 47% e 40% da variância nos dados.

Para analisar a influência da metodologia escolhida sobre os erros de predição das outras propriedades, foi realizada uma nova PCA, sem a presença dos teores de Parafinas e Naftênicos.

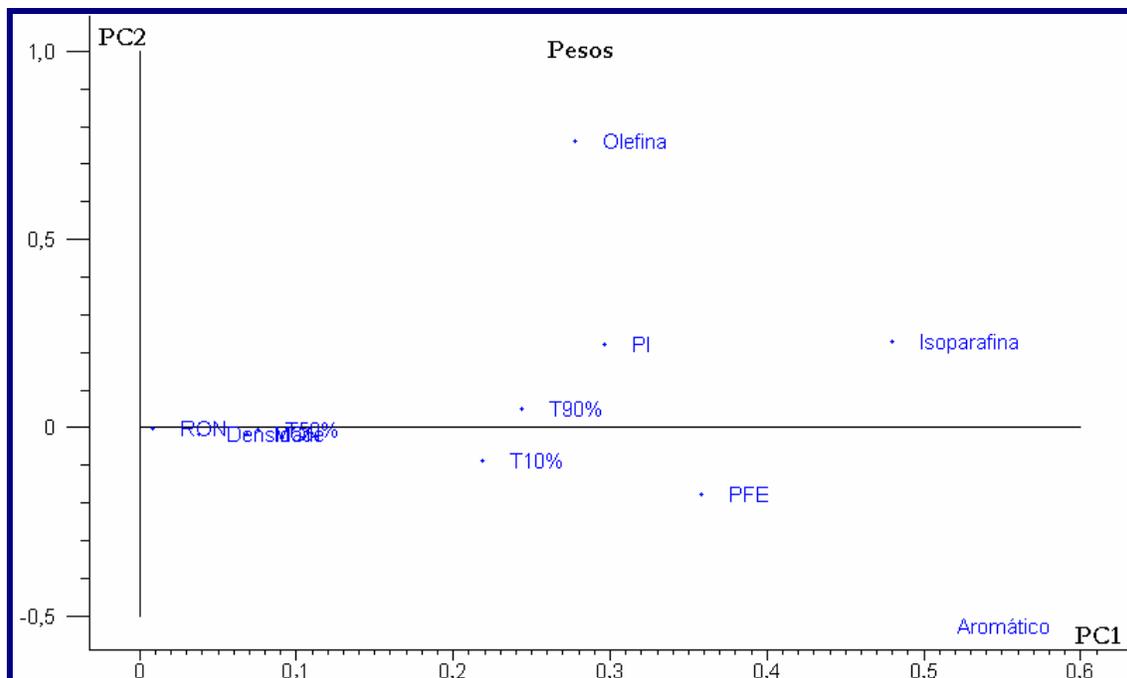


Figura 1.15 – Gráfico dos pesos obtido para os valores relativos de RMSEP das propriedades, excluindo Parafinas e Naftênicos. PC1 e PC2 descrevem, respectivamente, 55% e 17% da variância nos dados.

No novo gráfico dos pesos (**Figura 1.15**), em que PC1 e PC2 descrevem juntas 72% da variância, pode-se observar, ao longo de PC1, três grupos de variáveis: Isoparafina e Aromático, que têm os maiores pesos, PFE, PI, Olefina, T 90% e T 10%, com valores intermediários, e as variáveis restantes, todas com pesos inferiores a 0,1. Como esses pesos são todos positivos, deve-se esperar que as combinações mais favoráveis sejam as que tiverem os escores menos positivos no eixo PC1. Nos gráficos dos escores (**Figuras 1.16 e 1.17**), pode-se observar que as conclusões são bastante semelhantes às relativas a Parafinas e Naftênicos. O tipo de pré-processamento e de técnica de calibração tem importância secundária, ao passo que o uso da região NIR A e a escolha das variáveis pelo algoritmo genético levam a valores de RMSEP menores.

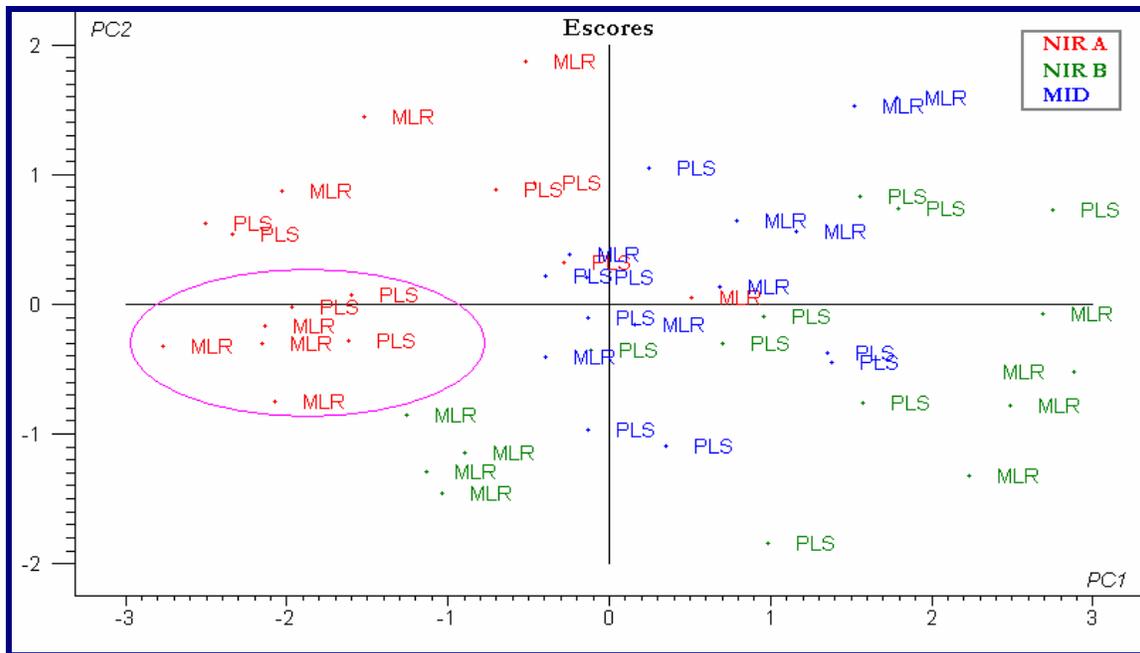


Figura 1.16 – Gráfico dos escores em PC1 e PC2 dos valores relativos de RMSEP, excluindo Parafinas e Naftênicos, identificadas de acordo com as faixas espectrais e os algoritmos de calibração. PC1 e PC2 descrevem, respectivamente, 55% e 17% da variância nos dados.

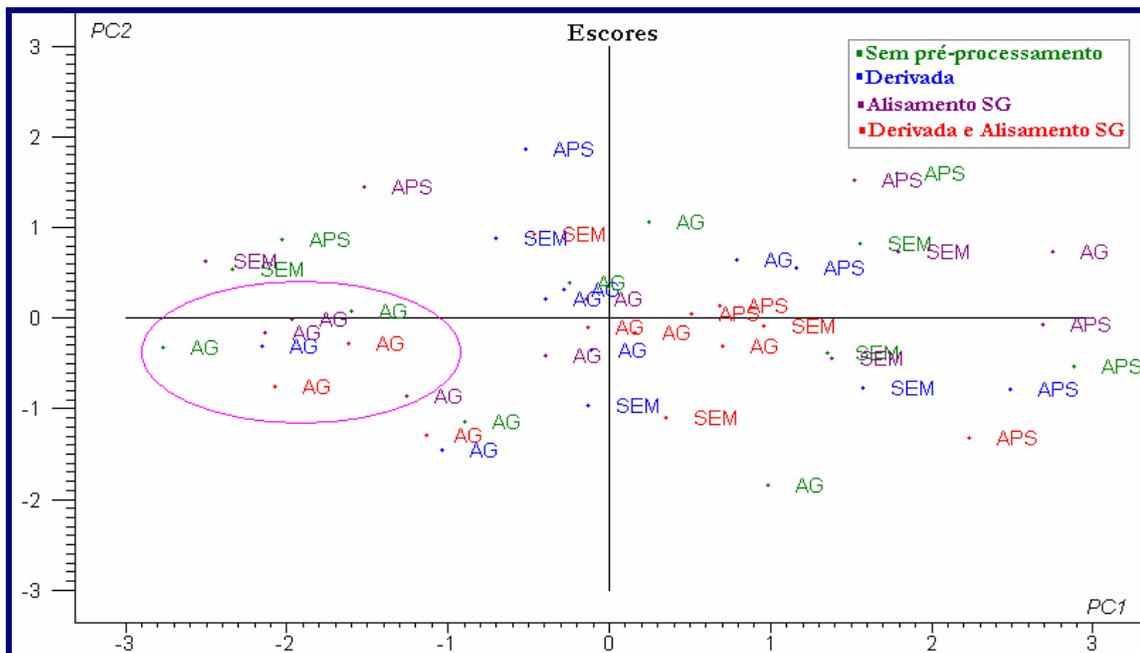


Figura 1.17 – Gráfico dos escores em PC1 e PC2 dos valores relativos de RMSEP excluindo Parafinas e Naftênicos identificados de acordo com os pré-processamentos e os algoritmos de seleção de variáveis. PC1 e PC2 descrevem, respectivamente, 55% e 17% da variância nos dados.

Considerando-se todas essas análises, sugere-se que uma condição de modelagem baseada em MLR, com seleção de variáveis por AG e sem pré-processamento dos dados na região NIR com caminho óptico de 1 mm (NIR A) fornece bons resultados de predição. A **Tabela 1.2** apresenta os valores de RMSEP obtidos nessas condições para todas as propriedades, em comparação com os valores das reprodutibilidades dos respectivos métodos padrão (ASTM).

Estes resultados indicam que todas as propriedades, com exceção da Densidade, apresentaram valores de RMSEP menores que a reprodutibilidade do método de referência. Para Densidade, este resultado é justificável, pois os dados de entrada dos modelos, isto é, os dados obtidos pelos métodos de referência, apresentam valores estimados para repetitividade e precisão intermediária maiores que o valor da reprodutibilidade do método padrão. Comparando-se esses valores de RMSEP àqueles apresentados na Tabela A do Anexo através de teste F, pode-se notar que estes não são estatisticamente diferentes (com 95% de confiança) dos “menores RMSEP’s”. As curvas obtidas dos valores previstos versus observados para amostras do conjunto de predição podem ser visualizadas na **Figura A.1** apresentada nos Anexos. As amostras em geral, apresentam-se bem distribuídas ao longo da reta de regressão não havendo indícios de erro sistemático, e as correlações apresentaram valores entre 0,90 e 0,99.

Tabela 1.2 – Valores de RMSEP para cada propriedade considerando a condição de modelagem mais favorável.

Propriedades	Reprod. ASTM	RMSEP	No. de Variáveis Selecionadas
Densidade (g/ml)	0,0005	0,0009	24
PIE (°C)	8,50	0,85	23
T 10% (°C)	7,00	0,88	17
T 50% (°C)	8,50	0,23	23
T 90% (°C)	9,50	1,72	25
PFE (°C)	10,5	3,67	19
MON	2,00	0,31	24
RON	0,70	0,28	24
Parafinas (%)	3,90	0,66	25
Isoparafinas (%)	ND	0,65	25
Aromáticos (%)	2,35	0,67	20
Naftênicos (%)	1,17	0,22	24
Olefinas (%)	1,94	1,07	12

## PARTE 2 - TRANSFERÊNCIA DE MODELOS DE CALIBRAÇÃO

### 2.1 – REVISÃO DA LITERATURA

A construção e implementação de modelos de calibração multivariada <sup>(50)</sup>, como dito anteriormente, passa por várias etapas cada uma das quais é determinante para as subsequentes.

O uso de um modelo desenvolvido em um equipamento (primário ou principal) com espectros medidos em um outro equipamento (secundário), ou sob novas condições ambientais ou instrumentais exige muito cuidado, devido a diferenças entre as respostas instrumentais de diferentes equipamentos e a mudanças na resposta instrumental ao longo do tempo em um mesmo equipamento. Tudo isto pode levar o modelo a fornecer previsões errôneas.

Um modelo pode tornar-se inválido se variações não incluídas na etapa de calibração forem introduzidas no novo espectro. Existem basicamente três situações que conduzem a esta condição <sup>(64)</sup>: (1) mudanças na constituição química e/ou física das amostras (viscosidade, tamanho de partículas, textura da superfície, etc.); (2) mudanças no ambiente do equipamento: variações de temperatura e umidade, por exemplo, podem provocar deslocamentos das bandas de absorção e mudanças não lineares nas suas intensidades <sup>(65)</sup>; (3) mudanças na função de resposta do equipamento, (o que faz com que o espectro apresente o mesmo perfil, mas tenha intensidades diferentes), que podem ser causadas por substituição de partes do equipamento, envelhecimento da fonte, instabilidade do detector, ou ainda pela aquisição dos espectros em um equipamento diferente.

Caso as fontes de variação sejam identificadas, elas podem ser incorporadas ao modelo utilizando-se novas amostras durante a calibração. Estratégias baseadas em técnicas de planejamento experimental como *Instrument matching* (Equiparação de instrumentos) <sup>(66)</sup> e Atualização dos modelos <sup>(67)</sup> também podem ser adotadas. Tais soluções, que podem ser utilizadas antes da implementação do modelo,

dependem de o analista ter condições de prever futuras fontes de variação e controlar os parâmetros experimentais.

Para evitar a recalibração, ou quando não se consegue identificar as fontes de variação, duas estratégias de transferência têm sido empregadas <sup>(64)</sup>: (a) adaptação do modelo de calibração ou das respostas espectrais (padronização) e (b) aumento da robustez do modelo de calibração, através de pré-processamento dos dados para eliminar informações irrelevantes e aumentar a seletividade para os parâmetros de interesse.

### **2.1.1 - Adaptação do Modelo de Calibração e das Respostas Espectrais - Métodos de Padronização**

Dois equipamentos podem fornecer diferentes respostas instrumentais para a mesma amostra por muitas razões. Se os dois equipamentos forem semelhantes, pode-se distinguir dois tipos principais de problema: deslocamentos de comprimentos de onda e mudanças nos valores de absorbância. Para estes casos, foram concebidos alguns métodos de padronização para corrigir apenas um desses problemas, e outros para corrigir ambos separadamente ou simultaneamente. Quando equipamentos de diferentes tipos são comparados, pode-se esperar todo tipo de modificação dos dados.

Nos métodos de padronização, a função de resposta do equipamento secundário é modificada para tornar-se parecida com a função de resposta do equipamento primário no qual o modelo de calibração foi desenvolvido. Isto pode ser feito através de manipulações matemáticas nas respostas espectrais ou nos valores previstos.

#### **2.1.1.1 - Padronização das Respostas Espectrais**

Estes métodos utilizam medidas espectrais de várias amostras (amostras de transferência) registradas nos equipamentos envolvidos na transferência, e não precisam de seus valores de referência. Se  $\mathbf{x}_p$  (1 x q) e  $\mathbf{x}_s$  (1 x p) são os espectros de

uma amostra de transferência registrados nos equipamentos primário e secundário, respectivamente, a padronização das respostas espectrais procura relacioná-los por meio de uma transformação linear,

$$x_p = b + x_s F \quad (2.1)$$

onde,  $\mathbf{b}$  (1 x q) é um vetor de deslocamentos da linha de base (*offsets*) e  $\mathbf{F}$  é uma matriz de transformação de dimensão (p x q). Os valores de  $\mathbf{b}$  e  $\mathbf{F}$  são estimados a partir de um conjunto de  $n_T$  amostras de transferência cujos espectros são registrados em ambos os equipamentos. Com essas estimativas, pode-se obter, a partir de um espectro registrado para uma nova amostra no equipamento secundário, o seu espectro ajustado,  $\mathbf{x}^a$ , de forma a assemelhar-se ao que seria obtido no equipamento primário:

$$x^a = \hat{b} + x_s \hat{F} \quad (2.2)$$

Se os dois equipamentos forem iguais, a matriz  $\mathbf{F}$  é quadrada. Se, além disto, ela também for diagonal, apenas o elemento de  $(\mathbf{x}_s)_j$  irá contribuir para o cálculo do elemento  $(\mathbf{x}^a)_j$  do novo espectro. Cada elemento do novo espectro terá, portanto uma dependência linear simples em relação ao elemento correspondente do espectro original, ou seja, a matriz  $\mathbf{F}$  ajusta separadamente as absorvâncias a cada comprimento de onda <sup>(68)</sup>. Se os valores não nulos de  $\mathbf{F}$  estiverem localizados numa paralela abaixo da diagonal principal, cada elemento ajustado será afetado pelo elemento correspondente ao comprimento de onda seguinte. Isto corrigirá diferenças de deslocamento de comprimentos de onda se as escalas dos equipamentos estiverem deslocadas por uma mesma distância entre pontos sucessivos. Caso não estejam, o deslocamento será modelado pela matriz  $\mathbf{F}$  com mais elementos não nulos, em paralelas à diagonal principal. Dependendo de sua forma, portanto, a matriz  $\mathbf{F}$  pode corrigir tanto diferenças de absorvâncias quanto diferenças de deslocamentos de comprimento de onda. O ajuste pode ser feito por:

padronização direta (DS), padronização direta em etapas (PDS) ou padronização univariada.

#### 2.1.1.1.1 - Padronização Direta (DS)

Este método foi proposto por Wang *et al*<sup>(69)</sup> para transformar o espectro medido no equipamento secundário de forma que ele parecesse ter sido medido no equipamento primário. As matrizes das respostas obtidas em ambos os equipamentos para as amostras de transferência relacionam-se pela matriz de transformação  $F$ .

$$X_p = X_s \cdot F \quad (2.3)$$

Sendo  $X_p$  e  $X_s$  as matrizes das respostas obtidas para as amostras de transferência nos equipamentos primário e secundário, respectivamente.

Se os dois equipamentos forem iguais, a matriz de transformação é quadrada e suas colunas descrevem, para os respectivos comprimentos de onda, as correlações entre as absorvâncias dos espectros primários e secundários. Se  $X_s$  for invertível,  $F$  pode ser dada por:

$$F = (X_s)^{-1} \cdot X_p, \quad (2.4)$$

mas isto raramente ocorre porque, como costuma haver menos amostras que variáveis,  $X_s$  é retangular. Nesses casos,  $F$  pode ser calculada utilizando-se métodos multivariados tais como regressão por componentes principais (PCR) ou mínimos quadrados parciais (PLS) e  $X_p$  e  $X_s$  devem ser projetados sobre o espaço coluna de  $X_p$ , calculado por decomposição em valores singulares. Os escores obtidos pelas projeções são empregados no lugar das matrizes das respostas originais, tanto na padronização como na posterior calibração.

Como já comentado no item anterior, a padronização do novo espectro é obtida a partir da matriz  $F$  calculada através da relação:

$$x^a = x_s \cdot F \quad (2.5)$$

Assim, o vetor resposta de uma amostra desconhecida, medida no equipamento secundário,  $x_s$ , é padronizado para o vetor resposta esperado no equipamento primário,  $x^a$ .

Este método apresenta duas desvantagens. Em primeiro lugar, as amostras de transferência precisam ser analisadas nos dois equipamentos exatamente nas mesmas condições, pois deve-se garantir que a matriz  $F$  represente verdadeiramente diferenças instrumentais e não mudanças na composição química das amostras. A segunda desvantagem é que o número de amostras de transferência é sempre muito menor que o número de constantes para ajustar o espectro inteiro, o que pode levar a sobreajuste (*overfitting*) na estimativa de  $F$ .

#### 2.1.1.1.2 - Padronização Direta em Etapas (PDS)

Este método, também proposto por Wang *et al*<sup>(69)</sup>, admite que as correlações espectrais sejam limitadas a pequenas regiões. Na padronização DS todo o espectro do equipamento secundário é utilizado para ajustar cada comprimento de onda do equipamento primário, mas na padronização PDS cada comprimento de onda primário é relacionado a comprimentos de onda contidos em uma pequena região do espectro obtido no equipamento secundário, utilizando-se uma janela espectral móvel em torno de um determinado comprimento de onda. Isto minimiza o risco de sobreajuste, já que o número de constantes a serem calculadas para estimar  $F$  é reduzido.

No método PDS, a relação entre as matrizes das respostas para as amostras de transferência também é dada pela **Equação 2.3**, porém para cada janela calcula-se um vetor de regressão  $b$ , por PCR ou PLS. Estes vetores são então dispostos em uma matriz diagonal  $F$ , dada por:

$$F = \text{diag}(b_1', b_2', \dots, b_j', \dots, b_k') \quad (2.6)$$

onde  $k$  é o número de comprimentos de onda. Dessa forma, qualquer nova amostra pode ser padronizada através da **Equação 2.5** como se estivesse sendo medida no equipamento primário.

A largura da janela móvel deve ser otimizada, o que pode ser feito comparando-se as posições dos picos e larguras de banda dos espectros obtidos nos dois equipamentos. A janela pode ser simétrica, o que é desejável quando os deslocamentos de comprimentos de onda são desprezíveis, ou assimétrica, quando há deslocamentos significativos de comprimentos de onda em torno de cada variável. A janela móvel pode ocasionar efeitos de borda, que podem ser corrigidos truncando-se as extremidades do espectro ou estimando-as por extrapolação.

O método PDS é empregado como método de referência, por permitir correção simultânea de diferenças de absorbância, deslocamentos de comprimentos de onda e alargamento de picos. Algumas modificações do algoritmo original já foram propostas, dentre elas a correção de diferenças aditiva entre espectros <sup>(70)</sup>, que é semelhante a incluir um termo de *offset* na equação do PDS e a correção de variações contínuas de temperatura <sup>(65)</sup>. No entanto, o PDS apresenta problemas para estimar os postos locais dos modelos PCR/PLS, pois o número de autovalores retidos é pré-determinado e baseado em valores de referência. A dificuldade para estimar os postos locais é atribuída a ruídos de alta frequência nos dados, que podem ser amenizados pré-processando-se o espectro antes da transferência. Para isto, Wang *et al* <sup>(71)</sup> utilizaram derivadas, Chen *et al* <sup>(72)</sup> transformadas de Fourier, Gemperline *et al* <sup>(73)</sup> recorreram a uma PCA para decompor e reconstituir o espectro apenas com fatores relevantes, e Tan e Brown <sup>(74)</sup> combinaram decomposição em coeficientes *wavelet* (coeficientes de aproximação – baixa frequência, e de detalhes - alta frequência) com DS e PDS para transferência.

### 2.1.1.1.3 - Padronização Univariada

Este método, proposto por Shenk e Westerhaus <sup>(75)</sup>, inicialmente destinava-se a ajustar diferenças de intensidades, e admitia que não houvesse deslocamento dos

comprimentos de onda entre os equipamentos. Posteriormente, introduziu-se uma correção para essa possível diferença, e o novo método foi patenteado <sup>(76)</sup>. Inicialmente, cada comprimento de onda  $i$  do equipamento primário é correlacionado aos comprimentos de onda contidos em uma janela em torno de  $i$  no equipamento secundário. Uma função quadrática é ajustada a esses dados, sendo o seu máximo,  $j$ , equivalente ao comprimento de onda correspondente do equipamento primário. Finalmente, a intensidade espectral é ajustada através de uma equação linear:

$$x_{iP} = a_i + b_i \cdot x_{jS}, \quad (2.7)$$

onde,  $\mathbf{x}_{i,p}$  é o vetor resposta do comprimento de onda  $i$  no instrumento primário;  $\mathbf{x}_{j,s}$  o vetor resposta do comprimento de onda do secundário equivalente ao primário, e  $a_i$  e  $b_i$  são respectivamente, *offset* e inclinação, estimados a partir das amostras de transferência.

Bouveresse *et al* <sup>(77)</sup> propuseram uma modificação na correção de intensidades, para o caso de as amostras de transferência não serem da mesma natureza que as amostras a serem analisadas. No método proposto por Forina *et al* <sup>(78)</sup>, as diferenças de comprimentos de onda são corrigidas ajustando-se a função quadrática às distâncias Euclidianas em vez dos coeficientes de correlação, que são mais sujeitos a ruídos.

Com qualquer um desses métodos, a padronização univariada corrige diferenças de comprimentos de onda e de intensidades, mas não alargamentos de picos e outros efeitos, que exigem uma correção por métodos multivariados.

### 2.1.1.2 - Padronização dos Valores Previstos

O ajuste dos valores previstos é comumente feito por uma transformação linear e univariada conhecida pelo acrônimo SBC <sup>(79)</sup> (*slope/bias correction*). As amostras de transferência têm suas propriedades  $y$  medidas e previstas em ambos os equipamentos utilizando-se o modelo de calibração desenvolvido no

equipamento primário, e os resultados são usados para ajustar a equação de regressão,

$$y_P = a + b \cdot y_S, \quad (2.8)$$

cujos coeficientes linear ( $a$ ) e angular ( $b$ ) corrigem respectivamente deslocamentos da linha de base e de inclinação devidos à transferência. Esta correção só funciona bem se os equipamentos envolvidos forem idênticos. Caso contrário, as diferenças entre as respostas instrumentais serão mais complexas e será necessário recorrer a outros métodos de padronização, como o PDS, por exemplo. Bouveresse *et al* <sup>(79)</sup> desenvolveram um procedimento para determinar quando a correção SBC pode ser empregada, usando um teste F para comparar as variâncias residuais obtidas nos equipamentos.

### 2.1.2 - Aumento da Robustez do Modelo

No desenvolvimento de um modelo de calibração, o critério de robustez é crítico. Um modelo de calibração é considerado robusto quando a resposta de novas amostras não é significativamente afetada por variações não calibradas. A robustez de um modelo de calibração pode ser influenciada pela escolha dos equipamentos envolvidos na transferência, pelo pré-processamento dos espectros e pela técnica de calibração.

#### 2.1.2.1 - Modelos Globais

Para se obter um modelo global, é necessário incluir na modelagem todas as variações esperadas (por exemplo, preparação das amostras, nas condições ambientais e nos parâmetros instrumentais) em uma larga faixa de condições experimentais, a fim de minimizar ou até mesmo evitar variações não calibradas. Para tanto, o analista deve ser capaz de prever as possíveis fontes de variações, bem como a extensão da influência dessas fontes nos dados. Como isto é difícil de determinar, os modelos globais normalmente têm grandes dimensões e, portanto

são mais complexos que os modelos locais, além de apresentarem predições menos precisas <sup>(67,80)</sup>. Wulfred *et al* <sup>(81)</sup> estudaram a influência dos efeitos da temperatura nos espectros e nos modelos de calibração multivariado. Concluíram que esses efeitos conduzem a relações não lineares entre as concentrações e as absorbâncias. Tais efeitos podem ser minimizados com modelos PLS ou PCR, ou por técnicas de modelagem não linear <sup>(82)</sup>.

### 2.1.2.2 - Pré-processamentos

Os métodos de padronização são úteis quando as amostras de transferência podem ser medidas em todos os equipamentos envolvidos na transferência do modelo de calibração, o que não é problema quando os equipamentos estão próximos e as amostras são quimicamente estáveis. Quando isto não é possível, pode-se aplicar métodos de pré-processamento para tentar remover variabilidades indesejáveis no espectro, causadas por efeitos de espalhamento ou variações entre equipamentos, sem que a informação química seja afetada. Os pré-processamentos mais utilizados são as derivadas <sup>(34,35)</sup> e a correção de sinal multiplicativa - MSC <sup>(13)</sup> discutidos na Seção 1.3.1.

Os filtros de resposta de impulso finito <sup>(83)</sup> (FIR – *Finite Impulse Response filters*) foram propostos como método de transferência de calibração para o caso de as diferenças aditivas e de inclinação variarem com o comprimento de onda. Neste método, análogo ao MSC por etapas, escolhe-se um espectro de referência, que pode ser o espectro de uma determinada amostra, ou o espectro médio de todas elas, e ajustam-se os demais espectros de forma a se assemelharem com o espectro de referência escolhido. O ajuste é feito utilizando-se uma janela móvel que varre todo o espectro e ajustando-se o comprimento de onda no centro da janela da mesma forma que no MSC.

O método FIR tem sua robustez limitada por introduzir variações espúrias no espectro transferido. Tais variações surgem do cálculo dos modelos de regressão local, onde as correlações entre os espectros da amostra e o de referência são quase

linearmente independentes. Tan *et al*<sup>(84)</sup> propuseram uma melhoria no método FIR, que torna a transferência mais robusta e produz espectros livres de variações espúrias. Tanto o FIR quanto o MSC têm aplicação limitada devido à perda de informação química que ocorre quando os espectros das amostras são forçados a se tornarem parecidos com um espectro de referência.

A correção de sinal ortogonal<sup>(85-88)</sup> (OSC, de *Orthogonal Signal Correction*) que foi desenvolvida para reduzir o número de fatores em regressões multivariadas, passou a ser também utilizada para transferir calibrações entre instrumentos, por evitar perda de informação durante o pré-processamento dos espectros<sup>(89)</sup>. O objetivo da OSC é reduzir a variação em  $\mathbf{X}$  que não esteja relacionada (isto é, seja ortogonal) a  $\mathbf{Y}$ , simplificando assim o modelo. Na transferência de calibração, os vetores ortogonais a  $\mathbf{Y}$  e comuns aos equipamentos envolvidos são removidos o que torna o modelo mais transferível. Quando utilizada com calibração multivariada, a OSC é calculada apenas com amostras de calibração. Para fins de transferência, a correção deve ser calculada para um conjunto de amostras medidas nos equipamentos envolvidos, a fim de permitir que todas as fontes de variação sejam consideradas<sup>(89,91)</sup>.

Um modelo OSC construído exclusivamente com amostras medidas no equipamento primário dificilmente será válido para a correção de amostras medidas no equipamento secundário, já que variações espectrais entre equipamentos costumam ser significativas. As amostras selecionadas para calcular o modelo OSC de transferência não precisam ser medidas em ambos os equipamentos, mas devem ser representativas das variações instrumentais.

A Transformada *Wavelet*<sup>(92)</sup> (WT) tem sido usada como método de processamento alternativo para transferência de calibração<sup>(93,94)</sup>. Na WT os sinais originais são decompostos em coeficientes *wavelet* com diferentes frequências, o que permite a remoção de certos valores de frequência. Em geral os coeficientes de baixa frequência estão relacionados a diferenças instrumentais, que correspondem à informação irrelevante para a análise. Além de ser mais versátil e dispensar o uso de

amostras de transferência, a transformada *wavelet* é útil na remoção de variações de linhas de base <sup>(95)</sup> entre os conjuntos de calibração e teste, e contribui para melhorar a relação sinal/ruído. Andrew e Fearn <sup>(96)</sup> propuseram recentemente um método chamado de Transferência por Projeções Ortogonais (TOP, de *Transfer by Orthogonal Projections*), que utiliza informação sobre a variação espectral entre instrumentos para fazer uma calibração robusta e transferível. Para tanto, os espectros de um pequeno conjunto de amostras são registrados em  $n$  equipamentos, e sua média é calculada para se obter um espectro médio para cada equipamento. Em seguida, efetua-se uma análise de componentes principais da matriz formada pelos espectros médios obtendo-se  $n-1$  componentes, que representam as principais direções de variação entre equipamentos no espaço espectral. A matriz dos dados espectrais para o conjunto de calibração  $\mathbf{X}$  é ortogonalizada contra algumas ou todas essas componentes, removendo-se certo número de dimensões  $d$  que melhor descrevem a variabilidade entre equipamentos. A escolha do valor de  $d$  é uma questão ainda em aberto. Pode-se utilizar todas as componentes ( $d = n-1$ ), ou selecionar as que forneçam os melhores resultados, a julgar pelo desempenho da calibração. Andrew e Fearn <sup>(94)</sup> utilizaram o número de componentes principais necessário para descrever 98-99% da variação.

O método TOP é semelhante à OSC, pois ambos removem dimensões dos dados espectrais antes da calibração. A diferença é que o TOP utiliza informações do conjunto de transferência para remover dimensões que interferem na transferência, enquanto o OSC só remove dimensões ortogonais a  $\mathbf{y}$ .

Cada pré-processamento opera eliminando uma fonte específica de variação, podendo levar a modelos mais parcimoniosos <sup>(80)</sup>. Os métodos descritos, em geral, são eficientes quando o problema se deve a variações nas intensidades das bandas espectrais, os equipamentos envolvidos são muito parecidos e as diferenças instrumentais são pequenas.

### 2.1.2.3 - Seleção de Variáveis

A seleção de variáveis insensíveis ou menos sensíveis a variações nas condições experimentais melhora a robustez dos modelos de calibração, além de aumentar a sua capacidade preditiva <sup>(97)</sup>. Existem várias razões práticas para isto. Por exemplo, algumas regiões espectrais podem não mostrar uma grande variação para o parâmetro de interesse, os espectros podem conter ruídos heteroscedásticos, pode haver comprimentos de onda que contenham absorbâncias que não são linearmente correlacionadas ao parâmetro de interesse, bem como pode haver comprimentos de onda que tenham uma correlação indireta com este parâmetro.

Swierenga *et al* <sup>(98)</sup> testaram a capacidade de transferência de modelos PLS para determinar o teor de água em tabletes utilizando, recozimento simulado (SA, de *simulated annealing*) para selecionar um subconjunto das variáveis originais que preservasse a capacidade de predição do modelo quando este fosse transferido para outro equipamento. Nessas condições, obtiveram resultados melhores dos que os produzidos a partir do método DS.

### 2.1.3 - Seleção das Amostras de Transferência

A escolha do conjunto de amostras que serão lidas nos equipamentos envolvidos na transferência de calibração (amostras de transferência ou de padronização) é importante para o sucesso do processo. Para evitar variações espúrias no espectro transferido, as amostras devem ser representativas do espaço experimental e descrever as diferenças instrumentais. Se isto não acontecer, os parâmetros de transferência calculados levarão a padronizações duvidosas.

Existem várias idéias, princípios e estratégias para a seleção de amostras de transferência <sup>(69,77,93,99)</sup>. Wang *et al* <sup>(69)</sup> selecionaram amostras de acordo com seus valores de alavancagem (*leverage*), que representa uma medida de quão extremo um dado ponto ou variável está comparado à maioria. A seleção inicia-se com a amostra de maior alavancagem, que é removida do conjunto de dados e

ortogonalizada contra as demais amostras. A seqüência é repetida até que o número desejado de amostras seja extraído.

Bouveresse *et al*<sup>(99)</sup> compararam a seleção do conjunto de amostras de transferência por três algoritmos diferentes, o clássico Kennard – Stone (KS)<sup>(100)</sup>, o baseado nos valores de alavancagem, e o método baseado na maximização da menor distância entre pontos (MSID). Os melhores resultados foram obtidos utilizando-se KS e MSDI.

O algoritmo clássico é o de Kennard - Stone<sup>(100)</sup>, que começa selecionando as duas amostras com a maior distância Euclidiana. A amostra com a menor distância dessas duas amostras selecionadas é retida, e o procedimento é repetido até que um determinado número de amostras seja selecionado.

Não há nenhum critério sobre o número ótimo de amostras para realizar a padronização. O que se pode fazer é examinar os resultados de validação e escolher para uma série de padrões o número de amostras que minimize os erros de predição. Uma outra opção seria utilizar amostras ou padrões genéricos, o que resolveria o problema de degradação das amostras com o tempo. É necessário, porém, que os espectros desses padrões apresentem o mesmo perfil que o espectro das amostras de teste, para que não haja variações espectrais. Pode-se ainda, selecionar o número de amostras de padronização considerando-se o número de amostras disponíveis para a transferência.

Dantas Filho *et al*<sup>(101)</sup> propuseram o algoritmo APSA para seleção de amostras de calibração, que também pode, com adaptações, ser utilizado para transferência. O APSA é uma versão do APS modificada para realizar seleção de amostras. A matriz das respostas instrumentais é transposta, o que faz com que suas colunas tornem-se as amostras de calibração, e o APS é usado para selecionar um subconjunto de amostras minimamente redundantes. Comparado ao algoritmo clássico KS, o APSA fornece resultados comparáveis ou ligeiramente melhores.

## 2.2 – OBJETIVO

Nesta segunda parte deste trabalho, propõe-se uma estratégia para construção de modelos MLR robustos às diferenças entre dois equipamentos. A estratégia utiliza o APS para selecionar variáveis (APSV) que minimizem o valor de um erro conjunto que considera não só o conjunto de amostras de teste do equipamento principal, mas também um pequeno conjunto de amostras medidas no equipamento secundário (amostras de transferência).

## 2.3. FUNDAMENTAÇÃO TEÓRICA

### 2.3.1 – Notação

As seguintes abreviaturas serão utilizadas:

- RMSEP** Erro quadrático médio de predição.
- RMSEP<sub>p</sub>** RMSEP para o conjunto de predição do equipamento primário utilizando o modelo calculado nesse equipamento.
- RMSEP<sub>s</sub>** Idêntico ao anterior, só que no equipamento secundário.
- RMSEP<sub>p</sub><sup>s</sup>** RMSEP para o conjunto de predição do equipamento secundário utilizando o modelo calculado no equipamento primário.
- RMSEP<sub>PT</sub><sup>s</sup>** RMSEP para o conjunto de predição do equipamento secundário utilizando o modelo calculado no primário, com técnicas de transferência.

### 2.3.2 - Critério de Robustez

Para melhorar a robustez do conjunto de variáveis selecionadas pelo APSV, o critério de seleção mencionado anteriormante (RMSEV) foi substituído por um critério que também leva em conta o erro de predição para o conjunto de dados de transferência, como sugerido por Swirenga *et al*<sup>(98)</sup>. O novo critério é a minimização do valor do erro  $E$  dado por:

$$E = \frac{1}{2}(RMSEV + RMSET), \quad (2.9)$$

onde, RMSET é o erro médio quadrático de predição para as amostras do conjunto de dados de transferência medidos no equipamento primário, definido como:

$$RMSET = \sqrt{\frac{1}{K_t} \sum_{k=1}^{K_t} (y_t^k - \hat{y}_t^k)^2} \quad (2.10)$$

Onde,  $y_t^k$  e  $\hat{y}_t^k$  são os valores de referência e previsto para o parâmetro de interesse na  $K$ -ésima amostra de transferência e,  $K_t$  é o número de amostras de transferência. Desta forma, a seleção de variáveis considerará não apenas o poder de predição do modelo (medido pelo RMSEV), mas também sua robustez (medida pelo RMSET) atribuindo a mesma importância a estes dois critérios, independentemente do número de amostras de teste e de transferência utilizados.

### 2.3.3 - Seleção das Amostras de Transferência

Duas técnicas para a seleção das amostras de transferência foram consideradas. A primeira é o algoritmo clássico Kennard – Stone (KS), que seleciona um conjunto de dados representativos que consiste de amostras separadas pelas maiores distâncias Euclidianas. A segunda foi proposta recentemente e consiste em uma versão, adaptada às amostras, do APS, denominada APSA, que opera na matriz transposta das respostas instrumentais  $\mathbf{X}_{cal}^T$ . Neste caso, o procedimento de seleção das amostras é muito parecido com o processo de seleção de variáveis. Considerando que o APS visa selecionar colunas com pequena colinearidade de uma dada matriz, basta aplicar o algoritmo a  $\mathbf{X}_{cal}^T$ . Dessa maneira, as colunas da matriz tornam-se amostras de calibração, de forma que o APSA selecionará um subconjunto de amostras minimamente redundantes.

É importante notar que o APSA utiliza a calibração MLR para comparar diferentes subconjuntos de amostras de acordo com o critério RMSEV. Torna-se,

portanto necessário tratar do problema de mau condicionamento que surge devido ao número de amostras ser menor que o número de variáveis. Uma forma de evitar este problema é utilizar a decomposição em valores singulares <sup>(44)</sup>. Outra possibilidade é empregar um procedimento simples de seleção de variáveis contido no software Matlab <sup>(63)</sup> para reduzir o número de variáveis a um número menor que o de amostras, de forma a obter uma matriz de calibração quadrada. Todavia, não importa qual seja a técnica utilizada, observou-se que o APSA fornece melhores resultados se o procedimento de seleção de amostras for precedido de uma seleção de variáveis preliminar pelo APSV <sup>(101)</sup>. Esta seleção preliminar é baseada apenas no critério RMSEV.

Deve-se ressaltar que o APSV será utilizado apenas uma vez se as amostras de transferência forem selecionadas pelo algoritmo KS, mas será utilizado duas vezes se elas forem selecionadas pelo APSA. Inicialmente o APS é utilizado para escolher um subconjunto de variáveis representativas no conjunto de dados do equipamento primário de acordo com o critério RMSEV, sem utilizar amostras de transferência. A matriz de respostas instrumentais  $\mathbf{X}_{cal}$  é então reduzida, mantendo apenas as colunas associadas àquelas variáveis representativas, antes de ser transposta e submetida ao APSA para seleção de amostras de transferência. Uma vez que as amostras de transferência tenham sido selecionadas, a matriz  $\mathbf{X}_{cal}$  é restabelecida e o APSV é aplicado uma segunda vez, para escolher um novo conjunto de variáveis de acordo com o critério de erro definido pela **Equação 2.9**. Escolhendo amostras de transferência de acordo com este critério admite-se que as amostras de calibração são apropriadas para construir um modelo com boa capacidade de predição e representativo do conjunto de calibração.

## 2.4 - EXPERIMENTAL

### 2.4.1 - Conjuntos de Dados

O primeiro conjunto consiste em 103 amostras de gasolinas C comum, coletadas nos postos de gasolina da região metropolitana dos estados de Pernambuco e Alagoas. O parâmetro de interesse foi a temperatura para 90% de amostra destilada,  $T_{90\%}$  que, como dito anteriormente, é importante para caracterizar os componentes da gasolina com altos pontos de ebulição. Os valores deste parâmetro para o conjunto de amostras foram determinados previamente pelo método de referência (ASTM D 86).

Os equipamentos principal e secundário são espectrômetros FT-IR Perkin Elmer modelo Spectrum GX (**Figura 1.1**) com sonda de ATR acoplada. O equipamento principal se encontra no Laboratório de Combustíveis - UFPE, e o equipamento secundário pertence ao Laboratório de Automação em Química Analítica e Quimiometria do Departamento de Química – UFPB. Os espectros foram obtidos nos dois equipamentos em temperatura ambiente média de 22°C, com resolução espectral de 4  $\text{cm}^{-1}$  e média de 16 varreduras, na faixa de 2500 a 15380  $\text{nm}$ . O equipamento principal tem sua sonda de ATR fixa no compartimento da amostra e é operado em um laboratório conveniado à ANP. As condições de temperatura e umidade são rigorosamente monitoradas e o acesso é restrito a pessoas autorizadas. Já no equipamento secundário a sonda de ATR é móvel, as condições ambientais são controladas, porém menos rigorosas, e o equipamento é operado por diversos usuários.

O segundo conjunto de dados contém os espectros de 80 amostras de milho obtidos em dois diferentes equipamentos na faixa de 1100 a 1498  $\text{nm}$ , junto com o teor de umidade determinado para cada amostra. Esses dados de milho encontram-se disponíveis no endereço <http://www.eigenvector.corn/Data/Corn/>, tendo sido consultados em 10/06/2004, e foram utilizados em artigo recente sobre transferência de calibração <sup>(94)</sup>. Neste trabalho os equipamentos 1 e 2 deste conjunto de dados são designados como principal e secundário, respectivamente.

### 2.4.2 - Pré-processamento dos Dados

Inicialmente foi feita uma análise PLS preliminar do conjunto de dados das gasolinas para escolher os conjuntos de calibração, teste e predição. Com base na análise dos escores, o conjunto de amostras foi dividido em 63 amostras de calibração, 20 de teste e 20 de predição. As amostras de teste e de predição foram selecionadas de forma a cobrir a variabilidade em **X** e **Y**, evitando-se os pontos extremos desses espaços. Para evitar problemas devidos a variações sistemáticas na linha de base, utilizou-se o espectro derivado depois de um alisamento por filtros Savitzky-Golay com polinômio de 2ª ordem e janela de 21 pontos. Além disso, variáveis com baixa relação sinal/ruído foram eliminadas, descartando-se os comprimentos de onda cujas intensidades, no espectro derivado, não excedessem 2% da intensidade máxima de todo o conjunto de dados. O conjunto dos dados de milho também foi submetido aos mesmos processos de escolha das amostras e pré-processamentos das variáveis. Isto resultou em 40 amostras de calibração, 20 de teste e 20 de predição. Neste caso, não foi feito o descarte de variáveis com baixa relação sinal/ruído.

### 2.4.3 - Modelagem e Validação

Os modelos MLR, a seleção de variáveis/amostras pelo APS e a seleção pelo KS foram feitos com o software Matlab<sup>(63)</sup>, através de programas previamente desenvolvidos pelo Profo. Dr. Roberto Kawakami da Divisão de Engenharia Eletrônica - ITA. A primeira derivada dos espectros com alisamento por filtros Savitzky-Golay foi também calculada no Matlab, utilizando-se ferramentas de processamento de sinais. Os cálculos para o método PDS foram feitos com o pacote de ferramentas PLS Toolbox para Matlab. A janela do PDS foi ajustada, variando-se seu tamanho de 1 até 7 pontos para os dados de gasolinas, e de 1 até 13 pontos para os dados de milho. Os melhores resultados foram obtidos com janelas de 3 pontos para gasolinas e 9 pontos para os dados de milho. Neste processo, utilizou-se a tolerância sugerida pelo PLS Toolbox ( $10^{-2}$ ). Os modelos PLS foram construídos

utilizando-se o software Unscrambler 7.6 <sup>(62)</sup>. O número de fatores nos modelos PLS foi determinado pelo próprio Unscrambler, com base nas variâncias residuais de **X** e **Y** no conjunto de teste. Na discussão dos resultados, foi feito o cálculo da significância estatística das diferenças entre os valores de RMSEP com teste F a 95 % de confiança.

## 2.5 - RESULTADOS E DISCUSSÃO

Os perfis dos espectros dos equipamentos primário e secundário para os conjuntos de dados de gasolinas e de milho são apresentados nas **Figuras 2.1** e **2.2** respectivamente.

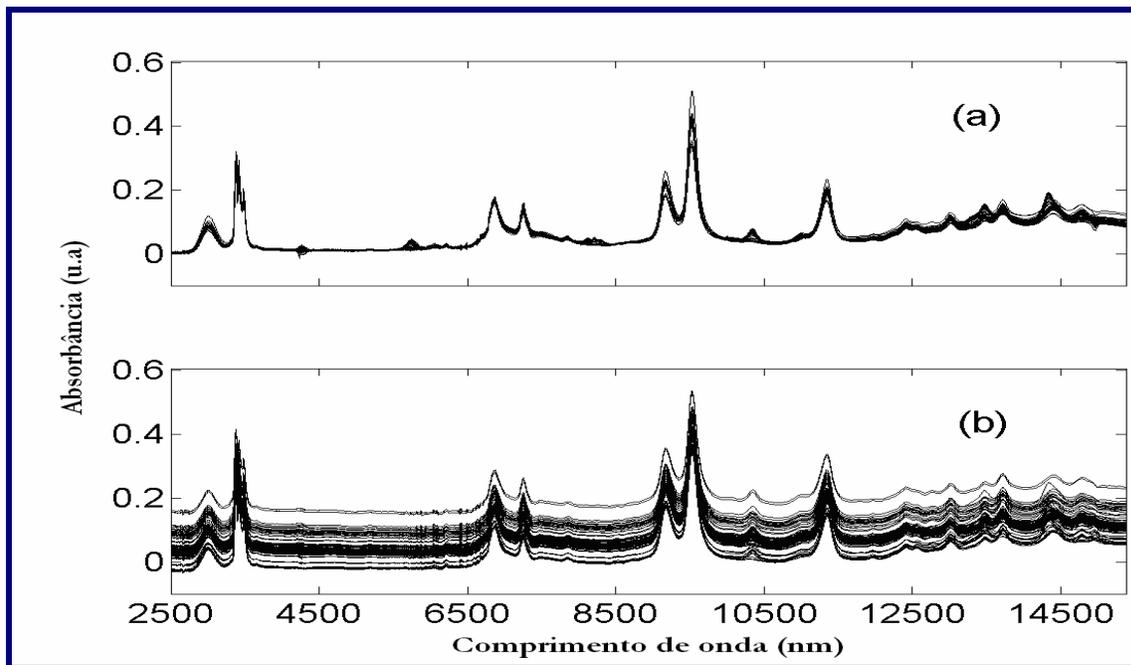


Figura 2.1 – Espectro original das amostras de gasolinas obtidos nos equipamentos primário (a) e secundário (b).

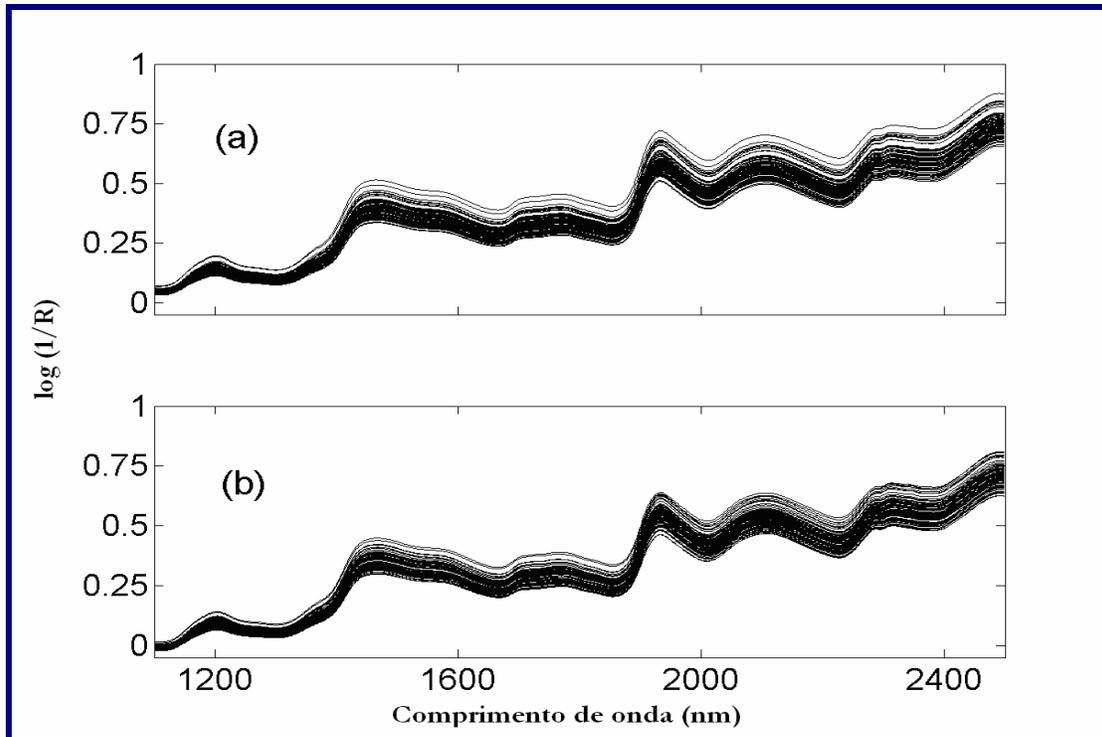


Figura 2.2 – Espectro original das amostras de milho obtidos nos equipamentos primário (a) e secundário (b).

Os dados espectrais das amostras de gasolinas foram submetidos aos pré-processamentos descritos na seção anterior, daí resultando os espectros mostrados na **Figura 2.3**, com 2013 variáveis que foram centradas na média antes dos procedimentos de modelagem.

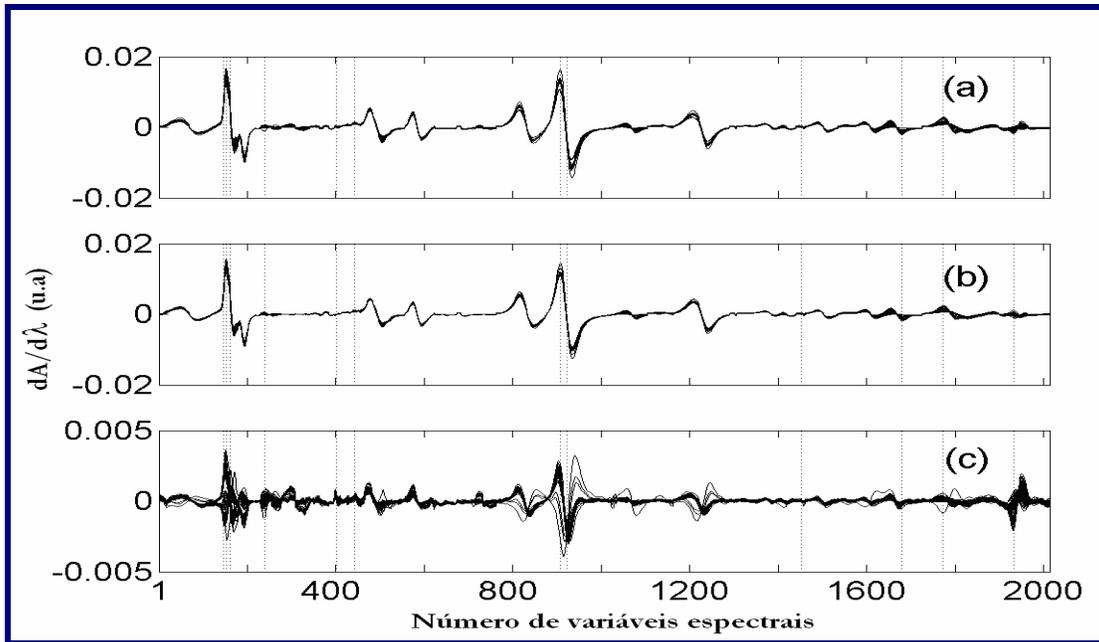


Figura 2.3 – Espectros derivados das amostras de gasolinas depois da pré-seleção das regiões espectrais: (a) primário, (b) secundário e (c) diferenças entre os espectros medidos no primário e no secundário.

Para os dados espectrais de milho, depois dos pré-processamentos, os espectros resultantes contêm 680 variáveis, e seu perfil é mostrado na **Figura 2.4**.

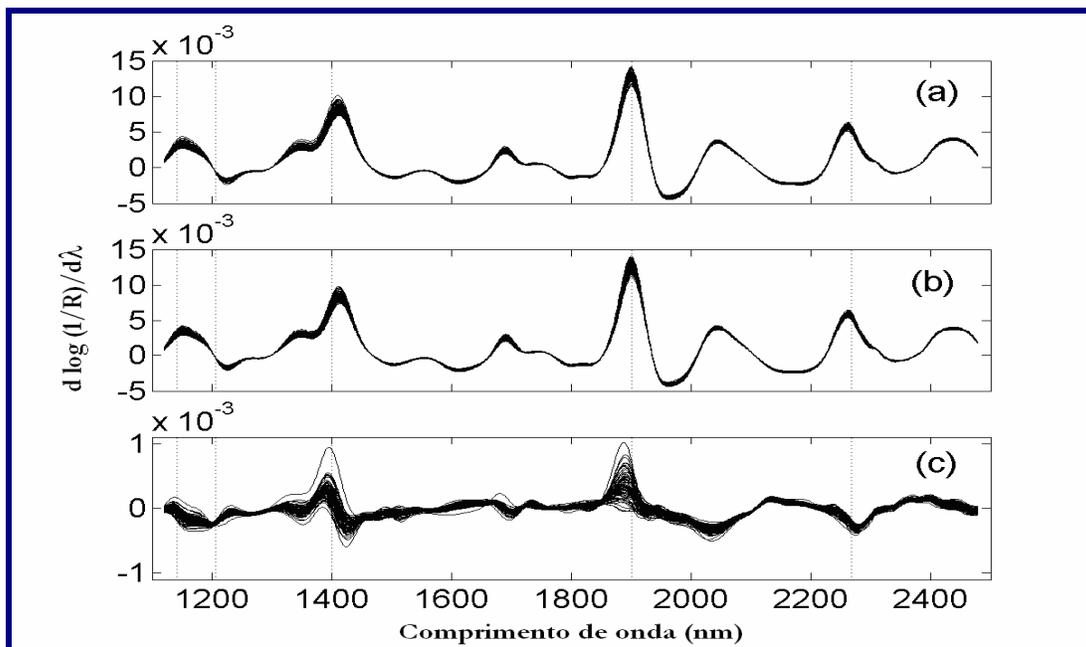


Figura 2.4 – Espectros derivados das amostras de milho: (a) primário, (b) secundário e (c) diferenças entre os espectros medidos no principal e no secundário.

Os desempenhos de predição da calibração PLS padrão e do protocolo APSV-MLR para os dois conjuntos de dados nos equipamentos primário e secundário são comparados na tabela abaixo.

**Tabela 2.1 – Valores dos RMSEP's obtidos pelos métodos PLS e APSV-MLR.**

<b>RMSEP</b>	<b>PLS</b>	<b>APSV-MLR</b>
<i>Dados de gasolinas (<math>^{\circ}C</math>)</i>		
RMSEP <sub>p</sub> <sup>P</sup>	2,4 (5) <sup>a</sup>	2,0 (21) <sup>b</sup>
RMSEP <sub>s</sub> <sup>S</sup>	2,3 (4)	2,2 (7)
RMSEP <sub>p</sub> <sup>S</sup>	10,1 (5)	3,3 (21)
<i>Dados de milho (% m/m)</i>		
RMSEP <sub>p</sub> <sup>P</sup>	0,067 (4)	0,013 (22)
RMSEP <sub>s</sub> <sup>S</sup>	0,170 (3)	0,136 (28)
RMSEP <sub>p</sub> <sup>S</sup>	1,499 (4)	1,494 (22)

<sup>a</sup> Número de fatores selecionados no PLS; <sup>b</sup> Número de variáveis selecionadas pelo APSV

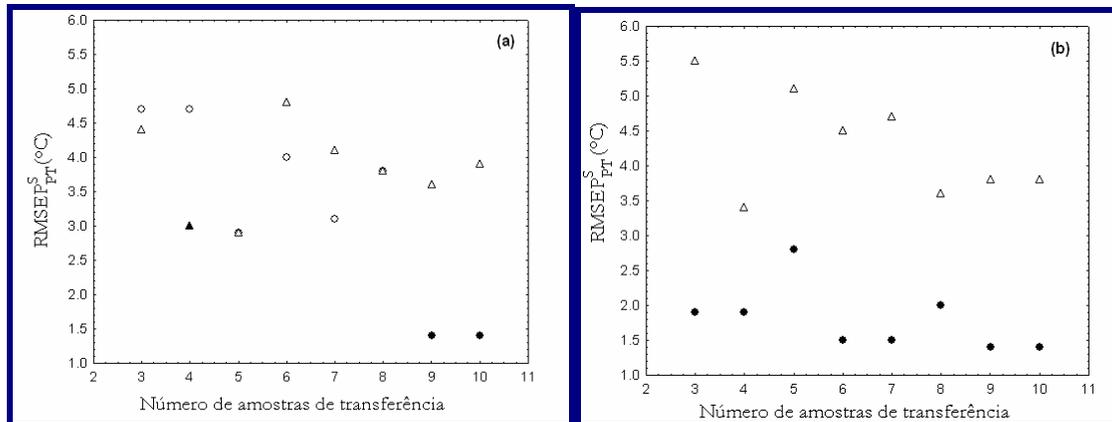
Para os dados de gasolinas, o modelo APSV-MLR é menos sensível a mudanças nas características das respostas instrumentais. O RMSEP do equipamento secundário quando o modelo desenvolvido no primário é aplicado (RMSEP<sub>p</sub><sup>S</sup>) é menor para o APSV-MLR (3,3) do que para o PLS (10,1).

Nos dados de milho, o RMSEP<sub>p</sub><sup>P</sup> é consideravelmente menor para APSV-MLR (0,013) do que para o PLS (0,067), mas os valores do RMSEP<sub>s</sub><sup>S</sup> (recalibração completa no equipamento secundário) dos dois métodos são estatisticamente indistinguíveis: 0,170 para o PLS e 0,136 para o APSV-MLR. Além disso, não existe diferença significativa entre os dois métodos quando se considera os valores do RMSEP<sub>p</sub><sup>S</sup> (erro de predição do equipamento secundário quando o modelo calculado no primário é aplicado).

Os resultados do procedimento de modelagem robusta APSV-MLR proposto, assim como os de PDS-PLS, são apresentados nas **Figuras 2.5 e 2.6**, para T 90% em gasolinas e teor de umidade em milho, respectivamente. Nestes

gráficos, os valores de  $RMSEP_{PT}^S$  são apresentados em função do número de amostras de transferência (de 3 a 10) utilizado na seleção de variáveis robustas pelo APSV-MLR e na padronização pelo PDS. Os resultados mostrados nas **Figuras 2.5a** e **2.6a** foram obtidos utilizando-se KS para a seleção de amostras de transferência, enquanto os resultados das **Figuras 2.5b** e **2.6b** foram obtidos utilizando-se o APSA.

O melhor resultado para gasolinas ( $RMSEP_{PT}^S = 1,4$ ) foi obtido com o APSV-MLR, com nove ou dez amostras selecionadas pelo KS ou pelo APSA. Para ilustrar, a **Figura 2.7a** apresenta a curva do valor do erro (E) utilizado no APSV-MLR para a determinação do melhor número de variáveis robustas, calculado a partir de nove amostras de transferência selecionadas pelo APSA. O ponto mínimo local em 12 variáveis foi escolhido porque as melhorias obtidas pelo aumento do número de variáveis (18 ou 21, por exemplo) são desprezíveis. As 12 variáveis assim selecionadas estão indicadas pelas linhas verticais pontilhadas na **Figura 2.3**.



**Figura 2.5** – Comparação entre os resultados do APSV-MLR (O) e PDS-PLS (Δ) para T 90% nos dados de gasolina utilizando-se dois métodos para seleção de amostras: (a) KS e (b) APSA. Os símbolos pretos indicam que os valores correspondentes são significativamente menores (ao nível de 95% de confiança) que os valores alternativos.

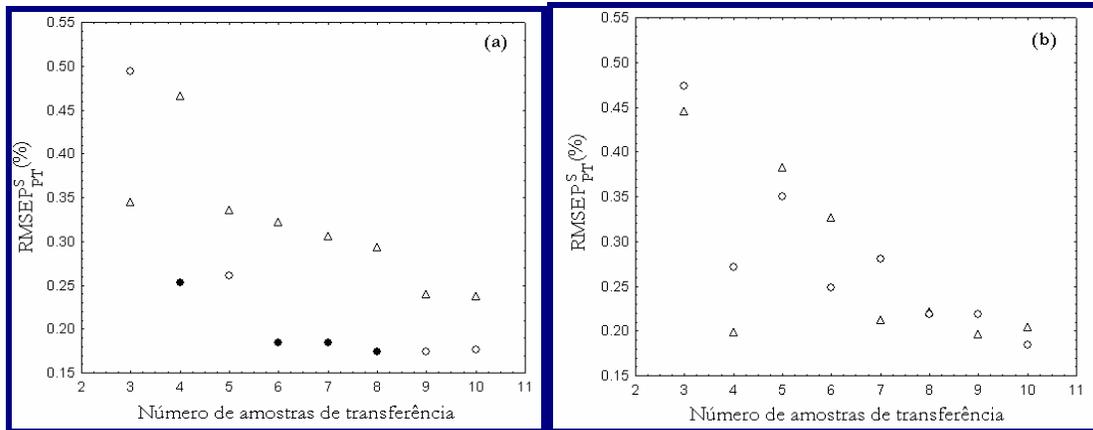


Figura 2.6 – Comparação entre os resultados do APSV-MLR (O) e PDS-PLS ( ) para teor de umidade nos dados de milho utilizando-se dois métodos para seleção de amostras: (a) KS e (b) APSA. Os símbolos pretos indicam que os valores correspondentes são significativamente menores (ao nível de 95% de confiança) que os valores alternativos.

Para os dados de milho, o melhor resultado ( $RMSEP_{PT}^s = 0,174$ ) foi obtido utilizando-se o APSV-MLR com nove amostras de transferência selecionadas pelo KS. A **Figura 2.7b** apresenta a curva do valor do erro (E) utilizado pelo APSV-MLR na seleção de variáveis robustas. Neste caso, o valor do erro depois do primeiro ponto de mínimo local é sempre maior. As cinco variáveis selecionadas estão indicadas por linhas verticais pontilhadas na **Figura 2.4**.

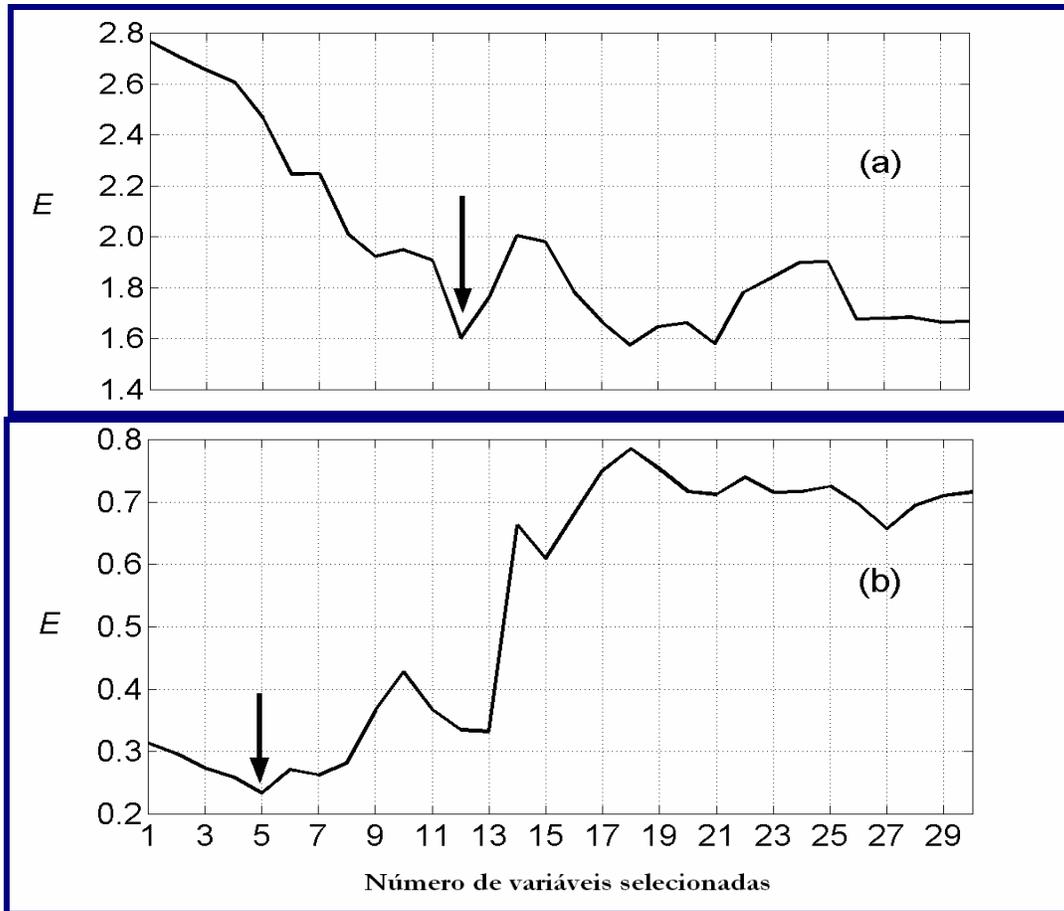


Figura 2.7 –Valor do erro (E) em função do número de variáveis selecionadas pelo APSV considerando nove amostras de transferência. (a) Dados de gasolina com amostras de transferência selecionadas pelo APSA. (b) Dados de milho com amostras de transferência selecionadas pelo KS. As setas indicam o ponto de mínimo utilizado para escolher o número de variáveis.

Uma comparação dos valores de  $RMSEP_{PT}^S$  das Figuras 2.5 e 2.6 com os valores de  $RMSEP_p^S$  da Tabela 2.1 mostra que a precisão da predição do modelo PLS é consideravelmente melhor quando o PDS é aplicado ao espectro obtido no equipamento secundário. A seleção de variáveis baseada no critério de robustez da Equação 2.10 também melhora os resultados do APSV-MLR. Embora para os dois conjuntos de dados a diminuição do  $RMSEP_p^S$  para  $RMSEP_{PT}^S$  seja estatisticamente significativa, ela é mais expressiva para o conjunto de dados de milho, onde se obtém uma redução do  $RMSEP_p^S = 1.497$  para  $RMSEP_{PT}^S = 0.174$  (melhor resultado). Vale ressaltar que os valores de  $RMSEP_{PT}^S$  obtidos para este

conjunto de dados são comparáveis aos obtidos por outros pesquisadores em artigo recente <sup>(96)</sup>.

As **Figuras 2.5** e **2.6** podem ser interpretadas como 32 comparações (3-10 amostras de transferência, dois métodos para seleção de amostras, dois conjuntos de dados) entre os desempenhos do PDS-PLS e APSV-MLR (com critério de robustez). Para os dados de gasolinas (**Figura 2.5**) o APSV-MLR é significativamente melhor que PDS-PLS em 10 das 16 comparações. Nas seis comparações restantes, os dois métodos são estatisticamente equivalentes em cinco casos. Para os dados de milho os resultados são menos discrepantes, e o desempenho é estatisticamente o mesmo em 12 das 16 comparações. Nos outros quatro casos, entretanto, a comparação é favorável ao APSV-MLR. De maneira geral, 15 das 32 comparações apresentam diferenças estatisticamente significativa, e todas, com exceção de uma, os resultados de predição favorecem o método APSV-MLR proposto.

## CONCLUSÕES E PROPOSTAS FUTURAS

---

Na primeira etapa deste trabalho, uma análise de componentes principais com os valores relativos de RMSEP, para as diferentes combinações de faixa espectral, pré-processamentos e técnicas de calibração, mostrou que uma condição de modelagem baseada em MLR, com seleção de variáveis por AG e sem pré-processamento dos dados na região NIR com caminho óptico de 1 mm (1600 – 2500 nm) fornece bons resultados de predição para todas as propriedades estudadas. Nesta condição, com exceção da propriedade Densidade, os valores de RMSEP obtidos foram menores que a reprodutibilidade do método de referência.

Na etapa de transferência de modelos de calibração, os resultados obtidos, em termos de erro de predição no equipamento secundário, para os modelos MLR robustos e PLS-PDS foram comparáveis e levemente melhores para o APSV-MLR. O APSV-MLR apresenta uma vantagem de não requerer que as mesmas amostras de transferência sejam medidas nos equipamentos primário e secundário. Isto pode ser uma vantagem decisiva se tais medidas são feitas para conjuntos de dados obtidos separadamente, seja no espaço ou tempo. Por esta razão, sugere-se que a estratégia proposta seja uma alternativa válida para a técnica que é normalmente utilizada, PLS-PDS.

Como propostas futuras espera-se avaliar:

- a robustez do critério proposto para um número maior de equipamentos semelhantes;
- o uso da transformada *wavelet* como pré-processamento, a fim de obter melhorias na capacidade de transferência dos modelos;
- as diferentes técnicas de transferência, inclusive a estratégia proposta, envolvendo equipamentos de diferentes fabricantes.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

---

1. American Society for Testing and Materials – ASTM. Annual Book of ASTM Standards, v. 05.02, D 4052 - Density and Relative Density of Liquids by Digital Density Meter. American Society for Testing and Materials, Philadelphia, PA. 2000a.
2. Associação Brasileira de Normas Técnicas – ABNT NBR 14065 – Determinação da massa específica e da densidade relativa pelo densímetro digital. Rio de Janeiro, 1998.
3. American Society for Testing and Materials – ASTM. Annual Book of ASTM Standards, v. 05.01, D 86 - Distillation of Petroleum Products. American Society for Testing and Materials, Philadelphia, PA. 2000a.
4. Associação Brasileira de Normas Técnicas – ABNT NBR 9619 – Determinação das propriedades de destilação. Rio de Janeiro, 1998.
5. American Society for Testing and Materials – ASTM. Annual Book of ASTM Standards, v. 05.01, D 2699 – Standards method for research octane of spark-ignition engine fuel. American Society for Testing and Materials, Philadelphia, PA. 2000a.
6. American Society for Testing and Materials – ASTM. Annual Book of ASTM Standards, v. 05.01, D 2700 - Standards method for research motor of spark-ignition engine fuel. American Society for Testing and Materials, Philadelphia, PA. 2000a.
7. American Society for Testing and Materials – ASTM. Annual Book of ASTM Standards, v. 05.01, D 1319 – Hydrocarbons types in liquid petroleum products by fluorescent indicator adsorption. American Society for Testing and Materials, Philadelphia, PA. 2000a.

8. Associação Brasileira de Normas Técnicas – ABNT NBR 14932 – Produtos líquidos do petróleo - Determinação dos tipos de hidrocarbonetos pelo indicador de adsorção por fluorescência. Rio de Janeiro, 1998.
9. BOKOBZA, L.; Near infrared spectroscopy, **Journal of Near Infrared Spectroscopy**, v. 6, p. 3-17, 1998.
10. BARTON, F. E.; Theory and principles of near infrared spectroscopy, **Proceedings of the 10<sup>th</sup> International NIR Conference**, Spectroscopy Europe, p. 12-15, 2002.
11. PASQUINI, C.; Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications, **Journal of the Brazilian Chemical Society**, v. 14, p. 198-219, 2003.
12. BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B.; Chemometrics – A practical guide, New York, Ed. John Wiley & Sons, Inc., 1988.
13. NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T.; A user-friendly guide to Multivariate Calibration and Classification, Chichester - UK, NIR Publications, 2002.
14. GARCIA-ALVAREZ, M.; HUIDOBRO, J.F.; HERMIDA, M.; RODRIGUEZ-OTERO, J. L.; Major components of honey analysis by near-infrared transreflectance spectroscopy, **Journal of the Agriculture and Food Chemistry**, v. 48, p. 5154-5158, 2000.
15. BARTHUS, R.C.; POPPI, R. J.; Determination of the total unsaturation in vegetable oils by Fourier transform Raman spectroscopy and multivariate calibration, **Vibrational Spectroscopy**, v. 26, p. 99-105, 2001.
16. FIDENCIO, P. H.; POPPI, R. J.; ANDRADE, J. C.; Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy, **Analytica Chimica Acta**, v. 453, p. 125-134, 2002.
17. CLEVE, E.; BACH, E.; SCHOLLMAYER, E.; Using chemometric methods and NIR spectrophotometry in the textile industry, **Analytica Chimica Acta**, v. 420, p. 163-167, 2000.

18. HAZEN, K. H.; ARNOLD, M. A.; SMALL, G. W.; Measurement of glucose and other analyte in undiluted human serum with near infrared transmission spectroscopy, **Analytica Chimica Acta**, v. 371, p. 255-267, 1998.
19. SPAREN, A.; MALM, M.; JOSEFSON, M.; FOLESTAD, S.; JOHANSON, J.; Light leakage effects with different sample holder geometries in quantitative near-infrared transmission spectroscopy of pharmaceutical tablets, **Applied Spectroscopy**, v. 56, p. 586-592, 2002.
20. SCAFI, S. H. F.; PASQUINI, C.; Identification of counterfeit drugs using near-infrared spectroscopy, **Analyst**, v. 126, p. 2218-2224, 2001.
21. United States Pharmacopoeia, USP-24-NF 19, United States Pharmacopoeial Convention, Rockville, 2000.
22. PARISI, A. F.; NOGUEIRAS, L.; PRIETO, H.; On-line determination of fuel quality parameters using near-infrared spectrometry with fiber optics and multivariate calibration, **Analytica Chimica Acta**, v. 238, p. 95-100, 1990.
23. KIM, M.; LEE, Y. H.; HAN, C.; Real-time classification of petroleum products using near-infrared spectra, **Computers and Chemical Engineering**, v. 24, p. 513-517, 2000.
24. CHOQUETTE, S. J.; CHESLER, S. N.; DUEWER, D. L.; WANG, S. O'HAVER, T. C.; Identification and quantitation of oxygenates in gasoline ampules using Fourier transform near-infrared and Fourier transform Raman spectroscopy, **Analytical Chemistry**, v. 68, p. 3525-3533, 1996.
25. GUCHARDI, R.; FILHO, P. A. C.; POPPI, R. J.; PASQUINI, C.; Determination of ethanol and methyl tert-butyl ether (MTBE) in gasoline by NIR-AOTF-based spectroscopy and multiple linear regression with variables selected by genetic algorithm, **Journal of Near Infrared Spectroscopy**, v. 6, p. 333-339, 1998.
26. BOHÁCS, G.; OVÁDI, Z.; SALGÓ, A.; Prediction of gasoline properties with near infrared spectroscopy, **Journal of Near Infrared Spectroscopy**, v. 6, p. 341-348, 1998.

27. BOSCHETTI C. E.; OLIVIERI, A. C.; A new genetic algorithm applied to the near infrared analysis of gasolines, **Journal of Near Infrared Spectroscopy**, v. 12, p. 85-91, 2004.
28. ANREUS, E. L.; GARRIGUES, S.; GUARDIA, M.; Vapour generation-Fourier transform infrared spectrometric determination of benzene, toluene and methyl tert.-butyl ether in gasolines, **Analytica Chimica Acta**, v.333, p. 157-165, 1996.
29. FODOR, G. E.; KOHL, K. B.; MASON, R. L.; Analysis of gasolines by FT-IR spectroscopy, **Analytical Chemistry**, v. 68, p. 23-30, 1996.
30. ANDRADE, J. M.; SÁNCHEZ, M. S.; SARABIA, L. A.; Applicability of high-absorbance MIR spectroscopy in industrial quality of reformed gasolines, **Chemometrics and Intelligent Laboratory Systems**, v. 46, p. 41-55, 1999.
31. PASADAKIS, N.; SOURLIGAS, S.; FOTEINOPOULOS, CH.; Prediction of the distillation profile and cold properties of diesel fuels using mid-IR spectroscopy and neural networks, **Fuel**, v. 85, p. 1131-1137, 2006.
32. TRAN, C. D.; Visualizing chemical composition and reaction kinetics by the near infrared multispectral imaging technique, **Journal of Near Infrared Spectroscopy**, v. 8, p. 87-99, 2000.
33. TRAN, C. D.; POLITI, M.; Visualizing chemical compositions and kinetics of sol-gel by near-infrared multispectral imaging, **Analytical Chemistry**, v. 74, p. 1604-1610, 2002.
34. FELÍCIO, C. C.; BRÁS, L. P.; LOPES, J. A.; CABRITA, L.; MENESES, J. C., Comparison of PLS algorithms in gasoline and gas oil parameter monitoring with MIR and NIR, **Chemometrics and Intelligent Laboratory Systems**, v. 78, p. 74-80, 2005.
35. SAVITZKY, A.; GOLAY, M. J. E.; Smoothing and differentiation of data by simplified least squares procedures, **Analytical Chemistry**, v.36, p. 1627-1639, 1964.
36. HIBBERT, D. B.; Genetic Algorithms in Chemistry, **Chemometrics and Intelligent Laboratory Systems**, v.19, p. 277-293, 1993.

37. EFRON, B.; An Introduction to the Bootstrap (Hardcover), Publisher: Chapman & Hall/CRC, May 15, 1994.
38. 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 spectroscopic multicomponent analysis, **Chemometrics and Intelligent Laboratory Systems**, v. 57, p. 65-73, 2001.
39. GALVÃO, R. K. H.; PIMENTEL, M. F.; YONEYAMA, T.; VISANI, V.; Aspects of the successive projections algorithm for variable selection in multivariate calibration applied to plasma emission spectrometric, **Analytica Chimica Acta**, v.443, p. 107-115, 2001.
40. BREITKREITZ, M. C.; RAIMUNDO JR, I.; ROHWEDDER, J. J. R.; PASQUINI, C.; DANTAS FILHO, H. A.; JOSÉ, G. E.; ARAÚJO, M. C. U.; Determination of total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration, **Analyst**, v. 128, p. 1204-1207, 2003.
41. COELHO, C. J.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; PIMENTEL, M. F.; SILVA, E. C.; A solution to the wavelet transform optimization problem in multicomponent analysis, **Chemometrics and Intelligent Laboratory Systems**, v. 66, p. 205-217, 2003.
42. COELHO, C. J.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; PIMENTEL, M. F.; SILVA, E. C.; A linear semi-infinite programming strategy for constructing optimal wavelet transforms in multivariate calibration problems, **Journal of Chemical Information and Computer Sciences**, v. 43, p. 928-933, 2003.
43. 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 Y data, **Chemometrics and Intelligent Laboratory Systems**, v. 70, p. 1-10, 2004.
44. XU, L.; ZHANG, W. J.; Comparison of different methods for variable selection, **Analytica Chimica Acta**, v. 446, p. 477-483, 2001.

45. 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.
46. HELLAND, I. S.; Some theoretical aspects of partial least squares regression, **Chemometrics and Intelligent Laboratory Systems**, v. 58, p. 97-107, 2001.
47. WOLD, S.; SJOSTROM, M.; ERIKSSON, L.; PLS- regression: a basic tool of chemometrics, **Chemometrics and Intelligent Laboratory Systems**, v. 58, p. 109-130, 2001.
48. WOLD, S.; TRYGG, J.; BERLUND, A.; ANTTI, H.; Some recent developments in PLS modeling, **Chemometrics and Intelligent Laboratory Systems**, v. 58, p. 131-150, 2001.
49. BRERETON, R. G.; Introduction to multivariate calibration in analytical chemistry, **Analyst**, v. 125, p. 2125-2154, 2000.
50. SWIERENGA, H.; WEIJER, A. P.; WIJK, R. J.; BUYDENS, L. M. C.; Strategy for constructing robust multivariate calibration models, **Chemometrics and Intelligent Laboratory Systems**, v. 49, p. 1-17, 1999.
51. VAN LEEUWEN, J. A.; JONKER, R. J.; GILL, R.; Octane number prediction based on gas chromatographic analysis with non-linear regression techniques, **Chemometrics and Intelligent Laboratory Systems**, v.25, p. 325-340, 1994.
52. CÔCCO, L. C.; YAMAMOTO, C. I.; MEIEN, O. F.; Study of correlations for physicochemical properties of Brazilian gasoline, **Chemometrics and Intelligent Laboratory Systems**, v. 76 , p.55-63 , 2004.
53. WIENK, D.; KATEMAN, G.; Adaptive resonance theory based artificial neural networks for treatment of open-category problems in chemical pattern recognition – application to UV-Vis and IR spectroscopy, **Chemometrics and Intelligent Laboratory Systems**, v.23, p. 309-329, 1994.
54. EGHBALDAR, A.; FORREST, T. P.; CABROL-BASS, D.; Development of neural networks for identification of structural features from mass spectral data, **Analytica Chimica Acta**, v.359, p. 283-301, 1998.

55. MEUSINGER, R.; MOROS, R.; Determination of quantitative structure-octane rating relationships of hydrocarbons by genetic algorithms, **Chemometrics and Intelligent Laboratory Systems**, v.46, p. 67-78, 1999.
56. FLECHER, P. E.; WELCH, W. T.; ALBIN, S.; COOPER, J. B.; Determination of octane numbers and Reid vapor pressure in commercial gasoline using dispersive fiber-optic Raman spectroscopy, **Spectrochimica Acta Part A**, v. 53, p. 199-206, 1997.
57. COOPER, J. B.; WISE, K. L.; GROVES, J.; WELCH, W. T.; Determination of octane numbers and Reid vapor pressure of commercial petroleum fuels using FT-Raman spectroscopy and partial least-squares regression analysis, **Analytical Chemistry**, v. 67, p. 4096-4100, 1995.
58. 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.
59. MACHO, S.; LARRECHI, M. S.; Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry, **Trends in Analytical Chemistry**, v.21, p. 799-806, 2002.
60. COOPER, J. B.; WISE, K. L.; WELCH, W. T.; SUMNER, M. B.; WILT, B. K.; BLEDSOE, R. R.; Comparison of Near-IR, Raman, and Mid-IR spectroscopies for the determination of BTEX in petroleum fuels, **Applied Spectroscopy**, v. 51, p. 1613-1620, 1997.
61. SANTOS Jr., V. O.; OLIVEIRA, F. C. C.; LIMA, D. G.; PETRY, A. C.; GARCIA, E.; SUAREZ, P. A. Z.; RUBIM, J. C.; A comparative study of diesel analysis by FTIR, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis, **Analytica Chimica Acta**, v. 547, p. 188-196, 2005.
62. The Unscrambler 7.6, CAMO ASA Nedre Vollgate 8, N-0158 OSLO, Norway.
63. Matlab 6.1 User's Guide, The Mathworks, Natick, MA, USA.

64. 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.
65. WULFERD, F.; KOK, W. T.; NOORD, O. E.; SMILDE, A. K.; Correction of temperature-induced spectral variation by continuous piecewise direct standardization, **Analytical Chemistry**, v. 72, p. 1639-1644, 2000.
66. ADHIHETTY, I. S.; MCGUIRE, J. A.; WANGMANEERAT, B.; NIEMCZYK, T. M.; HAALAND, D. M.; Achieving transferable multivariate spectral calibration models: demonstration with infrared spectra of thin-film dielectrics on silicon, **Analytical Chemistry**, v. 63, p. 2329-2338, 1991.
67. DESPAGNE, F.; MASSART, D. L.; CHABOT, P.; Development of a robust calibration model for nonlinear in-line process data, **Analytical Chemistry**, v. 72, p. 1657-1665, 2000.
68. FEARN, T.; Standardization and calibration transfer for near infrared instruments: a review, **Journal of Near Infrared Spectroscopy**, v. 9, p. 229-244, 2001.
69. WANG, Y.; VELTKAMP, D. J.; KOWALSKI, B. R.; Multivariate Instrument Standardization, **Analytical Chemistry**, v. 63, p. 2750-2758, 1991.
70. WANG, Z. Y.; DEAN, T.; KOWALSKY, B. R.; Additive background correction in multivariate instrument standardization, **Analytical Chemistry**, v. 67, p. 2379-2385, 1995.
71. WANG, Y.; LYSAGHT, M. J.; KOWALSKY, B. R.; Improvement of multivariate calibration through instrument standardization, **Analytical Chemistry**, v. 64, p. 562-564, 1992.
72. CHEN, C. S.; BROWN, C. W.; LO, S. C.; Calibration transfer from sample cell to fiber-optic probe, **Applied Spectroscopy**, v. 51, p. 744-748, 1997.
73. GEMPERLINE, P. J.; CHO, J. H.; ALDRIDGE, P. K.; SEKULIC, S. S.; Near-IR detection of polymorphism and process-related substances, **Analytical Chemistry**, v. 68, p. 997-1002, 1996.

74. TAN, H. W.; BROWN, S. D.; Wavelet hybrid direct standardization of near-infrared multivariate, **Journal of Chemometrics**, v. 15, p. 647-663, 2001.
75. SHENK, J. S.; WESTERHAUS, M. O.; TEMPLETON JR, W. C.; Calibration transfer between near infrared reflectance spectrophotometers, **Crop Science**, v. 25, p. 159-161, 1985.
76. SHENK, J. S., WESTERHAUS M.O.; *US Patent No. 4866644* (12 september 1989).
77. BOUVERESSE, E.; MASSART, D. L.; DARDENNE, P.; Modified algorithm for standardization of near-infrared spectrometric instruments, **Analytical Chemistry**, v. 67, p. 1381-1389, 1995.
78. FORINA, M.; CASOLINO, C.; Improvement of slope bias correction for the intercalibration of different instruments, **Quimica Analitica**, v. 19, p. 49-59, 1999.
79. BOUVERESSE, E.; HARTMANN, C.; MASSART, D. L.; LAST, I. R.; PREBBLE, K. A.; Standardization of near-infrared spectrometric instruments, **Analytical Chemistry**, v. 68, p. 982-990, 1996.
80. NOORD, O. E.; The influence of data preprocessing on the robustness and parsimony of multivariate calibration models, **Chemometrics and Intelligent Laboratory Systems**, v. 23, p. 65-70, 1994.
81. WULFERT, F.; KOK, W. T.; SMILDE, A. K.; Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models, **Analytical Chemistry**, v.70, p. 1761-1767, 1998.
82. CENTNER, V.; VERDU-ANDRES, J.; WALCZAK, B.; JOUAN, R. D.; DESPAGNE, F.; PASTI, L.; POPPI, R.; MASSART, D. L.; NOORD, O. E.; Comparison of multivariate calibration techniques applied to experimental NIR data sets, **Applied Spectroscopy**, v.54, p.608-623, 2000.
83. BLANK, T. B.; SUM, S. T.; BROWN S. D.; MONFRE, S. L.; Transfer of near-infrared multivariate calibrations without standards, **Analytical Chemistry**, v. 68, p. 2987-2995, 1996.

84. TAN, H.; SUM, S. T.; BROWN, S. D.; Improvement of a Standard-free method for near-infrared calibration transfer, **Applied Spectroscopy**, v. 56, p. 1098-1106, 2002.
85. WOLD, S.; ANTTI, H.; LINDGREN, F.; OHMAN, J.; Orthogonal signal correction of near-infrared spectra, **Chemometrics and Intelligent Laboratory Systems**, v. 44, p. 175-185, 1998.
86. FEARN, T.; On orthogonal signal correction, **Chemometrics and Intelligent Laboratory Systems**, v.50, p. 47-52, 2000.
87. WESTERHUIS, J. A.; JONG, S.; SMILDE, A. K.; Direct orthogonal signal correction, **Chemometrics and Intelligent Laboratory Systems**, v. 56, p. 13-25, 2001.
88. PIERNA, J. A. F.; MASSART, D. L.; NOORD, O. E.; RICOUX, Ph.; **Chemometrics and Intelligent Laboratory Systems**, v. 55, p.101-108, 2001.
89. SJOBLÖM, J.; SVENSSON, O.; JOSEFSON, M.; KULLBERG, H.; WOLD, S.; **Chemometrics and Intelligent Laboratory Systems**, v. 44, p. 229-244, 1998.
90. WOODY, N. A.; FEUDALE, N. R.; MYLES, A. J.; BROWN, S. D.; Transfer of multivariate calibrations between four near-infrared spectrometers using orthogonal signal correction, **Analytical Chemistry**, v. 76, p. 2595-2600, 2004.
91. BARBOZA, F. D.; POPPI, R. J.; Determination of alcohol content in beverages using short-wave near-infrared spectroscopy and temperature correction by transfer calibration procedures, **Analytical and Bioanalytical Chemistry**, v. 377, p. 695-701, 2003.
92. LEUNG, A. K.; CHAU, F.; GAO, J.; A review on applications of wavelet transform techniques in chemical analysis: 1989-1997, **Chemometrics and Intelligent Laboratory Systems**, v. 43, p. 165-184, 1998.
93. PARK, K. S.; KO, Y. H.; LEE, H.; JUN, C. H.; CHUNG, H.; KU, M. S.; Near-infrared spectral data transfer using independent standardization samples: a case

- study on the trans-alkylation process, **Chemometrics and Intelligent Laboratory Systems**, v. 55, p. 53-65, 2001.
94. YOON, J.; LEE, B.; HAN, C.; Calibration transfer of near-infrared spectra based on compression of wavelet coefficients, **Chemometrics and Intelligent Laboratory Systems**, v. 64, p. 1-14, 2002.
95. TAN, H. W.; BROWN, S. D.; Wavelet analysis applied to removing non-constant, varying spectroscopic background in multivariate calibration, **Journal of Chemometrics**, v. 16, p. 228-240, 2002.
96. ANDREW, A.; FEARN, T.; Transfer by orthogonal projection: making near-infrared calibrations robust to between-instrument variation, **Chemometrics and Intelligent Laboratory Systems**, v. 72, p. 51-56, 2004.
97. SWIERENGA, H.; HAANSTRA, W. G.; WEIJER, A. P.; BUYDENS, L. M. C.; Comparison of two different approaches toward model transferability in NIR spectroscopy, **Applied Spectroscopy**, v. 52, p. 7-16, 1998.
98. SWIERENGA, H.; GROOT, P. J.; WEIJER, A. P.; DERKSEN, M. W. J.; BUYDENS, L. M. C.; Improvement of PLS model transferability by robust wavelength selection, **Chemometrics and Intelligent Laboratory Systems**, v. 41, p. 237-248, 1998.
99. BOUVERESSE, E.; MASSART, D. L.; Improvement of the piecewise direct standardization procedure for the transfer of NIR spectra for multivariate calibration, **Chemometrics and Intelligent Laboratory Systems**, v. 32, p. 201-213, 1996.
100. KENNARD, R. W.; STONE, L. A.; **Technometrics**, v. 11, 1969.
101. DANTAS FILHO, H. A.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; SILVA, E. C.; SALDANHA, T. C. B.; JOSÉ, G. E.; PASQUINI, C.; RAIMUNDO JR, I.; ROHWEDDER, J. J. R.; A strategy for selecting calibration samples for multivariate modeling, **Chemometrics and Intelligent Laboratory Systems**, v. 72, p. 83-91, 2004.

Tabela A - Valores absolutos dos RMSEP's obtidos para cada propriedade.

Pré-processamento	Calibração	Faixa	Seleção	Pré-proces.	Densidade	PIE	T10%	T50%	T90%	PFE	MON	RON	P	I	A	N	O
1. NIR 1 spp	PLS	NIRA	SEM	SEM	<b>0,0008</b>	1,07	0,97	0,3	1,46	4,18	0,51	0,31	0,69	0,82	0,63	0,27	1,16
2. NIR 1 der	PLS	NIRA	SEM	DER	0,0015	1,47	0,83	0,32	1,73	3,83	0,35	0,28	0,8	0,98	0,84	0,26	1,3
3. NIR 1 sg	PLS	NIRA	SEM	SG	0,0009	1,09	<b>0,73</b>	0,3	1,61	4,08	0,43	0,31	0,65	0,8	0,62	0,28	1,16
4. NIR 1 der sg	PLS	NIRA	SEM	DER SG	0,0009	1,5	0,85	0,32	1,84	4,27	0,29	0,28	0,97	1,14	0,81	0,26	1,26
5. NIR 1 sel spp	PLS	NIRA	AG	SEM	0,0012	1,08	0,86	0,32	2,35	4,56	0,36	0,35	0,63	1,09	0,69	0,3	1,05
6. NIR 1 sel der	PLS	NIRA	AG	DER	0,0013	1,56	0,83	0,48	1,53	4,34	0,32	0,29	0,58	1,05	0,91	0,24	1,22
7. NIR 1 sel sg	PLS	NIRA	AG	SG	0,0017	0,99	1,18	0,32	2,06	<b>3,49</b>	0,47	0,29	0,74	1,04	0,67	0,26	1,03
8. NIR 1 sel der sg	PLS	NIRA	AG	DER SG	0,0019	1,19	0,96	0,36	2,8	5,86	0,47	0,27	<b>0,54</b>	0,78	0,72	24	1,07
9. NIR 10 spp	PLS	NIRB	SEM	SEM	0,0024	1,58	1,43	0,65	4,43	7,52	0,48	0,34	0,66	1,18	0,92	0,73	1,37
10. NIR 10 der	PLS	NIRB	SEM	DER	0,0021	1,54	1,4	0,5	3,65	5,7	0,45	0,32	0,85	1,48	1,04	0,71	1,04
11. NIR 10 sg	PLS	NIRB	SEM	SG	0,0024	1,65	1,34	0,66	4,6	7,58	0,8	0,37	0,82	1,18	0,96	0,63	1,37
12. NIR 10 der sg	PLS	NIRB	SEM	DER SG	0,0024	1,48	1,3	0,68	3,9	5,57	0,81	0,35	0,65	1,33	0,93	0,67	1,13
13. NIR 10 sel spp	PLS	NIRB	AG	SEM	0,0015	1,07	1,28	0,56	2,2	5,81	0,37	0,29	0,73	1,33	1,14	0,48	0,98
14. NIR 10 sel der	PLS	NIRB	AG	DER	0,0018	1,2	1,09	0,41	2,23	5,61	0,42	0,34	0,56	1,23	0,88	0,48	1,09
15. NIR 10 sel sg	PLS	NIRB	AG	SG	0,0025	1,57	1,29	0,48	4,7	7,37	0,47	0,36	0,88	1,36	1,11	0,46	1,45
16. NIR 10 sel der sg	PLS	NIRB	AG	DER SG	0,0028	1,41	1,56	0,35	2,12	6,27	0,34	0,32	0,85	1,31	0,92	0,66	1,12
17. MID spp	PLS	MID	SEM	SEM	0,0021	1,02	0,98	0,43	2,17	5,37	0,42	0,29	0,69	1,36	1,15	0,73	1,28
18. MID der	PLS	MID	SEM	DER	0,0028	1,08	1,02	0,36	2,73	5,39	0,44	0,28	0,8	1,03	0,99	0,39	1,09
19. MID sg	PLS	MID	SEM	SG	0,0021	1,04	0,98	0,43	2,12	5,34	0,42	0,26	0,69	1,36	1,16	0,67	1,27
20. MID der sg	PLS	MID	SEM	DER SG	0,0027	1,12	1,01	0,36	2,48	6,04	0,48	0,28	0,64	1,05	1,06	0,44	1,12
21. MID sel spp	PLS	MID	AG	SEM	0,0014	1,2	1,01	0,34	2,72	4,82	0,42	0,25	1,58	1,22	0,88	0,42	1,37
22. MID sel der	PLS	MID	AG	DER	0,0016	1,26	0,94	0,33	2,72	7,24	0,36	0,3	1,1	1,01	0,79	0,3	1,2
23. MID sel sg	PLS	MID	AG	SG	0,002	1,44	1,32	0,38	2,46	6,19	0,45	0,29	0,72	1,04	0,82	0,54	1,19
24. MID sel der sg	PLS	MID	AG	DER SG	0,0019	1,15	0,86	0,37	2,27	5,42	0,33	0,34	1,36	1,34	0,86	0,42	1,1
25. NIR 1 ag spp	MLR	NIRA	AG	SEM	0,0009	<b>0,85</b>	0,88	0,23	1,72	4,65	0,31	0,28	0,66	<b>0,65</b>	0,67	<b>0,22</b>	1,07
26. NIR 1 ag der	MLR	NIRA	AG	DER	0,001	0,91	0,91	0,23	<b>0,97</b>	4,18	<b>0,26</b>	0,32	0,62	0,82	0,76	0,25	1,1
27. NIR 1 ag sg	MLR	NIRA	AG	SG	0,0008	0,9	0,92	<b>0,17</b>	1,66	4,04	0,37	0,3	0,84	0,96	0,7	0,22	1,05

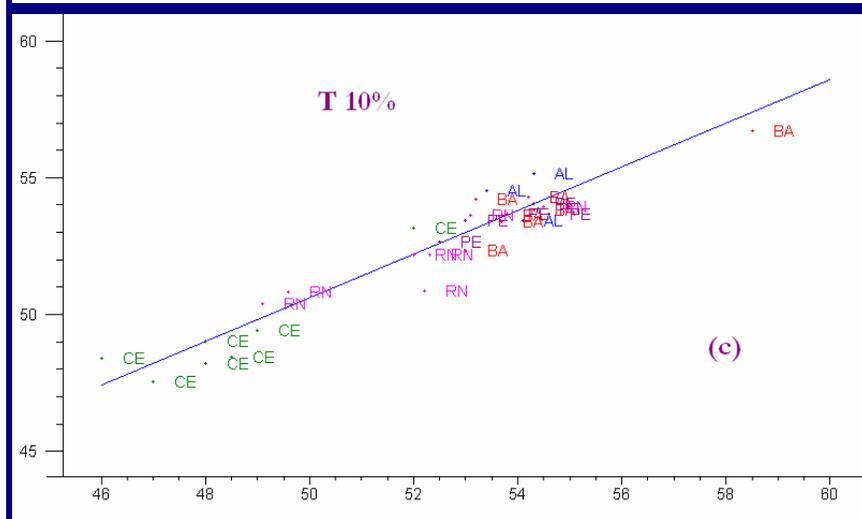
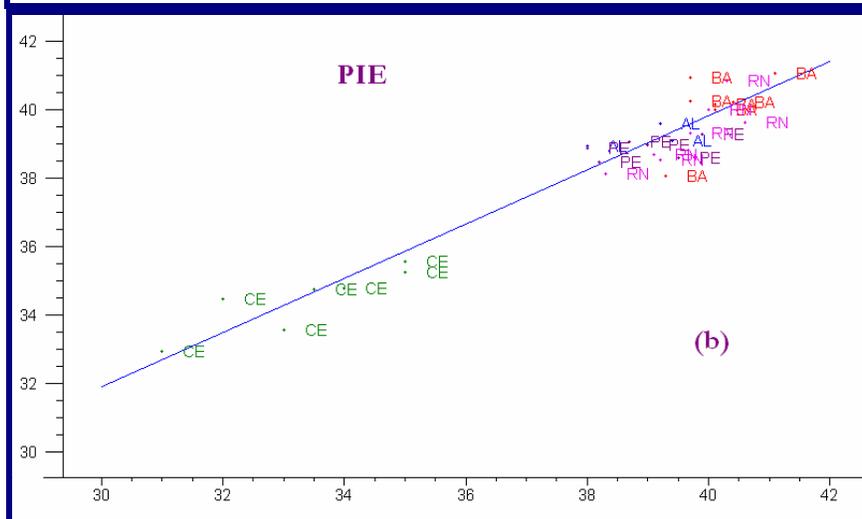
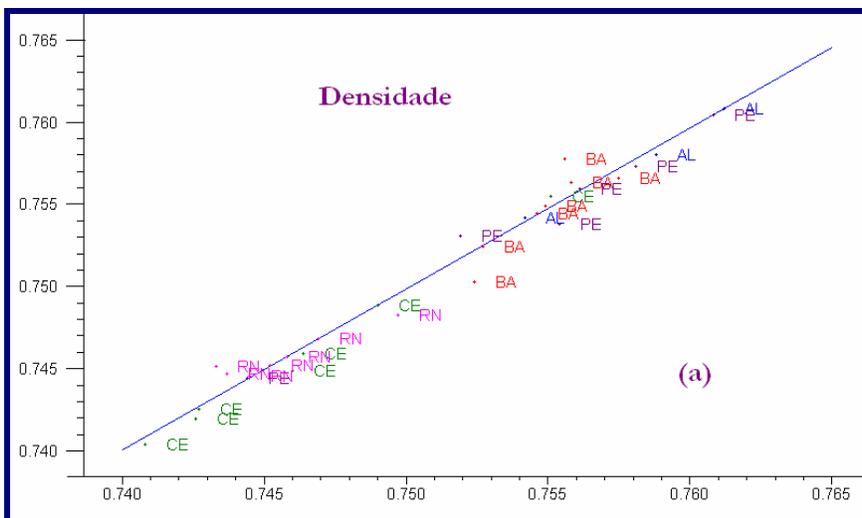
28. NIR 1 ag der sg	MLR	NIRA	AG	DER SG	0,0009	0,93	0,96	0,26	1,55	4,3	0,35	0,32	0,65	0,74	0,8	0,23	1,05
29. NIR 1 aps spp	MLR	NIRA	APS	SEM	0,0011	1,07	0,99	0,29	1,92	4,76	0,37	0,32	0,96	0,96	0,59	0,31	1,18
30. NIR 1 aps der	MLR	NIRA	APS	DER	0,0012	1,28	0,88	0,34	1,58	3,85	0,36	0,32	0,78	1,27	0,74	0,3	1,4
31. NIR 1 aps sg	MLR	NIRA	APS	SG	0,0011	1,15	0,85	0,26	1,89	4,25	0,42	0,34	1,17	1,26	<b>0,58</b>	0,32	1,21
32. NIR 1 aps der sg	MLR	NIRA	APS	DER SG	0,0014	1,31	1,11	0,32	1,87	4,13	0,38	0,33	1,07	1,27	1,01	0,3	1,23
33. NIR 10 ag spp	MLR	NIRB	AG	SEM	0,0014	1,01	1,1	0,37	2,15	5,24	0,33	0,31	0,59	1,01	0,89	0,32	0,99
34. NIR 10 ag der	MLR	NIRB	AG	DER	0,0016	1,06	1,06	0,33	2,18	5,22	0,35	0,29	0,63	0,99	0,89	0,4	<b>0,92</b>
35. NIR 10 ag sg	MLR	NIRB	AG	SG	0,0017	1,1	1,06	0,34	2,08	4,44	0,39	0,28	0,58	1,01	0,83	0,28	0,98
36. NIR 10 ag der sg	MLR	NIRB	AG	DER SG	0,0016	0,99	1,06	0,26	2,21	4,72	0,32	0,28	0,58	0,95	0,9	0,34	0,97
37. NIR 10 aps spp	MLR	NIRB	APS	SEM	0,0024	1,71	2,01	0,54	3,23	9,65	0,8	0,33	0,64	1,21	1,13	0,58	1,29
38. NIR 10 aps der	MLR	NIRB	APS	DER	0,0028	1,66	1,44	0,49	3,73	8,51	0,69	0,42	0,93	1,52	1,07	0,48	1,09
39. NIR 10 aps sg	MLR	NIRB	APS	SG	0,002	1,8	1,55	0,49	3,55	9,28	0,84	0,34	0,61	1,18	1,11	0,57	1,34
40. NIR 10 aps dersg	MLR	NIRB	APS	DER SG	0,0024	1,43	1,29	0,44	2,65	9,41	0,84	0,38	0,59	1,34	1,14	0,95	1,11
41. MID ag spp	MLR	MID	AG	SEM	0,0017	1,43	0,92	0,31	2,6	4,49	0,42	0,3	1,14	1,17	0,85	0,43	1,18
42. MID ag der	MLR	MID	AG	DER	0,0018	1,51	0,88	0,36	2,84	7,91	0,35	0,29	1,22	1,45	0,8	0,33	1,19
43. MID ag sg	MLR	MID	AG	SG	0,0019	1,56	1,02	0,35	2,39	5,97	0,39	0,26	0,77	1,01	0,85	0,51	1,06
44. MID ag der sg	MLR	MID	AG	DER SG	0,002	1,04	0,8	0,33	2,61	6,01	0,34	0,4	1,49	1,25	0,93	0,42	1,18
45. MID aps spp	MLR	MID	APS	SEM	0,0022	1,7	1,39	0,38	2,6	5,29	0,4	0,31	1,15	1,55	0,93	0,53	1,42
46. MID aps der	MLR	MID	APS	DER	0,0019	1,19	1,22	0,42	3,69	4,85	0,58	0,31	1,06	1,71	0,89	0,4	1,18
47. MID aps sg	MLR	MID	APS	SG	0,0021	1,23	1,29	0,45	3,23	5,76	0,4	<b>0,23</b>	1,06	1,54	0,91	0,72	1,45
48. MID aps der sg	MLR	MID	APS	DER SG	0,002	1,11	1,05	0,44	2,86	5,41	0,42	0,33	1,34	1,22	1	0,41	1,29

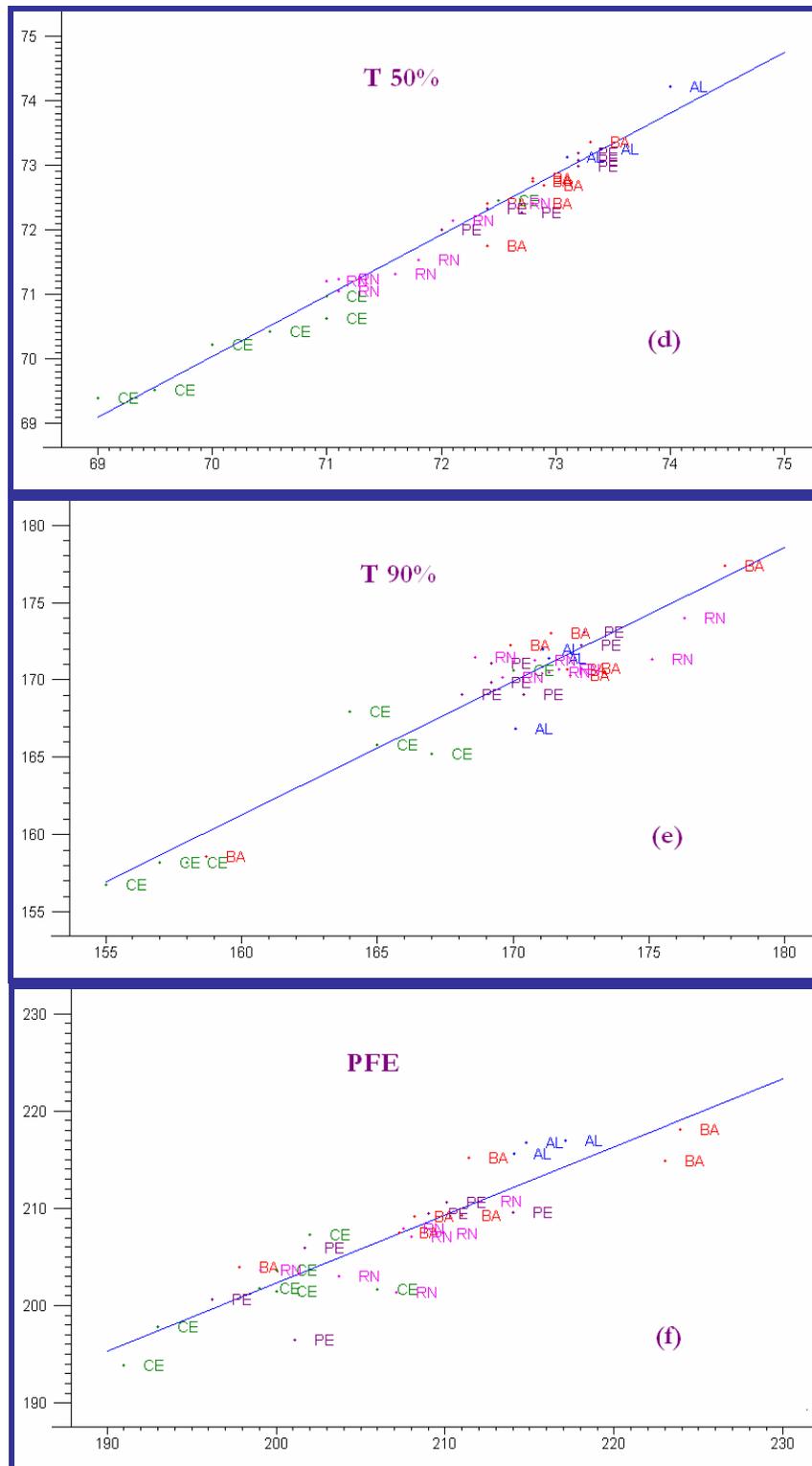
Em **vermelho** o menor valor de RMSEP para cada propriedade.

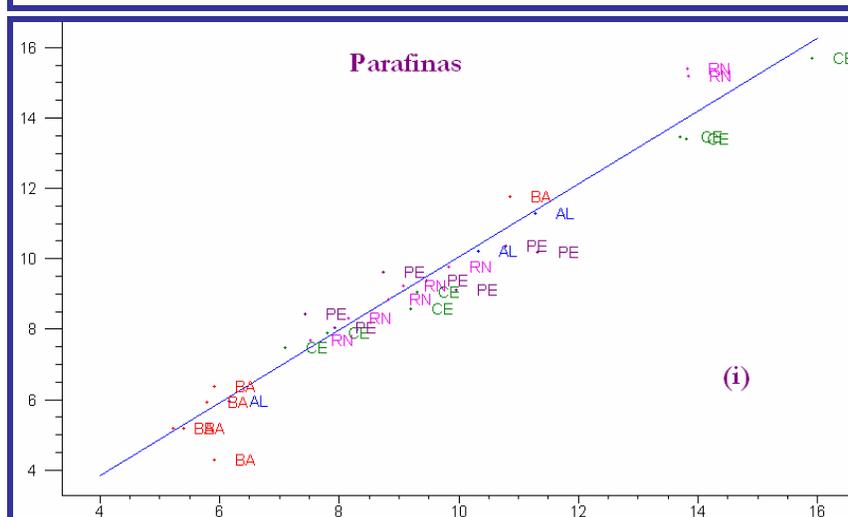
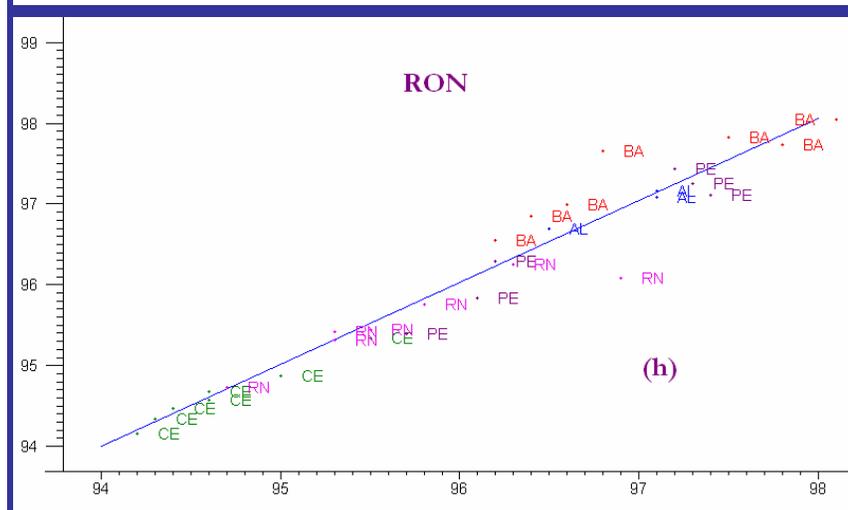
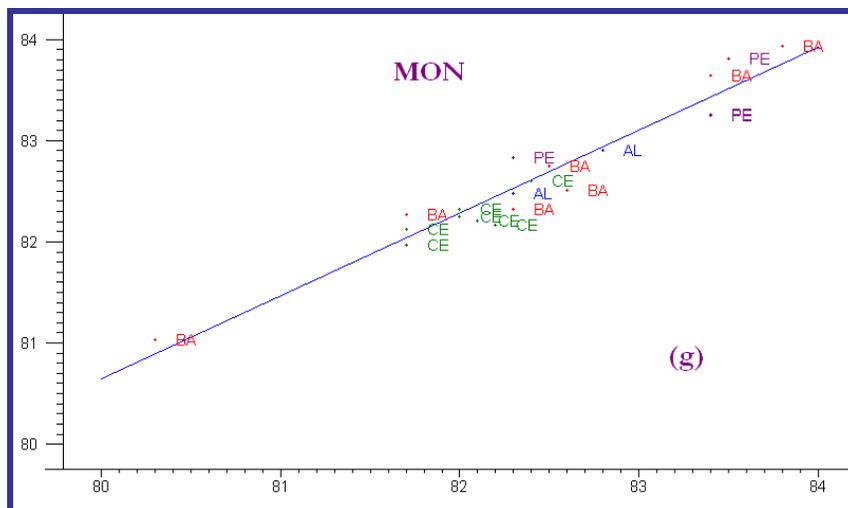
Tabela B - Valores relativos dos RMSEP's obtidos para cada propriedade.

Pré-processamento	Calibração	Faixa	Seleção	Pré-proces.	Densidade	PIE	T10%	T50%	T90%	PFE	MON	RON	P	I	A	N	O
1. NIR 1 spp	PLS	NIRA	SEM	SEM	0,11	2,73	1,84	0,41	0,87	2,03	0,62	0,32	7,82	3,34	4,33	3,88	7,20
2. NIR 1 der	PLS	NIRA	SEM	DER	0,2	3,75	1,58	0,44	1,04	1,86	0,42	0,29	9,07	3,99	5,77	3,74	8,07
3. NIR 1 sg	PLS	NIRA	SEM	SG	0,12	2,78	1,39	0,41	0,96	1,98	0,52	0,32	7,37	3,25	4,26	4,03	7,20
4. NIR 1 der sg	PLS	NIRA	SEM	DER SG	0,12	3,83	1,61	0,44	1,10	2,07	0,35	0,29	11,00	4,64	5,57	3,74	7,82
5. NIR 1 sel spp	PLS	NIRA	AG	SEM	0,16	2,76	1,63	0,44	1,41	2,21	0,44	0,36	7,14	4,43	4,74	4,32	6,52
6. NIR 1 sel der	PLS	NIRA	AG	DER	0,17	3,98	1,58	0,66	0,92	2,11	0,39	0,30	6,58	4,27	6,25	3,45	7,57
7. NIR 1 sel sg	PLS	NIRA	AG	SG	0,23	2,53	2,24	0,44	1,23	1,69	0,57	0,30	8,39	4,23	4,60	3,74	6,39
8. NIR 1 sel der sg	PLS	NIRA	AG	DER SG	0,25	3,04	1,82	0,50	1,68	2,84	0,57	0,28	6,12	3,17	4,95	3,45	6,64
9. NIR 10 spp	PLS	NIRB	SEM	SEM	0,32	4,03	2,72	0,90	2,65	3,65	0,58	0,35	7,48	4,80	6,32	10,50	8,50
10. NIR 10 der	PLS	NIRB	SEM	DER	0,28	3,93	2,66	0,69	2,19	2,77	0,54	0,33	9,64	6,02	7,15	10,21	6,45
11. NIR 10 sg	PLS	NIRB	SEM	SG	0,32	4,21	2,54	0,91	2,76	3,68	0,97	0,38	9,30	4,80	6,60	9,06	8,50
12. NIR 10 der sg	PLS	NIRB	SEM	DER SG	0,32	3,78	2,47	0,94	2,34	2,70	0,98	0,36	7,37	5,41	6,39	9,64	7,01
13. NIR 10 sel spp	PLS	NIRB	AG	SEM	0,2	2,73	2,43	0,78	1,32	2,82	0,45	0,30	8,28	5,41	7,84	6,91	6,08
14. NIR 10 sel der	PLS	NIRB	AG	DER	0,24	3,06	2,07	0,57	1,34	2,72	0,51	0,35	6,35	5,00	6,05	6,91	6,77
15. NIR 10 sel sg	PLS	NIRB	AG	SG	0,33	4,01	2,45	0,66	2,82	3,58	0,57	0,37	9,98	5,53	7,63	6,62	9,00
16. NIR 10 sel der sg	PLS	NIRB	AG	DER SG	0,37	3,60	2,96	0,48	1,27	3,04	0,41	0,33	9,64	5,33	6,32	9,50	6,95
17. MID spp	PLS	MID	SEM	SEM	0,28	2,60	1,86	0,60	1,30	2,61	0,51	0,30	7,82	5,53	7,90	10,50	7,94
18. MID der	PLS	MID	SEM	DER	0,37	2,76	1,94	0,50	1,64	2,62	0,53	0,29	9,07	4,19	6,80	5,61	6,77
19. MID sg	PLS	MID	SEM	SG	0,28	2,65	1,86	0,60	1,27	2,59	0,51	0,27	7,82	5,53	7,97	9,64	7,88
20. MID der sg	PLS	MID	SEM	DER SG	0,36	2,86	1,92	0,50	1,48	2,93	0,58	0,29	7,26	4,27	7,28	6,33	6,95
21. MID sel spp	PLS	MID	AG	SEM	0,19	3,06	1,92	0,47	1,63	2,34	0,51	0,26	17,91	4,96	6,05	6,04	8,50
22. MID sel der	PLS	MID	AG	DER	0,21	3,22	1,78	0,46	1,63	3,51	0,44	0,31	12,47	4,11	5,43	4,32	7,45
23. MID sel sg	PLS	MID	AG	SG	0,27	3,68	2,51	0,53	1,47	3,00	0,54	0,30	8,16	4,23	5,64	7,77	7,39
24. MID sel der sg	PLS	MID	AG	DER SG	0,25	2,93	1,63	0,51	1,36	2,63	0,40	0,35	15,42	5,45	5,91	6,04	6,83
25. NIR 1 ag spp	MLR	NIRA	AG	SEM	0,12	2,17	1,67	0,32	1,03	2,26	0,37	0,29	7,48	2,64	4,60	3,16	6,64
26. NIR 1 ag der	MLR	NIRA	AG	DER	0,13	2,32	1,73	0,32	0,58	2,03	0,31	0,33	7,03	3,34	5,22	3,60	6,83
27. NIR 1 ag sg	MLR	NIRA	AG	SG	0,11	2,30	1,75	0,24	0,99	1,96	0,45	0,31	9,52	3,91	4,81	3,16	6,52
28. NIR 1 ag der sg	MLR	NIRA	AG	DER SG	0,12	2,37	1,82	0,36	0,93	2,09	0,42	0,33	7,37	3,01	5,50	3,31	6,52
29. NIR 1 aps spp	MLR	NIRA	APS	SEM	0,15	2,73	1,88	0,40	1,15	2,31	0,45	0,33	10,88	3,90	4,05	4,46	7,32
30. NIR 1 aps der	MLR	NIRA	APS	DER	0,16	3,27	1,67	0,47	0,95	1,87	0,43	0,33	8,84	5,17	5,08	4,32	8,69

31. NIR 1 aps sg	MLR	NIRA	APS	SG	0,15	2,94	1,61	0,36	1,13	2,06	0,51	0,35	13,26	5,13	3,99	4,60	7,51
32. NIR 1 aps der sg	MLR	NIRA	APS	DER SG	0,19	3,34	2,11	0,44	1,12	2,00	0,46	0,34	12,13	5,17	6,94	4,32	7,63
33. NIR 10 ag spp	MLR	NIRB	AG	SEM	0,19	2,58	2,09	0,51	1,29	2,54	0,40	0,32	6,69	4,11	6,12	4,60	6,14
34. NIR 10 ag der	MLR	NIRB	AG	DER	0,21	2,71	2,01	0,46	1,31	2,53	0,42	0,30	7,14	4,03	6,12	5,76	5,71
35. NIR 10 ag sg	MLR	NIRB	AG	SG	0,23	2,81	2,01	0,47	1,25	2,15	0,47	0,29	6,58	4,11	5,70	4,03	6,08
36. NIR 10 ag der sg	MLR	NIRB	AG	DER SG	0,21	2,53	2,01	0,36	1,32	2,29	0,39	0,29	6,58	3,86	6,18	4,89	6,02
37. NIR 10 aps spp	MLR	NIRB	APS	SEM	0,32	4,36	3,82	0,75	1,94	4,68	0,97	0,34	7,26	4,92	7,77	8,34	8,01
38. NIR 10 aps der	MLR	NIRB	APS	DER	0,37	4,24	2,73	0,68	2,23	4,13	0,83	0,44	10,54	6,18	7,35	6,91	6,76
39. NIR 10 aps sg	MLR	NIRB	APS	SG	0,27	4,59	2,94	0,68	2,13	4,50	1,02	0,35	6,92	4,80	7,63	8,20	8,32
40. NIR 10 aps dersg	MLR	NIRB	APS	DER SG	0,32	3,65	2,45	0,61	1,59	4,57	1,02	0,39	6,69	5,45	7,84	13,67	6,89
41. MID ag spp	MLR	MID	AG	SEM	0,23	3,65	1,75	0,43	1,56	2,18	0,51	0,31	12,92	4,76	5,84	6,19	7,32
42. MID ag der	MLR	MID	AG	DER	0,24	3,85	1,67	0,50	1,70	3,84	0,42	0,30	13,83	5,90	5,50	4,75	7,39
43. MID ag sg	MLR	MID	AG	SG	0,25	3,98	1,94	0,48	1,43	2,90	0,47	0,27	8,73	4,11	5,84	7,34	6,58
44. MID ag der sg	MLR	MID	AG	DER SG	0,27	2,65	1,52	0,46	1,56	2,92	0,41	0,42	16,89	5,08	6,39	6,04	7,32
45. MID aps spp	MLR	MID	APS	SEM	0,29	4,34	2,64	0,53	1,56	2,57	0,48	0,32	13,04	6,30	6,39	7,62	8,81
46. MID aps der	MLR	MID	APS	DER	0,25	3,04	2,32	0,58	2,21	2,35	0,70	0,32	12,02	6,96	6,12	5,75	7,32
47. MID aps sg	MLR	MID	APS	SG	0,28	3,14	2,45	0,62	1,93	2,80	0,48	0,24	12,02	6,26	6,25	10,36	9,00
48. MID aps der sg	MLR	MID	APS	DER SG	0,27	2,83	1,9	0,61	1,71	2,63	0,51	0,34	15,19	4,96	6,87	5,90	8,01









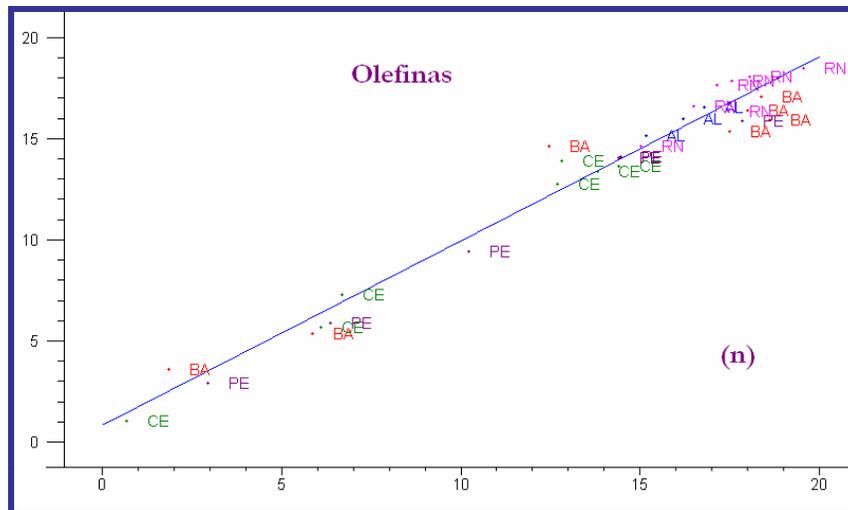


Figura A.1 – Gráfico dos valores previstos x observados das amostras de previsão (a) Densidade, (b) PIE, (c) T 10%, (d) T 50%, (e) T 90%, (f) PFE, (g) MON, (h) RON, (i) Parafinas (j) Isoparafinas, (l) Aromáticos, (m) Naftênicos, (n) Olefinas.