



UFSM

Dissertação de Mestrado

**DETERMINAÇÃO DO NÚMERO DE ACIDEZ TOTAL EM RESÍDUO
DE DESTILAÇÃO ATMOSFÉRICA E DE VÁCUO DO PETRÓLEO
EMPREGANDO A ESPECTROSCOPIA NO INFRAVERMELHO
(ATR-FTIR) E CALIBRAÇÃO MULTIVARIADA**

Graciele Parisotto

PPGQ

Santa Maria, RS - Brasil

2007

**DETERMINAÇÃO DO NÚMERO DE ACIDEZ TOTAL EM RESÍDUO
DE DESTILAÇÃO ATMOSFÉRICA E DE VÁCUO DO PETRÓLEO
EMPREGANDO A ESPECTROSCOPIA NO INFRAVERMELHO
(ATR-FTIR) E CALIBRAÇÃO MULTIVARIADA**

por

Graciele Parisotto

Dissertação apresentada ao Programa de Pós-Graduação em Química,
Área de Química Analítica, da Universidade Federal de Santa Maria (RS),
como requisito parcial para obtenção do grau de **Mestre em Química**

Santa Maria, RS - Brasil

2007

Universidade Federal de Santa Maria
Centro de Ciências Naturais e Exatas
Departamento de Química
Programa de Pós-Graduação em Química

A Comissão Examinadora abaixo assinada, aprova a Dissertação de Mestrado

**DETERMINAÇÃO DO NÚMERO DE ACIDEZ TOTAL EM RESÍDUO DE
DESTILAÇÃO ATMOSFÉRICA E DE VÁCUO DO PETRÓLEO EMPREGANDO A
ESPECTROSCOPIA NO INFRAVERMELHO (ATR-FTIR) E
CALIBRAÇÃO MULTIVARIADA**

elaborada por

GRACIELE PARISOTTO

Como requisito parcial para a obtenção do grau de

Mestre em Química

COMISSÃO EXAMINADORA

Érico Marlon de Moraes Flores – Orientador (UFSM-RS)

Marco Flôres Ferrão (UNISC-RS)

Julio César Magalhães Dias (CENPES-RJ)

Santa Maria, 14 de dezembro de 2007.

AGRADECIMENTOS

Ao **Programa de Pós-Graduação em Química** da Universidade Federal de Santa Maria e ao **Setor de Química Industrial e Ambiental** pela possibilidade de execução deste trabalho.

Ao Prof. Dr. **Érico M. M. Flores**, pela orientação deste trabalho e também pelo exemplo de profissionalismo, pelas oportunidades de crescimento profissional e pessoal, pelo convívio e pela amizade.

Ao Prof. Dr. **Marco F. Ferrão**, pela co-orientação deste trabalho, pela disponibilização dos programas computacionais necessários para o desenvolvimento deste trabalho na UNISC, pela amizade e convívio e mais uma vez pela oportunidade de crescimento profissional e pessoal.

Ao Dr. **Julio César Magalhães Dias**, pela participação na banca examinadora e pelo apoio e sugestões importantes para a melhoria do trabalho.

Aos Profs. Drs. **Valderi L. Dressler e José Neri G. Paniz**, pelos ensinamentos e pela contribuição durante o desenvolvimento das atividades de mestrado, amizade e exemplo profissional.

Ao **CENPES/PETROBRAS**, em especial a **Regina Célia Lourenço Guimarães** e **Maria de Fátima P. dos Santos**, pelo apoio no desenvolvimento deste trabalho.

A todos os integrantes do Setor de Química Industrial e Ambiental do Departamento de Química da UFSM, pelo convívio, amizade e colaboração e, em especial, ao **Edson I. Muller, Fabiana B. Silva, Paola A. Mello, Fabiane R. Bartz** e **Juliana S. F. Pereira** pela colaboração direta para a execução deste trabalho.

Ao **Ademir** e **Valéria** da secretária do PPGQ pela atenção.

Ao **Ricardo**, pelo apoio incondicional, companheirismo e amizade. A seus pais Gelson e Celi, ao meu cunhado Fábio e sua esposa Daniela pelo apoio e momentos de descontração.

Em especial, agradeço a minha **família** pelo suporte e compreensão incondicional, aos meus pais Arlindo e Marilene, aos meus irmãos Audrei, Murilo e Icaro e minhas cunhadas Viviane e Marília.

“Dedico esta conquista a todas as pessoas que me apoiaram, em especial, a minha família e ao meu noivo Ricardo.”

SUMÁRIO

LISTA DE FIGURAS	<i>ix</i>
LISTA DE TABELAS	<i>xii</i>
LISTA DE ABREVIATURAS E SÍMBOLOS	<i>xiv</i>
RESUMO	<i>xvi</i>
ABSTRACT	<i>xvii</i>
1. INTRODUÇÃO	1
2. REVISÃO DA LITERATURA	4
2.1. Petróleo	4
2.1.1. Generalidades.....	4
2.1.2. Propriedades físicas.....	5
2.1.3. Processamento do petróleo	6
2.1.4. Ácidos naftênicos	10
2.2. Espectroscopia no infravermelho associada à calibração	
multivariada	15
2.2.1. Espectroscopia no infravermelho	15
2.2.1.1. Espectroscopia no infravermelho com transformada de	
Fourier	17
2.2.1.2. Reflexão total atenuada	19
2.2.2. Análise multivariada: quimiometria.....	24
2.2.2.1. Tratamento e pré-processamento dos dados.....	25
2.2.2.2. Análise por componentes principais.....	27
2.2.2.3. Análise por agrupamentos hierárquicos.....	29
2.2.2.4. Métodos de calibração multivariada.....	30
2.2.2.4.1. Mínimos quadrados parciais.....	31

2.2.2.4.2. Critérios para avaliação de modelos de regressão	33
2.2.2.4.3. Técnicas de seleção de variáveis em modelos de regressão	36
2.3. Aplicação da espectroscopia no infravermelho combinada com os métodos multivariados para análise de amostras petroquímicas.....	37
3. MATERIAIS E MÉTODOS	40
3.1. Instrumentação	40
3.2. Reagentes	40
3.3. Amostras	41
3.4. Programas de computação	41
3.5. Titulação potenciométrica - Determinação do número de acidez total.....	41
3.6. Espectroscopia no infravermelho - Análise por ATR	42
3.6.1. Aquisição dos espectros	42
3.7. Construção dos modelos - Análise multivariada	44
3.7.1. Detecção de amostras anômalas e seleção das amostras dos conjuntos de calibração e previsão	46
3.7.2. Desenvolvimento dos modelos para a determinação do TAN por ATR-FTIR.....	46
3.7.2.1. Testes preliminares e obtenção do modelo global PLS	47
3.7.2.2. Seleção de variáveis (número de onda).....	47
3.7.2.3. Avaliação dos modelos obtidos.....	47
3.8. Comparação da norma ASTM D 664-04 com a metodologia proposta	48
4. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	49
4.1. Construção dos modelos - Análise multivariada.....	49
4.1.1. Detecção de amostras anômalas e seleção das amostras dos conjuntos de calibração e previsão	49
4.1.2. Desenvolvimento dos modelos para a determinação do TAN por ATR-FTIR.....	53
4.1.2.1. Testes preliminares e obtenção do modelo global PLS	53

4.1.2.2. Determinação do número de variáveis latentes e seleção de variáveis na construção dos modelos	55
4.1.2.2.1. Modelos iPLS	58
4.1.2.2.2. Modelos siPLS	63
4.1.2.2.3. Modelos biPLS	68
4.2. Comparação da norma ASTM D 664-04 com a metodologia proposta	73
5. CONCLUSÃO	75
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	77

LISTA DE FIGURAS

Figura 1.	Possíveis estruturas de ácidos naftênicos.....	11
Figura 2.	Esquema do interferômetro de Michelson	18
Figura 3.	Representação da reflexão interna em um elemento de ATR.....	20
Figura 4.	Representação da propagação da radiação infravermelha através do elemento de ATR	21
Figura 5.	Matriz de dados gerada a partir de um espectro	24
Figura 6.	Princípios da PCA	27
Figura 7.	Esquema da interpretação de um modelo de PCA com uma PC: a) conjunto de dados com dez objetos e três variáveis, b) primeira PC calculada através do conjunto de dados	28
Figura 8.	Seleção do ponto ótimo de LVs que melhor descreve o conjunto de dados multivariados.....	34
Figura 9.	Acessório de HATR acoplado ao espectrômetro de infravermelho: a) acessório sem amostra, para obtenção do sinal de referência, b) acessório com amostra	43
Figura 10.	Perfil dos espectros das amostras de RAT e RV obtidos por ATR-FTIR	44
Figura 11.	Principais etapas desenvolvidas para a obtenção do melhor modelo de calibração	45

- Figura 12.** Gráfico de diagnóstico de amostras anômalas com 95% de confiança para o conjunto total de amostras RAT e RV. As amostras destacadas foram excluídas do conjunto de dados 51
- Figura 13.** Dendograma obtido pela HCA, para a seleção das amostras dos conjuntos de calibração e previsão 52
- Figura 14.** Distribuição das amostras de calibração e previsão para a determinação do TAN de amostras de RAT e RV 53
- Figura 15.** Ilustração do perfil dos espectros das amostras de RAT e RV, depois de aplicada a primeira derivada por filtro de Savitzki-Golay 55
- Figura 16.** Gráfico utilizado para a seleção das VLs em função do RMSECV 56
- Figura 17.** Gráfico utilizado para a seleção das VLs em função do RMSECV 56
- Figura 18.** Representação da divisão das variáveis (número de onda), aplicado para as amostras de RAT e RV 57
- Figura 19.** Valores de referência do TAN vs. valores preditos pelo modelo iPLS60, usando o intervalo 41, para as amostras de RAT e RV 61
- Figura 20.** Valores de RMSECV para o modelo global (linha tracejada) e para os intervalos (barras) na determinação do TAN por ATR-FTIR. Os números correspondem às VLs para cada intervalo 62
- Figura 21.** Erro relativo (%) para as 13 amostras (RAT e RV) do conjunto de previsão dos dados na determinação do TAN por ATR-FTIR utilizando o modelo iPLS selecionado. 63
- Figura 22.** Valores de referência do TAN vs. valores preditos pelo modelo siPLS20, usando os intervalos 7, 14 e 20, para as amostras de RAT e RV 65

- Figura 23.** Valores de RMSECV para o modelo global (linha tracejada) e para os intervalos (barras) na determinação do TAN por ATR-FTIR. Os números correspondem às VLs para cada intervalo..... 66
- Figura 24.** Erro relativo (%) para as 13 amostras (RAT e RV) do conjunto de previsão dos dados na determinação do TAN por ATR-FTIR utilizando o modelo siPLS selecionado 67
- Figura 25.** Valores de referência do TAN vs. valores preditos pelo modelo biPLS20, usando os intervalos 7, 14, 15, 16 e 20, para as amostras de RAT e RV 69
- Figura 26.** Valores de RMSECV para o modelo global (linha tracejada) e para os intervalos (barras) na determinação do TAN por ATR-FTIR. Os números correspondem às VLs para cada intervalo 70
- Figura 27.** Erro relativo (%) para as 13 amostras (RAT e RV) do conjunto de previsão dos dados na determinação do TAN por ATR-FTIR utilizando o modelo biPLS selecionado..... 72
- Figura 28.** Gráfico de correlação entre os resultados da norma ASTM D 664-04 vs. biPLS, obtido através da regressão linear por OLS, com um grau de 95% de confiança..... 74

LISTA DE TABELAS

Tabela 1.	Proporção dos principais constituintes presentes no petróleo.	5
Tabela 2.	Principais produtos obtidos do processamento do petróleo e faixas de destilação de suas respectivas frações.....	8
Tabela 3.	Faixas de composição habitual de petróleo bruto e resíduos.....	10
Tabela 4.	Quantidade de amostra recomendada em função do número de acidez.....	13
Tabela 5.	Região espectral do infravermelho.....	16
Tabela 6.	Principais características das diferentes composições dos elementos de ATR disponíveis.....	23
Tabela 7.	Instrumentação e regiões empregadas na aquisição dos espectros por ATR	43
Tabela 8.	Variância percentual e variância percentual acumulada em função do número de CPs utilizadas.....	50
Tabela 9.	Parâmetros estatísticos dos modelos globais PLS para o TAN, em mg KOH g ⁻¹ , para amostras de RAT e RV.....	54
Tabela 10.	Parâmetros estatísticos dos modelos iPLS, empregando a MSC como tratamento e autoescalado (A) como pré-processamento, para o TAN, em mg KOH g ⁻¹ , para amostras de RAT e RV	58
Tabela 11.	Parâmetros estatísticos dos modelos iPLS, empregando a MSC como tratamento e centrado na média (CM) como pré-processamento, para o TAN, em mg KOH g ⁻¹ , para amostras de RAT e RV	59

Tabela 12. Parâmetros estatísticos dos modelos iPLS, empregando a primeira derivada (D) como tratamento e autoescalado (A) como pré-processamento, para o TAN, em mg KOH g ⁻¹ , para amostras de RAT e RV	59
Tabela 13. Parâmetros estatísticos dos modelos iPLS, empregando a primeira derivada (D) como tratamento e centrado na média (CM) como pré-processamento, para o TAN, em mg KOH g ⁻¹ , para amostras de RAT e RV	60
Tabela 14. Parâmetros estatísticos dos modelos siPLS, empregando a primeira derivada (D) como tratamento e centrado na média (CM) como pré-processamento, para o TAN, em mg KOH g ⁻¹ , para amostras de RAT e RV	64
Tabela 15. Atribuições das regiões selecionadas para o melhor modelo siPLS, aplicado para amostras de RAT e RV.....	67
Tabela 16. Parâmetros estatísticos dos modelos biPLS, empregando a primeira derivada (D) como tratamento e centrado na média (CM) como pré-processamento, para o TAN, em mg KOH g ⁻¹ , para amostras de RAT e RV	68
Tabela 17. Atribuições das regiões selecionadas para o melhor modelo biPLS, aplicado para amostras de RAT e RV.....	71
Tabela 18. Resultado previsto e erro percentual apresentado pelos melhores modelos desenvolvidos empregando os métodos de seleção de variáveis iPLS, siPLS e biPLS, para a determinação do TAN (mg KOH g ⁻¹) por ATR-FTIR de amostras de RAT e RV	73

LISTA DE ABREVIATURAS E SÍMBOLOS

1. API, do inglês *American Petroleum Institute*.
2. ASTM, do inglês *American Society for Testing and Materials*.
3. ATR, reflexão total atenuada, do inglês *attenuated total reflection*.
4. ATR-FTIR, espectroscopia no infravermelho médio com transformada de Fourier e reflexão total atenuada, do inglês *Fourier transform infrared spectroscopy with attenuated total reflection*.
5. biPLS, mínimos quadrados parciais por exclusão, do inglês *backward interval partial least squares*.
6. CM, centrado na média.
7. CP, componente principal.
8. DTGS, detector de sulfato de triglicina deuterada.
9. FIR, infravermelho distante, do inglês *far IR*.
10. FT, transformada de Fourier, do inglês *Fourier transform*.
11. FTIR, espectroscopia no infravermelho com transformada de Fourier, do inglês *Fourier transform infrared spectroscopy*.
12. GLP, gás liquefeito de petróleo.
13. H/C, razão entre hidrogênio e carbono.
14. HCA, análise por agrupamentos hierárquicos, do inglês *hierarchical cluster analysis*.
15. HATR, reflexão total atenuada horizontal, do inglês *horizontal attenuated total reflection*.
16. iPLS, mínimos quadrados parciais por intervalo, do inglês *interval partial least squares*.
17. IR, espectroscopia na região do infravermelho, do inglês *infrared spectroscopy*.
18. MLR, regressão linear múltipla, do inglês *multiple lineal regression*.
19. MSC, correção do espalhamento de luz, do inglês *multiplicative scatter correction*.
20. NA, ácidos naftênicos, do inglês, *naphthenic acids*.
21. NIR, infravermelho próximo, do inglês *near IR*.

22. OLS, regressão dos mínimos quadrados ordinários, do inglês *ordinary least-squares regression*.
23. P.A., para análise.
24. PCA, análise por componentes principais, do inglês *principal component analysis*.
25. PCR, regressão por componentes principais, do inglês *principal component regression*.
26. PLS, mínimos quadrados parciais, do inglês *partial least squares*.
27. R^2 , coeficiente de determinação.
28. RAT, resíduo atmosférico (resíduo de destilação do petróleo sob pressão atmosférica).
29. RMSE, raiz quadrada dos erros médios, do inglês *root mean square error*.
30. RMSEC, raiz quadrada do erro médio da calibração, do inglês *root mean square error of calibration*.
31. RMSECV, raiz quadrada do erro médio da validação cruzada, do inglês *root mean square error of cross-validation*.
32. RMSEP, raiz quadrada do erro médio de previsão do inglês *root mean square error of prediction*.
33. RMSEV, raiz quadrada do erro médio da validação, do inglês *root mean square error of validation*.
34. RSEP erro padrão de predição relativo, do inglês *relative standard error of prediction*.
35. RV, resíduo de vácuo (resíduo de destilação do petróleo a vácuo).
36. siPLS, mínimos quadrados parciais por sinergismo de intervalos, do inglês *synergy partial least squares*.
37. TAN, número de acidez total, do inglês *total acid number*.
38. VL, variáveis latentes.

RESUMO

O petróleo, além de hidrocarbonetos, pode apresentar uma série de contaminantes que podem levar a sua desqualificação, tais como os ácidos naftênicos (NA), os quais são classificados entre os principais causadores de corrosão nas refinarias. O controle da acidez total em produtos do petróleo é feito através da determinação do número de acidez total (TAN) e é padronizada pela norma ASTM D 664-04, baseada na titulação potenciométrica em meio não-aquoso. Como alternativa a metodologia oficial, o TAN para amostras de resíduo da destilação atmosférica (RAT) e de resíduo da destilação a vácuo (RV) do petróleo, foi determinado utilizando a espectroscopia no infravermelho médio com transformada de Fourier e reflectância total atenuada (ATR-FTIR), associada a métodos quimiométricos. Para o desenvolvimento dos modelos de calibração, foi feita a detecção de amostras anômalas através da análise por componentes principais (PCA), seguida da seleção das amostras para os conjuntos de calibração e previsão, com o auxílio da análise por agrupamentos hierárquicos (HCA). O conjunto de calibração consistiu de 44 amostras e o de previsão de 13, totalizando 16 amostras de RAT e 41 de RV. Os modelos de calibração foram desenvolvidos através do emprego de três métodos de seleção de variáveis: o método dos mínimos quadrados parciais por intervalo (iPLS), o método dos mínimos quadrados parciais por sinergismo de intervalos (siPLS) e o método dos mínimos quadrados parciais por exclusão (biPLS). Foram avaliados diferentes tratamentos e pré-processamentos para o desenvolvimento dos modelos. O tratamento da primeira derivada com o filtro de Savitzky-Golay e os dados centrados na média produziram os melhores modelos bem como o algoritmo biPLS, onde o espectro foi dividido em 20 intervalos e 5 intervalos foram combinados (2992 a 2826, 1823 a 1657, 1656 a 1490, 1489 a 1323, 821 a 655 cm^{-1}). Este modelo forneceu erro quadrático médio de validação cruzada (RMSECV) e erro quadrático médio de previsão (RMSEP) de 0,1649 mg KOH g^{-1} e de 0,1642 mg KOH g^{-1} , respectivamente, apresentando coeficiente de determinação de 0,9819 e um erro médio de 22,46%. O método analítico para a determinação do TAN em amostras de RAT e RV do petróleo proposto possibilita uma análise rápida e relativamente de baixo custo, sendo de fácil implementação no ambiente industrial.

ABSTRACT

The crude oil might contain a series of contaminants that can to produce undesired properties. Naphthenic acids (NA), for example, perhaps promotes corrosion in refineries. The determination of total acid number (TAN) in petroleum products is recommended by ASTM standard D 664-04, based on potentiometric titration in non aqueous media. As an alternative to official methodology, in this work, TAN in atmospheric residue and vacuum residue samples of petroleum distillation was determined using middle infrared spectroscopy with attenuated total reflection (ATR-FTIR) in association to chemometric methods. For the development of calibration models, at first, was made the detection of outliers samples using principal component analysis (PCA). After, the selection of samples for calibration and prediction set was made by hierarchical cluster analysis (HCA). Calibration set was consisted of 44 samples and prediction one by 13, summarizing 16 samples of atmospheric residue and 41 samples of vacuum residue. Calibration models were developed using three variable selection models: method of interval partial least squares (iPLS), method of synergy partial least squares (siPLS) and method of backward interval partial least squares (biPLS). Different treatments and pre-processing were evaluated for models development also. The treatment based on first derivative with Savitzky-Golay filter and the data centered in the media produced the best models using biPLS. Spectra were divided in 20 intervals and, finally, 5 intervals were combined (2992 till 2826, 1823 till 1657, 1656 till 1490, 1489 till 1323 and 821 till 655 cm^{-1}). This model produced a root mean square error of cross-validation (RMSECV) of 0.1649 mg KOH g^{-1} and a root mean square error of prediction (RMSEP) of 0.1642 mg KOH g^{-1} , showing a coefficient of determination of 0.9819 and a medium error of 22.5%. The analytical method for TAN determination allows fast analysis and relatively low cost in these samples, being of easy application in industry environment.

1. INTRODUÇÃO

O petróleo é de extrema importância na sociedade moderna pois, além de prover os diversos derivados petroquímicos é utilizado como combustível em indústrias e nos meios de transporte.^{76c} Na indústria petroquímica há um maior interesse no petróleo leve, pois este é de mais fácil processamento, acarretando em menor custo operacional. Entretanto, nos últimos anos, as reservas mundiais de petróleos leves, vêm diminuindo e, cada vez mais, torna-se necessário o processamento de petróleos pesados.¹¹ No Brasil, em especial, o petróleo bruto tem procedência de muitos poços distribuídos ao longo do território brasileiro que produzem, predominantemente, petróleo pesado.²⁹

Deve-se destacar que o refino de petróleos pesados e extra-pesados leva à produção de elevadas quantidades de resíduo da destilação atmosférica (RAT) e resíduo da destilação a vácuo (RV). Estes resíduos requerem temperaturas de processamento maiores, podendo provocar maior corrosão das tubulações, já que esta aumenta com o incremento da temperatura. Tanto o RAT quanto o RV apresentam alta viscosidade, mesmo sob aquecimento, e apresentam-se como sólidos a temperatura ambiente, o que dificulta seu processamento e manipulação.^{20,51}

Além de hidrocarbonetos, o petróleo pode ser constituído de contaminantes e compostos que desqualificam as suas propriedades, tais como compostos de enxofre e ácidos naftênicos (NA).^{76d} Por sua vez, os NA são caracterizados por uma mistura de hidrocarbonetos naftênicos (ou cicloparafínicos)

^{76c,d} Speight, J. G.; *Handbook of Petroleum Analysis*, 2001, 11-18, 25-41.

¹¹ Bathia, S.; Sharma, D. K.; *Pet. Sci. Technol.* 24 (2006) 1125-1159.

²⁹ Falla, F. S.; *et al.*, *Pet. Sci. Eng.* 51 (2006) 127-137.

²⁰ Chung, H.; Ku, M.; *Appl. Spectrosc.* 54 (2000) 239-245.

⁵¹ Kraft, W. W.; *Ind. Eng. Chem.* 40 (1948) 807-819.

com cadeia lateral alquílica e grupamento carboxílico.^{58b} A presença de NA é uma das principais causas de corrosão nas tubulações do processo de refino e ocorre nas unidades de destilação, onde as temperaturas situam-se na faixa de 220 a 400 °C.⁴⁷ Consequentemente, óleos com altas concentrações de NA são considerados de má qualidade e possuem baixo valor de mercado.⁸⁶

Atualmente, o controle da acidez total em produtos do petróleo é feito através da norma ASTM D 664-04,⁶ que utiliza a titulação potenciométrica em meio não aquoso. Esta é expressa em miligramas de hidróxido de potássio necessários para neutralizar 1 grama de amostra. Alguns trabalhos descrevem que petróleos com acidez maior que 0,5 mg KOH g⁻¹ já podem ser considerados prejudiciais para as tubulações das refinarias.^{8,75} Entretanto, tem sido relatado que a norma ASTM D 664-04 pode apresentar problemas quanto à reprodutibilidade, quanto a detecção do ponto de equivalência, massa de amostra utilizada, e possibilidade de precipitação de asfaltenos.³⁵ Além disso, devido à utilização de solventes, são gerados resíduos que aumentam a quantidade de efluentes prejudiciais, quando em contato com o meio ambiente.

A espectroscopia na região do infravermelho (IR), com a utilização da transformada de Fourier (FT) e acessórios de reflexão, fornecem algumas vantagens em relação aos demais métodos analíticos, como a redução do uso de solventes, menor pré-tratamento da amostras e maior rapidez na análise. Com o advento dos microprocessadores digitais e de técnicas computacionais avançadas, os dados gerados pelo IR passaram a ser tratados com ferramentas quimiométricas, possibilitando a análise de misturas complexas, através do tratamento de dados multivariados, correlacionando inúmeras variáveis simultaneamente. Os métodos de análise multivariada podem ser divididos em métodos de análise exploratória, de classificação e de regressão e têm por objetivo melhorar a qualidade dos resultados e facilitar a sua interpretação, tornando possível o tratamento de dados mais complexos do ponto de vista matemático e

^{58b} Matar, S.; *Chemistry of Petrochemical Process*, 2000, 130-131.

⁴⁷ Jayaraman, A.; Singh, H.; Lefebvre, Y.; *Rev. Inst. Fr. Pet.* 41 (1986) 265-274.

⁸⁶ Zhang, A.; *et al.*, *Appl. Catal. A* 303 (2006) 103-109.

⁶ Annual Book of ASTM Standards, D 664-04, 2004.

⁸ Babaian-Kibala, E.; *et al.*, *Mater. Perform.* 32 (1993) 50-55.

⁷⁵ Slavcheva, E.; Shone, B.; Turnbull, A.; *Br. Corr. J.* 34 (1999) 125-131.

³⁵ Fuhr, B.; *et al.*, *Energy Fuels* 21 (2007) 1322-1324.

estatístico.¹⁵ Como conseqüência, tem-se a vantagem de utilizar, simultaneamente, uma quantidade maior de informações, para a obtenção dos dados analíticos.¹⁹

Na literatura podem ser encontrados trabalhos que utilizam a espectroscopia na região do infravermelho e ferramentas quimiométricas para determinar diversos parâmetros do petróleo e seus produtos, como a determinação da composição, da densidade e a estabilidade das emulsões do petróleo.^{2,7,29,42,48,65}

Desta forma, devido à aplicabilidade das ferramentas quimiométricas, neste trabalho é proposta a determinação do número de acidez total em resíduo da destilação atmosférica e em resíduo da destilação a vácuo do petróleo, utilizando a espectroscopia no infravermelho médio com transformada de Fourier e reflexão total atenuada (ATR-FTIR) combinada às técnicas de calibração multivariada. Para tal procedimento foram utilizados três algoritmos de seleção de variáveis, o método dos mínimos quadrados parciais por intervalo,⁶¹ o método dos mínimos quadrados parciais por sinergismo de intervalos⁶⁰ e o método dos mínimos quadrados parciais por exclusão.⁵⁶

Cabe ressaltar que, apesar das possíveis vantagens, não foram encontradas na literatura aplicações destes métodos quimiométricos para a determinação do número de acidez total (TAN) em frações de petróleo, especialmente, para o RAT e o RV, que é o foco do presente trabalho.

¹⁵ Breton, R. G.; *Analyst* 125 (2000) 2125-2154.

¹⁹ Chau, F.; et al., *Chemometrics: From basics to wavelet transform*, 2004, 1-21.

² Akhlaq, M. S.; *J. Pet. Sci. Eng.* 22 (1999) 229-235.

⁷ Aske, N.; Kallevik, H.; Sjöblom, J.; *J. Pet. Sci. Eng.* 36 (2002) 1-17.

²⁹ Falla, F. S.; et al., *Pet. Sci. Eng.* 51 (2006) 127-137.

⁴² Hannisdal, A.; Hemmingsen, P. V.; Sjöblom, J.; *Ind. Eng. Chem. Res.* 44 (2005) 1349-1357.

⁴⁸ Kallevik, H.; Kvalheim, O. M.; Sjöblom, J.; *J. Colloid Interface Sci.* 225 (2000) 494-504.

⁶⁵ Pasquini, C.; Bueno, A. F.; *Fuel* 86 (2007) 1927-1934.

⁶¹ Norgaard, L.; Saudland, A.; *Appl. Spectrosc.* 54 (2000) 413-419.

⁶⁰ Munck, L.; et al., *Anal. Chim. Acta* 446 (2001) 171-186.

⁵⁶ Leardi, R.; Norgaard, L.; *J. Chemometrics* 18 (2004) 486-497.

2. REVISÃO DA LITERATURA

2.1. Petróleo

2.1.1. Generalidades

O petróleo pode ser classificado de acordo com os hidrocarbonetos que o compõem em: parafínicos, naftênicos (ou cicloparafínicos) e aromáticos. Os parafínicos contêm, principalmente, hidrocarbonetos saturados, sem anéis aromáticos, com cadeia linear ou ramificada. Geralmente, a proporção de hidrocarbonetos parafínicos diminui com o aumento do ponto de ebulição da respectiva fração. Por exemplo, a gasolina é composta por cerca de 80% de constituintes parafínicos, enquanto que nos óleos lubrificantes estes são cerca de 30%.^{76e}

Os constituintes naftênicos (assim chamados por sua presença no petróleo ocorrer a partir da nafta) são constituídos em sua maioria por hidrocarbonetos saturados com um ou mais anéis e cadeias laterais. A principal variação estrutural é o número de anéis presentes na molécula e, normalmente, as frações de petróleo apresentam cinco ou seis anéis, sendo estes mais estáveis. Hidrocarbonetos naftênicos podem representar até, aproximadamente, 60% do total de hidrocarbonetos presentes no petróleo.^{76e}

A presença de compostos aromáticos é comum em todos os petróleos. Geralmente, apresentam um ou mais anéis, como benzeno, naftaleno e fenantreno, podendo apresentar em sua estrutura também cadeias parafínicas e anéis naftênicos. Em geral, quando há o aumento do peso molecular, ocorre o aumento da proporção de hidrocarbonetos aromáticos.^{76e}

^{76e} Speight, J. G.; *Handbook of Petroleum Analysis*, 2001, 31-39.

Além de hidrocarbonetos, o petróleo pode apresentar alguns heteroátomos como, nitrogênio, oxigênio, enxofre, além constituintes metálicos, particularmente vanádio, níquel, ferro e cobre. De acordo com sua origem, podem ocorrer diferentes características relativas à cor, odor, ponto de ebulição, densidade e viscosidade.^{4,76a}

As variações nas propriedades do petróleo ocorrem, devido às proporções em que os diferentes constituintes estão presentes, o que depende da sua origem, da idade do reservatório e da profundidade da extração. Na Tabela 1 são apresentadas as proporções dos principais constituintes geralmente encontrados para o petróleo bruto.^{76d}

Tabela 1. Proporção dos principais constituintes presentes no petróleo.^{76d}

Constituinte (%)	Proporção
Carbono	83 - 87
Hidrogênio	10 - 14
Nitrogênio	0,1 - 2,0
Oxigênio	0,05 - 1,5
Enxofre	0,05 - 6,0
Metais	< 0,1

2.1.2. Propriedades físicas

Em função de sua composição e de suas propriedades físicas (ponto de ebulição e densidade) o petróleo pode ser classificado em leve ou pesado.^{76a} Petróleo leve ou petróleo convencional é rico em constituintes de baixos pontos de ebulição. Já o petróleo pesado contém grandes proporções de moléculas de alto ponto de ebulição, compostos aromáticos e heteroátomos.^{77a} Geralmente, petróleos leves apresentam 97% (m/m) de hidrocarbonetos, enquanto os petróleos pesados podem apresentar cerca de somente 50% (m/m) de hidrocarbonetos.^{76d} O

⁴ Annual Book of ASTM Standards, D 4175-96, 1996.

^{77a} Speight, J. G.; Handbook of Petroleum Product Analysis, 2002, 2.

^{76a,d} Speight, J. G.; Handbook of Petroleum Analysis, 2001, 1, 8, 25-41, 72-75.

petróleo pesado é mais viscoso que o petróleo leve e tem uma baixa mobilidade nos reservatórios, o que dificulta sua extração, bem como seu processamento.^{76b}

Além da caracterização da composição, o petróleo extraído das diferentes reservas da superfície terrestre deve, também, ser caracterizado quanto a suas propriedades físicas, como o ponto de ebulição das frações que o compõem, densidade ou °API e a viscosidade.⁵ A forma mais usual de classificar o petróleo e seus produtos é a densidade (ou gravidade específica), que significa a massa de líquido por unidade de volume a 15 °C e 101,325 kPa (com unidade de medida em kg cm⁻³), baseado no grau API, segundo o *American Petroleum Institute Gravity*, que pode ser definido pela equação 1.³

$$^{\circ}API = \frac{141,5}{d_{15,56^{\circ}C}} - 131,5 \quad (1)$$

onde, $d_{15,56^{\circ}C}$ é a densidade relativa do petróleo a 15,56 °C.^{77b} Valores inferiores a 20 °API, indicam que o petróleo possui densidade elevada (óleo pesado), geralmente na faixa de 10 a 15 °API. Quando os valores são superiores a 20 °API, isto indica que o petróleo possui densidade menor (neste caso, o petróleo é classificado como óleo leve).^{76g} O °API pode ser estendido para outras frações do petróleo, tais como os resíduos de destilação. Nesse caso, para o resíduo atmosférico são encontrados, normalmente, valores de 10 a 15 °API e para o resíduo de vácuo de 2 a 8 °API.^{76g}

2.1.3. Processamento do petróleo

Uma vez que, o petróleo pesado e seus resíduos não apresentam as mesmas características correspondentes a petróleos leves, os processos convencionais de refino não são adequados para o tratamento dessas cargas pesadas. Assim é essencial o conhecimento de seus constituintes, para adequar os processos químicos e físicos a que estes serão submetidos.^{76d}

^{76b,d,g} Speight, J. G.; *Handbook of Petroleum Analysis*, 2001, 8, 25-41, 72-75.

⁵ Annual Book of ASTM Standards, D 445-97, 1997.

³ Annual Book of ASTM Standards, D 1298-99, 1999.

^{77b} Speight, J. G.; *Handbook of Petroleum Product Analysis*, 2002, 34-38.

De um modo geral, uma refinaria é composta por redes integradas de unidades de processos com o objetivo de produzir uma variedade de produtos do petróleo. Cada refinaria tem sua própria linha de processamento do petróleo, levando em conta as características do petróleo disponível para o processamento, o tipo de equipamento disponível, a necessidade dos produtos formados e os custos envolvidos. De acordo com essas necessidades, o processo de refino pode ser dividido, basicamente, em três etapas: separação, conversão e acabamento.^{77c}

No processo de separação ocorre a divisão do petróleo em diferentes frações, de acordo com sua composição. Na etapa de conversão ocorre a produção de derivados do petróleo, geralmente por alteração da estrutura carbônica ou da função química dos compostos envolvidos. Na última etapa, o tratamento, ocorre a purificação das frações do petróleo por uma variedade de processos que, essencialmente, são responsáveis pela remoção de impurezas.^{77c}

A destilação é a principal unidade de processamento da refinaria e, através dela, o petróleo bruto é separado em diferentes frações, de acordo com seus pontos de ebulição para, após, submeter cada fração destilada a uma unidade de processamento definida, para que sejam atendidas as especificações finais de cada produto.³⁷ Na Tabela 2 são apresentados alguns dos produtos obtidos do processamento do petróleo e as faixas de destilação de suas respectivas frações.

^{77c} Speight, J. G.; Handbook of Petroleum Product Analysis, 2002, 51-58.

³⁷ Gary, J. H.; Hadwerk, G. E.; *Petroleum refining: technology and economics*, 2001, 37-57.

Tabela 2. Principais produtos obtidos do processamento do petróleo e faixas de destilação de suas respectivas frações.^{77b}

Produto	Faixa de composição	Faixa de destilação (°C)
Gás natural	C ₁ a C ₄	- 61 a -1
Gás liquefeito de petróleo	C ₃ a C ₄	- 42 a -1
Gasolina	C ₅ a C ₁₇	- 1 a 216
Nafta	C ₄ a C ₁₂	36 a 302
Querosene e diesel	C ₈ a C ₁₈	126 a 258
Combustível de aviação	C ₈ a C ₁₆	126 a 287
Óleo combustível	C ₁₂ a >C ₂₀	216 a 421
Cera	C ₁₇ a >C ₂₀	302 a >343
Óleo lubrificante	>C ₂₀	>343
Asfalto	>C ₂₀	>343
Coque	>C ₅₀	>1000

Antes da destilação atmosférica, o petróleo é aquecido e enviado a uma unidade dessalgadora, onde é feita a remoção de sal e água. Após a etapa de dessalgação, o petróleo é levado a uma torre de pré-fracionamento, para a separação de gás combustível, gás liquefeito de petróleo (GLP) e nafta.^{37,58a}

Através de destilação à pressão atmosférica, vários produtos são obtidos de acordo com seus pontos de ebulição, como nafta leve, média e pesada, querosene, diesel leve e pesado, e demais frações que destilem até temperaturas em torno de 399 °C. O resíduo deste processo, denominado de resíduo atmosférico, pode ser utilizado como óleo combustível ou transferido para a unidade de destilação à vácuo.^{37,58a,75}

O RAT resultante da destilação atmosférica alimenta uma torre de destilação a vácuo. O vácuo é aplicado, porque as temperaturas de ebulição diminuem com a redução da pressão evitando o craqueamento térmico das

^{77b} Speight, J. G.; Handbook of Petroleum Product Analysis, 2002, 34-38.

³⁷ Gary, J. H.; Hadwerk, G. E.; *Petroleum refining: technology and economics*, 2001, 37-57.

^{58a} Matar, S.; *Chemistry of Petrochemical Process*, 2000, 49-111.

⁷⁵ Slavcheva, E.; Shone, B.; Turnbull, A.; *Br. Corr. J.* 34 (1999) 125-131.

moléculas de alto ponto de ebulição, perdas dos produtos gerados, além de evitar danos aos equipamentos. A pressão de operação é de, aproximadamente, 25 a 40 mmHg e são aplicadas temperaturas na faixa de 400 a 450 °C. Nesta etapa da destilação é promovida a separação da maioria dos compostos pesados, como o asfalto, e um resíduo conhecido como resíduo de vácuo.^{37,58a,75}

Após os processos de destilação, são aplicados os processos de conversão por aquecimento térmico ou por ação de catalisadores (craqueamento térmico e craqueamento catalítico, respectivamente). Nestes processos ocorrem modificações do número de átomos de carbono por molécula, alteração da razão hidrogênio-carbono ou modificação da estrutura molecular do material. Este processo é importante porque permite obter frações leves, como nafta e gás liquefeito de petróleo a partir de frações pesadas, como gasóleo pesado ou resíduo atmosférico. Também podem ser aplicadas etapas de hidroprocessamento, onde podem ser removidos os contaminantes na presença de hidrogênio com auxílio de catalisadores (hidrotratamento) e decomposição térmica na presença de hidrogênio (hidrocraqueamento).^{77c}

No processo de destilação do petróleo, são gerados resíduos que são constituídos de moléculas que não volatilizam facilmente. Estes resíduos têm concentrações elevadas de impurezas, como, enxofre, nitrogênio, oxigênio, níquel, vanádio e outros elementos traços. Estas impurezas caracterizam propriedades únicas ao petróleo, exigindo atenção especial, para que se possa tornar estes derivados do petróleo passíveis de comercialização.⁷¹

O resíduo obtido da destilação a vácuo é constituído, basicamente, por compostos com ponto de ebulição superior a 550 °C. Este resíduo apresenta baixa razão H/C, alta viscosidade e, para petróleos pesados, pode representar até 50% (m/m) da carga original.^{49,87}

Na Tabela 3 é possível observar a composição do petróleo bruto e seus resíduos.

³⁷ Gary, J. H.; Hadwerk, G. E.; *Petroleum refining: technology and economics*, 2001, 37-57.

^{58a} Matar, S.; *Chemistry of Petrochemical Process*, 2000, 49-111.

⁷⁵ Slavcheva, E.; Shone, B.; Turnbull, A.; *Br. Corr. J.* 34 (1999) 125-131.

^{77c} Speight, J. G.; *Handbook of Petroleum Product Analysis*, 2002, 51-58.

⁷¹ Robbins, W. K.; *Metals, nitrogen and sulfur in petroleum residue, analysis of*, 2006, 1-11.

⁴⁹ Khan, Z. H.; *et al., Fuel Sci. Technol. Int'L.* 12 (1994) 1413-1424.

⁸⁷ Zhao, L.; *et al., Energy Fuels* 20 (2006) 2067-2071.

Tabela 3. Faixas de composição habitual de petróleo bruto e resíduos.⁷¹

Elemento	Petróleo bruto	Resíduo atmosférico	Resíduo de vácuo
Carbono (%)	83 - 87	86 - 88	83 - 86
Hidrogênio (%)	10 - 14	10 - 13	9 - 13
Enxofre (%)	0,04 - 6,0	0,05 - 7,0	0,1 - 12
Nitrogênio (%)	0,02 - 1,0	0,05 - 1,2	0,1 - 1,5
Oxigênio (%)	0,05 - 1,5	0,2 - 2,5	0,5 - 3,0
Níquel (ppm)	0,1 - 100	0,5 - 200	1 - 300
Vanádio (ppm)	0,1 - 500	0,2 - 800	0,5 - 1000

2.1.4. Ácidos naftênicos

Um dos principais problemas relacionados ao refino do petróleo é a corrosão associada aos constituintes ácidos.⁵³ Para melhor entendimento da origem desta corrosão deve-se ter conhecimento das espécies responsáveis por tal ação, como, água, enxofre, sais inorgânicos de cloro e, em especial, os ácidos naftênicos (NA).^{46,79} A corrosão por ácidos naftênicos é assim chamada tendo em vista que nas temperaturas empregadas no refino, os ácidos naftênicos se tornam potencialmente corrosivos, principalmente para as tubulações do processo de refino.⁷⁵

De maneira geral, pequenas quantidades de NA estão presentes no petróleo e suas frações, sendo normalmente encontrados valores de até 3%. A acidez naftênica é expressa em miligramas de hidróxido de potássio suficientes para neutralizar a acidez de um grama de óleo. A medida de acidez é chamada de número de acidez total (TAN) do óleo.^{6,75} Normalmente, a corrosão por ácidos naftênicos ocorre quando o petróleo e suas frações apresentam acidez maior que $0,5 \text{ mg KOH g}^{-1}$.^{8,75} Outros estudos indicam que o limiar para que ocorra a corrosão

⁷¹ Robbins, W. K.; Metals, nitrogen and sulfur in petroleum residue, analysis of, 2006, 1-11.

⁵³ Laredo, G. C.; *et al.*, *Fuel* 83 (2004) 1689-1695.

⁴⁶ Jayaraman, A.; Saxena, R. C.; *Corros. Prevent. Contr.* 42 (1995) 123-131.

⁷⁹ Tomczyk, N. A.; *et al.*, *Energy Fuels* 15 (2001) 1498-1504.

⁷⁵ Slavcheva, E.; Shone, B.; Turnbull, A.; *Br. Corr. J.* 34 (1999) 125-131.

⁶ Annual Book of ASTM Standards, D 664-04, 2004.

⁸ Babaian-Kibala, E.; *et al.*, *Mater. Perform.* 32 (1993) 50-55.

é entre 1,5 a 2 mg KOH g⁻¹ de óleo.⁴⁵ Acredita-se que os NA contribuem com a corrosão das tubulações na faixa compreendida entre 220 a 400 °C.⁴⁷

Em geral, os NA são caracterizados por uma mistura de hidrocarbonetos naftênicos (ou cicloparafínicos) com cadeia lateral alquílica e grupamento carboxílico.^{58b} Apresentam fórmula geral R[CH₂]_n COOH, onde R é, usualmente, um anel de ciclopentano ou ciclohexano, e normalmente, n é maior que 12. Na Figura 1 são apresentadas algumas possíveis estruturas de ácidos naftênicos.^{75,86}

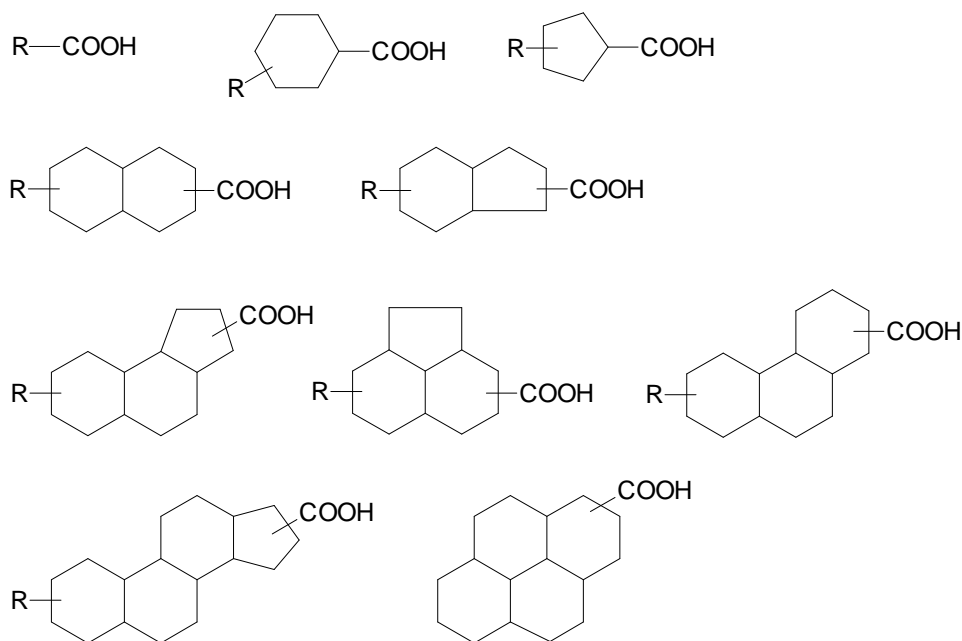


Figura 1. Possíveis estruturas de ácidos naftênicos.⁵³

Ácidos naftênicos de baixo peso molecular possuem comumente de 8 a 12 carbonos, contendo anéis de ciclopentanos ou ciclohexanos com uma cadeia lateral carboxialquílica e são, normalmente, encontrados em produtos da destilação média, como querosene e gasóleo. Em contrapartida, NA que apresentam ponto

⁴⁵ Humphries, M.; Sorrel, G.; *Mater. Perform.* 15 (1976) 13-21.

⁴⁷ Jayaraman, A.; Singh, H.; Lefebvre, Y.; *Rev. Inst. Fr. Pet.* 41 (1986) 265-274.

^{58b} Matar, S.; *Chemistry of Petrochemical Process*, 2000, 130-131.

⁷⁵ Slavcheva, E.; Shone, B.; Turnbull, A.; *Br. Corr. J.* 34 (1999) 125-131.

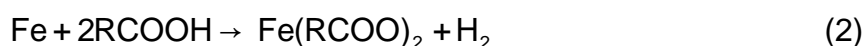
⁸⁶ Zhang, A.; *et al., Appl. Catal. A* 303 (2006) 103-109.

⁵³ Laredo, G. C.; *et al., Fuel* 83 (2004) 1689-1695.

de ebulição elevados são ácidos monocarboxílicos com 14 a 19 carbonos, com uma média de 2,6 anéis, e geralmente presentes nos óleos lubrificantes.^{58b}

Variações significativas de corrosividade podem ser obtidas para óleos com os mesmos valores de TAN, dependendo da estrutura química dos ácidos naftênicos. Sobretudo, o TAN expressa uma medida de acidez e não a acidez relativa aos ácidos naftênicos.⁷⁵

A quantidade de enxofre presente no petróleo é outro fator importante na corrosão por ácidos naftênicos, principalmente devido à competição entre os dois tipos de processos: o ataque dos ácidos naftênicos e do sulfeto de hidrogênio, de acordo com as reações 2, 3 e 4.⁷⁵



A reação 2 representa o ataque dos ácidos naftênicos sobre a tubulação metálica, enquanto que a reação 3 representa a corrosão por sulfeto de hidrogênio. Uma diferença significativa é que, o produto da corrosão, o naftenato de ferro é muito solúvel em óleo, deixando a superfície da tubulação livre para que a reação de corrosão continue ocorrendo, enquanto que o sulfeto de ferro tende a formar um filme protetor que se deposita sobre a tubulação. Já a reação 4 representa o caso onde o sulfeto de hidrogênio reage com o naftenato de ferro produzindo sulfeto de ferro, que precipita no óleo. Esta reação regenera a acidez naftênica.⁷⁵ Dessa forma, petróleos com TAN relativamente altos e baixa quantidade de enxofre, serão mais corrosivos a altas temperaturas que aqueles com os mesmos ácidos naftênicos e alta quantidade de enxofre. Estima-se que para o enxofre formar uma camada protetora, o petróleo precisa ter de 2 a 3% de enxofre.⁴⁷

A norma que padroniza a determinação do número de acidez total é a ASTM D 664-04.⁶ Esta norma utiliza a titulação potenciométrica em meio não-aquoso, com eletrodo íon-seletivo com solução interna de cloreto de lítio em etanol,

^{58b} Matar, S.; Chemistry of Petrochemical Process, 2000. 130-131.

⁷⁵ Slavcheva, E.; Shone, B.; Turnbull, A.; *Br. Corr. J.* 34 (1999) 125-131.

⁴⁷ Jayaraman, A.; Singh, H.; Lefebvre, Y.; *Rev. Inst. Fr. Pet.* 41 (1986) 265-274.

⁶ Annual Book of ASTM Standards, D 664-04, 2004.

apropriado para este meio. A titulação em solvente não-aquoso é utilizada quando, os reagentes ou produtos são insolúveis em água, os reagentes ou produtos podem reagir com a água e quando o titulado é uma base ou um ácido muito fraco para ser titulado em água.⁴³

A norma ASTM D664-04 determina constituintes ácidos em produtos de petróleo e lubrificantes, solúveis ou quase solúveis, em misturas de tolueno e propan-2-ol. O método é aplicado na determinação de ácidos com constante de dissociação na água maior que 10^{-9} . Sais com constante de hidrólise maiores que 10^{-9} reagem como ácidos interferindo na determinação de acidez total.⁶

Os constituintes de caráter ácido determinados no método incluem: ácidos orgânicos (entre eles os ácidos naftênicos) e inorgânicos, ésteres, compostos fenólicos, lactonas, resinas, sais de metais pesados, sais de amônio, outras bases fracas, sais ácidos de ácidos polibásicos e inibidores e detergentes, também são medidos como ácidos.⁶

Segundo a norma ASTM D 664-04, para se realizar o procedimento de titulação deve-se pesar a quantidade de amostra em função do número de acidez, de acordo com a Tabela 4, e proceder a solubilização em 125 mL de uma mistura de solventes contendo 50% de tolueno, 45% de álcool iso-propílico e 5% de água.

Tabela 4. Quantidade de amostra recomendada em função do número de acidez.⁶

Número de acidez	Massa (g)
0,05 a <1,0	20,0 ± 2,0
1,0 a <5,0	5,0 ± 0,5
5 a <20	1,0 ± 0,1
20 a <100	0,25 ± 0,02
100 a <260	0,10 ± 0,01

A robustez da norma foi estudada por Fuhr *et al.*³⁵ variando diferentes parâmetros a fim de verificar as possíveis influências no valor de TAN e na precipitação de asfaltenos. Foram estudados os seguintes parâmetros: proporção

⁴³ Harris, C. D.; *Análise Química Quantitativa*, 1999, 225-248.

⁶ Annual Book of ASTM Standards, D 664-04, 2004.

³⁵ Fuhr, B.; *et al.*, *Energy Fuels* 21 (2007) 1322-1324.

dos solventes, concentrações do titulante, quantidade de amostra titulada e a relação do TAN com a quantidade de ácidos naftênicos presentes. A amostra foi previamente aquecida, e após, diluída em tolueno até completa dissolução (1 a 2 g de amostra). Em seguida, completou-se o com o solvente indicado pela norma e titulou-se a amostra num prazo máximo de 30 minutos. Foi observado que a concentração do titulante e a quantidade de amostra pesada não influenciam no valor de TAN, bem como possíveis precipitações de asfaltenos parecem não afetar os resultados. Também, foi observado que em proporções de solvente onde a presença de tolueno era majoritária não era observada a precipitação de asfaltenos. Quanto a quantidade de acidez presente nas amostras, excelentes recuperações foram obtidas (utilizando solução de referência para ácidos naftênicos, Fluka), confirmando que os resultados de TAN são devido a ácidos orgânicos (naftênicos), pois em óleos mais leves o TAN pode ser referente a outras espécies, como sulfeto de hidrogênio. O pré-tratamento da amostra possibilitou o uso da proporção de solventes recomendada pela norma ASTM. Ainda, os resultados mostraram que quantidades de amostra abaixo de 1 g podem resultar em instabilidade na leitura de alguns eletrodos, enquanto que a utilização de 5 g pode resultar no encobrimento do eletrodo, impedindo uma resposta adequada. A baixa reprodutibilidade dos resultados utilizando a norma ASTM D664-04 foi observada em relação a oito laboratórios. Após as mudanças sugeridas, utilizando o pré-tratamento, foi obtida uma boa reprodutibilidade.

Poucos trabalhos envolvendo a determinação do TAN foram encontrados na literatura. Laredo *et al.*⁵³ determinaram e verificaram a variação do TAN e a quantidade de enxofre ao longo da curva de destilação, através das normas recomendadas pela ASTM, para caracterizar dois óleos mexicanos. Para isso, inicialmente, foram destiladas as frações do óleo para, depois, separar os ácidos naftênicos presentes utilizando-se a cromatografia iônica e espectrometria de massa para sua identificação. Os autores concluíram que, aparentemente, um dos óleos era mais corrosivo que o outro, possivelmente, devido a presença de ácidos naftênicos de baixo peso molecular e baixa quantidade de enxofre, observadas em temperaturas abaixo de 200 °C. Este comportamento indica a falta da camada de proteção fornecida pelo enxofre.

⁵³ Laredo, G. C.; *et al.*, *Fuel* 83 (2004) 1689-1695.

No que diz respeito a utilização da espectroscopia no infravermelho em estudos envolvendo a acidez em produtos de petróleo, a literatura também é escassa. Zhang *et al.*⁸⁶ verificaram a eficiência da descarboxilação catalítica dos componentes carboxílicos presentes em óleos utilizando óxido de magnésio. A efetividade da reação de descarboxilação foi monitorada com o auxílio da espectroscopia de infravermelho. Ao longo da reação a acidez foi acompanhada através da determinação do TAN. Para melhor observação das mudanças do estiramento COOH ($\sim 1700\text{ cm}^{-1}$) dos NA durante a reação, foi adicionada uma quantidade suficiente de um ácido carboxílico, já que a amostra tinha pouca quantidade dos mesmos.

Barth *et al.*¹⁰ utilizaram o TAN (através da titulação potencimétrica), a espectroscopia no infravermelho médio com reflexão difusa e a cromatografia para determinar as diferenças dos ácidos orgânicos de óleos biodegradados e não biodegradados de doze amostras de óleo. Os espectros no infravermelho destas amostras indicaram que ácidos extraídos de óleos biodegradados são constituídos, predominantemente, por ácidos carboxílicos alifáticos, enquanto que óleos não biodegradados apresentam, em sua maioria, ácidos com características fenólicas.

2.2. Espectroscopia no infravermelho associada à calibração multivariada

2.2.1. Espectroscopia no infravermelho

A região do espectro eletromagnético correspondente ao infravermelho se estende de, aproximadamente, $12800\text{ a }10\text{ cm}^{-1}$ ($0,78\text{ }\mu\text{m a }1000\text{ }\mu\text{m}$). Esta região é dividida em: infravermelho distante ou longínquo (FIR, do inglês, *Far IR*), infravermelho médio (MIR, do inglês, *Mid IR*) e infravermelho próximo (NIR, do inglês, *Near IR*) (Tabela 5).⁷⁴

⁸⁶ Zhang, A.; *et al.*, *Appl. Catal. A* 303 (2006) 103-109.

¹⁰ Barth, T.; *et al.*, *Org. Geochem.* 35 (2004) 1513-1525.

⁷⁴ Skoog, D.; Holler, F. J.; Nieman, T. A.; *Principles of instrumental analysis*, 1998, 380-403.

Tabela 5. Região espectral do infravermelho.⁷⁴

Região	Faixa de número de onda (cm ⁻¹)	Faixa de comprimento de onda (μm)
Próximo	12800 a 4000	0,78 a 2,5
Médio	4000 a 200	2,5 a 50
Distante ou longínquo	200 a 10	50 a 1000

A radiação no infravermelho não possui energia suficiente para promover transições eletrônicas como a radiação ultravioleta e visível. A radiação no infravermelho está restrita a espécies moleculares que apresentam pequena diferença de energia entre os diversos estados vibracionais ou rotacionais. Portanto, para que uma molécula absorva radiação no infravermelho é necessário que exista alteração do momento de dipolo elétrico como consequência de seu movimento vibracional ou rotacional (o momento de dipolo é determinado pela magnitude da diferença de carga e a distância entre dois centros de carga), resultando em absorção da energia radiante.⁷⁴

As vibrações provocadas por absorção da radiação no infravermelho podem ser classificadas em deformações axiais (estiramentos) e deformações angulares. Uma vibração de deformação axial envolve uma mudança contínua na distância interatômica ao longo do eixo da ligação entre dois átomos, aumentando e diminuindo alternadamente, podendo ser de dois tipos: o estiramento simétrico (v_s) que provoca um movimento harmônico da distância interatômica e o estiramento assimétrico (v_{as}) que promove um movimento desarmônico da distância interatômica.⁷⁴

As vibrações de deformação angular são caracterizadas pela mudança do ângulo entre duas ligações e podem acontecer no plano (δ) ou fora do plano (γ) da molécula. São comuns quatro tipos de vibrações fundamentais, que apresentam denominações características (em relação ao movimento): deformação angular simétrica no plano (ρ , balanço) e fora do plano (ω , sacudida), e deformação assimétrica no plano (δ , tesoura) e fora do plano (τ , torção).⁷⁴

O funcionamento de um espectrômetro de infravermelho é similar ao de outros espectrômetros. A radiação em comprimentos de onda específica é gerada

⁷⁴ Skoog, D.; Holler, F. J.; Nieman, T. A.; *Principles of instrumental analysis*, 1998, 380-403.

e essa radiação entra em contato com a amostra, gerando absorções em comprimentos de onda específicos, correspondendo ao tipo de composto, ligação e interação existentes na amostra.⁴³ Até o início dos anos 1980, a maioria dos espectrômetros de infravermelho eram do tipo dispersivo. Estes espectrômetros estão em desuso, já que são extremamente lentos para gerar os espectros. Atualmente os espectrômetros são equipados com transformada de Fourier sendo que, seus principais aspectos serão abordados a seguir.^{9a}

2.2.1.1. Espectroscopia no infravermelho com transformada de Fourier

A utilização da espectroscopia no infravermelho médio, como técnica para análise quantitativa, deve-se à combinação da transformada de Fourier (FT) e da nova geometria dos espectrômetros com a utilização do interferômetro de Michelson. Estes fatores, aliados ao crescimento exponencial da informática, possibilitaram o desenvolvimento de novos algoritmos, tornando os espectrômetros mais rápidos e robustos.⁷⁴

O interferômetro (Figura 2) consiste, basicamente, de dois espelhos planos, posicionados perpendicularmente um ao outro, sendo um deles fixo e o outro móvel. Um espelho semitransparente (denominado divisor de feixes) é alinhado com a fonte de radiação. Este divisor tem a função de transmitir 50% da radiação para o espelho móvel e 50% é refletida no espelho fixo. Em seguida, os dois raios são refletidos por esses espelhos, retornando ao divisor de feixes, onde se recombina e sofre interferência. Novamente, 50% da radiação que chega ao divisor de feixes é refletida de volta em direção à fonte. O raio que emerge do divisor de feixes em direção à amostra, e em seguida ao detector, é denominado radiação transmitida.⁸⁰

⁴³ Harris, C. D.; *Análise Química Quantitativa*, 1999, 225-248.

^{9a} Barbosa, L. C. A.; *Espectroscopia no infravermelho: na caracterização de compostos orgânicos*, 2007, 25-43.

⁷⁴ Skoog, D.; Höller, F. J.; Nieman, T. A.; *Principles of instrumental analysis*, 1998, 380-403.

⁸⁰ Voort, F. R.; *Food Res. Int.* 25 (1992) 397-403.

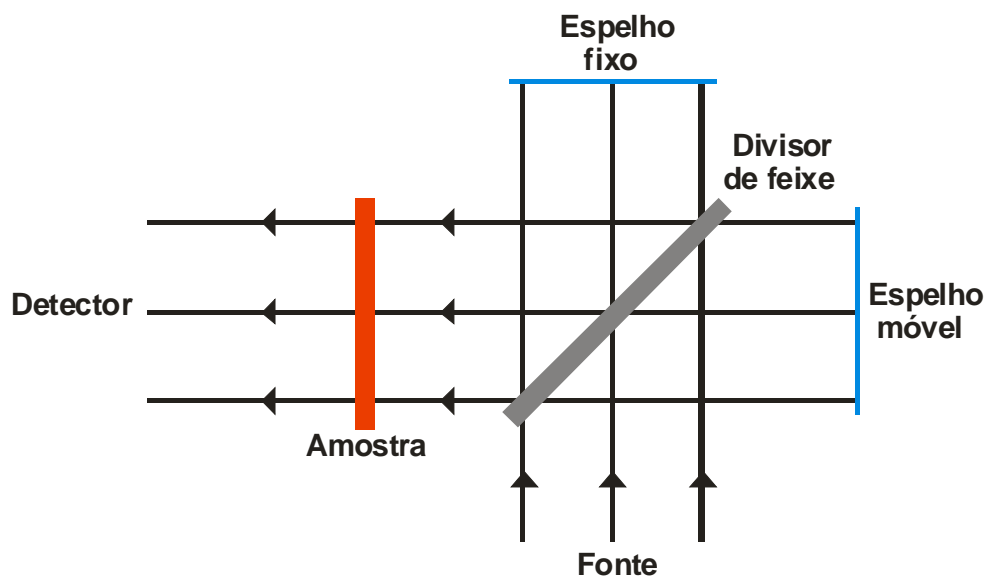


Figura 2. Esquema do interferômetro de Michelson.⁸⁰

Desta forma, para obtenção de um espectro no infravermelho utilizando o método interferométrico, primeiro adquire-se o sinal de fundo (*background*), obtendo um interferograma sem a amostra e em seguida a amostra é posicionada entre o interferômetro e o detector. A amostra absorverá alguns comprimentos de onda, reduzindo, assim, as intensidades destes. Posteriormente, utiliza-se um procedimento matemático, a FT para a obtenção de um espectro convencional, passível de interpretação.^{9a,80} Dentre as principais vantagens da FTIR podem ser citadas, a melhoria da razão sinal-ruído numa dada resolução, a obtenção de espectros em poucos segundos, espectros com maior resolução e maior precisão nas medidas de número de onda, até $0,01 \text{ cm}^{-1}$.⁴³

Os métodos mais usados para obtenção de espectros no infravermelho envolvem transmissão e reflexão. Em ambos os casos, existem vários acessórios disponíveis comercialmente para a obtenção de espectros no infravermelho de amostras sólidas, líquidas e gasosas.^{9a}

O método de transmissão é o mais antigo e tido como clássico na obtenção de espectros no infravermelho. Neste método a radiação passa através da amostra, sendo parte absorvida e parte transmitida, onde as amostras devem ser colocadas em uma célula, cuja janela seja transparente na região de trabalho do infravermelho. Para fins quantitativos, este método pode apresentar dificuldades

⁸⁰ Voort, F. R.; *Food Res. Int.* 25 (1992) 397-403.

^{9a} Barbosa, L. C. A.; *Espectroscopia no infravermelho: na caracterização de compostos orgânicos*, 2007, 25-43.

⁴³ Harris, C. D.; *Análise Química Quantitativa*, 1999, 225-248.

de utilização, na etapa de confecção das pastilhas, pois o material utilizado nas janelas dessas células, deve ser compatível com a amostra e com a análise que se pretende realizar.^{9a}

Quanto às técnicas de reflexão, estas vêm aumentando sua aplicabilidade, particularmente, porque elas não envolvem processos morosos de preparo de amostras, sendo úteis tanto para análises qualitativas quanto quantitativas, facilitando o estudo de amostras de difícil preparo para análise por transmissão. Elas permitem a obtenção de espectros da maior parte das substâncias sólidas e líquidas. As técnicas de reflexão mais utilizadas para obtenção de espectros são: reflexão especular, reflexão difusa e reflexão total atenuada.^{9a} Tendo em vista, que para a realização deste trabalho foi utilizado o acessório de reflexão total atenuada, a seguir serão abordadas características relevantes a esta técnica.

2.2.1.2. Reflexão total atenuada

A reflexão total atenuada (ATR), também conhecida por espectroscopia de reflexão interna, é uma técnica de amostragem rápida que requer mínima preparação da amostra. Além disso, permite a obtenção de espectros de materiais espessos e fortemente absorventes, difíceis de serem analisados por espectroscopia por transmissão, possibilitando a obtenção de espectros por reflexão com boa qualidade e tornando possível sua utilização em diversas aplicações analíticas.⁶²

A técnica surgiu da observação de que quando a radiação no infravermelho atravessa um prisma, produzido com material de alto índice de refração (seleneto de zinco ou germânio, entre outros materiais) sob determinado ângulo, ocorre reflexão total dessa energia. A luz infravermelha refletida pelo prisma consegue interagir com a superfície de materiais colocados em íntimo contato com este prisma. Nas frequências em que tal material absorve energia ocorre a atenuação de intensidade, responsável pela geração do espectro. A principal vantagem da técnica é que, por refletir a radiação, praticamente dispensa preparo de amostra, permitindo a análise de sólidos e de líquidos.⁵⁹ Dessa forma, a técnica de ATR se

^{9a} Barbosa, L. C. A.; *Espectroscopia no infravermelho: na caracterização de compostos orgânicos*, 2007, 25-43.

⁶² Olinger, J. M.; Griffiths, P. R.; *Anal. Chem.* 60 (1988) 2427-2435.

⁵⁹ Mirabella, F. M.; *Appl. Spectrosc. Rev.* 21 (1985) 45-178.

caracteriza pelas múltiplas reflexões da radiação infravermelha que ocorrem no interior de cristais, de materiais com alto índice de refração, interagindo apenas com a amostra que estiver superficialmente no cristal.³⁰

A Figura 3 ilustra o fenômeno da reflexão interna através do emprego de um cristal adequado ATR. Este cristal deve ser composto de um material com alto índice de refração, para que somente uma pequena parte do feixe de radiação incidente seja refletido ao atingir o cristal.³⁰

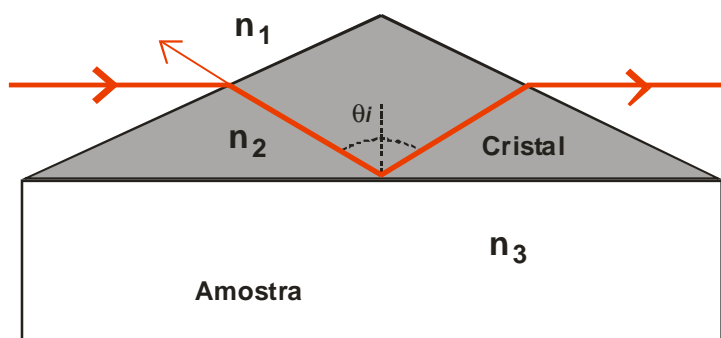


Figura 3. Representação da reflexão interna em um elemento de ATR.³⁰

As técnicas espectroscópicas por reflexão interna²⁸ só podem ser aplicadas sob certas condições. No momento em que a radiação entra em um prisma com um alto índice de refração em relação ao meio externo (cristal de ATR) a radiação será totalmente refletida internamente, o que pode ser observado através da Figura 3, na qual o feixe de radiação encontra-se no interior do elemento de ATR. Desta forma, a técnica baseia-se no fato de que, embora ocorra reflexão interna total na interface dos dois meios a radiação, na verdade, penetra determinada distância no meio menos denso (amostra). Essa radiação denominada onda evanescente pode ser parcialmente absorvida, colocando-se uma amostra em contato com o meio mais denso (prisma ou elemento de reflexão interna) (Figura 4).

Para um índice de refração tal que $n_2 > n_3$, aplicando-se a lei de Snell o ângulo de refração é imaginário para valores do ângulo de incidência tal que satisfaça a equação 5.

$$\theta_i > \text{sen}^{-1}(n_3 / n_2) \quad (5)$$

³⁰ Ferrão, M. F.; Tese de Doutorado. Universidade Estadual de Campinas, Campinas-SP, 2000.

²⁸ Fahrenfort, J.; *Spectrochim. Acta* 17 (1961) 698-709.

O ângulo acima do qual esta onda refratada deixa de ser real é denominado de ângulo crítico (θ_c) e é expresso conforme equação 6.

$$\theta_i = \text{sen}^{-1}(n_3/n_2) \quad (6)$$

Nestas condições o feixe é totalmente refletido na interface n_3/n_2 .

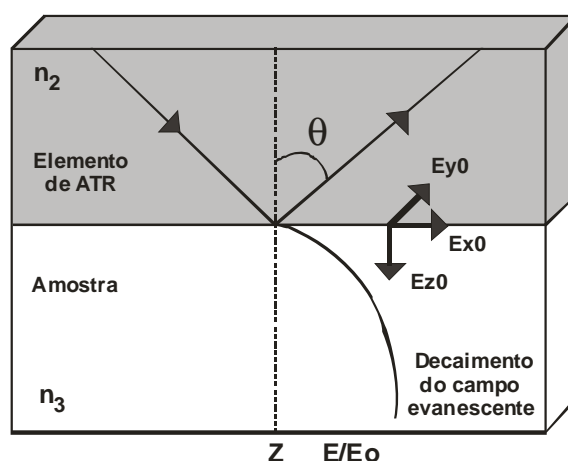


Figura 4. Representação da propagação da radiação infravermelha através do elemento de ATR.⁵⁹

A profundidade de penetração da radiação no segundo meio pode ser quantitativamente definida em termos da distância requerida para que a intensidade do campo elétrico decaia de um fator de $1/e$ do valor inicial, sendo calculada conforme equação 7.^{9b}

$$D_p = \frac{\lambda}{2\pi n_1 \sqrt{\text{sen}^2 \theta \left(n_3 / n_2 \right)^2}} \quad (7)$$

Em que:

D_p = profundidade de penetração, usualmente em μm ;

λ = comprimento de onda, usualmente em μm ;

n_1 = índice de refração do elemento de reflexão interna;

⁵⁹ Mirabella, F. M.; *Appl. Spectrosc. Rev.* 21 (1985) 45-178.

^{9b} Barbosa, L. C. A.; *Espectroscopia no infravermelho: na caracterização de compostos orgânicos*, 2007, 38-41.

n_3/n_2 = razão entre os índices de refração da amostra n_3 e do elemento de reflexão interna n_2 ;

θ = ângulo de incidência da radiação.

A equação 7 mostra que a profundidade de penetração depende do comprimento de onda da radiação no meio com índice de refração mais alto (elemento de ATR).

Geralmente, a profundidade de penetração está na faixa de 0,05 a 0,2 λ , e com isso o caminho óptico varia de 0,25 a 4 μm , dependendo do comprimento de onda (isto é, proporcional a profundidade) e do material do elemento de ATR. Desta forma, qualquer material que esteja em contato com o cristal de ATR pode absorver a radiação incidente atenuando sua intensidade, dando origem ao espectro infravermelho. Em função disto, a composição da superfície pode ser estudada e o espectro torna-se independente da espessura da amostra.³¹

Dois parâmetros controláveis são o ângulo de incidência e a escolha do elemento de ATR. A espessura efetiva e a profundidade de penetração aumentam, quando o ângulo aproxima-se do ângulo crítico, podendo-se variar a profundidade de penetração, ajustando o ângulo de incidência. Contudo, o número de reflexões também varia com o ângulo de incidência.⁵⁹

Existem inúmeros elementos de ATR e uma variedade de acessórios desenvolvidos para obter espectros de materiais líquidos, sólidos e viscosos para uma larga faixa de tipos de amostras, incluindo filmes, resíduos, papel, revestimentos sobre papel, pós, tintas, tecidos, espumas, minerais, vidros, etc. Alguns possibilitam o aquecimento da amostra e os mais recentes permitem a análise em fluxo. Na Tabela 6 são apresentadas as diferentes composições dos elementos de ATR disponíveis, e características a eles relacionadas, sendo determinante para sua escolha a amostra que se deseja analisar. Um dos materiais mais comumente utilizados é o seleneto de zinco.⁵⁹

³¹ Ferrão, M. F.; *Tecno-lógica* 5 (2001) 65-85.

⁵⁹ Mirabella, F. M.; *Appl. Spectrosc. Rev.* 21 (1985) 45-178.

Tabela 6. Principais características das diferentes composições dos elementos de ATR disponíveis.*

Composição	Dureza, kg/mm ²	Faixa espectral, cm ⁻¹	Índice de reflexão	Profundidade de penetração 45°, 1000 cm ⁻¹ , μ	Faixa de pH
AMTIR	170	11000 - 630	2,5	1,7	1 - 9
Diamante / KRS-5	5700	30000 - 250	2,4	2	1 - 14
Diamante / ZnSe	5700	30000 - 525	2,4	2	1 - 14
Ge	780	5500 - 570	4,0	0,66	1 - 14
Si / ZnSe	1150	8900 - 550	3,4	0,85	1 - 12
Si	1150	8900 - 1500 475 - 40	3,4	0,85	1 - 12
ZnSe	120	15000 - 520	2,4	2	5 - 9

* Adaptado da Ref. 68.

Um problema geralmente encontrado é a dificuldade de se obter boa reprodutibilidade no contato da amostra, principalmente, com o elemento de ATR. Em amostras sólidas, este efeito é observado na variação da intensidade das bandas com a pressão aplicada. Com o incremento da pressão, a eficiência de contato é aumentada e, conseqüentemente, as intensidades das bandas também aumentam.¹⁸ A área de contato entre o cristal e a amostra é um fator que também influencia na intensidade das bandas. Para medidas quantitativas, deve-se colocar toda a área do cristal em contato com a amostra para garantir uma boa reprodutibilidade. Entretanto, irregularidades na superfície da amostra podem dificultar um contato mais efetivo. Por outro lado, este efeito pode ser útil, por fornecer informações sobre as propriedades ou condições da superfície de um material, incluindo identificação, modificações físicas e adsorção superficial.⁵⁹

⁶⁸ PIKE Technologies, <http://www.piketech.com>, acessado em 23/09/07.¹⁸ Carlsson, D. J.; Wiles, D. M.; *Can. J. Chem.* 48 (1970) 45-178.⁵⁹ Mirabella, F. M.; *Appl. Spectrosc. Rev.* 21 (1985) 45-178.

2.2.2. Análise multivariada: quimiometria

O termo quimiometria foi proposto em meados de 1970 por Wold e Kowalski para descrever as técnicas e operações associadas com a manipulação e interpretação de dados químicos. A quimiometria é uma ferramenta para a automação laboratorial e está relacionada com a análise multivariada.^{19,82,83} De um modo geral, a análise multivariada refere-se aos métodos estatísticos e matemáticos que analisam, simultaneamente, múltiplas medidas de um objeto sob investigação, seja ele de caráter químico ou não.⁴¹

As ferramentas quimiométricas, aliadas a tecnologia computacional e instrumental, viabilizam a manipulação de dados de absorbância espectral associados a mais de uma frequência ao mesmo tempo. Estes, por sua vez, podem estar relacionados com a propriedade de interesse e permitem a modelagem de sistemas onde estão sobrepostas informações espectrais ou cromatográficas de vários componentes. Desta forma, uma matriz de dados pode ser construída a partir de um vetor de medidas instrumentais, onde “n” linhas correspondem a diferentes amostras e as “m” colunas às variáveis (absorbância de um composto em diferentes números de onda),⁵⁴ conforme ilustrado na Figura 5.

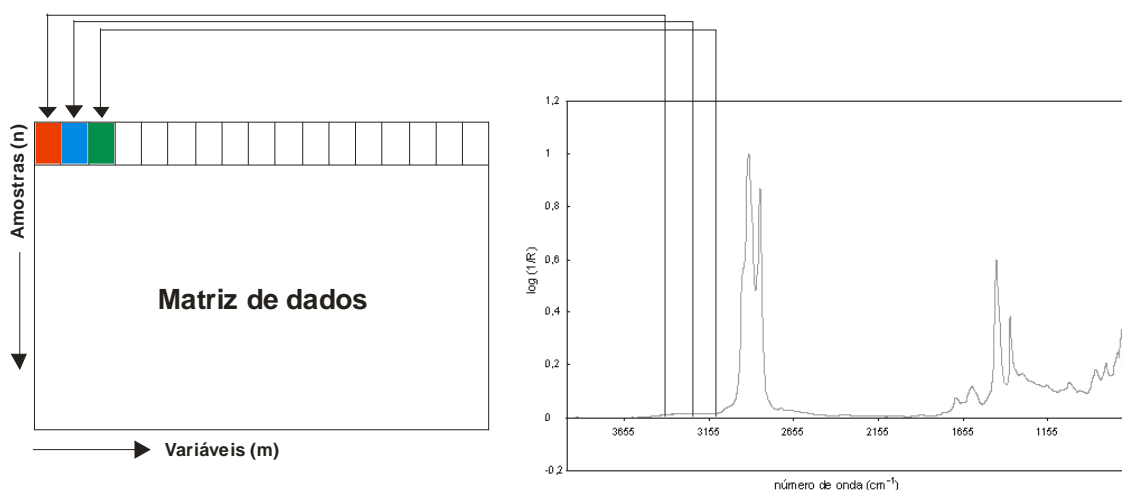


Figura 5. Matriz de dados gerada a partir de um espectro.

¹⁹ Chau, F.; et al., *Chemometrics: From basics to wavelet transform*, 2004, 1-21.

⁸² Wold, S.; *Chemom. Intell. Lab. Syst.* 30 (1995) 109-115.

⁸³ Wold, S.; *J. Pharm. Biomed. Anal.* 9 (1991) 589-596.

⁴¹ Hair, J. F.; et al., *Análise multivariada de dados*, 2005, 23-29.

⁵⁴ Lavine, B. K.; *Clustering and classification of analytical data*, 2000, 9689-9710.

Desta forma, é possível modelar propriedades físicas e químicas de sistemas simples ou complexos a partir de seus dados espectroscópicos, para amostras que possuam matrizes diversificadas, como por exemplo, o estudo de formulações farmacêuticas e derivados do petróleo.^{36,54,66,81} Por sua vez, os métodos de análise multivariada podem ser divididos como métodos de análise exploratória, classificação e de regressão.¹⁹

Os métodos de análise exploratória são aqueles que têm como meta reduzir a dimensionalidade do conjunto de dados, descrevendo a variação ou dispersão de um determinado conjunto. Desta forma é possível descrever e identificar grupos de amostras dentro dos dados, agrupando-os de modo a permitir a identificação das semelhanças. Entre os métodos de análise exploratória, podem ser citados a análise por componentes principais (PCA) e análise por agrupamentos hierárquicos (HCA).²³

Por outro lado, o método de calibração multivariada desenvolve um modelo matemático que correlaciona a propriedade de interesse de uma espécie com os sinais instrumentais. Entre os métodos de calibração podem ser citados a regressão por componentes principais (PCR), o método dos mínimos quadrados parciais (PLS) e o da regressão linear múltipla (MLR).¹⁵

Tendo em vista que o método da análise por componentes principais (PCA), da análise por agrupamentos hierárquicos (HCA) e dos mínimos quadrados parciais (PLS) foram utilizados no presente trabalho, estes serão abordados de forma preferencial.

2.2.2.1. Tratamento e pré-processamento dos dados

Anteriormente a aplicação das ferramentas quimiométricas ao conjunto de dados a ser investigado, podem ser necessários, transformações dos dados espectrais originais, pois estes podem não ter uma distribuição adequada para a

³⁶ Gabrielsson, J.; Lindberg, N; Lundstedt, T.; *J. Chemom.* 16 (2002) 141-160.

⁵⁴ Lavine, B. K.; Clustering and classification of analytical data, 2000, 9689-9710.

⁶⁶ Pasquini, C.; *J. Braz. Chem. Soc.* 14 (2003) 198-219.

⁸¹ Willians, P. C.; Commercial near-infrared reflectance analyzers, 1987.

¹⁹ Chau, F.; *et al.*, *Chemometrics: From basics to wavelet transform*, 2004, 1-21.

²³ Correia, P. R. M.; Ferreira, M. M. C.; *Quim. Nova* 30 (2007) 481-487.

¹⁵ Breton, R. G.; *Analyst* 125 (2000) 2125-2154.

análise, dificultando a extração de informações úteis e interpretação dos mesmos. Basicamente, essas transformações consistem de tratamentos e pré-processamentos dos dados espectrais.³²

O tratamento dos dados tem por objetivo remover variações sistemáticas não desejadas ao espectro, como, mudanças na linha de base, efeitos de espalhamento e fatores externos, não controláveis.^{16a}

- Correção do espalhamento de luz (MSC):³⁹ Tem por finalidade avaliar a regressão linear entre os espectros e o espectro de referência, retendo os resíduos bem como as informações químicas.
- Normalização: Consiste em utilizar os dados adquiridos em uma mesma faixa de amplitude de sinal.
- Primeira e segunda derivadas:⁷³ A aplicação das derivadas remove uma parte da deformação da linha de base, e melhora a definição de bandas que se encontram sobrepostas em uma mesma região espectral, revelando picos de pequena absorbância.

O pré-processamento é aplicado quando se deseja comparar variáveis com diferentes dimensões e consiste basicamente em:³⁹

- Centrar os dados na média: A média das intensidades é calculada para cada comprimento de onda e subtrai-se cada intensidade do respectivo valor médio.
- Autoescalar os dados: Significa centrar os dados na média e dividi-los pelo respectivo desvio padrão, sendo um para cada comprimento de onda.

³² Ferreira, M. M. C.; *et al.*, *Quim. Nova* 22 (1999) 724-731.

^{16a} Brereton, R. G.; *Chemometrics: Data analysis for the laboratory and chemical plant*, 2003, 120-142.

³⁹ Geladi, P.; MacDougall, D.; Martens, H.; *Appl. Spectrosc.* 39 (1985) 491-500.

⁷³ Savitzky, A.; Golay, M. J. E.; *Anal. Chem.* 36 (1964) 1627-1632.

³⁹ Geladi, P.; MacDougall, D.; Martens, H.; *Appl. Spectrosc.* 39 (1985) 491-500.

2.2.2.2. Análise por componentes principais

Segundo Wold *et al.*⁸⁵ a análise por componentes principais foi introduzida na química por Malinowski por volta de 1960 e é, provavelmente, a técnica de análise multivariada mais antiga que se conhece.⁵⁴ A PCA^{15,27,33,85} é um método de decomposição onde uma matriz de dados X relaciona um conjunto de variáveis à diversos experimentos (amostras). Essas novas variáveis são denominadas componentes principais (CP) e são ortogonais entre si, sendo que os dados (variáveis) são centrados na média (no novo sistema de coordenadas). Essas novas matrizes são produtos de vetores chamados *scores* e *loadings*, como podem ser observados na Figura 6.

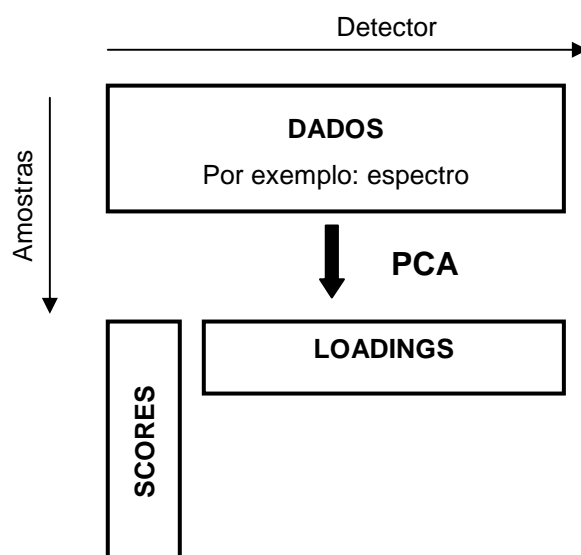


Figura 6. Princípios da PCA.¹⁵

O valor de *scores* (t_h) é dado pela projeção de cada ponto (amostra) no novo sistema de eixos (CPs), medindo a distância θ (através da distância Euclidiana, que será abordada posteriormente) entre o centro do ponto e sua projeção. Por exemplo, se um conjunto de dados contém dez objetos, o mesmo número de *scores* é obtidos para a primeira CP. O valor de *loadings* (p'_h) é obtido entre os cossenos do ângulo (θ) entre a linha (CP) e cada eixo e determina a

⁸⁵ Wold, S.; Esbensen, K.; Geladi, P.; *Chemom. Intell. Lab. Syst.* 2 (1987) 37-52.

⁵⁴ Lavine, B. K.; Clustering and classification of analytical data, 2000, 9689-9710.

¹⁵ Breerton, R. G.; *Analyst* 125 (2000) 2125-2154.

²⁷ Einax, J. W.; Zwanziger, H. W.; Geiß, S.; *Chemometrics in environmental analysis*, 1997, 153-178.

³³ Ferreira, M. M. C.; *J. Braz. Chem. Soc.* 13 (2002) 742-753.

⁸⁵ Wold, S.; Esbensen, K.; Geladi, P.; *Chemom. Intell. Lab. Syst.* 2 (1987) 37-52.

influência de cada variável. Por exemplo, se o conjunto de dados contém três variáveis, este irá possuir três valores de *loadings*. Um esquema da interpretação da PCA é mostrado na Figura 7.²⁷

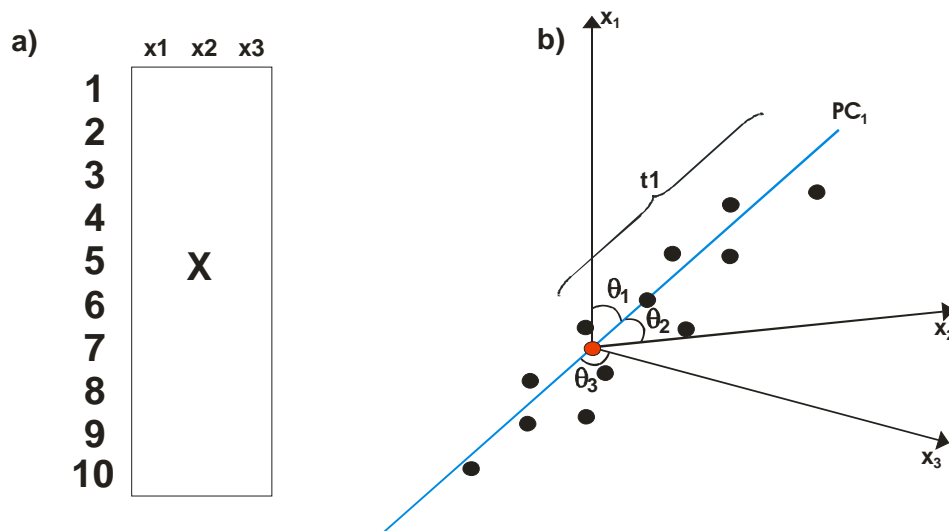


Figura 7. Esquema da interpretação de um modelo de PCA com uma PC: a) conjunto de dados com dez objetos e três variáveis, b) primeira PC calculada através do conjunto de dados.³⁸

Quando a CP_1 é calculada, permanece uma variância não explicada chamada de matrix E, conforme equação 8:²⁷

$$X = TP^t + E \quad (8)$$

Depois que a CP_1 é calculada, a próxima CP é obtida com a matriz residual E, que contém a variância não explicada pela CP_1 , e assim sucessivamente, conforme equação 9:

$$X = t_1p_1^t + t_2p_2^t + \dots + t_h p_h^t + E \quad (9)$$

A utilização da PCA tem por objetivo reduzir a dimensionalidade do conjunto de dados original, a partir de combinações lineares dessas variáveis, preservando a maior quantidade de informação possível (variância). Esta análise

²⁷ Einax, J. W.; Zwanziger, H. W.; Geiß, S.; *Chemometrics in environmental analysis*, 1997, 153-178.

³⁸ Geladi, P.; Kowalski, B. R.; *Anal. Chim. Acta* 185 (1986) 1-17.

vetorial permite descrever a variação ou dispersão de um determinado conjunto de dados.⁵⁴

2.2.2.3. Análise por agrupamentos hierárquicos

A análise por agrupamentos hierárquicos é uma técnica que tem o objetivo de descrever e identificar grupos de amostras dentro dos dados originais, agrupando-os de modo que permitam a identificação das semelhanças entre os grupos de variáveis. Através da aplicação desta técnica pretende-se maximizar a semelhança intergrupo e a diferença entre eles.^{14,27} Segundo Lavine⁵⁴ a HCA é uma técnica empregada em muitas áreas, como biologia, geologia, química, entre outras.

O resultado é apresentado na forma gráfica e é denominado dendograma. Assemelha-se a uma estrutura em forma de árvore e a faixa de similaridade varia de zero a um.^{22,27,54} Inicialmente, é calculada a distância entre dois vetores (amostras), d_{ab} , no espaço n-dimensional (variáveis), normalmente, através da distância Euclidiana, conforme equação 10.

$$d_{ab} = \sqrt{\sum_{i=1}^m (x_{ai} - x_{bi})^2} \quad (10)$$

A distância entre as amostras pode variar com o tipo e o número de medidas, desta forma, é comum transforma-las em um padrão de similaridade, conforme equação 11.

$$S_{ab} = 1 - \frac{d_{ab}}{d_{\text{máx}}} \quad (11)$$

⁵⁴ Lavine, B. K.; Clustering and classification of analytical data, 2000, 9689-9710.

¹⁴ Bratchell, N.; *Chemom. Intell. Lab. Syst.* 6 (1989) 105-125.

²⁷ Einax, J. W.; Zwanziger, H. W.; Geiß, S.; *Chemometrics in environmental analysis*, 1997, 153-178.

²² Cooper, J. B.; *et al.*, *Appl. Spectrosc.* 51 (1997) 1613-1620.

²⁷ Einax, J. W.; Zwanziger, H. W.; Geiß, S.; *Chemometrics in environmental analysis*, 1997, 153-178.

Sendo que, d_{ab} corresponde à distância entre os pontos a e b, e $d_{\text{máx}}$ é a maior distância de um par de amostras no conjunto de dados.

Após as distâncias entre os pares de amostras serem calculados, agrupamentos sucessivos de acordo com suas similaridades são formados, até não haver mais similaridade entre eles. Utilizam-se vários métodos para calcular a similaridade entre os grupos formados, como a construção simples, flexível e incremental. Neste trabalho será empregada a construção incremental, definida na equação 12.⁷⁰

$$d_{ab \rightarrow c} = \frac{[(n_a + n_c)d_{ac}^2 + (n_b + n_c)d_{bc}^2 - n_c d_{ab}^2]^{1/2}}{n_a + n_b + n_c} \quad (12)$$

Onde n_i é o número de amostras no grupo i.

2.2.2.4. Métodos de calibração multivariada

Os métodos de calibração multivariada permitem o tratamento de dados complexos do ponto de vista matemático e estatístico, correlacionando os sinais obtidos (intensidade, por exemplo) com o analito de interesse.¹⁵

Métodos de calibração multivariada têm sido cada vez mais utilizados em química analítica, principalmente quando os componentes presentes em uma mistura necessitam ser determinados, mas a informação analítica disponível não apresenta seletividade, ou seja, não é possível identificar os componentes individuais de cada espécie a partir da resposta instrumental, devido a complexidade da matriz da amostra. A base da calibração multivariada é estabelecer uma relação entre duas matrizes de dados químicos, quando houver uma dependência entre as propriedades que descrevem cada uma delas.¹⁷

Em métodos de calibração multivariada, um modelo pode ser entendido como uma simplificação de um sistema sob investigação, sendo uma representação matemática de um sistema complexo, dentro de uma região limitada

⁷⁰ Pirouette, *Multivariate Data Analysis*, versão 3,11, Infometrix.

¹⁵ Brereton, R. G.; *Analyst* 125 (2000) 2125-2154.

¹⁷ Brown, S. D.; *Appl. Spectrosc.* 49 (1995) 14A-31A.

(necessariamente a região investigada), ou seja, é uma aproximação útil para descrever um conjunto de dados reais.⁴⁰

O desenvolvimento do método de calibração consiste de duas etapas: a calibração e a predição. A etapa de construção do modelo de calibração começa com a seleção de um conjunto de amostras cuidadosamente escolhidas para que sejam representativas de toda a região a ser modelada. Estas amostras (conjunto de calibração) serão utilizadas na construção de um modelo apropriado para relacionar as respostas instrumentais com a informação desejada. Durante essa etapa dois fatores são considerados cruciais: o número de componentes principais ou número de variáveis latentes, e detecção de amostras anômalas (*outliers*).³² Após, deve ser realizada a etapa de validação, verificando a capacidade preditiva do modelo. A validação consiste em testar o modelo com amostras externas, das quais se tem conhecimento prévio das propriedades (ou concentrações) que se desejam medir.^{15,23}

2.2.2.4.1. Mínimos quadrados parciais

A técnica dos mínimos quadrados parciais (PLS) foi desenvolvida na década de 70 por Wold,⁸⁴ e é considerada a principal técnica de regressão para dados multivariados.^{16b}

A regressão por PLS é uma extensão da PCA, pois a construção de um modelo PLS ocorre de maneira similar ao PCA. Porém, além da matriz de dados X ser projetada num novo sistema de coordenadas, também, a matriz dependente Y é decomposta da mesma maneira, simultaneamente. Portanto, o PLS utiliza as respostas analíticas, bem como as informações de interesse, para capturar a variância dos dados da matriz X e da matriz dependente Y, através de suas decomposições sucessivas e simultâneas, correlacionando-as. Uma função é então obtida de forma que descreva a variância entre as matrizes X e Y, bem como, maximize a correlação entre elas. Deste primeiro ajuste nos dados é gerada

⁴⁰ Gemperline, P. J.; *Practical guide to chemometrics*, 2006, 140-145.

³² Ferreira, M. M. C.; *et al.*, *Quim. Nova* 22 (1999) 724-731.

¹⁵ Brereton, R. G.; *Analyst* 125 (2000) 2125-2154.

²³ Correia, P. R. M.; Ferreira, M. M. C.; *Quim. Nova* 30 (2007) 481-487.

⁸⁴ Wold, S.; *Technometrics* 20 (1978) 397-405.

^{16b} Brereton, R. G.; *Chemometrics: Data analysis for the laboratory and chemical plant*, 2003, 297-306.

a primeira variável latente (VL) do PLS, também chamadas de fatores. O conjunto de variáveis latentes no novo sistema de coordenadas é então usado para a calibração. Desta forma, o PLS é um método de covariância máxima, visto que, o seu principal objetivo é prever as variáveis Y a partir das variáveis X.^{15,38}

Portanto, a construção de um modelo PLS consiste de uma regressão entre os *scores* das matrizes X e Y em uma soma de “h” variáveis latentes. O modelo PLS pode ser definido através de relações externas, que correlacionam, individualmente, as matrizes X e Y (conforme as equações 13 e 14), enquanto, as internas correlacionam ambas as matrizes.³⁸

$$X = TP' + E = \sum t_h p_h' + E \quad (13)$$

$$Y = UQ' + F = \sum u_h q_h' + F \quad (14)$$

onde, T e U são matrizes de *scores* das matrizes X e Y, respectivamente; P e Q são as matrizes de *loadings* das matrizes X e Y, respectivamente; e E e F são os resíduos, respectivamente. A relação interna dos *scores* das matrizes X e Y é obtida através do coeficiente de regressão linear, como descrito na equação 15:

$$u_h = b_h t_h \quad (15)$$

sendo $b_h = u_h' t_h / t_h' t_h$, para h variáveis latentes. Os valores de b_h são agrupados na matriz diagonal B (matriz identidade), que contém os coeficientes de regressão entre a matriz de *scores* U de Y e a matriz de *scores* T de X. A melhor relação linear possível entre os *scores* das matrizes é obtida através de pequenas rotações das variáveis latentes das matrizes X e Y. A matriz Y pode ser calculada através das informações contidas em u_h (equação 5), conforme a equação 16:

$$Y = TBQ' + F \quad (16)$$

¹⁵ Brereton, R. G.; *Analyst* 125 (2000) 2125-2154.

³⁸ Geladi, P.; Kowalski, B. R.; *Anal. Chim. Acta* 185 (1986) 1-17.

onde T são os *scores* da matriz X , B é a matriz identidade de b_h , Q' são os *loadings* da matriz Y e F é a matriz residual de Y . Desta forma, a concentração das novas amostras é prevista a partir dos novos *scores* de X , dado por T^* , substituídos na equação 17:

$$Y = T^*BQ' \quad (17)$$

Alguns critérios como, o número de variáveis latentes (VLs) e a avaliação da raiz quadrada dos erros médios (RMSE, do inglês, *Root mean square error*) devem ser avaliados para se obter o melhor modelo referente aos dados sob investigação. A seguir serão abordados aspectos importantes referentes a estes critérios.

2.2.2.4.2. Critérios para avaliação de modelos de regressão

Na quimiometria, os dados consistem, principalmente, de informação e ruído e esta ferramenta pode ser usada para extrair informações relevantes dos dados, criando um modelo matemático. No entanto, este modelo deverá sempre ser validado antes de ser aplicado.⁴⁰

Habitualmente, os métodos de calibração multivariada utilizam diferentes estratégias para determinar o nível do erro de predição de um modelo. Para que um modelo produza resultados satisfatórios, este deve explicar o máximo da variância dos dados com um mínimo de VLs, minimizando a influência do ruído para obtenção de um modelo de fácil interpretação.⁴⁰

Portanto, um modelo contendo um número de VLs inferior ao ideal resultará em subajuste, pois não irá considerar na totalidade a informação contida no conjunto de dados. Por outro lado, quanto maior o número de VLs, também aumentará o ruído e os erros de modelagem resultando em sobreajuste. Deste modo, o número ótimo de VLs a ser usado para descrever um modelo será o número de VLs que corresponda ao ponto no qual a diminuição do erro (produzido pelo aumento da complexidade do modelo) é compensado pelo aumento de erro

⁴⁰ Gemperline, P. J.; *Practical guide to chemometrics*, 2006, 140-145.

de superavaliação. Na Figura 8, este ponto é representado pelo ponto mínimo na curva, e para o qual um erro mínimo na previsão é produzido.⁴⁰

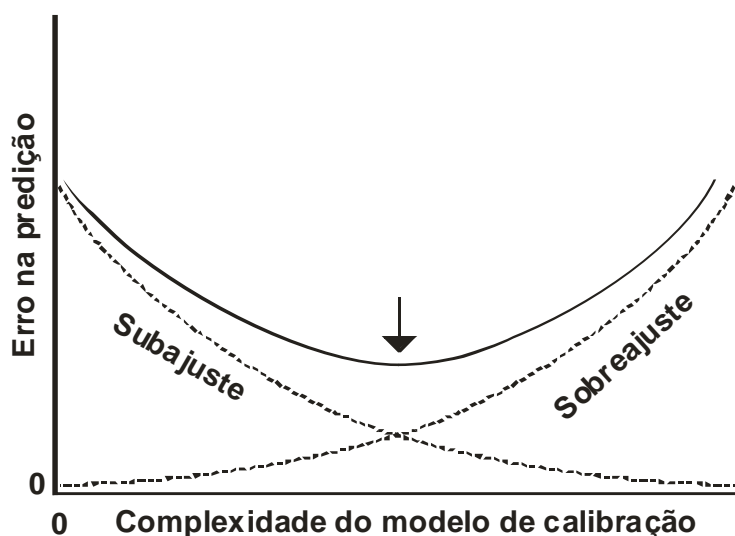


Figura 8. Seleção do ponto ótimo de LVs que melhor descreve o conjunto de dados multivariados.^{30,40}

O número de VLs pode ser calculado através da avaliação da raiz quadrada dos erros médios (RMSE), conforme equação 18. Este erro pode ser avaliado através do conjunto de calibração (RMSEC), de validação (RMSEV) e, também, através da validação-cruzada (RMSECV).^{16c,26}

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (18)$$

onde, y_i é o valor de referência para a i -ésima amostra, e \hat{y}_i é a previsão para o valor desta amostra, sendo n igual ao número total de amostras modeladas.

Outro critério, tão importante quanto a determinação do número de VLs para avaliar um modelo de regressão, é a detecção de amostras anômalas (do inglês, *outliers*). Ao verificar a qualidade do conjunto de calibração, deve-se

⁴⁰ Gemperline, P. J.; Practical guide to chemometrics, 2006, 140-145.

³⁰ Ferrão, M. F.; Tese de Doutorado. Universidade Estadual de Campinas, Campinas-SP, 2000.

^{16c} Brereton, R. G.; Chemometrics: Data analysis for the laboratory and chemical plant, 2003, 313-323.

²⁶ Eastment, H. T.; Krzanowski, W. J.; *Technometrics* 24 (1982) 73-77.

assegurar de que as amostras formam um conjunto homogêneo, removendo-se aquelas amostras que são anômalas. Para isto, uma ferramenta utilizada é a PCA através da análise do gráfico do resíduo de *Student*, o qual permite analisar o resíduo de cada amostra para o total de CPs selecionado, conforme equação 19.^{32,70}

$$\text{resíduo de Student}_i = \frac{f_i}{\text{RMSE}(1 - h_i)^{1/2}} \quad (19)$$

onde, h_i é o número de CPs para a i -ésima amostra, e f_i é o resíduo (erro) da propriedade de interesse para a i -ésima amostra. Normalmente, os programas computacionais utilizam o teste t para determinar se o valor de resíduo de *Student* para dada amostra é elevado, baseado no conjunto de calibração, com um mínimo de 95% de confiança.

A raiz quadrada do erro médio de previsão (RMSEP), conforme equação 18,^{16c} é usada para avaliar quantitativamente a habilidade preditiva de um modelo de calibração multivariada, possibilitando a comparação com o erro proporcionado pelo método de referência. Para comparar a exatidão entre os modelos de calibração multivariada, com base nos valores previstos por cada modelo em relação ao método de referência, é avaliado o erro percentual médio (conforme equação 20).⁵⁷

$$\text{Erro médio(\%)} = \frac{\sum_{i=1}^n \left(\frac{|y_1 - \hat{y}_i|}{y_1} \times 100 \right)}{n} \quad (20)$$

Sendo assim, para avaliar a aplicabilidade do modelo desenvolvido, todos estes parâmetros podem ser empregados de maneira comparativa em relação ao método de referência.

³² Ferreira, M. M. C.; *et al.*, *Quim. Nova* 22 (1999) 724-731.

⁷⁰ Pirouette, *Multivariate Data Analysis*, versão 3,11, Infometrix.

^{16c} Brereton, R. G.; *Chemometrics: Data analysis for the laboratory and chemical plant*, 2003, 313-323.

⁵⁷ Maspoch, S.; Blanco, M.; Coello, J.; Iturriaga, H.; Pagès, J.; *Anal. Chim. Acta* 384 (1999) 207-214.

2.2.2.4.3. Técnicas de seleção de variáveis em modelos de regressão

No PLS a calibração multivariada pode ser realizada utilizando a informação de toda faixa espectral de trabalho para construir um modelo de regressão correlacionando com a propriedade de interesse. No entanto, considerando o grande número de variáveis fornecidas por toda a faixa espectral, algumas destas variáveis podem interferir na modelagem, além de tornar o tratamento dos dados mais lento. Portanto, para melhorar o desempenho de técnicas de calibração multivariada, tem sido utilizados procedimentos apropriados para a seleção das regiões espectrais associadas.⁶¹

Existem vários critérios para avaliar a escolha da região espectral. Normalmente, a escolha da região a ser modelada depende da experiência do analista (das regiões, habitualmente excluídas), e do seu conhecimento do sistema sob investigação. Para avaliar as regiões possivelmente escolhidas, parâmetros como, avaliação da correlação da região selecionada (R^2) com a o método de referência, análise da distribuição espectral da relação sinal-ruído, avaliando a raiz quadrada dos erros médios de validação cruzada (RMSECV), bem como a avaliação da raiz quadrada dos erros médios de previsão (RMSEP), podem ser utilizados.⁶⁴

Recentemente, alguns algoritmos foram desenvolvidos com o propósito de pesquisar e selecionar as regiões espectrais, localizando o conjunto de variáveis que podem resultar nos melhores modelos. Neste caso, existem alguns algoritmos, dentre os quais se destacam, os algoritmos genéticos^{24,25,55} e, mais recentemente, o método dos mínimos quadrados parciais por intervalo (iPLS).⁶¹ A técnica de seleção de variáveis permite a eliminação de informação não relevante, como por exemplo, bandas que não contenham nenhuma informação das espécies ou propriedades a serem analisadas e a amplitude da razão sinal-ruído.⁶⁴

O método iPLS⁶¹ é uma extensão do PLS, no qual divide-se o conjunto de dados (espectros) em um determinado número de subintervalos iguais.

⁶¹ Norgaard, L.; *et al.*, *Appl. Spectrosc.* 54 (2000) 413-419.

⁶⁴ Oliveira, F. C. C.; *et al.*, *Quim. Nova* 27 (2004) 218-225.

²⁴ Costa, P. A.; *et al.*, *Anal. Chim. Acta* 433 (2001) 111-117.

²⁵ Costa, P. A.; Poppi, R. J.; *Quim. Nova* 22 (1999) 405-411.

⁵⁵ Leardi, R.; *J. Chemometrics* 14 (2000) 643-655.

A otimização dos subintervalos iPLS consiste de duas etapas: selecionar o número de intervalos empregados e escolher a amplitude de cada intervalo (tamanho). Os modelos obtidos com estes intervalos espectrais são avaliados através do RMSECV, do RMSEP e do R^2 . A detecção de *outliers*, o número de LVs e os erros do modelos, são determinados, igualmente, como em um modelo PLS convencional. A diferença consiste, apenas, na divisão do conjunto de dados em subintervalos iguais. Para cada intervalo é construído um modelo PLS, sendo os resultados apresentados na forma gráfica para facilitar a comparação com toda a faixa espectral. O método é planejado para dar uma visão geral dos dados e pode ser útil para selecionar as variáveis mais representativas na construção de um modelo de calibração adequado. Porém, o método iPLS indica a região que está contida a informação, sendo uma aproximação univariada, pois não fornece sinergismo das regiões espectrais envolvidas.⁵⁶

Para relacionar os subintervalos, a fim de obter melhores habilidades preditivas, pode ser utilizado, também, o algoritmo dos mínimos quadrados parciais por sinergismo de intervalos (siPLS),⁶⁰ uma extensão do iPLS. Este possibilita selecionar a melhor combinação de intervalos, combinando 2 a 2, 3 a 3 e até 4 a 4 sub-regiões do espectro fornecendo, geralmente, melhores coeficientes de determinação e os menores erros de predição do que o iPLS.

Ainda, pode ser utilizado o método dos mínimos quadrados parciais por exclusão (biPLS).⁵⁶ Neste método o algoritmo iPLS é aplicado aos dados e, através da avaliação do valor de RMSECV, regiões espectrais não relevantes são excluídas do modelo.

2.3. Aplicação da espectroscopia no infravermelho combinada com os métodos multivariados para análise de amostras petroquímicas

Tendo em vista que, apesar de não ter sido encontrado na literatura aplicações destes métodos quimiométricos para a determinação do número de acidez total em frações de petróleo, especialmente, para RAT e RV, diversos parâmetros físico-químicos podem ser determinados por espectroscopia no

⁵⁶ Leardi, R.; Norgaard, L.; *J. Chemometrics* 18 (2004) 486–497.

⁶⁰ Munck, L.; *et al.*, *Anal. Chim. Acta* 446 (2001) 171-186.

infravermelho combinado a métodos multivariados, sendo alguns destacados a seguir.

Uma metodologia para classificação de produtos derivados de petróleo em tempo real, foi proposta por Kim *et al.*⁵⁰ empregando a espectroscopia NIR. Neste trabalho foram classificadas amostras de gasolina, querosene, gasóleo e nafta, com boa precisão.

Borin e Poppi¹³ utilizaram PCA e cartas de controle associadas a espectroscopia MIR para detectar o tipo e a qualidade de óleos lubrificantes de motores a gasolina e diesel. A PCA foi utilizada para classificar o tipo do lubrificante em mineral, sintético e semi-sintético. Além de classificar, a metodologia permitiu avaliar as condições de uso dos óleos estudados.

Com o objetivo de desenvolver modelos com maior poder de predição, alguns trabalhos têm proposto a associação de ferramentas quimiométricas de classificação e regressão aliadas as espectroscopia no infravermelho médio e próximo. Estas associações possibilitaram, por exemplo, classificar e prever a origem geográfica de amostras de petróleo,¹ determinar porções saturadas, aromáticas, resinas e asfaltenos em petróleo pesado,⁴² determinar propriedades do diesel,⁷² determinar benzeno, tolueno, etilbenzeno, o-xileno, m-xileno e p-xileno em amostras de gasolina,²² determinar a quantidade de água presente no petróleo,⁶⁹ caracterizar grupos funcionais de óleos ricos em asfaltenos,² prever a estabilidade de emulsões de amostras de petróleo,^{7,34,48} classificar petróleo através do °API, sal e água,^{29,65} determinar propriedades químicas e físicas da nafta⁵² e

⁵⁰ Kim, M.; Lee, Y.; Han, C.; *Comput. Chem. Eng.* 24 (2000) 513-517.

¹³ Borin, A.; Poppi, R. J.; *J. Braz. Chem. Soc.* 15 (2004) 570-576.

¹ Abbas, O.; *et al.*, *Appl. Spectrosc.* 60 (2006) 304-314.

⁴² Hannisdal, A.; Hemmingsen, P. V.; Sjöblom, J.; *Ind. Eng. Chem. Res.* 44 (2005) 1349-1357.

⁷² Santos, V. O.; *et al.*, *Anal. Chim. Acta* 547 (2005) 188-196.

²² Cooper, J. B.; *et al.*, *Appl. Spectrosc.* 51 (1997) 1613-1620.

⁶⁹ Pironon, J.; *et al.*, *J. Geochem. Explor.* 69-70 (2000) 663-668.

² Akhlaq, M. S.; *J. Pet. Sci. Eng.* 22 (1999) 229-235.

⁷ Aske, N.; Kallevik, H.; Sjöblom, J.; *J. Pet. Sci. Eng.* 36 (2002) 1-17.

³⁴ Fossen, M.; *et al.*, *J. Dispersion Sci. Technol.* 26 (2005) 227-241.

⁴⁸ Kallevik, H.; Kvalheim, O. M.; Sjöblom, J.; *J. Colloid Interface Sci.* 225 (2000) 494-504.

²⁹ Falla, F. S.; *et al.*, *Pet. Sci. Eng.* 51 (2006) 127-137.

⁶⁵ Pasquini, C.; Bueno, A. F.; *Fuel* 86 (2007) 1927-1934.

⁵² Ku, M.; Chung, H.; *Appl. Spectrosc.* 53 (1999) 557-564.

verificar a presença de adulterantes em combustíveis, como gasolina, diesel/biodiesel.^{63,78}

Poucos trabalhos sobre resíduo atmosférico e de vácuo utilizando a espectroscopia no infravermelho e métodos de calibração multivariada são encontrados na literatura. Chung e Ku²⁰ realizaram um estudo comparativo entre as espectroscopias no infravermelho próximo e médio (na análise por infravermelho médio foi utilizado um acessório de reflexão total atenuada) e a espectroscopia Raman, aliadas ao PLS para a determinação do γ API de amostras de RAT. Após aquisição dos espectros foi verificado que o NIR forneceu melhor reprodutibilidade espectral, maior razão sinal-ruído e menor erro de predição. Recentemente, Coelho *et al.*²¹ também empregaram a espectroscopia no infravermelho médio utilizando reflexão para caracterizar RV.

Aplicando o algoritmo iPLS para a seleção de variáveis, Borin e Poppi¹² desenvolveram uma metodologia para a determinação de contaminantes como gasolina, etilenoglicol e água em óleo lubrificante. As amostras passaram apenas por um processo de homogeneização e os espectros no infravermelho médio foram adquiridos usando um acessório de ATR.

Nesse sentido, a espectroscopia no infravermelho aliada as técnicas multivariadas de análise pode ser uma alternativa a metodologias oficiais, tendo em vista vantagens como mínimo preparo e manipulação da amostra, redução de tempo, redução do volume de solventes tóxicos utilizados, reduzindo custos e resultando em menor risco ao analista e menor geração de resíduos, preservando o meio-ambiente. Outra vantagem é o fato de que a partir de um único conjunto de espectros, diversos parâmetros de qualidade podem ser estimados simultaneamente.⁷

⁶³ Oliveira, F. C. C.; *et al.*, *Anal. Chim. Acta* 587 (2007) 194-199.

⁷⁸ Teixeira, L. S. G.; *et al.*, *Fuel* 8 (2007) 346-352.

²⁰ Chung, H.; Ku, M.; *Appl. Spectrosc.* 54 (2000) 239-245.

²¹ Coelho, R. R.; *et al.*, *Fuel Process. Technol.* 87 (2006) 325-333.

¹² Borin, A.; Poppi, R. J.; *Vib. Spectrosc.* 37 (2005) 27-32.

⁷ Aske, N.; Kallevik, H.; Sjöblom, J.; *J. Pet. Sci. Eng.* 36 (2002) 1-17.

3. MATERIAIS E MÉTODOS

3.1. Instrumentação

As determinações de acidez através do método de referência ASTM D 664-04, foram feitas em um titulador automático Metrohm (modelo Titrand 836, <http://www.metrohm.com>, Suíça) equipado com eletrodo íon-seletivo para meio não aquoso (Metrohm).

Os espectros de infravermelho foram obtidos empregando um espectrômetro PerkinElmer (modelo Spectrum One[®] FTIR, <http://www.perkinelmer.com>, EUA), equipado com acessório de reflexão total atenuada horizontal (HATR) PIKE Technologies, (HATR[®], Horizontal Attenuated Total Reflection, <http://www.piketech.com>, EUA) com cristal de seleneto de zinco. Argônio foi utilizado como gás de purga (99,998% de pureza, White Martins, <http://www.whitemartins.com.br>, Brasil).

As amostras foram pesadas empregando uma balança analítica Shimadzu (modelo AY220, <http://www.shimadzu.com.br>, Brasil), com resolução de 0,0001 g e carga máxima de 220 g. O aquecimento das amostras foi feito utilizando-se uma estufa convencional com circulação de ar (modelo 400/2ND, <http://www.novaetica.com.br>, Brasil).

3.2. Reagentes

A água utilizada foi previamente destilada e deionizada em uma coluna trocadora de íons convencional (condutividade máxima de $0,6 \mu\text{S cm}^{-1}$) e purificada em um sistema Milli-Q (Milli-Q[®], <http://www.millipore.com>, EUA), com resistividade final de $18,2 \text{ M}\Omega \text{ cm}$.

Os solventes utilizados na determinação de acidez por titulação potenciométrica foram: propan-2-ol P.A, 0,78 kg L⁻¹, tolueno P.A, 0,87 kg L⁻¹ (Vetec, www.vetecquimica.com.br, Brasil). Também, foram utilizados, hidróxido de potássio P.A. (Merck, www.merck.de, Alemanha) e biftalato de potássio P.A. (Vetec, www.vetecquimica.com.br, Brasil). Querosene comercial também foi utilizado para auxiliar na limpeza dos materiais.

3.3. Amostras

As amostras de resíduo da destilação atmosférica e resíduo da destilação a vácuo do petróleo foram doadas pelo Centro de Pesquisas e Desenvolvimento Leopoldo Américo Miguez de Mello (CENPES, Petrobrás). As amostras foram armazenadas em frascos de vidro âmbar à temperatura ambiente.

3.4. Programas de computação

Para o tratamento dos dados foram utilizados dois programas computacionais: o Pirouette[®] versão 3.11 (<http://www.infometrix.com>, EUA) do qual se utilizou os algoritmos PCA e HCA; e o MATLAB[®] versão 6.5 (The Math Works, <http://www.mathworks.com>, EUA), sendo usado o pacote “iToolbox” versão 2 desenvolvido por Lars Norgaard (iToolbox for MATLAB, <http://www.models.kvl.dk>, EUA), para a seleção das variáveis, com o objetivo de desenvolver os modelos iPLS, siPLS e biPLS.

3.5. Titulação potenciométrica – Determinação do número de acidez total

O número de acidez total foi determinado por titulação potenciométrica nas amostras de RAT e RV, segundo a norma ASTM D 664-04 *Standard Test Method for Acid Number of Petroleum Products by Potenciometric Titration*.⁶ Inicialmente, preparou-se a solução de hidróxido de potássio em propan-2-ol. A norma ASTM foi adaptada, conforme Fuhr *et al.*³⁵ no que diz respeito ao pré-tratamento da amostra

⁶ Annual Book of ASTM Standards, D 664-04, 2004.

³⁵ Fuhr, B.; *et al.*, *Energy Fuels* 21 (2007) 1322-1324.

e a quantidade de amostra utilizada. Estas adaptações a norma foram adotadas, pois as amostras de RAT e RV são de difícil manuseio e solubilização.

Hidróxido de potássio foi dissolvido em propan-2-ol por 10 minutos para total dissolução, sob fervura branda. Esta solução foi deixada em repouso por 2 dias e, em seguida, procedeu-se a filtração da mesma. O sobrenadante foi padronizado com biftalato de potássio (padrão primário) em solução aquosa. A padronização do hidróxido de potássio foi feita diariamente uma vez que o KOH não é totalmente solúvel em propan-2-ol.

Em seguida, preparou-se a mistura de solventes nas seguintes proporções: 45% de propan-2-ol, 50% tolueno e 5% de água purificada. Previamente, as amostras de RAT e RV (aproximadamente 1 g) foram solubilizadas em 25 mL de tolueno em copo de béquer de 150 mL e completou-se a 125 mL com a mistura de solventes. Por fim, as amostras de RAT e RV foram potenciometricamente tituladas com solução de KOH $0,1 \text{ mol L}^{-1}$ em propan-2-ol, sendo o processo de detecção obtido por inflexão.

3.6. Espectroscopia no infravermelho – Análise por ATR

Os espectros das amostras de RAT e RV foram obtidos em espectrômetro de infravermelho com transformada de Fourier e reflexão total atenuada (ATR-FTIR) conforme parâmetros especificados a seguir.

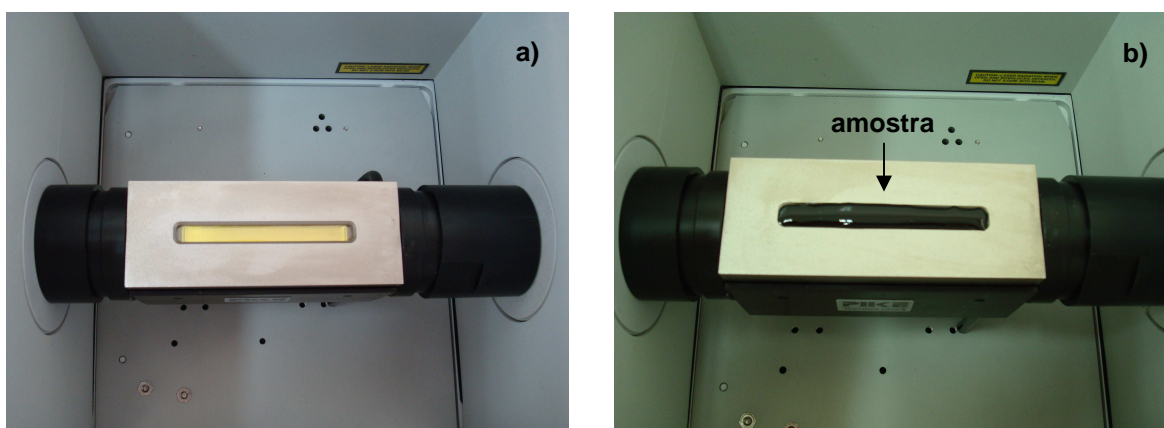
3.6.1. Aquisição dos espectros

Para a aquisição dos espectros, as amostras foram previamente aquecidas em estufa a $100 \text{ }^{\circ}\text{C}$ e transferidas, imediatamente, com o auxílio de uma espátula para a superfície do cristal de ATR. Na Tabela 7 são apresentados os parâmetros utilizados na aquisição dos espectros.

Tabela 7. Instrumentação e regiões empregadas na aquisição dos espectros por ATR.

Parâmetros	Condições
Região espectral	650 a 4000 cm^{-1}
Detector	DTGS – detector de sulfato de triglicina deuterada
Divisor de feixe	cristal selado de KBr
Varreduras	08
Resolução	2 cm^{-1}
Acessório	Reflexão total atenuada horizontal (PIKE)
Aplicativo	Spectrum v 5.0.1

Anteriormente à obtenção do espectro de cada amostra, foi obtido um espectro de referência (*background*). Para tal, foi adquirido um espectro sem amostra na superfície do cristal de ATR (Figura 9a). Após cada espectro de referência, as amostras foram dispostas no compartimento do acessório sobre o cristal e submetidas à análise (Figura 9b). O equipamento foi continuamente purgado com argônio para eliminar qualquer contribuição da umidade do ar e do CO_2 na aquisição dos espectros.

**Figura 9.** Acessório de HATR acoplado ao espectrômetro de infravermelho: a) acessório sem amostra, para obtenção do sinal de referência, b) acessório com amostra.

Foram coletados os espectros das amostras de RAT e RV em duplicata, os quais foram inicialmente normalizados, e em seguida obteve-se o espectro médio para cada amostra (não foi aplicado nenhum teste estatístico entre os espectros), cujo perfil dos espectros pode ser observado na Figura 10.

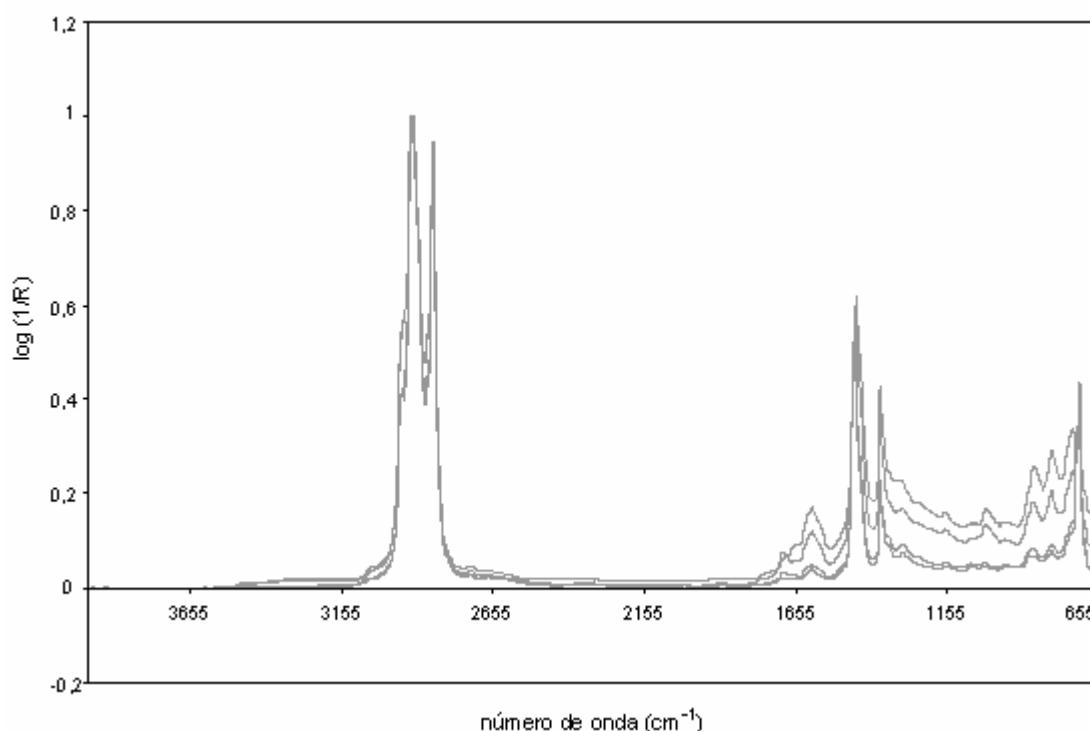


Figura 10. Perfil dos espectros das amostras de RAT e RV obtidos por ATR-FTIR.

Após a aquisição do espectro de cada amostra, o acessório de HATR foi cuidadosamente limpo com querosene e tolueno, sendo seco com ar quente, por aproximadamente 1 minuto.

3.7. Construção dos modelos – Análise multivariada

Antes da construção dos modelos, foi feita a detecção de amostras anômalas e a seleção das amostras que constituíram os conjuntos de calibração e previsão, através da PCA e da HCA, disponíveis no software Pirouette[®] versão 3.11. Posteriormente, foi construído o modelo global utilizando PLS empregando toda a faixa espectral (3340 números de onda). Por fim, foram construídos os modelos de calibração através de 3 diferentes algoritmos: iPLS, siPLS e o biPLS, disponíveis no pacote “iToolbox” versão 2 (MATLAB[®]). Na Figura 11 está ilustrado o fluxograma das etapas utilizadas para o desenvolvimento dos modelos.

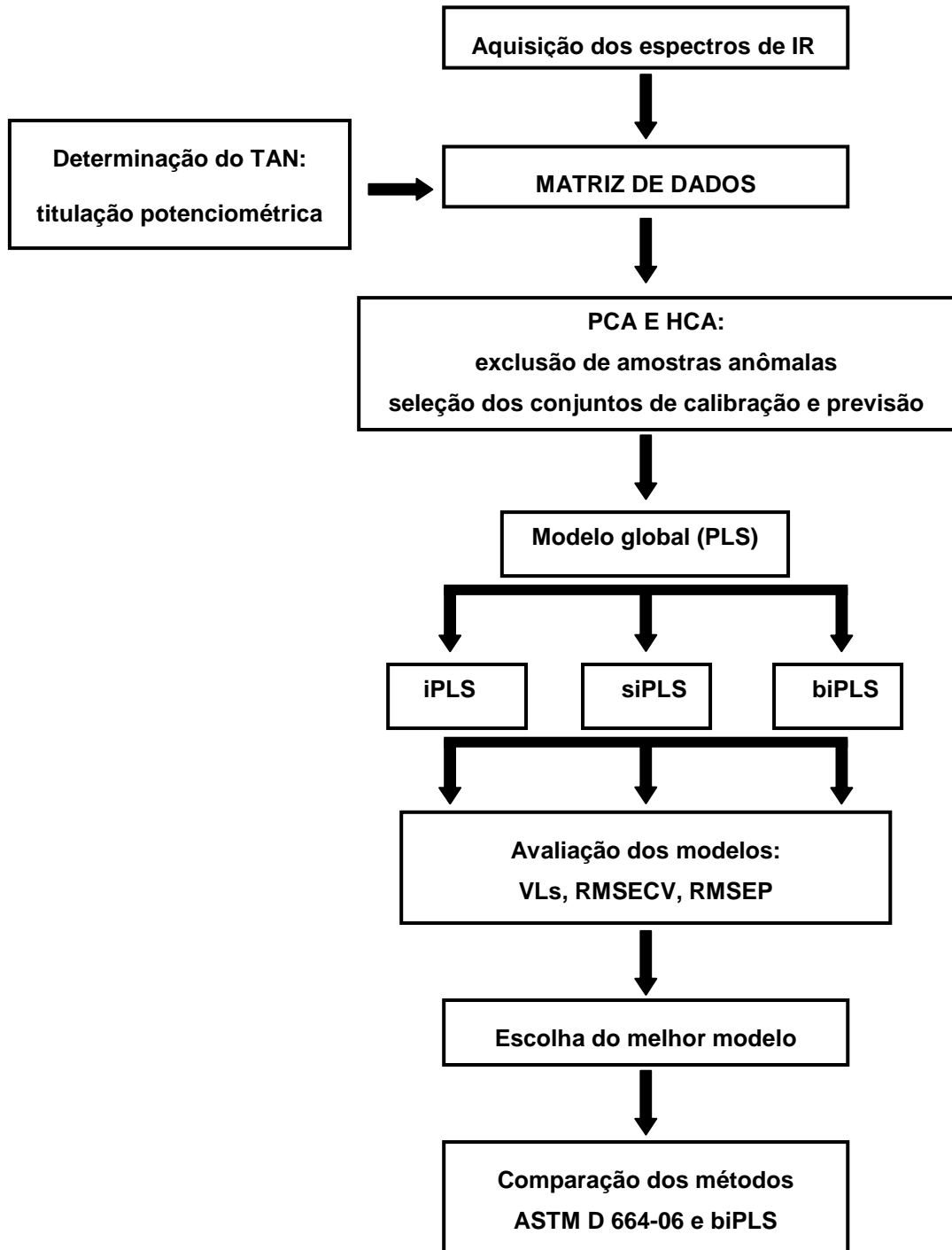


Figura 11. Principais etapas desenvolvidas para a obtenção do melhor modelo de calibração.

3.7.1. Detecção de amostras anômalas e seleção das amostras dos conjuntos de calibração e previsão

Para assegurar a qualidade dos conjuntos de calibração e previsão, inicialmente, foi feita a detecção de amostras anômalas. Para tal procedimento, foi utilizada a PCA, através da análise do gráfico dos resíduos de Student vs. a distância de Mahalanobis,⁷⁰ o qual permite verificar a influência de cada amostra em particular no modelo. Amostras com resíduo alto e/ou grande distância de Mahalanobis foram descartadas.

A etapa de seleção dos conjuntos de calibração e previsão foi feita através da similaridade das amostras (dendograma), fornecido pela HCA. Dentre os grupos similares fornecidos pela HCA, foi selecionada, no mínimo, uma amostra para compor o conjunto de previsão. Também foi critério para seleção destas amostras em cada grupo a informação contida na variável Y, referente ao TAN das amostras. Sendo assim, foram selecionadas amostras com valores alternados dentro da faixa de variação do TAN, porém garantindo que as amostras com valores extremos de TAN de cada grupo estejam presentes no conjunto de calibração.

3.7.2. Desenvolvimento dos modelos para a determinação do TAN por ATR-FTIR

O desenvolvimento dos modelos de calibração para a determinação do TAN em amostras de RAT e RV foi dividido em 3 etapas: na primeira etapa foram testados diferentes tratamentos e pré-processamentos na construção de modelos utilizando PLS e todas as variáveis independentes (item 3.7.2.1.); em seguida foram utilizados os modelos iPLS, siPLS e biPLS para a seleção dos números de onda correlacionados com o TAN (item 3.7.2.2.); e, finalmente, foi feita a escolha dos melhores modelos a partir dos valores de RMSECV, RMSEP e R^2_{cal} (item 3.7.2.3.).

⁷⁰ Pirouette, *Multivariate Data Analysis*, versão 3,11, Infometrix.

3.7.2.1. Testes preliminares e obtenção do modelo global PLS

Os modelos de calibração foram construídos a partir de 16 amostras de RAT e 41 amostras de RV com valores do TAN compreendidos entre 0,08 e 3,50 mg KOH g⁻¹.

Foram utilizados todos os números de onda e foram testados como recurso de tratamento dos dados, a primeira derivada e o MSC. Como pré-processamentos foram avaliados os dados centrados na média e autoescalados. Foram obtidos quatro modelos, dos quais escolheu-se o modelo com menores valores de RMSECV e RMSEP e R^2_{cal} mais próximo de 1, para comparar com os modelos obtidos por iPLS, siPLS e biPLS.

3.7.2.2. Seleção de variáveis (número de onda)

As variáveis espectrais foram selecionadas através do emprego dos métodos de seleção de variáveis, iPLS, siPLS e biPLS. Nesta etapa, tanto nos modelos iPLS, quanto nos modelos siPLS e biPLS, os espectros foram divididos em 10, 20, 40 e 60 intervalos. Quando da aplicação do siPLS, foram combinados até 3 intervalos.

A rotina destes algoritmos gera informações gráficas, indicando o número de variáveis latentes usadas em cada intervalo e os valores de RMSECV. O subintervalo que apresenta o menor valor de RMSECV deve ser selecionado, para posterior comparação com valores de RMSECV de modelos construídos com a informação de toda a faixa espectral (modelo global PLS).⁶¹

3.7.2.3. Avaliação dos modelos obtidos

Os modelos foram avaliados em termos do número de variáveis latentes (VLs), RMSECV, RMSEP⁶¹ e, por fim, através do coeficiente de determinação (R^2_{cal}). As regiões com valores de RMSECV menores que o modelo global PLS foram selecionadas para a construção dos diferentes modelos sendo, posteriormente, comparadas com a eficiência do modelo global. O desempenho dos modelos escolhidos foi avaliado através do erro percentual médio de previsão.

⁶¹ Norgaard, L.; *et al.*, *Appl. Spectrosc.* 54 (2000) 413-419.

3.8. Comparação da norma ASTM D 664-04 com a metodologia proposta

Primeiramente, a presença de erro sistemático na metodologia proposta foi avaliada através da análise por regressão dos mínimos quadrados ordinários (OLS) utilizando o software XLSTAT versão 2007.8.03. Utilizou-se, também, a regressão ortogonal com procedimento de Mandel.⁴⁴

⁴⁴ Hartmann, C.; *et al.*, *Anal. Chim. Acta* 338 (1997) 19-40.

4. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

4.1. Construção dos modelos – Análise multivariada

A seguir, serão discutidos os parâmetros avaliados para o desenvolvimento dos modelos de calibração para a determinação do número de acidez total em amostras de resíduo da destilação atmosférica e do resíduo da destilação a vácuo do petróleo.

4.1.1. Detecção de amostras anômalas e seleção das amostras dos conjuntos de calibração e previsão

Para a modelagem quimiométrica, a matriz de dados (variáveis X e Y) foi inicialmente tratada no aplicativo Pirouette[®] 3.11 (Infometrix), para a detecção de amostras anômalas e para a seleção das amostras dos conjuntos de calibração e previsão. As regiões onde foi evidenciada pouca informação espectral, bem como aquelas com grande ruído (devido ao vapor d'água e o dióxido de carbono), foram previamente excluídas. Portanto, nesta etapa os modelos foram construídos utilizando as regiões espectrais compreendidas entre 650 a 1999 cm^{-1} e 2401 a 3999 cm^{-1} , sendo utilizado os dados centrados na média, como pré-processamento.

A PCA foi aplicada ao conjunto de dados, a fim de identificar as amostras anômalas de RAT e de RV presentes, para eliminar qualquer influência negativa nos modelos de calibração. Na Tabela 8, é apresentada a variância percentual dos dados atribuídas a cada componente principal (CP) utilizada, bem como a variância percentual acumulada em função do aumento do número de CPs utilizadas.

Tabela 8. Variância percentual e variância percentual acumulada em função do número de CPs utilizadas.

Número de CPs	Variância (%)	Variância acumulada (%)
1	87,53	87,53
2	8,85	96,37
3	1,17	97,54
4	0,71	98,25
5	0,65	98,89
6	0,28	99,17
7	0,24	99,41
8	0,13	99,54
9	0,12	99,66
10	0,08	99,74

Com base na Tabela 8, foi verificado que 2 CPs são suficientes para descrever a informação global presente no conjunto de dados modelados. O modelo apresenta uma variância acumulada de 96,4%, representado a variação contida no conjunto de dados. Assim, baseado nas 2 CPs selecionadas, as amostras anômalas presentes no conjunto de dados total, foram detectadas através da análise do gráfico apresentado na Figura 12, que relaciona os resíduos de Student vs. a distância de Mahalanobis.⁷⁰ Amostras com resíduo alto e/ou grande distância de Mahalanobis foram excluídas do conjunto de dados, em função da diferença apresentada em relação as demais amostras, como as destacadas na Figura 12.

⁷⁰ Pirouette, *Multivariate Data Analysis*, versão 3,11, Infometrix.

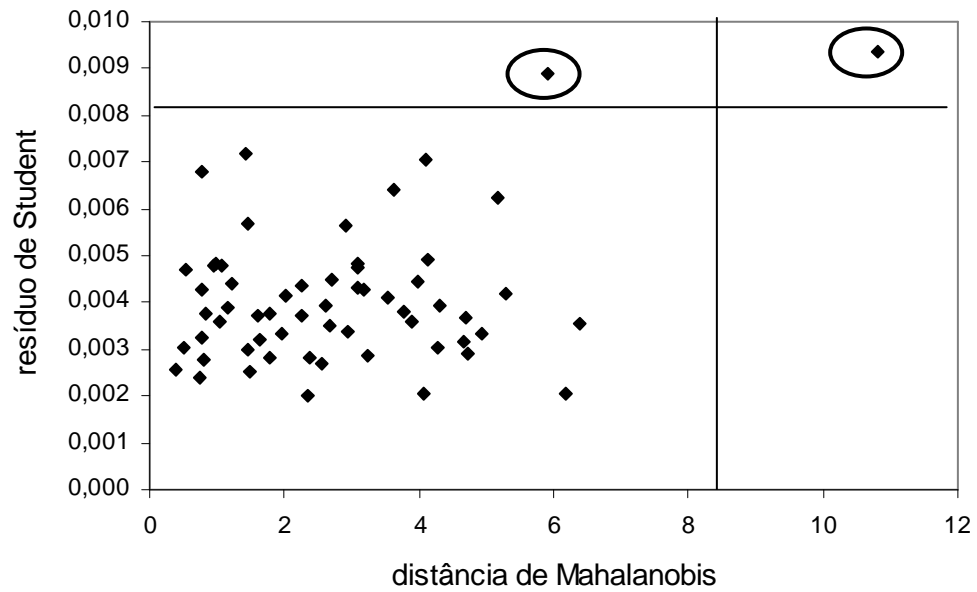


Figura 12. Gráfico de diagnóstico de amostras anômalas com 95% de confiança para o conjunto total de amostras RAT e RV. As amostras destacadas foram excluídas do conjunto de dados.

Após a exclusão das amostras anômalas presentes no conjunto de dados, selecionaram-se as amostras que constituíram os conjuntos de calibração e previsão. A seleção das amostras dos conjuntos de calibração e previsão foi baseada no critério de similaridade entre os espectros das amostras, sendo feita a partir da análise através do dendograma obtido pela HCA (Figura 13).

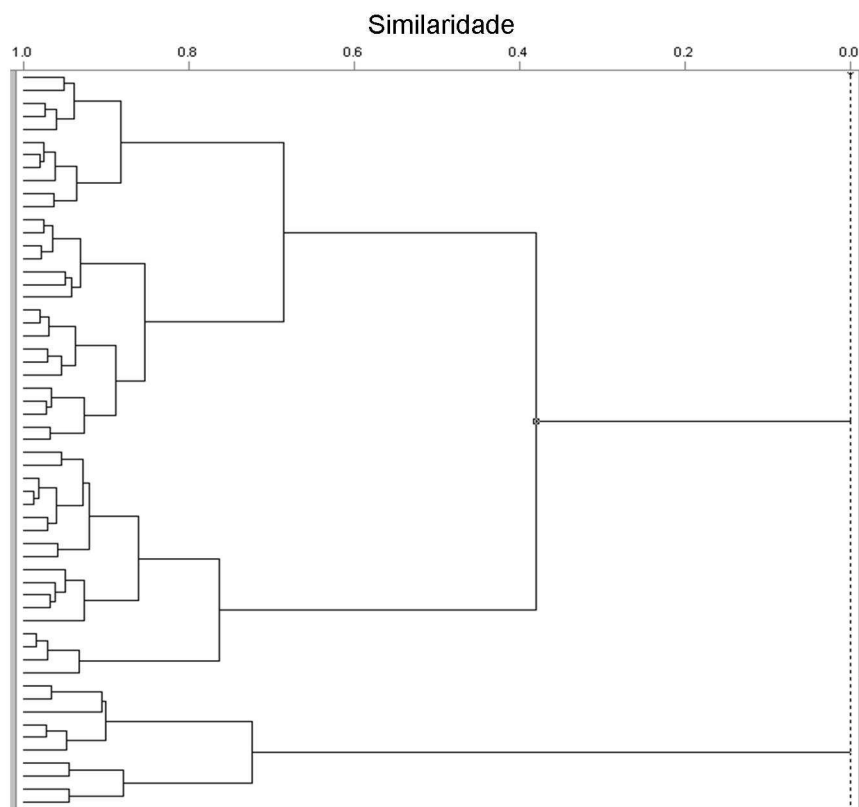


Figura 13. Dendrograma obtido pela HCA, para a seleção das amostras dos conjuntos de calibração e previsão.

Além da utilização da análise da HCA para a etapa de seleção das amostras que constituíram os conjuntos de calibração e previsão, foi utilizada também a informação contida na variável Y, referente ao TAN das amostras. Assim, levando em consideração esses dois fatores (HCA e TAN) as amostras de RAT e RV foram distribuídas de modo a garantir que a maior variabilidade presente nos dados esteja contemplada tanto no conjunto de calibração quanto no conjunto de previsão.¹⁵ Teve-se o cuidado para que os valores extremos do TAN estejam no conjunto de calibração, conforme a Figura 14. Desta forma, o conjunto de calibração foi composto por 44 amostras, as quais se encontram dentro da faixa de concentração de 0,08 até 3,50 mg KOH g⁻¹. O conjunto de previsão foi constituído por 13 amostras, sendo que suas concentrações variaram de 0,13 a 2,46 mg KOH g⁻¹.

¹⁵ Breton, R. G.; *Analyst* 125 (2000) 2125-2154.

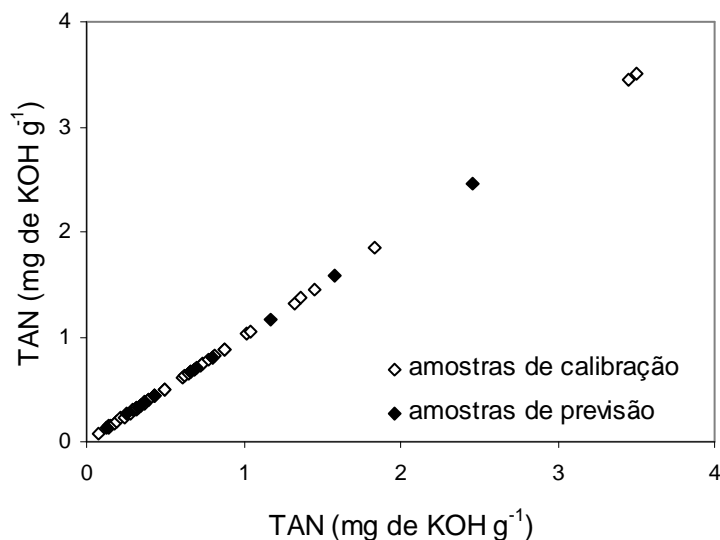


Figura 14. Distribuição das amostras de calibração e previsão para a determinação do TAN de amostras de RAT e RV.

4.1.2. Desenvolvimento dos modelos para a determinação do TAN por ATR-FTIR

4.1.2.1. Testes preliminares e obtenção do modelo global PLS

Inicialmente, a escolha do tratamento e do pré-processamento mais adequados para o desenvolvimento dos modelos foi avaliada através da modelagem de toda a faixa espectral (modelo global), utilizando o algoritmo dos mínimos quadrados parciais (PLS).

Dentre os tratamentos avaliados, estão a aplicação da correção do espalhamento de luz (MSC)³⁹ e a primeira derivada por filtro de Savitzky-Golay.⁷³ A MSC visa a correção do efeito do espalhamento de luz presente em espectros obtidos por técnicas de reflexão e, a primeira derivada, empregando onze pontos e um polinômio de ordem 2, tem o objetivo de corrigir a linha de base, além de acentuar as informações espectrais presentes. Para facilitar a visualização dos resultados nas tabelas foi utilizada a sigla MSC, para a correção do espalhamento

³⁹ Geladi, P.; MacDougall, D.; Martens, H.; *Appl. Spectrosc.* 39 (1985) 491-500.

⁷³ Savitzky, A.; Golay, M. J. E.; *Anal. Chem.* 36 (1964) 1627-1632.

da luz e a letra D para a identificação da aplicação da primeira derivada por filtro de Savitzky-Golay.⁷³

Por outro lado, para os pré-processamentos avaliados, os dados foram autoescalados e centrados na média. Da mesma maneira, para facilitar a visualização dos resultados nas tabelas foram utilizadas as letras A e CM para a identificação dos pré-processamentos, autoescalado e centrado na média, respectivamente.

Analisando-se os dados da Tabela 9, escolheu-se como modelo global o PLS (D, CM) que utiliza o tratamento da primeira derivada e, como pré-processamento, os dados foram centrados na média. Estes apresentaram o melhor coeficiente de determinação (R^2_{cal}) e valores de RMSECV e RMSEP na mesma ordem de grandeza, indicando que os conjuntos de calibração e previsão possuem erros associados semelhantes.

Tabela 9. Parâmetros estatísticos dos modelos globais PLS para o TAN, em mg KOH g⁻¹, para amostras de RAT e RV.

Modelo	TV ^a	Intervalos	VLs ^b	RMSECV (mg KOH g ⁻¹)	R^2_{cal}	RMSEP (mg KOH g ⁻¹)
PLS (MSC, A)	3340	todos	10	0,2336	0,9618	0,1832
PLS (MSC, CM)	3340	todos	10	0,2295	0,9678	0,2047
PLS (D, A)	3340	todos	5	0,2340	0,9756	0,1548
PLS (D, CM)	3340	todos	10	0,1550	0,9855	0,1889

^a TV: número total de variáveis, ^b VLs: variáveis latentes.

* Em negrito: modelo selecionado.

Na Figura 15 é ilustrado o perfil dos espectros após a aplicação da primeira derivada por filtro de Savitzki-Golay.⁷³

⁷³ Savitzky, A.; Golay, M. J. E.; *Anal. Chem.* 36 (1964) 1627-1632.

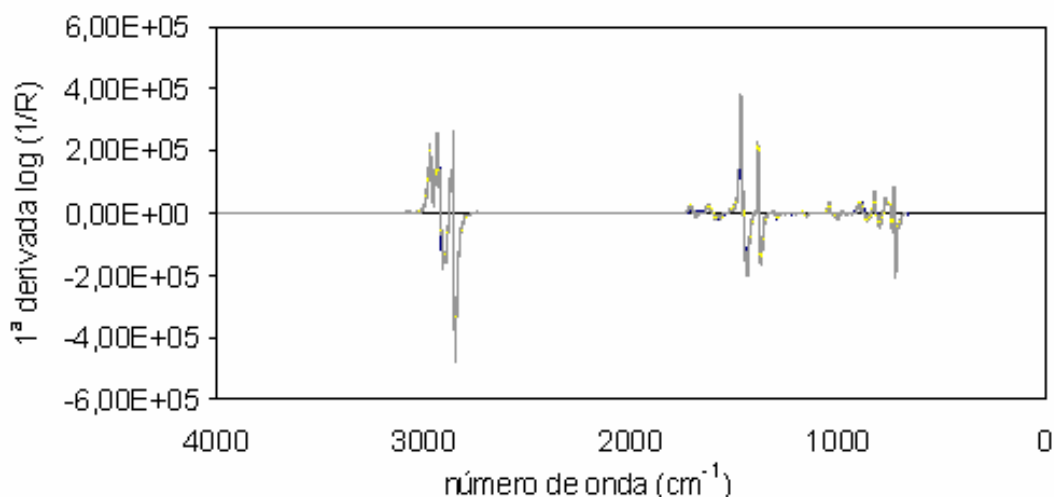


Figura 15. Ilustração do perfil dos espectros das amostras de RAT e RV, depois de aplicada a primeira derivada por filtro de Savitzki-Golay.⁷³

4.1.2.2. Determinação do número de variáveis latentes e seleção de variáveis na construção dos modelos

A escolha do melhor número de VLs dos modelos construídos foi baseada na avaliação do RMSECV, conforme descrito no item 2.2.2.4.2. Nos métodos de calibração multivariada, o número apropriado de VLs é aquele que melhor descreve a variabilidade do conjunto de dados, tendo por objetivo avaliar a dimensionalidade dos modelos desenvolvidos, sem que ocorra subajuste ou sobreajuste.⁴⁰ Segundo o pacote iToolbox, este deve ser escolhido através do menor valor de RMSECV.⁶¹ Como exemplo, na Figura 16 é demonstrado o gráfico do número de variáveis latentes vs. RMSECV.

⁷³ Savitzky, A.; Golay, M. J. E.; *Anal. Chem.* 36 (1964) 1627-1632.

⁴⁰ Gemperline, P. J.; *Practical guide to chemometrics*, 2006, 140-145.

⁶¹ Norgaard, L.; *et al.*, *Appl. Spectrosc.* 54 (2000) 413-419.

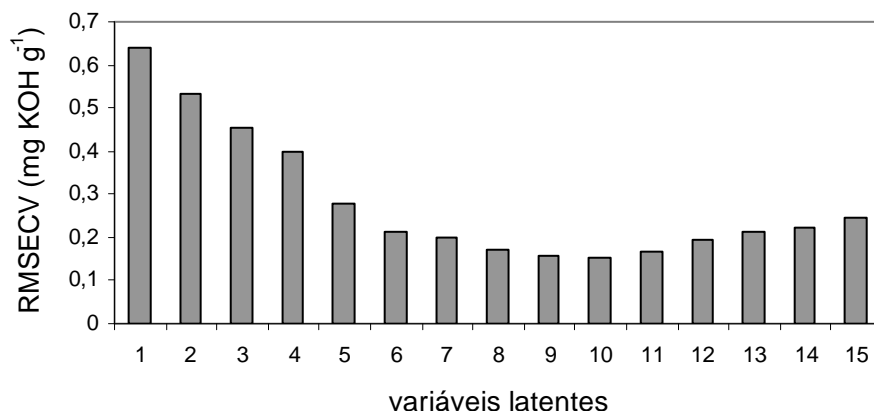


Figura 16. Gráfico utilizado para a seleção das VLs em função do RMSECV.

Com base no gráfico da Figura 16, o algoritmo determina 10 VLs, pois não há mais redução do RMSECV em relação ao valor de RMSECV da variável latente posterior. Porém, o número de VLs pode apresentar uma evolução monotônica a partir de determinadas VLs, como pode ser observado na Figura 17. Nesse caso, ao invés de utilizar 6 VLs, como indicado pelo pacote iToolbox, pode-se utilizar 5 VLs, pois essas 5 VLs apresentam um valor de RMSECV de $0,1989 \text{ mg KOH g}^{-1}$, não apresentando diferença significativa em relação a 6 VLs, que apresentam um valor de RMSECV de $0,1975 \text{ mg KOH g}^{-1}$.

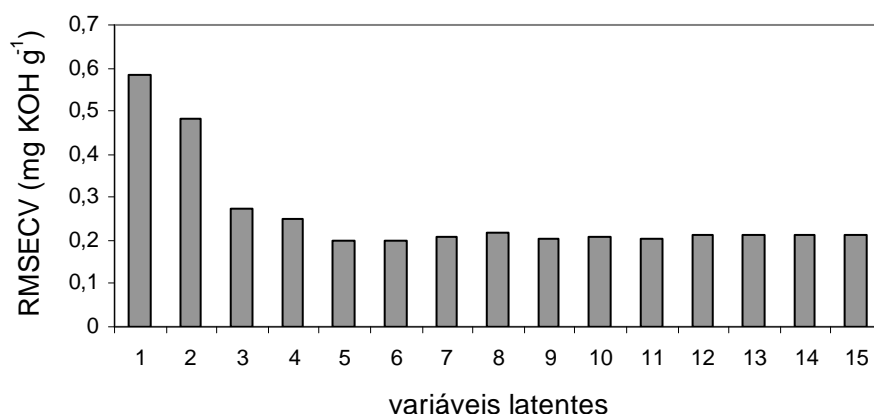


Figura 17. Gráfico utilizado para a seleção das VLs em função do RMSECV.

Os espectros das amostras de RAT e RV foram divididos pelos algoritmos iPLS, siPLS e biPLS em vários subintervalos, conforme o exemplo ilustrado na Figura 18.

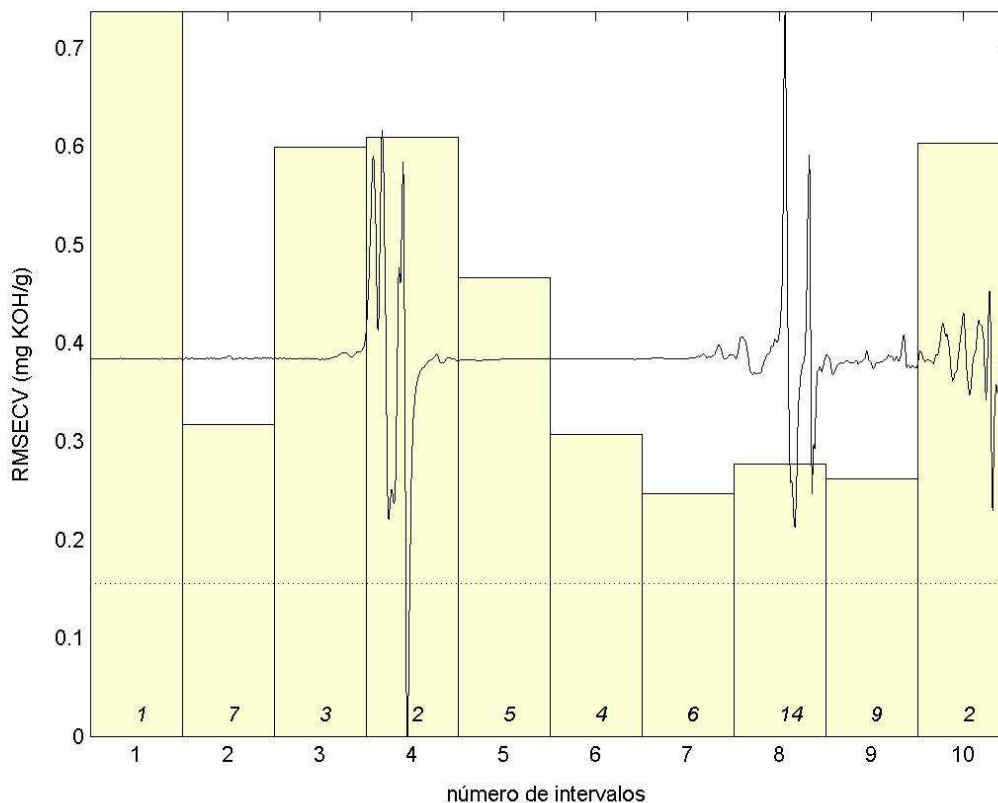


Figura 18. Representação da divisão das variáveis (números de onda), aplicado para as amostras de RAT e RV.

Neste exemplo, o espectro de primeira derivada foi dividido em 10 intervalos. A linha tracejada corresponde ao valor do RMSECV para o modelo global, enquanto que as barras correspondem ao RMSECV dos intervalos e o número na base da barra corresponde às variáveis latentes utilizadas para construir cada modelo. Assim, na Figura 18 o subintervalo 7 possui o menor RMSECV (em torno de $0,25 \text{ mg KOH g}^{-1}$) e indica a utilização de 6 variáveis latentes. Deve-se verificar a faixa de número de onda que está compreendida neste intervalo e neste caso, o intervalo 7 (Figura 18) compreende a faixa espectral de 1990 até 1657 cm^{-1} .

Dessa forma, os algoritmos de seleção de variáveis têm por objetivo utilizar o menor número de variáveis para descrever a propriedade de interesse, bem como identificar os sinais do espectro relacionados a esta propriedade.⁶¹

4.1.2.2.1. Modelo iPLS

O algoritmo iPLS além ser utilizado para a seleção de variáveis, foi utilizado, também, para confirmar o tratamento e o pré-processamento utilizado no modelo global (PLS). Nas Tabelas 10, 11, 12 e 13 são apresentados os melhores modelos construídos com o algoritmo iPLS, sendo o iPLS10 correspondente ao modelo no qual o espectro foi dividido em 10 intervalos, o iPLS20 em 20 intervalos e assim por diante.

Tabela 10. Parâmetros estatísticos dos modelos iPLS, empregando a MSC como tratamento e autoescalado (A) como pré-processamento, para o TAN, em mg KOH g⁻¹, para amostras de RAT e RV.

Modelos*	TV ^a	Intervalo	VLS ^b	RMSECV (mg KOH g ⁻¹)	R ² _{cal}	RMSEP (mg KOH g ⁻¹)
iPLS10	335	7	7	0,2455	0,9368	0,2160
iPLS20	167	12	4	0,2375	0,9277	0,2124
iPLS40	84	24	3	0,3602	0,8099	0,3293
iPLS60	57	36	2	0,4234	0,7371	0,2186

^a TV: número total de variáveis, ^b VLS: variáveis latentes.

* Em negrito: modelo selecionado.

⁶¹ Norgaard, L.; *et al.*, *Appl. Spectrosc.* 54 (2000) 413-419.

Tabela 11. Parâmetros estatísticos dos modelos iPLS, empregando a MSC como tratamento e centrado na média (CM) como pré-processamento, para o TAN, em mg KOH g⁻¹, para amostras de RAT e RV.

Modelos*	TV ^a	Intervalo	VLS ^b	RMSECV (mg KOH g ⁻¹)	R ² _{cal}	RMSEP (mg KOH g ⁻¹)
iPLS10	335	7	8	0,2103	0,9600	0,2315
iPLS20	166	14	6	0,3442	0,9046	0,1554
iPLS40	84	24	2	0,3624	0,7917	0,3150
iPLS60	57	36	2	0,4318	0,7330	0,2399

^a TV: número total de variáveis, ^b VLS: variáveis latentes.

* Em negrito: modelo selecionado.

Tabela 12. Parâmetros estatísticos dos modelos iPLS, empregando a primeira derivada (D) como tratamento e autoescalado (A) como pré-processamento, para o TAN, em mg KOH g⁻¹, para amostras de RAT e RV.

Modelos*	TV ^a	Intervalo	VLS ^b	RMSECV (mg KOH g ⁻¹)	R ² _{cal}	RMSEP (mg KOH g ⁻¹)
iPLS10	335	7	8	0,1527	0,9850	0,2287
iPLS20	167	14	4	0,2821	0,9041	0,2383
iPLS40	84	28	6	0,2527	0,9210	0,4136
iPLS60	57	41	6	0,2747	0,9282	0,1712

^a TV: número total de variáveis, ^b VLS: variáveis latentes.

* Em negrito: modelo selecionado.

Tabela 13. Parâmetros estatísticos dos modelos iPLS, empregando a primeira derivada (D) como tratamento e centrado na média (CM) como pré-processamento, para o TAN, em mg KOH g⁻¹, para amostras de RAT e RV.

Modelos*	TV ^a	Intervalo	VLs ^b	RMSECV (mg KOH g ⁻¹)	R ² _{cal}	RMSEP (mg KOH g ⁻¹)
iPLS10	335	7	5	0,2585	0,9182	0,3871
iPLS20	167	13	8	0,2561	0,9745	0,2348
iPLS40	84	28	5	0,2831	0,9515	0,4868
iPLS60	57	41	4	0,2622	0,9146	0,2422

^a TV: número total de variáveis, ^b VLs: variáveis latentes.

* Em negrito: modelo selecionado.

O melhor modelo utilizando iPLS, MSC e dados autoescalados é o modelo iPLS20 (em negrito) com coeficiente de determinação 0,9277, 4 VLs e valor de RMSEP de 0,2124 mg KOH g⁻¹ (Tabela 10). Já na Tabela 11 são apresentados os parâmetros estatísticos obtidos para os modelos iPLS, MSC e dados centrados na média. Observa-se um aumento no valor de RMSEP e VLs do modelo iPLS10 (em negrito) em relação ao modelo iPLS20 com os dados autoescalados (Tabela 10).

Através da análise das Tabelas 12 e 13, nas quais são mostrados os modelos iPLS com aplicação da primeira derivada e dados autoescalados e centrados na média, observa-se que os modelos, de uma forma geral, apresentam coeficientes de determinação melhores que os modelos com MSC e valores de RMSEP de mesma ordem de grandeza. Os melhores modelos utilizando a primeira derivada estão em negrito nas Tabelas 12 e 13, com valores de RMSEP iguais a 0,2383 e 0,2422 mg KOH g⁻¹, respectivamente. Sendo assim, o modelo da Tabela 13 (em negrito) que utiliza primeira derivada, dados centrados na média e apenas 57 variáveis foi escolhido.

Para a determinação do TAN por ATR-FTIR, utilizando o algoritmo iPLS, o melhor modelo iPLS é o modelo que divide a faixa espectral em 60 intervalos, conforme Tabela 13. Este modelo utiliza 4 variáveis latentes e apenas 57 variáveis (números de onda) para descrever a variabilidade presente nos dados. Cabe destacar que o modelo iPLS20 (Tabela 10) apresenta parâmetros estatísticos semelhantes ao modelo iPLS60 (Tabela 13), porém utiliza três vezes mais variáveis para desenvolver o modelo.

Quando comparado ao modelo global PLS (Tabela 9, em negrito), o modelo mostrado na Tabela 13 apresenta valor de RMSEP maior ($0,2422 \text{ mg KOH g}^{-1}$) do que o apresentado para o modelo global ($0,1889 \text{ mg KOH g}^{-1}$). Isto indica que na aplicação do método de seleção de variáveis iPLS para este conjunto de dados, não houve diminuição no erro de previsão, até o limite de divisão do espectro em 60 intervalos. Segundo Pataca *et al.*⁶⁷ a informação relevante pode estar distribuída por toda a faixa espectral e a seleção de uma única região poderia reduzir a informação necessária para a construção do modelo, o que, automaticamente, induz a um aumento do RMSEP. De um modo geral, o modelo criado com o intervalo 41 (Tabela 13) quando o espectro é dividido em 60 intervalos, reduziu o número de VLs utilizadas, bem como o número de variáveis totais quando comparado ao modelo global (de 3340 para 57). A Figura 19 mostra os valores do TAN obtidos para as amostras de RAT e RV através do método de referência vs. os valores preditos utilizando ATR-FTIR e o modelo iPLS60 selecionado.

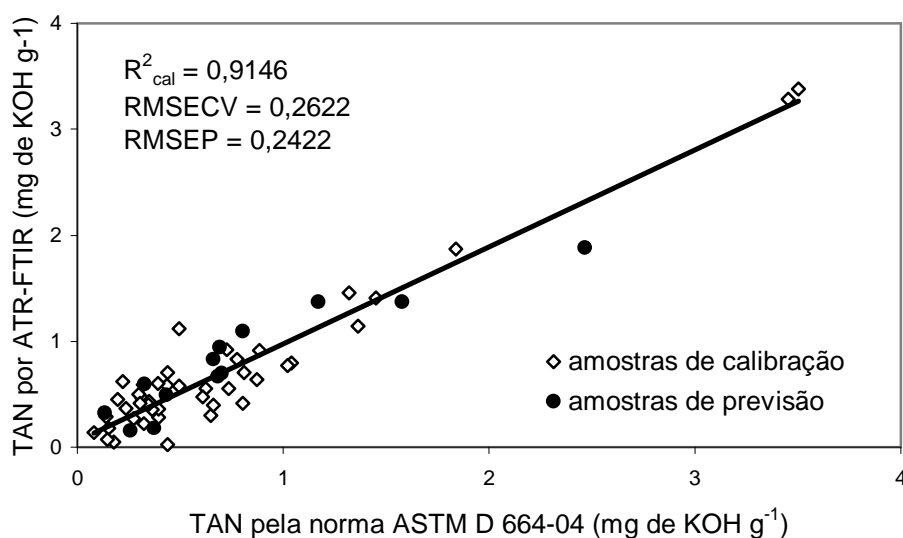


Figura 19. Valores de referência do TAN vs. valores preditos pelo modelo iPLS60, usando o intervalo 41, para as amostras de RAT e RV.

Na Figura 20 é mostrado o gráfico fornecido pelo software MATLAB[®], após divisão do espectro em 60 intervalos. A linha tracejada corresponde ao valor do RMSECV para o modelo global, as barras correspondem ao RMSECV dos

⁶⁷ Pataca, L. C. M.; *et al.*, *Talanta* 71 (2007) 1926-1931.

dar-se de forma aleatória, a amostra 7 apresenta um erro significativo, maior que 100%.

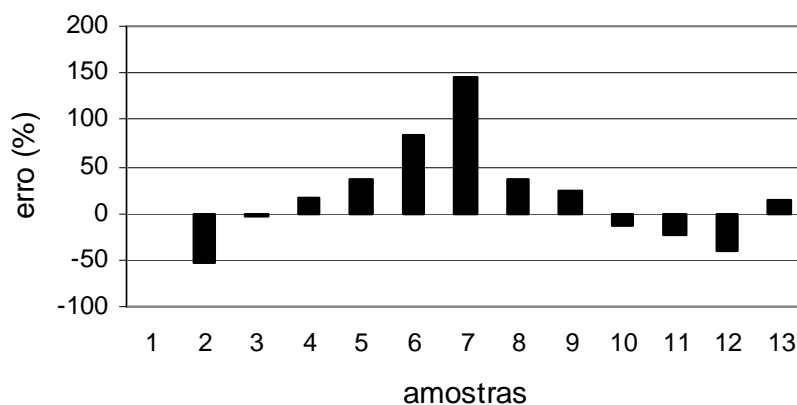


Figura 21. Erro relativo (%) para as 13 amostras (RAT e RV) do conjunto de previsão dos dados na determinação do TAN por ATR-FTIR utilizando o modelo iPLS selecionado.

4.1.2.2.2. Modelo siPLS

Uma vez que o melhor modelo obtido com o algoritmo iPLS utilizou primeira derivada e dados centrados na média, os modelos siPLS foram construídos utilizando estas ferramentas. Os parâmetros estatísticos obtidos para os modelos siPLS para a determinação do TAN por ATR-FTIR, estão apresentados na Tabela 14.

Tabela 14. Parâmetros estatísticos dos modelos siPLS, empregando a primeira derivada (D) como tratamento e centrado na média (CM) como pré-processamento, para o TAN, em mg KOH g⁻¹, para amostras de RAT e RV.

Modelos*	TV ^a	Intervalos	VLS ^b	RMSECV (mg KOH g ⁻¹)	R ² _{cal}	RMSEP (mg KOH g ⁻¹)
siPLS10	670	2 e 7	6	0,2249	0,9548	0,2265
siPLS10	1004	2, 7 e 10	8	0,1829	0,9709	0,2633
siPLS20	333	14 e 17	5	0,2174	0,9348	0,2460
siPLS20	500	7, 14 e 20	9	0,1519	0,9816	0,2111
siPLS40	168	28 e 39	5	0,1989	0,9550	0,3063
siPLS40	253	11, 29 e 33	10	0,1797	0,9819	0,2808
siPLS60	113	41 e 59	5	0,1982	0,9501	0,2267
siPLS60	170	21, 41 e 58	10	0,1514	0,9797	0,3228

^a TV: número total de variáveis, ^b VLS: variáveis latentes.

* Em negrito: modelo selecionado.

O melhor modelo obtido pelo algoritmo siPLS, dividiu o espectro em 20 subintervalos, combinados 3 a 3, selecionando os intervalos 7, 14 e 20 (Tabela 14, em negrito). Este modelo reduziu significativamente o número de variáveis utilizadas, quando comparado ao modelo global (de 3340 para 500) (Tabela 9, em negrito). Porém, apresenta um valor de RMSEP maior (0,2111 mg KOH g⁻¹) do que o apresentado para o modelo global (0,1889 mg KOH g⁻¹). Isto indica, igualmente como no melhor modelo iPLS, que não houve diminuição no erro de previsão, até o limite de divisão do espectro em 20 intervalos, mesmo combinando-os 3 a 3, acarretando em pequena perda da informação presente nos dados. Desta forma, é provável que a informação relevante possa estar distribuída em mais faixas espectrais. A Figura 22 mostra os valores do TAN obtidos para as amostras de RAT e RV através do método de referência vs. os valores preditos utilizando ATR-FTIR e o modelo siPLS20 selecionado.

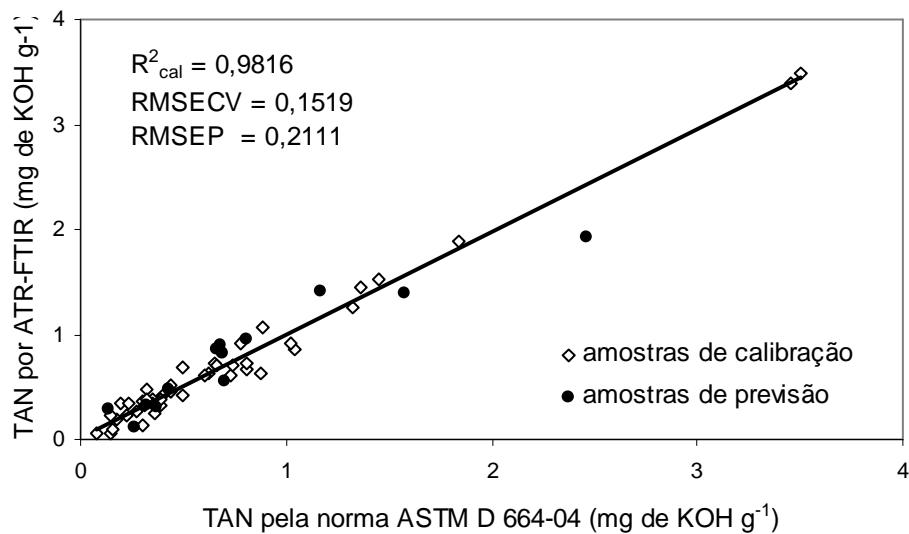


Figura 22. Valores de referência do TAN vs. valores preditos pelo modelo siPLS20, usando os intervalos 7,14 e 20, para as amostras de RAT e RV.

Na Figura 23 é mostrado o gráfico fornecido pelo software MATLAB[®], após divisão do espectro em 20 intervalos para o modelo iPLS20. A linha tracejada corresponde ao valor do RMSECV para o modelo global, as barras correspondem ao RMSECV dos intervalos e o número na base da barra corresponde às variáveis latentes utilizadas para construir cada modelo. Com base neste gráfico podemos observar quais intervalos e correspondentes sinais no espectro foram modelados no siPLS20.

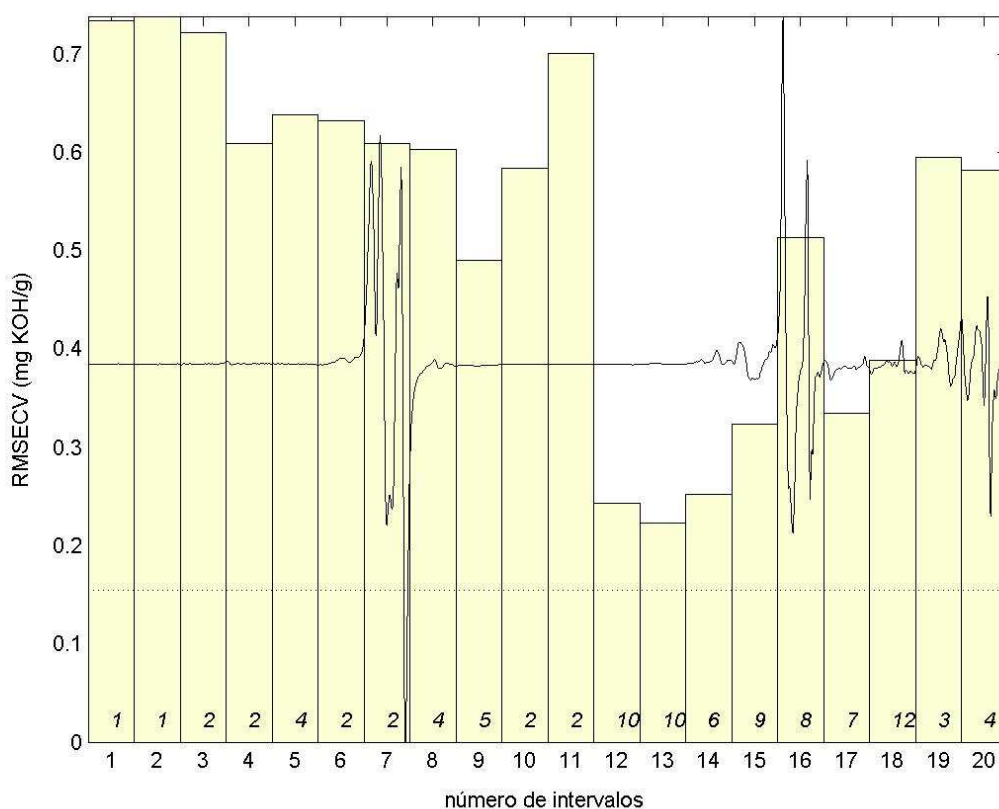


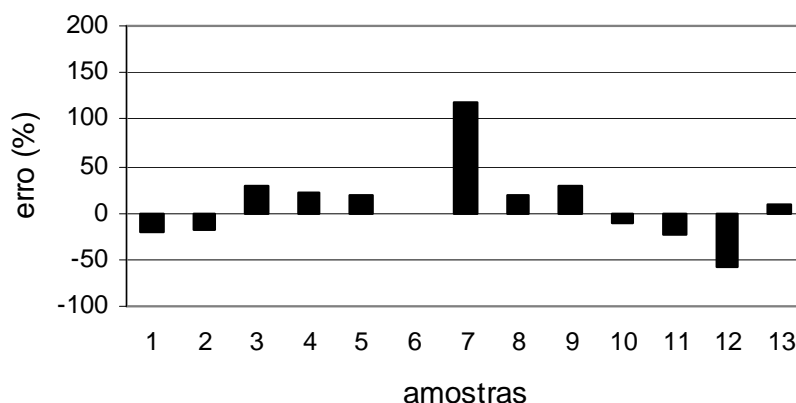
Figura 23. Valores de RMSECV para o modelo global (linha tracejada) e para os intervalos (barras) na determinação do TAN por ATR-FTIR. Os números correspondem às VLS para cada intervalo.

Os estiramentos atribuídos aos intervalos 7, 14 e 20 ilustrados na Figura 23, são listados na Tabela 15, indicando a faixa de absorção e a que parte da estrutura química são correspondentes. Estes grupamentos podem fazer parte da composição química do RAT e do RV.

Tabela 15. Atribuições das regiões selecionadas para o melhor modelo siPLS, aplicado para amostras de RAT e RV.⁹

Intervalo e região selecionada (cm ⁻¹)	Faixa de absorção (cm ⁻¹)*	Atribuição*	Possível vibração
7: 2992 - 2826	2975 - 2950	ν_{as} CH ₃	alcanos e grupos alquila
	2885 - 2865	ν_s CH ₃	
	2930 - 2920	ν_{as} CH ₂	
	2870 - 2840	ν_s CH ₂	
14: 1823 - 1657	1750 - 1725	ν C=O	ésteres e lactonas
	1725 - 1710	ν C=O	ácidos carboxílicos saturados, entre eles os ácidos naftênicos
	1730 - 1680	C ₂ O ₄ ⁻	oxalatos
20: 821 - 655	~ 800	ρ NH ₂ ⁺	sais de amônio
	760 - 730	γ C-H	derivados de naftaleno
	770 - 740		
	770 - 730	γ C-H	benzeno monossustituído
	710 - 690	γ C=C	

A Figura 24 ilustra o erro relativo (%) para as 13 amostras do conjunto de previsão. Igualmente, como no modelo iPLS selecionado, no melhor modelo siPLS a amostra 7 permaneceu com erro maior que 100%.

**Figura 24.** Erro relativo (%) para as 13 amostras (RAT e RV) do conjunto de previsão dos dados na determinação do TAN por ATR-FTIR utilizando o modelo siPLS selecionado.

⁹ Barbosa, L. C. A.; Espectroscopia no infravermelho: na caracterização de compostos orgânicos, 2007.

4.1.2.2.3. Modelo biPLS

Da mesma forma que nos modelos siPLS, utilizou-se somente a primeira derivada e os dados foram centrados na média. Os parâmetros estatísticos obtidos para os modelos biPLS para a determinação do TAN por ATR-FTIR, estão demonstrados na Tabela 16.

Tabela 16. Parâmetros estatísticos dos modelos biPLS, empregando a primeira derivada (D) como tratamento e centrado na média (CM) como pré-processamento, para o TAN, em mg KOH g⁻¹, para amostras de RAT e RV.

Modelos*	TV ^a	Intervalos	VLS ^b	RMSECV (mg KOH g ⁻¹)	R ² _{cal}	RMSEP (mg KOH g ⁻¹)
biPLS10	1139	4, 7, 8 e 10	9	0,1662	0,9811	0,1686
biPLS20	836	7, 14, 15, 16 e 20	9	0,1649	0,9819	0,1642
biPLS40	173	8, 14, 28, 29, 31,35, 37 e 40	10	0,1245	0,9936	0,2182
biPLS60	225	41, 46, 55 e 59	7	0,1902	0,9628	0,1977

^a TV: número total de variáveis, ^b VLS: variáveis latentes.

* Em negrito: modelo selecionado.

O melhor modelo construído utilizando o algoritmo biPLS, dividiu o espectro em 20 subintervalos, sendo selecionados os intervalos 7, 14, 15, 16 e 20 (Tabela 16, em negrito). O modelo forneceu um valor de RMSEP menor (0,1642 mg KOH g⁻¹) do que o apresentado para o modelo global (0,1889 mg KOH g⁻¹) (Tabela 9, em negrito), bem como, apresentou um valor de RMSECV na mesma ordem de grandeza do RMSEP. Isto indica que os conjuntos de calibração e previsão possuem erros semelhantes e que a faixa de trabalho do conjunto de calibração foi adequada para prever as concentrações do conjunto de previsão. Além disso, também reduziu significativamente o número de variáveis totais para a construção do modelo, quando comparado ao modelo global (de 3340 para 836).

Quando comparado ao modelo siPLS20 (Tabela 14, em negrito) os intervalos 15 e 16 incluíram a informação necessária para o desenvolvimento de um modelo mais robusto, com o mesmo número de VLS. A Figura 25 mostra os

valores do TAN obtidos para as amostras de RAT e RV através do método de referência vs. os valores preditos utilizando ATR-FTIR e o modelo biPLS20 selecionado.

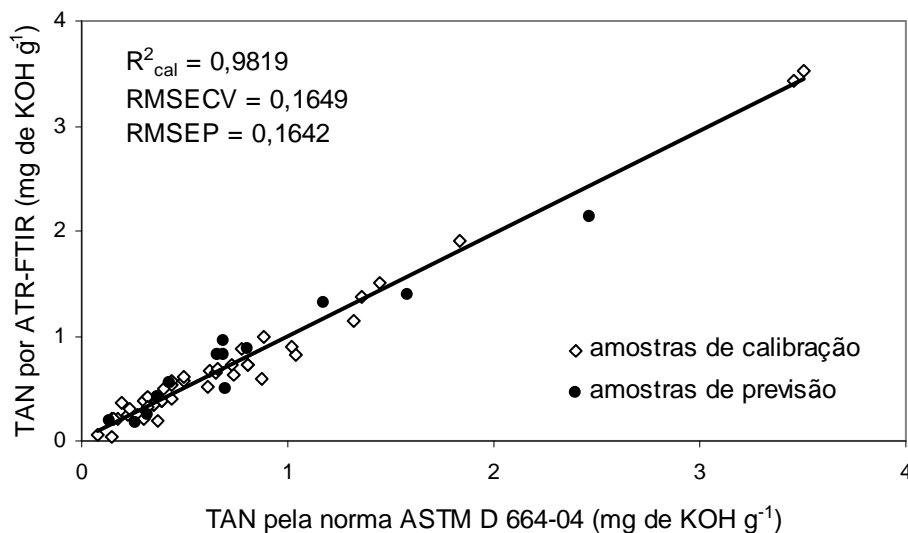


Figura 25. Valores de referência do TAN vs. valores preditos pelo modelo biPLS20 usando os intervalos 7, 14, 15, 16 e 20, para as amostras de RAT e RV.

Na Figura 26 é mostrado o gráfico fornecido pelo software MATLAB[®], após a divisão do espectro em 20 intervalos para o modelo iPLS20. A linha tracejada corresponde ao valor do RMSECV para o modelo global, as barras correspondem ao RMSECV dos intervalos e o número na base da barra corresponde às variáveis latentes utilizadas para construir cada modelo. Com base neste gráfico podemos observar quais intervalos e correspondentes sinais no espectro foram modelados no biPLS20.

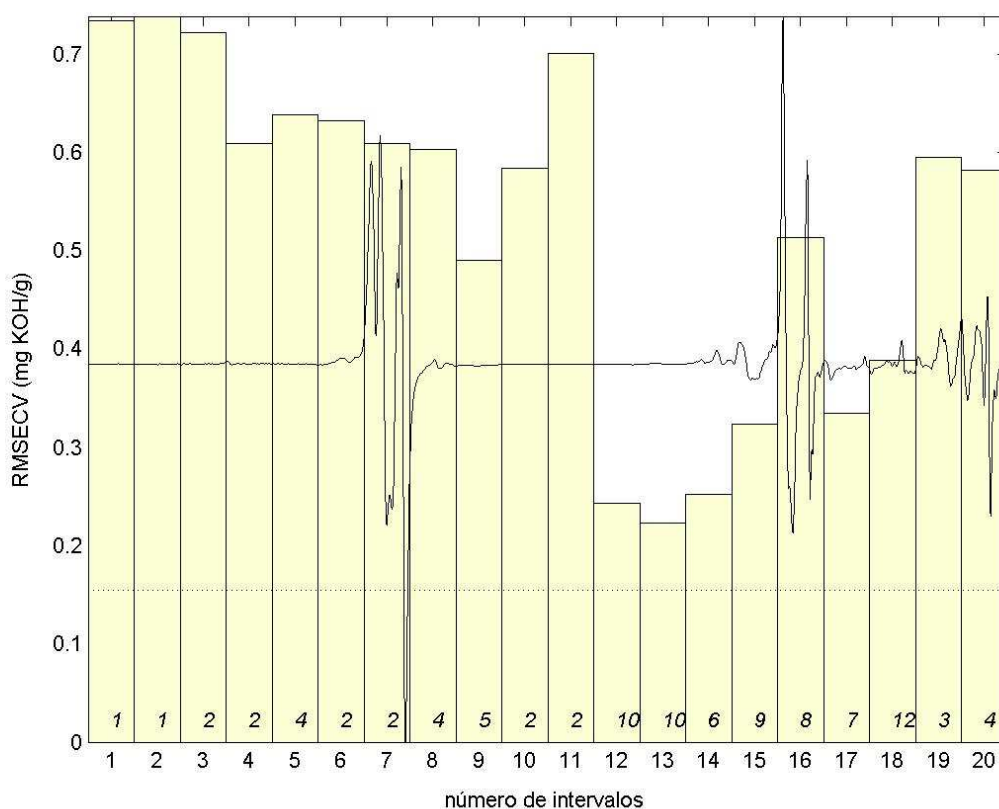


Figura 26. Valores de RMSECV para o modelo global (linha tracejada) e para os intervalos (barras) na determinação do TAN por ATR-FTIR. Os números correspondem às VLs para cada intervalo.

Os estiramentos atribuídos aos intervalos 7, 14, 15, 16 e 20 ilustrados na Figura 26, são listados na Tabela 17, indicando a faixa de absorção e a que parte da estrutura química são correspondentes. Estes grupamentos podem fazer parte da estrutura química do RAT e do RV.

Tabela 17. Atribuições das regiões selecionadas para o melhor modelo biPLS, aplicado para amostras de RAT e RV.⁹

Intervalo e região selecionada (cm ⁻¹)	Faixa de absorção (cm ⁻¹)*	Atribuição*	Possível vibração
7: 2992 - 2826	2975 - 2950	$\nu_{as} CH_3$	alcanos e grupos alquila
	2885 - 2865	$\nu_s CH_3$	
	2930 - 2920	$\nu_{as} CH_2$	
	2870 - 2840	$\nu_s CH_2$	
14: 1823 - 1657	1750 - 1725	$\nu C=O$	ésteres e lactonas
	1725 - 1710	$\nu C=O$	ácidos carboxílicos saturados, entre eles os ácidos naftênicos
	1730 - 1680	$C_2O_4^-$	oxalatos
15: 1656 - 1490	1650 - 1550	$\nu_{as} CO_2^-$	sais de ácidos carboxílicos
	1625 - 1560	$\delta_{as} NH_3^+$	sais de amônio
	1550 - 1500	$\delta_s NH_3^+$	
16: 1489 - 1323	1440 - 1335	$\nu_s CO_2^-$	sais de ácidos carboxílicos
	1400 - 1300	NO_2^-	nitritos
	1410 - 1340	NO_3^-	nitratos
	1465 - 1440	$\delta_{as} CH_3$	alcanos e grupos alquila
	1390 - 1370	$\delta_s CH_3$	
	1480 - 1440	$\delta_s CH_2$	
20: 821 - 655	~ 800	ρNH_2^+	sais de amônio
	760 - 730	$\gamma C-H$	derivados de naftaleno
	770 - 740		
	770 - 730	$\gamma C-H$	benzeno monossustituído
	710 - 690	$\gamma C=C$	

Com respeito ao erro percentual fornecido pelo melhor modelo biPLS selecionado, na Figura 27 é verificado de forma homogênea, que os erros atribuídos as amostras estão distribuídos aleatoriamente entre valores positivos e negativos, o que indica que o modelo não é tendencioso, além de ter reduzido significativamente o erro da amostra 7.

⁹ Barbosa, L. C. A.; Espectroscopia no infravermelho: na caracterização de compostos orgânicos, 2007.

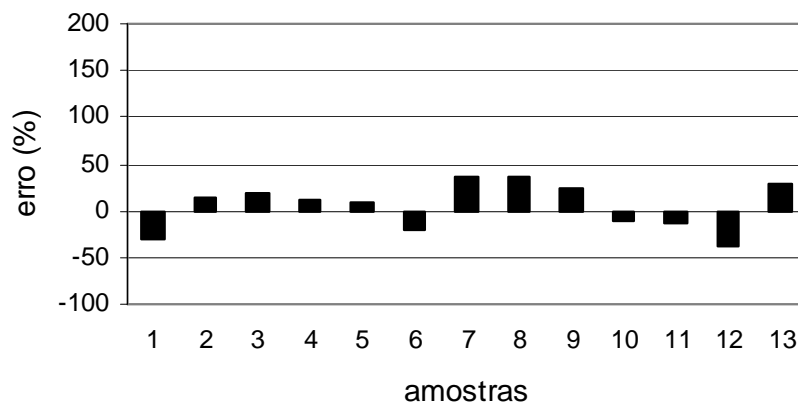


Figura 27. Erro relativo (%) para as 13 amostras (RAT e RV) do conjunto de previsão dos dados na determinação do TAN por ATR-FTIR utilizando o modelo biPLS selecionado.

Conforme Tabela 18, através do emprego do algoritmo de seleção de variáveis biPLS, foi possível obter um modelo adequado para a determinação do TAN por ATR-FTIR, obtendo-se um erro percentual médio de 22,46%.

Tabela 18. Resultado previsto e erro percentual apresentado pelos melhores modelos desenvolvidos empregando os métodos de seleção de variáveis iPLS, siPLS e biPLS, para a determinação do TAN (mg KOH g⁻¹) por ATR-FTIR de amostras de RAT e RV.

Referência (mg KOH g ⁻¹)	iPLS		siPLS		biPLS	
	Previsto (mg KOH g ⁻¹)	Erro* (%)	Previsto (mg KOH g ⁻¹)	Erro* (%)	Previsto (mg KOH g ⁻¹)	Erro* (%)
0,70	0,70	0,37	0,55	21,20	0,49	30,39
0,37	0,18	51,85	0,30	19,10	0,42	13,30
0,68	0,67	2,38	0,89	30,09	0,81	18,92
1,17	1,37	17,03	1,42	20,91	1,32	12,55
0,80	1,09	36,06	0,95	18,39	0,87	8,49
0,33	0,60	83,42	0,32	1,63	0,26	21,10
0,13	0,33	144,59	0,29	116,97	0,18	35,89
0,69	0,95	36,54	0,82	18,32	0,95	36,84
0,66	0,83	25,13	0,86	29,29	0,81	22,78
1,58	1,37	13,31	1,40	11,61	1,40	11,46
2,46	1,88	23,67	1,92	21,94	2,14	13,19
0,26	0,15	40,61	0,11	58,30	0,16	37,05
0,43	0,49	14,20	0,47	8,42	0,56	29,98
Erro médio (%)	37,63		28,94		22,46	

* Valores em módulo.

4.2. Comparação da norma ASTM D 664-04 com a metodologia proposta

Através da aplicação da regressão por OLS⁴⁴ e regressão ortogonal, não se observou erro sistemático quando comparados os valores obtidos pela metodologia proposta (biPLS), com os valores da norma ASTM D 664-04 (valor de F calculado menor que F tabelado).

⁴⁴ Hartmann, C.; *et al.*; *Anal. Chim. Acta* 338 (1997) 19-40.

A equação matemática obtida pelo OLS para correlacionar os métodos segue a equação de regressão 20. O gráfico de correlação pode ser visualizado na Figura 28.

$$D664-04 = -0,076 + 1,085 \times biPLS \quad (20)$$

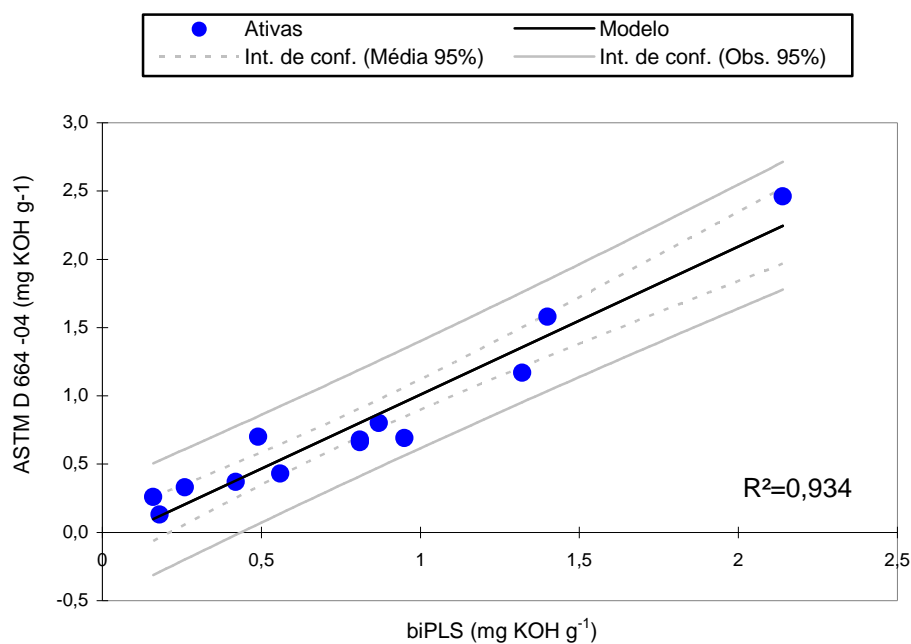


Figura 28. Gráfico de correlação entre os resultados da norma ASTM D 664-04 vs. biPLS, obtido através da regressão linear por OLS, com um grau de 95% de confiança.

O coeficiente de determinação encontrado conforme a equação de regressão 20 foi de 0,934, indicando que os resultados são equivalentes e que apresentam uma diferença cerca de 8,5% do método biPLS em relação a norma ASTM664-04.

5. CONCLUSÃO

A espectroscopia no infravermelho com transformada de Fourier e reflexão total atenuada (ATR-FTIR) combinada às técnicas quimiométricas, mostrou-se um método apropriado e rápido para a determinação do número de acidez total em amostras de resíduo da destilação atmosférica e da destilação a vácuo do petróleo. Esse método tem como vantagens: baixo tempo para a realização da análise; a não destruição da amostra, a facilidade no manuseio dos instrumentos; apresenta um baixo custo operacional frente a sua utilização diária em análises de rotina num ambiente industrial.

O modelo utilizando todas as variáveis independentes (modelo global com 3340 comprimentos de onda do espectro de infravermelho) forneceu um valor de RMSEP de 0,1889 mg KOH g⁻¹, coeficiente de determinação de 0,9855, utilizando 10 variáveis latentes (VLs). Porém, empregando-se o método de seleção de variáveis biPLS foi possível obter um modelo melhor, que apresentou um valor de RMSEP de 0,1642 mg KOH g⁻¹ (13% menor que o do modelo global), coeficiente de determinação de 0,9819, 9 variáveis latentes, utilizando somente 836 variáveis independentes (75% menos variáveis que o modelo global).

Assim, o método biPLS mostrou-se uma ferramenta eficiente na redução do número de variáveis independentes aumentando a exatidão e a precisão dos resultados obtidos em relação ao modelo global. Estes resultados podem ser considerados razoáveis visto que o erro médio, da ordem de 20%, pode ser atribuído às características das amostras e aos problemas apresentados pelo método de referência evidenciados por Fuhr *et al.*³⁵ tais como dificuldade de solubilização das amostras, precipitação de asfatenos e reprodutibilidade. Ainda, o número de variáveis latentes elevado, utilizado na modelagem do TAN, pode ser justificado pela complexidade das amostras de RAT e RV. Conforme a norma

³⁵ Fuhr, B.; *et al.*, *Energy Fuels* 21 (2007) 1322-1324.

ASTM D 664-04 muitos compostos presentes nas amostras são responsáveis pela propriedade em questão, sendo os sinais espectrais destes indispensáveis na construção de modelos multivariados que consigam prever o TAN.

A primeira derivada e os dados centrados na média foram as melhores ferramentas de tratamento e pré-processamento, respectivamente, fornecendo resultados mais exatos quando comparados a correção do espalhamento de luz e/ou autoescalamento dos dados.

6. REFERÊNCIAS BIBLIOGRÁFICAS

1. Abbas, O.; Dupuy, N.; Rebufa, C.; Vrielynck, L.; Kister, J.; Permanyer, A.; Prediction of source rock origin by chemometric analysis of Fourier transform infrared-attenuated total reflectance spectra of oil petroleum: evaluation of aliphatic and aromatic fractions by self-modeling mixture analysis, *Appl. Spectrosc.* 60 (2006) 304-314.
2. Akhlaq, M. S.; Characterization of the isolated wetting crude oil components with infrared spectroscopy, *J. Pet. Sci. Eng.* 22 (1999) 229-235.
3. Annual Book of ASTM Standards, ASTM D 1298-99, *Standard test method for density, relative density (specific gravity), or API gravity of crude petroleum and liquid petroleum products by hydrometer method*, 1999.
4. Annual Book of ASTM Standards, ASTM D 4175-96, *Standard terminology relating to petroleum, petroleum products, and lubricants*, 1996.
5. Annual Book of ASTM Standards, ASTM D 445-97, *Standard test method for kinematic viscosity of transparent and opaque liquids (the calculations of dynamic viscosity)*, 1997.
6. Annual Book of ASTM Standards, ASTM D 664-04, *Standard Test Method for Acid Number of Petroleum Products by Potenciometric Titration*, 2004.
7. Aske, N.; Kallevik, H.; Sjöblom, J.; Water-in-crude oil emulsion stability studied by critical electric field measurements. Correlation to physico-chemical parameters and near-infrared spectroscopy, *J. Pet. Sci. Eng.* 36 (2002) 1-17.
8. Babaian-Kibala, E.; Craig, H. L.; Rush, G. L.; Blanchard, K. V.; Rose, T. J.; Uehlein, B. L.; Quinter, R. C.; Summers, M. L.; Naphthenic acid corrosion in refinery settings, *Mater. Perform.* 32 (1993) 50-55.

9. Barbosa, L. C. A.; *Espectroscopia no infravermelho: na caracterização de compostos orgânicos*, Viçosa: Ed. UFV, 2007; (a) p. 25-43, (b) p. 38-41.
10. Barth, T.; Hoiland, S.; Fotland, P.; Askvik, K. M.; Pedersen, B. S.; Borgund, A. E.; Acidic compounds in biodegraded petroleum, *Org. Geochem.* 35 (2004) 1513-1525.
11. Bathia, S.; Sharma, D. K.; Emerging role of biorefining of heavier crude oils and integration of biorefining with petroleum refineries in the future, *Pet. Sci. Technol.* 24 (2006) 1125-1159.
12. Borin, A.; Poppi, R. J.; Application of mid infrared spectroscopy and iPLS for the quantification of contaminants in lubricating oil, *Vib. Spectrosc.* 37 (2005) 27-32.
13. Borin, A.; Poppi, R. J.; Multivariate quality control of lubricating oils using Fourier transform infrared spectroscopy, *J. Braz. Chem. Soc.* 15 (2004) 570-576.
14. Bratchell, N.; Cluster Analysis, *Chemom. Intell. Lab. Syst.* 6 (1989) 105-125.
15. Brereton, R. G.; Introduction to multivariate calibration in analytical chemistry, *Analyst* 125 (2000) 2125-2154.
16. Brereton, R. G.; *Chemometrics: Data analysis for the laboratory and chemical plant*, *Chemical Analysis Series*, Chichester: John Wiley & Sons, Ltd., 2003; (a) p. 120-142, (b) p. 297-306, (c) p. 313-323.
17. Brown, S. D.; Chemical systems under indirect observation: latent properties and chemometrics, *Appl. Spectrosc.* 49 (1995) 14A-31A.
18. Carlsson, D. J.; Wiles, D. M.; Surface studies by attenuated total reflection spectroscopy, *Can. J. Chem.* 48 (1970) 45-178.
19. Chau, F.; Liang, Y.; Gao, J.; Shao, X.; *Chemometrics: From basics to wavelet transform*, *Chemical Analysis Series*, New Jersey: John Wiley & Sons, Inc., 2004; p. 1-21.
20. Chung, H.; Ku, M.; Comparison of near-infrared, infrared, and Raman spectroscopy for the analysis of heavy petroleum products, *Appl. Spectrosc.* 54 (2000) 239-245.

21. Coelho, R. R.; Havell, I.; Monte, M. B. M.; Middea, A. L. S.; Characterization of aliphatic chains in vacuum residues (VRs) of asphaltenes and resins using molecular modelling and FTIR techniques, *Fuel Process. Technol.* 87 (2006) 325-333.
22. Cooper, J. B.; Wise, K. L.; Welch, W. T.; Summer, 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, *Appl. Spectrosc.* 51 (1997) 1613-1620.
23. Correia, P. R. M.; Ferreira, M. M. C.; Reconhecimento de padrões por métodos não supervisionados: explorando procedimentos quimiométricos para tratamento de dados analíticos, *Quim. Nova* 30 (2007) 481-487.
24. Costa, P. A.; Arruda, M. A. Z.; Mello, C.; Pereira-Filho, E. R.; Poppi, R. J.; Neuro-genetic approach for optimization of the spectrophotometric catalytic determination of cobalt, *Anal. Chim. Acta* 433 (2001) 111-117.
25. Costa, P. A.; Poppi, R. J.; Algoritmo genético em química, *Quim. Nova* 22 (1999) 405-411.
26. Eastment, H. T.; Krzanowski, W. J.; Cross-validatory choice of the number of components from a principal component analysis, *Technometrics* 24 (1982) 73-77.
27. Einax, J. W.; Zwanziger, H. W.; Geiß, S.; *Chemometrics in environmental analysis*, Weinheim: Wiley-VCH, 1997; p. 153-178.
28. Fahrenfort, J.; Attenuated total reflection, *Spectrochim. Acta* 17 (1961) 698-709.
29. Falla, F. S.; Larini, C.; Le Roux, G. A. C.; Quina, F. H.; Moro, L. F. L.; Nascimento, C. A. O.; Characterization of crude petroleum by NIR, *Pet. Sci. Eng.* 51 (2006) 127-137.
30. Ferrão, M. F.; *Aplicação de técnicas espectroscópicas de reflexão no infravermelho no controle de qualidade de farinha de trigo*, Tese de Doutorado, Instituto de Química. Universidade Estadual de Campinas, Campinas-SP, 2000.
31. Ferrão, M. F.; Técnicas de reflexão no infravermelho aplicadas na análise de alimentos, *Tecno-lógica* 5 (2001) 65-85.

32. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O.; Quimiometria I: calibração multivariada, um tutorial, *Quim. Nova* 22 (1999) 724-731.
33. Ferreira, M. M. C.; Multivariate QSAR, *J. Braz. Chem. Soc.* 13 (2002) 742-753.
34. Fossen, M.; Hemmingsen, P. V.; Hannisdal, A.; Sjöblom, J.; Kallevik, H.; Solubility parameters based on IR and NIR spectra: I. Correlation to polar solutes and binary systems, *J. Dispersion Sci. Technol.* 26 (2005) 227-241.
35. Fuhr, B.; Banjac, B.; Blackmore, T.; Rahimi, P.; Applicability of total acid number analysis to heavy oils and bitumens, *Energy Fuels* 21 (2007) 1322-1324.
36. Gabrielsson, J.; Lindberg, N; Lundstedt, T.; Multivariate methods in pharmaceutical applications. *J. Chemom.* 16 (2002) 141-160.
37. Gary, J. H.; Hadwerk, G. E.; *Petroleum refining: technology and economics*, 4th ed., New York: Marcel Dekker, Inc., 2001; p. 37-57.
38. Geladi, P.; Kowalski, B. R.; Partial least-squares: a tutorial, *Anal. Chim. Acta* 185 (1986) 1-17.
39. Geladi, P.; MacDougall, D.; Martens, H.; Linearization and scatter-correction for near-infrared reflectance spectra of meat, *Appl. Spectrosc.* 39 (1985) 491-500.
40. Gemperline, P. J.; Practical guide to chemometrics, 2nd ed., New York: CRC Press Taylor & Francis, 2006; p. 140-145.
41. Hair, J. F.; Anderson, R. E.; Tatham, R. L.; Black, W. C.; *Análise multivariada de dados*, 5^a ed., Porto Alegre: Bookman, 2005; p. 23-29.
42. Hannisdal, A.; Hemmingsen, P. V.; Sjöblom, J.; Group-type analysis of heavy crude oils using vibrational spectroscopy in combination with multivariate analysis, *Ind. Eng. Chem. Res.* 44 (2005) 1349-1357.
43. Harris, C. D.; *Análise Química Quantitativa*, 5^a ed., Rio de Janeiro: LTC Editora, 1999; p. 225-248.

-
44. Hartmann, C.; Smeyers, J.; Penninckx, W.; Massart, D. L.; Detection of bias in method comparison by regression analysis, *Anal. Chim. Acta* 338 (1997) 19-40.
 45. Humphries, M.; Sorrel, G.; Corrosion control in crude oil distillation units, *Mater. Perform.* 15 (1976) 13-21.
 46. Jayaraman, A.; Saxena, R. C.; Corrosion and its control in petroleum refineries – a review, *Corros. Prevent. Contr.* 42 (1995) 123-131.
 47. Jayaraman, A.; Singh, H.; Lefebvre, Y.; Naphthenic acid corrosion in petroleum refineries - a review, *Rev. Inst. Fr. Pet.* 41 (1986) 265-274.
 48. Kallevik, H.; Kvalheim, O. M.; Sjöblom, J.; Quantitative determination of asphaltenes and resins in solution by means of near-infrared spectroscopy. Correlations to emulsion stability, *J. Colloid Interface Sci.* 225 (2000) 494-504.
 49. Khan, Z. H.; Al-Assi, S. H.; Madouh, H. A.; Al-Muhareb, E.; Compositional analysis of a vacuum residue, *Fuel Sci. Technol. Int'L.* 12 (1994) 1413-1424.
 50. Kim, M.; Lee, Y.; Han, C.; Real-time classification of petroleum products using near-infrared spectra, *Comput. Chem. Eng.* 24 (2000) 513-517.
 51. Kraft, W. W.; Vacuum distillation of petroleum residues, *Ind. Eng. Chem.* 40 (1948) 807-819.
 52. Ku, M.; Chung, H.; Comparison of near-infrared and Raman spectroscopy for the determination of chemical and physical properties of naphtha, *Appl. Spectrosc.* 53 (1999) 557-564.
 53. Laredo, G. C.; López, C. R.; Álvarez, R. E.; Cano, J. L.; Naphthenic acids, total acid number and sulfur content profile characterization in Isthmus and Maya crude oils, *Fuel* 83 (2004) 1689-1695.
 54. Lavine, B. K.; Clustering and classification of analytical data in the Encyclopedia of Analytical Chemistry: instrumentation and applications, Chichester: John Wiley & Sons Ltd., 2000; p. 9689-9710.

-
55. Leardi, R.; Application of genetic algorithm-PLS for feature selection in spectral data sets, *J. Chemometrics* 14 (2000) 643-655.
 56. Leardi, R.; Norgaard, L.; Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions, *J. Chemometrics* 18 (2004) 486–497.
 57. Maspoch, S.; Blanco, M.; Coello, J.; Iturriaga, H.; Pagès, J.; Calibration in non-linear near infrared reflectance spectroscopy: a comparison of several methods, *Anal. Chim. Acta* 384 (1999) 207-214.
 58. Matar, S.; *Chemistry of Petrochemical Process*, 2nd ed., Houston: Gulf Publishing Company, 2000; (a) p. 49-111, (b) p. 130-131.
 59. Mirabella, F. M.; Internal reflection spectroscopy, *Appl. Spectrosc. Rev.* 21 (1985) 45-178.
 60. Munck, L.; Nielsen, J. P.; Moller, B.; Jacobsen, S.; Sondergaard, I.; Engelsen, S. B.; Norgaard, L.; Bro, R.; Exploring the phenotypic expression of a regulatory proteome-altering gene by spectroscopy and chemometrics, *Anal. Chim. Acta* 446 (2001) 171-186.
 61. Norgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L.; Engelsen, S. B.; Interval Partial Least-Squares Regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy, *Appl. Spectrosc.* 54 (2000) 413-419.
 62. Olinger, J. M.; Griffiths, P. R.; Quantitative effects of absorbing matrix on near-infrared diffuse reflectance spectra, *Anal. Chem.* 60 (1988) 2427-2435.
 63. Oliveira, F. C. C.; Brandão, C. R. R.; Ramalho, H. F.; Costa, L. A. F.; Suarez, P. A. Z.; Rubim, J. C.; Adulteration of diesel/biodiesel blends by vegetable oil as determined by Fourier transform (FT) near infrared spectrometry and FT-Raman spectroscopy, *Anal. Chim. Acta* 587 (2007) 194-199.
 64. Oliveira, F. C. C.; Souza, A. T. P. C.; Dias, J. A.; Dias, S. C. L.; Rubim, J. C.; A escolha da faixa espectral no uso combinado de métodos espectroscópicos e

- quimiométricos, *Quim. Nova* 27 (2004) 218-225.
65. Pasquini, C.; Bueno, A. F.; Characterization of petroleum using near-infrared spectroscopy: Quantitative modeling for the true boiling point curve and specific gravity, *Fuel* 86 (2007) 1927-1934.
66. Pasquini, C.; Near infrared spectroscopy: fundamentals practical aspects and analytical applications, *J. Braz. Chem. Soc.* 14 (2003) 198-219.
67. Pataca, L. C. M.; Borges, W.; Marcucci, M. C.; Poppi, R. J.; Determination of apparent reducing sugars, moisture and acidity in honey by attenuated total reflectance-Fourier transform infrared spectrometry, *Talanta* 71 (2007) 1926-1931.
68. PIKE Technologies, <http://www.piketch.com>, acessado em 23/09/07.
69. Pironon, J.; Thierry, R.; Teinturier, S.; Walgenwitz, F.; Water in petroleum inclusions: evidence from Raman and FT-IR measurements, PVT consequences, *J. Geochem. Explor.* 69-70 (2000) 663-668.
70. Pirouette, *Multivariate Data Analysis*, versão 3,11, Infometrix, Inc, Washington, EUA.
71. Robbins, W. K.; *Metals, nitrogen and sulfur in petroleum residue, analysis of*, Encyclopedia of Analytical Chemistry, New Jersey: John Wiley & Sons, 2006; p. 1-11.
72. Santos, 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 FTIT, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis, *Anal. Chim. Acta* 547 (2005) 188-196.
73. Savitzky, A.; Golay, M. J. E.; Smoothing and differentiation by simplified least squares procedures, *Anal. Chem.* 36 (1964) 1627-1632.
74. Skoog, D.; Holler, F. J.; Nieman, T. A.; *Principles of instrumental analysis*, Orlando: Harcourt Brace & Company, 1998; p. 380-403.
75. Slavcheva, E.; Shone, B.; Turnbull, A.; Review of naphthenic acid corrosion in oil refining, *Br. Corr. J.* 34 (1999) 125-131.

-
76. Speight, J. G.; *Handbook of Petroleum Analysis*, New York: John Wiley & Sons, Inc., 2001; (a) p. 1, (b) p. 8, (c) p. 11-18, (d) p. 25-41, (e) p. 31-39, (f) p. 51-58, (g) p. 72-75.
 77. Speight, J. G.; *Handbook of Petroleum Product Analysis*, New Jersey: John Wiley & Sons, Inc., 2002; (a) p. 2, (b) p. 34-38, (c) 51-58.
 78. Teixeira, L. S. G.; Oliveira, F. S.; Santos, H. C.; Cordeiro, P. W. L.; Almeida, S. Q.; Multivariate calibration in Fourier transform infrared spectrometry as a tool to detect adulterations in Brazilian gasoline, *Fuel* 8 (2007) 346-352.
 79. Tomczyk, N. A.; Winans, R. E.; Shinn, J. H.; Robinson, R. C.; On the nature and origin of acidic species in petroleum. 1. Detailed acid type distribution in a California crude oil, *Energy Fuels* 15 (2001) 1498-1504.
 80. Voort, F. R.; Fourier transform infrared spectroscopy applied to food analysis, *Food Res. Int.* 25 (1992) 397-403.
 81. Willians, P. C.; Commercial near-infrared reflectance analyzers. In: Near-infrared technology in the agricultural and food industries, *American Association of Cereal Chemists*, Saint Paul: by Phil Willians & Karl Norris, 1987.
 82. Wold, S.; Chemometrics, what do we mean with it, and what do we want from it?, *Chemom. Intell. Lab. Syst.* 30 (1995) 109-115.
 83. Wold, S.; Chemometrics, why, what and where to next?, *J. Pharm. Biomed. Anal.* 9 (1991) 589-596.
 84. Wold, S.; Cross-validatory estimation of the number of components in factor and principal components models, *Technometrics* 20 (1978) 397-405.
 85. Wold, S.; Esbensen, K.; Geladi, P.; Principal component analysis, *Chemom. Intell. Lab. Syst.* 2 (1987) 37-52.
 86. Zhang, A.; Ma, Q.; Wang, K.; Liu, X.; Shuler, P.; Tang, Y.; Naphthenic acid removal from crude oil through catalytic decarboxylation on magnesium oxide, *Appl. Catal. A* 303 (2006) 103-109.

87. Zhao, L.; Cheng, Z.; Ding, Y.; Yuan, P.; Lu, S.; Yuan, W.; Experimental study on vacuum residuum upgrading through pyrolysis in supercritical water, *Energy Fuels* 20 (2006) 2067-2071.