

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

**“Utilização da RMN Associada a Métodos Quimiométricos
na Caracterização de Petróleos e Derivados”.**

Emanuele Catarina da Silva Oliveira

Tese de Doutorado em Química

**VITÓRIA
2016**

Emanuele Catarina da Silva Oliveira

**“Utilização da RMN Associada a Métodos Quimiométricos
caracterização de Petróleos e Derivados”.**

Tese apresentada ao Programa de Pós-Graduação em Química do Centro de Ciências Exatas da Universidade Federal do Espírito Santo como requisito parcial para obtenção do título de Doutor em Química.

Orientador: Prof. Dr. Álvaro Cunha Neto

**VITÓRIA
2016**

**“Utilização da RMN Associada a Métodos Quimiométricos na
caracterização de Petróleos e Derivados”.**

Emanuele Catarina da Silva Oliveira

**Tese submetida ao Programa de Pós-Graduação em Química da Universidade
Federal do Espírito Santo como requisito parcial para a obtenção do grau de
Doutor em Química.**

Aprovado(a) em 19/12/2016 por:

**Prof. Dr. Álvaro Cunha Neto
Universidade Federal do Espírito Santo
Orientador**

**Prof. Dr. Valdemar Lacerda Junior
Universidade Federal do Espírito Santo**

**Prof. Dr. Paulo Roberto Filgueiras
Universidade Federal do Espírito Santo**

**Dra. Sonia Maria Cabral de Menezes
Centro de Pesquisas e Desenvolvimento Leopoldo Américo Miguez de Mello
(Cenpes – Petrobras/RJ)**

**Dr. Luiz Silvino Chinelatto Júnior
Centro de Pesquisas e Desenvolvimento Leopoldo Américo Miguez de Mello
(Cenpes – Petrobras/RJ)**

**Universidade Federal do Espírito Santo
Vitória, Dezembro de 2016**

AGRADECIMENTOS

Agradeço a Deus pela sabedoria, inteligência, saúde, por todas as conquistas até aqui e por renovar minhas forças a cada amanhecer.

Aos meus pais, José Ricardo de Oliveira e Gislene Aparecida da Silva Oliveira, a minha irmã, Emyli, que sempre estiveram do meu lado em todos os momentos da minha vida.

Ao meu orientador Professor Dr. Álvaro Cunha Neto pela paciência em me conduzir na execução deste trabalho, pelas valiosas críticas e sugestões, pelos ensinamentos, pelo exemplo de profissionalismo, pelas oportunidades de crescimento profissional e pessoal, pela amizade e pelo apoio constante.

Agradeço ao grupo de RMN de alto campo que ofereceram enorme ajuda nos momentos em que precisei, ouviram e discutiram minhas argumentações, em especial Christiane.

Ao professor Paulo R. Filgueiras por todo ensino de quimiometria.

Ao Labpetro/UFES, por todos os recursos que permitiram a elaboração deste trabalho.

À banca de qualificação, os professores Valdemar, Eustáquio e Rosangela, pelas correções, contribuições e sugestões

À banca de defesa que aceitaram fazer parte deste momento tão importante em minha vida: Sonia Cabral, Luiz Silvino, Valdemar Lacerda e Paulo R. Filgueiras.

Meus agradecimentos a: Julia Tristão pela amizade e por toda ajuda em quimiometria.

Aos grandes amigos: Roberta, Samanta e Suzy pelo apoio e grande amizade. À toda família LabPetro, da qual me orgulho de fazer parte há tantos anos.

Aos professores, colegas de trabalho e acima de tudo, amigos do Ifes Campus Venda Nova do Imigrante, por todo apoio, amizade e força que me deram, em especial Daniela, Fabricio, Dante, Leo, Rodrigo, Admildo, Lily, Robinho, Maira, Lucas, Jajá, Samantha, Bárbara, Fabiano S., Fabiano Caliman...

"Mesmo quando tudo parece desabar, cabe a mim
decidir entre rir ou chorar, ir ou ficar, desistir ou
lutar; porque descobri, no caminho incerto da vida,
que o mais importante é o decidir."

Cora Coralina

LISTA DE TABELAS

Tabela 1. Parâmetros moleculares médios de RMN de ^1H , de acordo com Hasan..	27
Tabela 2. Parâmetros moleculares médios de RMN de ^{13}C , de acordo com Hasan.	27
Tabela 3. Parâmetros moleculares médios de RMN de ^1H , de acordo com Poveda	28
Tabela 4. Parâmetros moleculares médios de RMN de ^{13}C , de acordo com Poveda	28
Tabela 5. Histórico da Quimiometria	30
Tabela 6. Métodos ASTM referentes às propriedades.....	54
Tabela 7. Conjuntos de variáveis utilizados na construção dos modelos.	56
Tabela 8. Propriedades físico-químicas selecionadas para os óleos estudados.	61
Tabela 9. Número de amostras utilizado na construção dos modelos para cada propriedade físico-química estudada.	66
Tabela 10. Resultados dos modelos PLS e MLR na determinação das propriedades físico-químicas gravidade API e ponto de fluidez de petróleo.	68
Tabela 11. Resultados dos modelos PLS e MLR na determinação da propriedade físico-química viscosidade cinemática a 50°C de petróleo.....	69
Tabela 12. Resultados dos modelos PLS e MLR na determinação do SARA de petróleo.	71
Tabela 13. Valores calculados para a validação dos modelos selecionados para gravidade API, por PLS e MLR.	73
Tabela 14. Coeficientes de regressão do modelo PLS, com 6 variáveis latentes e pré-processamento MSC, e coeficientes do modelo MLR, na determinação da gravidade API.....	75
Tabela 15. Resultados dos modelos PLS-DA para a determinação da classe das amostras de petróleo baseadas na gravidade API.....	77
Tabela 16. Matrizes de confusão obtidas na construção do modelo (ajuste) e na validação cruzada.	77
Tabela 17. Parâmetros estatísticos obtidos no ajuste, validação cruzada e teste para avaliação do modelo de classificação construído.....	78
Tabela 18. Valores calculados para a validação dos modelos selecionados para ponto de fluidez máximo, por PLS e MLR.	81

Tabela 19. Coeficientes de regressão do modelo PLS, com 6 variáveis latentes e pré-processamento MSC, e coeficientes do modelo MLR, na determinação do ponto de fluidez máximo.	82
Tabela 20. Valores calculados para a validação dos modelos selecionados para viscosidade cinemática a 50°C, por PLS e MLR.	85
Tabela 21. Coeficientes de regressão do modelo PLS, com 6 variáveis latentes e pré-processamento MSC, e coeficientes do modelo MLR, na determinação da viscosidade cinemática a 50°C.....	86
Tabela 22. Valores calculados para a validação dos modelos selecionados para saturados, aromáticos, resinas e asfaltenos, por PLS e MLR.	89
Tabela 23. Coeficientes de regressão do modelo PLS e modelo MLR, na determinação do teor de saturados, aromáticos, resinas e asfaltenos (continua).	90
Tabela 24. Frações típicas do petróleo.....	95
Tabela 25. Conjuntos de variáveis utilizados na construção dos modelos.	99
Tabela 26. Propriedades dos óleos originais utilizados nas destilações.....	100

LISTA DE FIGURAS

Figura 1. Movimento de precessão ao redor de B_0 .	22
Figura 2. Magnetização resultante após interação com o campo oscilante	23
Figura 3. Relaxação Transversal	25
Figura 4. Blindagem diamagnética dos núcleos	26
Figura 5. Número de artigos publicados sobre a utilização da RMN na caracterização de petróleo e derivados, segundo portal da CAPES, no período de 1981 a 2016.	29
Figura 6. Relação da Quimiometria com outras disciplinas (adaptada)	31
Figura 7. Representação da decomposição da matriz de dados X.	34
Figura 8. Espectros de três amostras de petróleo - a) de ^1H e b) de ^{13}C	65
Figura 9. Gráfico utilizado para a seleção das VL's em função do RMSECV para a previsão da densidade por RMN de ^1H e ^{13}C segundo Poveda ¹⁵ com pré-tratamento do tipo MSC.	67
Figura 10. Relação entre valores de referência (o) e previstos (*) da propriedade gravidade API dos modelos criados por (a) PLS e (b) MLR.	73
Figura 11. Relação entre os erros experimentais e calculados para a gravidade API.	74
Figura 12. Distribuição das amostras nas classes.	76
Figura 13. Gráfico de <i>scores</i> do modelo selecionado nas duas primeiras variáveis latentes.	78
Figura 14. Gráfico de <i>loadings</i> do modelo selecionado nas duas primeiras variáveis latentes.	79
Figura 15. Relação entre valores de referência (o) e previstos (*) da propriedade ponto de fluidez máximo dos modelos criados por (a) PLS e (b) MLR.	80
Figura 16. Relação entre os erros experimentais e calculados para o ponto de fluidez máximo.	82
Figura 17. Correlação entre a gravidade API e o ponto de fluidez de um conjunto de amostras de petróleo.	84
Figura 18. Relação entre valores de referência (o) e previstos (*) para $\log [\log (\text{viscosidade cinemática a } 50^\circ\text{C})]$ dos modelos criados por (a) PLS e (b) MLR	85
Figura 19. Relação entre valores de referência (o) e previstos (*) para os teores de saturados dos modelos criados por (a) PLS e (b) MLR.	87

Figura 20. Relação entre valores de referência (o) e previstos (*) para os teores de aromáticos dos modelos criados por (a) PLS e (b) MLR.....	87
Figura 21. Relação entre valores de referência (o) e previstos (*) para os teores de resinas dos modelos criados por (a) PLS e (b) MLR.....	88
Figura 22. Relação entre valores de referência (o) e previstos (*) para os teores de asfaltenos dos modelos criados por (a) PLS e (b) MLR.....	88
Figura 23. Relação entre os erros experimentais e calculados para os teores de saturados, aromáticos, resinas e asfaltenos.	90
Figura 24. Biplot da PCA utilizando parâmetros moleculares médios do espectro de RMN de ^1H de acordo com Hasan (1983) ¹⁶ . Os vetores em azul representam os <i>loadings</i> da PCA.....	101
Figura 25. Biplot da PCA utilizando parâmetros moleculares médios do espectro de RMN de ^{13}C de acordo com Hasan (1983) ¹⁶ . Os vetores em azul representam os <i>loadings</i> da PCA.....	102
Figura 26. Biplot da PCA utilizando parâmetros moleculares médios do espectro de RMN de ^1H e ^{13}C de acordo com Hasan (1983) ¹⁶ . Os vetores em azul representam os <i>loadings</i> da PCA.....	103
Figura 27. Biplot da PCA utilizando parâmetros moleculares médios do espectro de RMN de ^1H de acordo com Poveda (2012) ¹⁵ . Os vetores em azul representam os <i>loadings</i> da PCA.....	104
Figura 28. Biplot da PCA utilizando parâmetros moleculares médios do espectro de RMN de ^{13}C de acordo com Poveda (2012) ¹⁵ . Os vetores em azul representam os <i>loadings</i> da PCA.....	105
Figura 29. Biplot da PCA utilizando parâmetros moleculares médios do espectro de RMN de ^1H e ^{13}C de acordo com Poveda (2012) ¹⁵ . Os vetores em azul representam os <i>loadings</i> da PCA.....	106

LISTA DE ABREVIATURAS E SIGLAS

ANP – Agência Nacional de Petróleo Gás Natural e Biocombustíveis

API – do inglês, *American Petroleum Institute*

ASTM – Sociedade americana de testes e materiais (do inglês, *American Society for Testing and Materials*)

AUTO – Autoescalonamento

CENPES – Centro de Pesquisas e Desenvolvimento Leopoldo Américo Miguez de Mello

FID – Decaimento Livre de Indução (do inglês, *Free Induction Decay*).

FN – amostras falso negativas (do inglês, *false negative*)

FP – amostras falso positivas (do inglês, *false positive*)

FTIR – Infravermelho com transformada de Fourier (do inglês, *Fourier transform infrared*)

LabPetro – Laboratório de Pesquisa e Desenvolvimento de Metodologias para Análises de Petróleo

LD – Limite de detecção

LQ – Limite de quantificação

MLR – Regressão linear múltipla (do inglês, *Multiple Linear Regression*)

MSC – Correção do espalhamento multiplicativo (do inglês, *multiplicative scatter correction*)

NAT – Número de Acidez Total

NIR – Infravermelho próximo (do inglês, *Near infrared*)

NORM – Normalização

PC – Componentes principais

PCA – Análise por componentes principais (do inglês, *Principal Components Analysis*)

Petrobras – Petróleo Brasileiro SA

PEV – Pontos de Ebulição Verdadeiros

PF – Ponto de fluidez

PLS – Método de mínimos quadrados parciais (do inglês, *Partial Least Square*)

PLS-DA – Método de mínimos quadrados parciais com análise de discriminante (do inglês, *Partial Least Squares Discriminant Analysis*)

PNV – prevalência de valor negativo (do inglês, *prevalence of negative value*)

PPV – prevalência de valor positivo (do inglês, *prevalence of positive value*)

RF – ondas de radiofrequência

RMN – Ressonância Magnética Nuclear

RMSE – Raiz quadrada do erro quadrático médio (do inglês, *root mean square error*)

RMSEC – Raiz quadrada do erro quadrático médio de calibração (do inglês, *Root Mean Square Error of Calibration*)

RMSECV – Raiz quadrada do erro quadrático médio de validação cruzada (do inglês, *Root Mean Square Error of Cross Validation*)

RMSEP – Raiz quadrada do erro quadrático médio de previsão (do inglês, *Root Mean Square Error of Prediction*)

SAP – saturados, aromáticos e polares

SARA – saturados, aromáticos, resinas e asfaltenos

SNV – Variação normal padrão (do inglês, *standard normal variate*)

SVD - desvio padrão dos erros (do inglês, *Standard Deviation of the Validation Errors*)

TMS – tetrametilsilano

TN – amostras negativas corretamente identificadas (do inglês, *true negative*)

TP – amostras positivas corretamente identificadas (do inglês, *true positive*)

VL – Variável latente

LISTA DE SÍMBOLOS

- θ - Ângulo do desvio sofrido pela magnetização após a aplicação do pulso de RF
- B_{ef} - campo magnético efetivo
- B_0 - campo magnético estático
- B_i - campo magnético induzido
- R^2 – coeficiente de determinação
- R^2_p – coeficiente de determinação de predição
- b_{PLS} - coeficientes de regressão do modelo PLS
- b_n - coeficiente polinomial
- b_n^i - coeficiente polinomial ajustado para o i -ésimo vetor y e permutado
- b_n^* - coeficiente polinomial ajustado para os dados originais dos resíduos em função dos valores de referência
- W – coeficientes ponderados (pesos)
- M_x - componente da magnetização transversal no eixo x
- M_y - componente da magnetização transversal no eixo y
- k_B - constante de Boltzmann
- C – constante de calibração do viscosímetro
- T_1 - constante de tempo da relaxação Longitudinal
- T_2 - constante de tempo da relaxação transversal
- $\rho(60^\circ/60^\circ F)$ - densidade específica do óleo a $60^\circ F$ em relação à densidade da água na mesma temperatura
- E - energia de um estado de *spin* nuclear
- ε_i – erro aleatório independente
- $\sigma_{previsão}$ – erro total de medição do modelo de previsão
- β - estado de maior energia
- α - estado de menor energia
- t_{tab} e t_{calc} – estatística de teste calculado e tabelado, respectivamente
- $g(y_i)$ – função polinomial
- H_0 e H_1 – hipóteses do teste-t bicaudal
- M_0 - Magnetização resultante na direção do campo magnético externo
- X^* - matriz de dados para novas amostras
- X – matriz ($n \times m$) de n observações e m variáveis
- b_{MLR} – matriz dos coeficientes de regressão do modelo MLR

Y – matriz ou vetor com a propriedade de interesse

$\bar{y}_{previsão}$ – média dos valores de previsão

μ - momento magnético

n_{prev} – número de amostras de previsão (ou teste)

c – número de classes utilizadas no modelo de classificação

N_{β} - Número de núcleos no estado de maior energia β

N_{α} - Número de núcleos no estado de menor energia α

$N_{\alpha,\beta}$ - Número de núcleos nos estados energéticos

B_1 - Pulso de excitação

γ - razão giromagnética

R - ruído ou matriz de resíduos

I - *spin* nuclear

T – temperatura

t – tempo de escoamento em segundos

σ - tensor de blindagem

S – valores de *scores* da matriz X

U – valores de *scores* da matriz Y

y_{ref} - valores de referência de calibração

P – valores dos pesos (do inglês, *loadings*) da matriz X

Q - valores dos pesos (do inglês, *loadings*) da matriz Y

y_{prev} – valores estimados pelo procedimento de validação cruzada para n amostras do conjunto de calibração

$y_{médio}$ – valores médios

ν – viscosidade cinemática

RESUMO

Os produtos petrolíferos em geral são altamente complexos e é exigido um esforço considerável para a caracterização de suas propriedades químicas e físicas. Às vezes tem-se urgência no resultado de determinadas análises e isto fica prejudicado pela forma como as análises são feitas. Assim, a RMN (ressonância magnética nuclear), associada a técnicas quimiométricas vem gerando métodos alternativos para a caracterização e avaliação de propriedades físicas e químicas de petróleos e seus derivados com elevada exatidão, confiabilidade e rapidez. Desta forma, face às suas aplicabilidades, foi proposto neste trabalho a utilização da RMN como método espectroscópico e a partir das áreas de integração dos espectros de RMN de ^1H e ^{13}C associadas às ferramentas quimiométricas de calibração (método dos mínimos quadrados parciais, PLS e regressão linear múltipla, MLR) foram determinadas as seguintes propriedades em amostras de petróleo: Gravidade API, Ponto de Fluidez Máximo ($^{\circ}\text{C}$), Viscosidade Cinemática (mm^2/s a 50°C) e SARA (%m/m), avaliando, assim, a performance dos modelos obtidos. Essa avaliação se deu pela determinação e análise do coeficiente de determinação (R^2), de diversos erros calculados para os conjuntos de calibração e previsão. Os modelos foram, ainda, submetidos a testes estatísticos, e tiveram suas figuras de mérito calculadas. As propriedades químicas com influência direta no perfil espectral, tais como gravidade API, viscosidade cinemática e teores de saturados e aromáticos nos modelos SARA geraram modelos mais robustos e com melhor capacidade preditora, quando comparados aos modelos selecionados para a predição das propriedades restantes, ponto de fluidez máximo, resinas e asfaltenos. Utilizou-se também ferramenta de reconhecimento de padrão (método dos mínimos quadrados parciais com análise de discriminante, PLS-DA), a fim de classificar as amostras de petróleo em petróleos leves, médios ou pesados de acordo com a gravidade API. Neste trabalho foram ainda construídos modelos PCA (Análise de Componentes Principais) para análise exploratória de um conjunto de espectros de amostras de derivados de petróleo a fim de determinar a acidez do óleo original a partir de parâmetros estruturais das frações.

Palavras-chave: RMN. Quimiometria. Petróleo. Frações.

ABSTRACT

Petroleum products are in general highly complex and considerable effort is needed to characterize their chemical and physical properties. Frequently the results of several analyses are urgent and this is compromised by the way the analyses are carried out. Thus, NMR (Nuclear Magnetic Resonance) associated with chemometrics methods have been generating alternative methods to characterize and evaluate physical and chemical properties of petroleum and its derivatives with high precision, reliability and promptitude. Considering its applicability, this work proposes the use of NMR as spectroscopic method and from the integration areas of ^1H and ^{13}C NMR's spectra associated to the chemometric calibration tools (Partial Least Square Method, PLS, and Multiple Linear Regression, MLR) it were determined the following properties in petroleum samples: API gravity, Maximum Pour Point ($^{\circ}\text{C}$), Cinematic Viscosity (mm^2/s at 50°C) and SARA-saturated, aromatic, resins and asphaltenes ($\%m/m$), in order to evaluate the obtained models' performance. This evaluation was obtained through determination and analyses of the coefficient of determination (R^2) of several calculated errors for the calibration and prediction sets. The models were still exposed to statistical tests and their figures of merit calculated. The chemical properties with direct influence on the spectral profile, such as, API gravity, kinematic viscosity and saturated and aromatic contents in SARA models, had generated more robust models and with better predictive capacity, when compared to the models selected for the prediction of the remaining properties, maximum pour point, resins and asphaltenes. A standard recognition tool (partial least squares method with discriminant analysis, PLS-DA) was used to classify petroleum samples into light, medium or heavy oils according to API gravity. In this work, PCA (Principal Component Analysis) models were also constructed aiming the exploratory analysis of a set of spectra of oil derivative samples in order to determine acidity of the original oil considering the structural parameters of the fractions..

Keyword: NMR. Chemometrics. Crude oil. Fractions.

SUMÁRIO

1	Introdução	17
1.1	Petróleo	17
1.1.1	Origem do petróleo	17
1.1.2	Aspectos técnicos	17
1.2	Ressonância Magnética Nuclear	19
1.2.1	Histórico	19
1.2.2	Fundamentos Teóricos	21
1.2.3	Deslocamento Químico	25
1.3	Quimiometria	30
1.3.1	Histórico	30
1.3.2	Aspectos Gerais	30
1.3.3	Análise por Componentes Principais (PCA)	33
1.3.4	Método de Mínimos Quadrados Parciais (PLS)	35
1.3.5	Regressão Linear Múltipla (MLR)	37
1.3.6	Método de Mínimos Quadrados Parciais com Análise de Discriminante (PLS-DA)	38
1.3.7	Pré-processamento dos dados	39
1.3.8	Otimização dos modelos	41
2	Objetivos	47
2.1	Objetivos Gerais	47
3	Aplicação 1: Determinação da gravidade API, viscosidade cinemática, ponto de fluidez e SARA em petróleo por espectroscopia RMN e métodos multivariados de calibração e classificação.	48
3.1	Introdução	48
3.2	Objetivos	53
3.3	Metodologia	53
3.3.1	Obtenção das amostras	53
3.3.2	Determinação dos parâmetros físico-químicos	53
3.3.3	Análises de RMN	54
3.3.4	Tratamento Quimiométrico	55
3.4	Resultados e Discussões	61

3.5	Conclusões	92
4	Aplicação 2: Análise exploratória de frações de petróleo por RMN e PCA na investigação da acidez dos óleos originais	94
4.1	Introdução	94
4.2	Objetivos	96
4.3	Metodologia.....	97
4.3.1	Obtenção das amostras.....	97
4.3.2	Análises de RMN	97
4.3.3	Tratamento Quimiométrico.....	98
4.4	Resultados e Discussões	99
4.5	Conclusões	106
5	Conclusões Gerais	107
6	Referências	108
ANEXO 1	113

1 Introdução

1.1 Petróleo

1.1.1 Origem do petróleo

A origem do petróleo ainda é um assunto muito discutido e estudado. Porém, a teoria mais aceita atualmente é a da origem a partir da matéria orgânica depositada junto com os sedimentos.

Para a formação do petróleo é necessário um ambiente não oxidante, assim, pressupõe-se um ambiente de deposição composto de sedimentos de baixa permeabilidade, inibidor da ação da água circulante em seu interior. Após a incorporação da matéria orgânica ao sedimento, há um aumento da carga sedimentar e de temperatura, iniciando de fato o processo de formação do petróleo. Inicialmente a atividade bacteriana converte a matéria orgânica em querogênio. Em seguida, com o aumento da temperatura, têm-se novos processos de transformação, gerando os hidrocarbonetos líquidos e gasosos que constituem o petróleo. A rocha na qual este processo ocorre é chamada de rocha geradora. Em seguida, ocorre a migração dos fluidos, devido à expulsão da água das rochas geradoras que levam também o petróleo durante os processos de compactação. O petróleo então migra, percorrendo espaços por rochas permeáveis até que seja contido por uma armadilha geológica, onde se acumula. Para que ocorra a acumulação é necessário que haja uma combinação de uma rocha porosa e permeável - geralmente arenitos e calcarenitos - e de uma rocha de baixa permeabilidade e alta plasticidade, que funciona como uma barreira, interrompendo a migração do petróleo. A rocha porosa é dita rocha reservatório, enquanto que a impermeável é dita rocha selante. As rochas selantes geralmente são folhelhos e os evaporitos (sais).¹

1.1.2 Aspectos técnicos

O petróleo, também chamado óleo cru, pode ser definido como uma mistura complexa de ocorrência natural, constituída predominantemente de hidrocarbonetos (podendo chegar a mais de 90% de sua composição) e derivados orgânicos sulfurados (presentes como mercaptanas, sulfetos, tiofenos, etc), nitrogenados (na forma de piridina, pirrol, quinolina, porfirinas, etc), oxigenados (presentes como ácidos

carboxílicos e naftênicos, fenol, cresol) e organometálicos.² Os óleos obtidos de diferentes reservatórios de petróleo possuem características diferentes. Alguns são pretos, densos, viscosos, liberando pouco ou nenhum gás, enquanto que outros são castanhos ou bastante claros, com baixa viscosidade e densidade, liberando quantidade apreciável de gás. Outros reservatórios, ainda, podem produzir somente gás. Entretanto, todos eles possuem quantidades semelhantes dos elementos que os compõem.¹

Cada petróleo tem uma combinação única de moléculas, que definem suas propriedades físicas e químicas e, finalmente, seu comportamento durante o refino.³ Para que o potencial energético deste produto seja aproveitado ao máximo, deve-se submetê-lo a uma série de processos. Dessa forma, o óleo deve ser processado e transformado de maneira conveniente, com o propósito de obter-se a maior quantidade possível de produtos de maior qualidade e valor comercial. Atingir este objetivo, com o menor custo operacional, é a diretriz básica do refino.⁴

Os teores de cada tipo de composto presente no petróleo são responsáveis pela classificação dos diversos tipos de óleos, uma vez que essa divisão é feita de acordo com seus constituintes em relação aos demais constituintes existentes. A classificação do petróleo de acordo com seus constituintes interessa desde os geoquímicos até os refinadores. Os primeiros visam caracterizar o óleo para relacioná-lo à rocha-mãe e medir o seu grau de degradação. Os refinadores querem saber a quantidade das diversas frações que podem ser obtidas, assim como sua composição e propriedades físicas.¹

A caracterização do petróleo envolve além da determinação da composição, a investigação da adequação do óleo para o refino ou para o uso, tais como combustível ou como matéria-prima de outros segmentos industriais. Dessa forma, os dados gerados pelo conjunto de análises permitem especificar a natureza e qualidade do óleo.² A espectroscopia no infravermelho com transformada de Fourier (FTIR) e infravermelho próximo (NIR), em conjunto com técnicas de análise multivariada de dados, têm se mostrado muito eficientes, confiáveis e rápidas na determinação de propriedades de derivados médios, óleos crus, e misturas de hidrocarbonetos. NIR e FTIR têm sido utilizadas no estudo da qualidade dessas amostras com uma boa razoável resolução espectral e alta sensibilidade. Em comparação com a

espectroscopia no infravermelho, a Ressonância Magnética Nuclear (RMN) vem oferecendo uma boa resolução espectral, a qual apresenta ainda a vantagem de permitir a determinação não só qualitativa, mas também quantitativa, de uma larga faixa de tipos de óleos e amostras. Diferentes grupos funcionais aparecem em regiões espectrais bem resolvidos, e este fator tem estabelecido a RMN como uma técnica analítica altamente discriminatória e quantitativamente confiável.⁵ Assim, métodos de análise utilizando a RMN têm despertado grande interesse a fim de otimizar a caracterização de petróleos e derivados.^{3, 6, 7}

1.2 Ressonância Magnética Nuclear

1.2.1 Histórico

A Ressonância Magnética Nuclear (RMN) surge do fato de alguns núcleos apresentarem momento angular de *spin* intrínseco, tendo como consequência um momento de dipolo magnético. A existência de um momento de dipolo magnético para núcleos de hidrogênio foi estabelecida em 1933 por Gerlach e Stern, que observaram o efeito de um gradiente de campo magnético aplicado em um feixe de moléculas de hidrogênio. As primeiras experiências bem-sucedidas de RMN em amostras de fase condensada foram feitas em 1945 e publicadas em 1946, separadamente pelo grupo de Stanford liderada por Bloch, que observou os prótons na molécula de água, e um grupo de Harvard liderada por Purcell, que também observou prótons, mas em parafina sólida.

A blindagem magnética de um núcleo a partir do campo gerado pelos elétrons circundantes foi reconhecida por Lamb, que publicou um método para calcular o efeito em 1941. O objetivo destes cálculos foi para corrigir o efeito da blindagem nas frequências de ressonância observada nas experiências de feixe molecular e, conseqüentemente, para se obter o verdadeiro momento magnético nuclear. Ramsey estendeu esses cálculos para núcleos nas moléculas em 1949-1952, e neste mesmo período, o fenômeno foi observado na RMN de amostras em fase condensada, pela primeira vez nas ressonâncias de metais e sais metálicos por Knight em 1949, e no ano seguinte por Proctor e Yu, que observaram diferentes ressonâncias de ¹⁴N em nitrato de amônio, e por Dickinson, que relataram o mesmo fenômeno para ¹⁹F em vários compostos. Foi também nessa época que os efeitos produzidos pelo

acoplamento *spin-spin* foram observados pela primeira vez, sendo ao mesmo tempo observado por Hahn e Maxwell como uma modulação de um sinal de *spin-eco*, Hahn tendo descoberto o fenômeno de *spin-eco* em 1949.⁸ Porém foi só na década de 50 que a técnica realmente ganhou uma base mais sólida, quando se descobriu que a frequência de ressonância de um núcleo é influenciada pelo ambiente químico ao qual ele está submetido, caracterizando os deslocamentos químicos.

A década de 1950 também foi marcada por uma série de avanços no desenvolvimento instrumental da RMN, dentre eles o uso da rotação da amostra durante a análise a fim de minimizar os efeitos de inhomogeneidade do campo magnético, acarretando num substancial aumento na resolução, revelando finas separações provenientes do acoplamento *spin-spin*.⁹ Em 1952, todas as interações básicas importantes que afetam os espectros de RMN tinham sido demonstradas, e a sua relação com a estrutura molecular foi explicada. Havia ainda muitos novos efeitos a serem descobertos, e estes apareceram rapidamente quando os pioneiros começaram a explorar esta nova espectroscopia. Em 1953 Overhauser previu que deveria ser possível transferir polarização de *spin* de elétrons para núcleos. Os fenômenos de acoplamento *spin-spin* e deslocamento químico foram claramente explicados quando os campos magnéticos gerados pelos ímãs se tornaram suficientemente homogêneos. Houveram muitos desenvolvimentos em curso no período de 1955-1965. O sucesso dos pioneiros incentivou o desenvolvimento de espectrômetros comerciais, e isso proporcionou maior acesso a RMN para a comunidade científica em geral. O primeiro espectrômetro comercial (30 MHz para ^1H) foi comercializado pela *Varian Associates* em 1952, e muitos dos novos desenvolvimentos iniciais resultaram do departamento de pesquisa e desenvolvimento da *Varian*, rotação da amostra e travamento de frequência de campo são apenas dois exemplos que levaram a melhorias dramáticas na qualidade de espectros de alta resolução de líquidos.⁸

A inclusão de computadores durante a década de 1960 possibilitou a aquisição de inúmeros espectros repetidas vezes, que combinados geravam ganhos significativos em sensibilidade por meio da utilização da média do sinal, permitindo o uso de menores quantidades de amostra. Foi introduzido também, o processo conhecido como Transformada de Fourier (FT), como técnica mais rápida e eficiente para a obtenção do sinal adquirido, aumentando substancialmente a razão sinal/ruído, que

Andersen e Ernst realizaram na *Varian*.^{8, 9} No entanto, o primeiro espectrômetro comercial com FT foi comercializado pela Bruker em 1969. A Varian introduziu ímãs supercondutores em RMN, com um espectrômetro de 200 MHz para ^1H , produzido pela primeira vez em 1962, e cuja força do campo logo foi aumentada de modo a dar ressonância de prótons a 220 MHz.⁸ Esse aumento nos campos magnéticos acoplado ao procedimento de FT possibilitou a observação de uma variedade de núcleos com baixa abundância isotópica até então inacessíveis.⁹

Em 1971, Jeener sugeriu a ideia de espectroscopia multidimensional, e em 1973 Lauterbur publicou o seu método para geração de imagens de objetos através da aplicação de gradientes de campo magnético. Estes dois eventos estimularam Ernst e seus colaboradores a realizarem os primeiros experimentos bidimensionais, e começou uma nova era de desenvolvimento rápido na RMN, levando às diversas possibilidades de experimentos de espectroscopia de RMN e de imagem que estão disponíveis hoje.⁸

A evolução dos métodos continuou no decorrer dos anos 90, com o uso de gradientes de campo magnético pulsado, gerando espectros de melhor qualidade em tempos de aquisição menores, o que consolidou a RMN como técnica de elucidação estrutural. Além disso, ocorreu o desenvolvimento de métodos analíticos acoplados a RMN, como a cromatografia líquida.

A última década foi marcada pelo desenvolvimento das sondas criogênicas de alta sensibilidade, as quais aumentam significativamente a razão sinal/ruído. E a disponibilidade de ímãs blindados reduzindo a dispersão do campo magnético no ambiente onde se encontra o equipamento. A espectroscopia de RMN moderna é considerada altamente desenvolvida e tecnologicamente avançada.^{9, 10}

1.2.2 Fundamentos Teóricos

A espectroscopia RMN constitui um exemplo de interação da matéria com a radiação eletromagnética.¹⁰ Os núcleos possuem diversas propriedades físicas destacando-se, massa, carga elétrica, momento magnético (μ) e *spin* nuclear (I). Sendo o momento magnético e *spin* nuclear fundamentais para a ocorrência do fenômeno da RMN, os quais conferem ao núcleo uma propriedade denominada magnetismo, isto é, confere

ao núcleo um comportamento semelhante a pequenos ímãs.¹¹ Um núcleo pode assumir valores de *spin* nuclear de zero, semi-inteiro ou inteiro, de acordo com a quantidade de prótons e nêutrons que apresenta. Somente núcleos com valores de *spin* nuclear diferente de zero irão apresentar propriedades magnéticas, porque o momento de dipolo magnético está relacionado com o *spin* nuclear através da equação:

$$\vec{\mu} = \gamma \cdot \vec{I}, \quad (\text{Eq. 1})$$

Onde γ é a razão giromagnética, uma constante específica para cada isótopo, e informa o quão magnético um núcleo pode ser. A abundância natural dos isótopos também é um fator importante na RMN, pois a baixa sensibilidade - uma característica inerente à técnica - somada à baixa abundância isotópica de certos núcleos limita a análise dos mesmos.¹⁰

Na ausência de campo magnético, os momentos magnéticos dos núcleos estão espalhados aleatoriamente e apresentam uma magnetização resultante nula, pois os momentos magnéticos de cada núcleo se cancelam mutuamente.¹² Porém quando um núcleo é colocado em um grande campo magnético estático (B_0), ele começa a precessar. O eixo de rotação do núcleo durante o processo de precessão não é paralelo ao campo magnético, apresentando um ângulo em relação ao eixo do campo magnético B_0 . Desta forma ele rotaciona ao redor do seu próprio eixo e ao redor de B_0 , como pode ser visto na Figura 1:

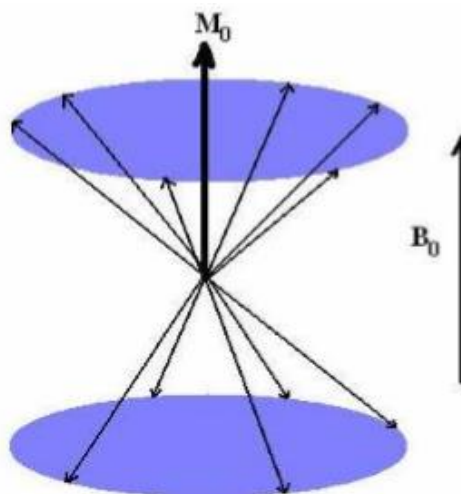


Figura 1. Movimento de precessão ao redor de B_0 .¹³

A figura anterior mostra uma representação visual do movimento de precessão ao redor de B_0 , mais especificamente precessão de Larmor, cuja frequência é chamada de frequência de Larmor, onde os *spins* de alta energia se cancelam com os de baixa energia. Desde que ainda exista mais *spins* de baixa energia, estes serão responsáveis pelo sinal do vetor magnetização resultante (M_0).¹³ Esse excesso populacional existente no estado de menor energia (α), quando comparado ao estado de maior energia (β) é definido pela distribuição de Boltzmann, conforme equação 2:

$$\frac{N_\alpha}{N_\beta} = e^{\Delta E/k_B T} \quad (\text{Eq. 2})$$

Onde $N_{\alpha,\beta}$ representa o número de núcleos em cada uma das orientações dos *spins*, E a energia em cada estado de *spin* nuclear, k_B a constante de Boltzmann e T a temperatura. A diferença entre os níveis de energia é pequena, por isso as diferenças populacionais correspondentes são igualmente pequenas. Por isso, a RMN é considerada uma técnica pouco sensível em relação às outras técnicas espectroscópicas. O excesso populacional, como já foi dito, dá origem a um vetor magnetização resultante.⁹ O conjunto de *spins* quando submetido a um campo oscilante B_1 (ondas de radiofrequência – RF), sofre uma alteração na distribuição de Boltzmann, e a magnetização resultante sofre um desvio de θ da direção z . Esse desvio causa o aparecimento de uma magnetização no plano xy , com componentes M_x e M_y (Figura 2), que induz uma corrente elétrica numa bobina receptora, esse sinal é gerado precisamente quando se verifica absorção da radiação eletromagnética.¹⁰

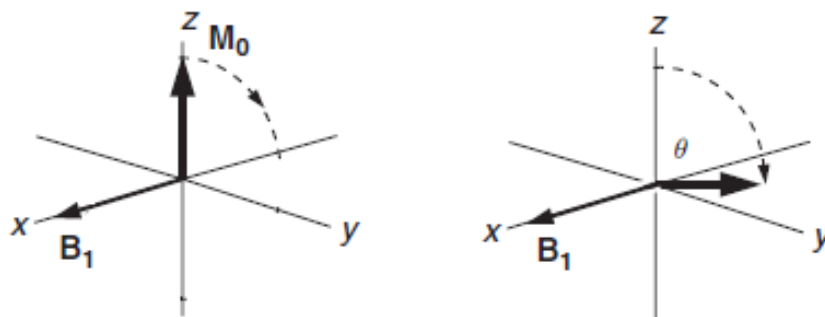


Figura 2. Magnetização resultante após interação com o campo oscilante.⁹

Instantes após o cessar do pulso de excitação (B_1), os momentos magnéticos retornam à posição de equilíbrio, no sentido do eixo z , perdendo assim a energia

transmitida pelo pulso de RF, esse retorno é regido pelo fenômeno de relaxação. O sinal gerado no plano transversal é denominado de FID (do inglês *Free Induction Decay*).⁹

- Relaxação Longitudinal

A recuperação da magnetização ao longo do eixo z, denominada relaxação longitudinal (T_1), corresponde ao retorno das populações de *spins* ao equilíbrio térmico inicial, por meio da perda da energia absorvida durante a excitação. A energia perdida pelos *spins* é transferida para o ambiente, sob a forma de calor, porém esse calor liberado não gera mudanças significativas de temperatura na amostra. Isso explica o fato desse tipo de relaxação também ser chamada de relaxação *spin-rede*. Sobre esse processo de recomposição da magnetização longitudinal (eixo z), tem-se três pontos importantes a salientar: o fenômeno é entálpico, com transferência de energia; a energia é dissipada na rede através de campos magnéticos oscilantes com frequências em torno da frequência de Larmor; e assume que a relaxação ocorra de forma exponencial.⁹

- Relaxação Transversal

Paralelamente à relaxação longitudinal, ocorre a magnetização transversal (T_2), caracterizada pela diminuição da magnetização no plano x-y chamada de relaxação transversal ou relaxação *spin-spin*. Em uma amostra esperar-se-ia que após a aplicação de B_1 , os vetores permanecessem estáticos, no entanto isso só se verificaria se o campo magnético experimentado por cada *spin* da amostra fosse exatamente o mesmo. Porém esse não é o caso, alguns *spins* experimentam um campo local, ligeiramente maior que a média total, fazendo com que tenham uma frequência de precessão maior que outros, isso resulta em uma defasagem dos vetores magnetização individuais, que acaba conduzindo a uma diminuição da magnetização resultante no plano transversal (Figura 3):

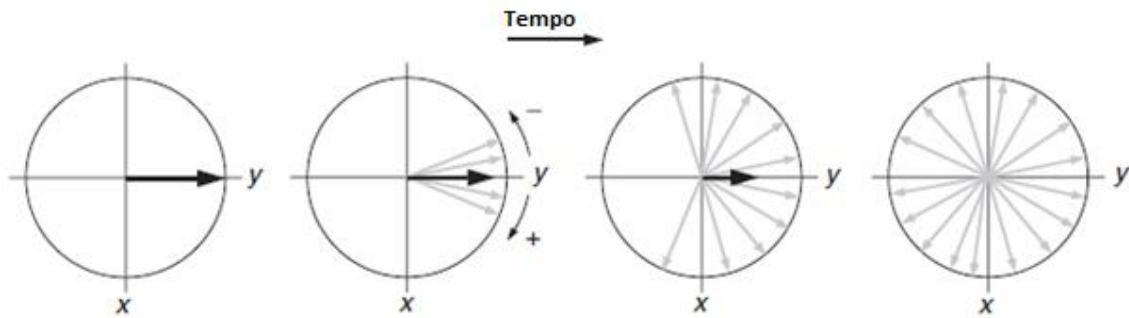


Figura 3. Relaxação Transversal.⁹

Diferentemente da relaxação longitudinal, tem natureza entrópica, e essas diferenças nas frequências dos núcleos da amostra podem ser originadas de duas fontes: devido à inhomogeneidade local proveniente do campo magnético externo ao longo da amostra, e como resultado das interações intermoleculares e intramoleculares na amostra, que representam o processo de relaxação “natural” ou “genuína”.⁹

1.2.3 Deslocamento Químico

Os núcleos não se encontram livres no espaço, mas estão envoltos por elétrons que são partículas carregadas. Sob influência de um campo magnético externo, os elétrons que se movimentam circularmente ao redor do núcleo, geram seus próprios campos magnéticos, os quais chamamos de campo magnético induzido (B_i), o qual é oposto ao campo magnético externo (B_0). Como resultado, enquanto os núcleos desblindados experimentam a magnitude completa do campo externo, os núcleos blindados experimentam o que chamamos de campo magnético efetivo (B_{ef}), caracterizado pela diferença entre o campo magnético externo e o induzido (Figura 4), conforme a equação 3:^{12, 14}

$$B_{ef} = B_0 - B_i, \quad (\text{Eq. 3})$$

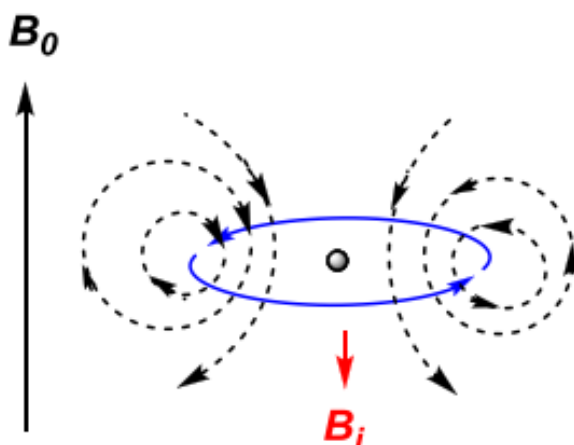


Figura 4. Blindagem diamagnética dos núcleos.¹⁴

Devido ao fato da intensidade do campo magnético induzido ser diretamente proporcional ao campo externo, pode-se definir uma constante de proporcionalidade, denominada tensor de blindagem (σ), que define o ambiente molecular do núcleo em estudo.^{10, 12} Sendo assim, o campo efetivo (B_{ef}) experimentado pelo núcleo equivale a:

$$B_i = \sigma \cdot B_0, \quad (\text{Eq. 4})$$

$$B_{ef} = B_0 - B_i \rightarrow B_{ef} = B_0(1 - \sigma) \quad (\text{Eq. 5})$$

Além do efeito de blindagem tornar a frequência de ressonância do núcleo menor do que na ausência de elétrons, esta frequência depende do ambiente químico local e, por isso, um espectro de RMN é útil na determinação estrutural de uma molécula.

A RMN tem se mostrado adequada na identificação e quantificação de diferentes tipos de hidrocarbonetos e pode fornecer informações sobre o conteúdo relativo de componentes alifáticos, olefínicos e aromáticos. Vários trabalhos, como citado anteriormente, têm reportado a RMN como técnica aplicável em estudos de constituintes de petróleo.⁵

Em geral, a literatura propõe parâmetros para classificar todos os grupos possíveis ou fragmentos estruturais presentes em compostos de petróleo e derivados, denominados parâmetros moleculares médios, os quais serão representados da seguinte maneira: X_{a-b} . Onde X representa o átomo caracterizado: C, carbono ou H, hidrogênio. Os subíndices *a* e *b* designam o grupo estrutural e o subgrupo,

respectivamente. O subíndice *a* é *Par* para grupos parafínicos, *Naf* para grupos naftênicos ou *Ar* para grupos aromáticos. Adicionalmente, o subíndice *b* pode ser α , β , CH, CH₂, para indicar a protonação do carbono ou grupos vizinhos ligados ao núcleo em estudo.¹⁵ Neste trabalho dois autores foram utilizados como base para a obtenção dos parâmetros moleculares médios das amostras em estudo, sendo eles Hasan (1983) e Poveda (2012), conforme parâmetros listados abaixo nas tabelas 1, 2, 3 e 4.^{15, 16}

Tabela 1. Parâmetros moleculares médios de RMN de ¹H, de acordo com Hasan.¹⁶

Região espectral (ppm) e representação	Parâmetro Molecular
6,0 – 9,0 (H _{Ar})	H aromático
0,5 – 4,0 (H _{Par, Naf})	H parafínico e naftênico
0,5 – 1,0 (H _{Par - γ})	H parafínicos em posição γ ou mais em relação ao sistema aromático
1,0 – 2,0 (H _{Par, Naf - β})	H parafínicos e naftênicos do tipo CH ₂ ligados a sistemas aromáticos na posição β e alquílicos terminais
2,0 – 4,0 (H _{Par, Naf - α})	H parafínico e naftênico ligados a carbonos saturados na posição α em relação ao sistema aromático

Tabela 2. Parâmetros moleculares médios de RMN de ¹³C, de acordo com Hasan.¹⁶

Região espectral (ppm) e representação	Parâmetro Molecular
0 – 70 (C _{Par, Naf})	C parafínico e naftênico
110 – 160 (C _{Ar})	C aromático
110 – 130 (C _{Ar - H} e C _{Ar - ar,ar,ar})	C aromático protonados e C em junção de três anéis aromáticos
128 – 136 (C _{Ar - B})	C aromático cabeça de ponte exceto em junção de três anéis
129 – 137 (C _{Ar - CH₃})	C aromático ligados a grupos metil
137 – 160 (C _{Ar - Alq})	C aromaticos ligados a grupos alquílicos exceto metilas

Tabela 3. Parâmetros moleculares médios de RMN de ^1H , de acordo com Poveda.¹⁵

Região espectral (ppm) e representação	Parâmetro Molecular
0,1 – 1,0 ($\text{H}_{\text{Par}} - \text{CH}_3, \gamma$)	H parafínico do tipo CH_3 , em posição γ ou mais em relação ao sistema aromático
1,0 – 1,5 ($\text{H}_{\text{Par}} - \text{CH}_3, \beta$ e $\text{H}_{\text{Par}} - \text{CH}_3$)	H parafínico ligados a sistemas aromáticos na posição β e alquílicos terminais
1,5 – 2,0 ($\text{H}_{\text{Naf}} - \text{CH}_2, \beta$)	H naftênico do tipo CH_2 ligados a sistemas aromáticos na posição β
2,0 – 4,5 ($\text{H}_{\text{Par}}, \text{Naf} - \text{CH}, \text{CH}_2, \text{CH}_3, \beta$)	H parafínico e naftênico do tipo CH, CH_2 e CH_3 , ligados a sistemas aromáticos na posição α
6,0 – 7,2 ($\text{H}_{\text{Ar}} - \text{ma}$)	H aromático ligado a anéis monoaromáticos
7,2 – 9,0 ($\text{H}_{\text{Ar}} - \text{pa}$)	H aromático ligado a C aromáticos de anéis di e poliaromáticos
9,0 – 12,0 ($\text{H}_{\text{Ald}}, \text{Ac}$)	H de aldeído e ac. Carboxílico

Tabela 4. Parâmetros moleculares médios de RMN de ^{13}C , de acordo com Poveda.¹⁵

Região espectral (ppm) e representação	Parâmetro Molecular
3,0 – 18,5 (C_{CH_3})	C do tipo CH_3
18,5 – 21,5 ($\text{C}_{\text{CH}_3, \alpha}$)	C metílico em posição α em relação ao anel aromático
21,5 – 50,0 ($\text{C}_{\text{Par}}, \text{Naf} - \text{CH}, \text{CH}_2$)	C parafínico e naftênicos do tipo CH e CH_2
50,0 – 60,0 ($\text{C}_{\text{Par}} - \text{CH}, \text{C}$)	C parafínico terciários (CH) e quaternários (C)
60,0 – 70,0 ($\text{C}_{\text{Par}} - \alpha, \text{OH}$)	C parafínico e naftênico em posição α ao OH
85,0 – 129,2 ($\text{C}_{\text{Ole}} - \text{CH}, \text{C}_{\text{Ar}} - \text{Naf}, \text{Alq}$ e $\text{C}_{\text{Ar}} - \text{ar}, \text{ar}, \text{ar}$)	C aromático olefínico protonado, C em sistemas benzenaftênicos e alquílicos e C em junção de três anéis aromáticos
129,2 – 137,0 ($\text{C}_{\text{Ar}} - \text{CH}_3$ e $\text{C}_{\text{Ar}} - \text{ar}, \text{ar}$)	C aromático ligado a grupos metil e C cabeça de ponte entre anéis aromáticos
137,0 – 160,0 ($\text{C}_{\text{Ar}} - \text{C}$)	C aromático não protonado
190,0 – 220,0 ($\text{C}_{\text{Ald}}, \text{Cet}$)	C de cetonas e aldeídos

De modo a sintetizar mais quantitativamente o que há de publicações sobre o uso da RMN na caracterização de petróleo e derivados, foram analisados artigos científicos que publicados no período entre 1981 e 2016 (outubro/16) na plataforma CAPES. Embora existam certamente mais artigos disponíveis sobre o assunto, os 105 artigos analisados demonstram a crescente tendência no uso dessa técnica espectroscópica na caracterização de petróleo. Existe um crescente interesse no tema, com grande aumento no número de artigos publicados ao longo dos anos no período analisado, conforme Figura 5.

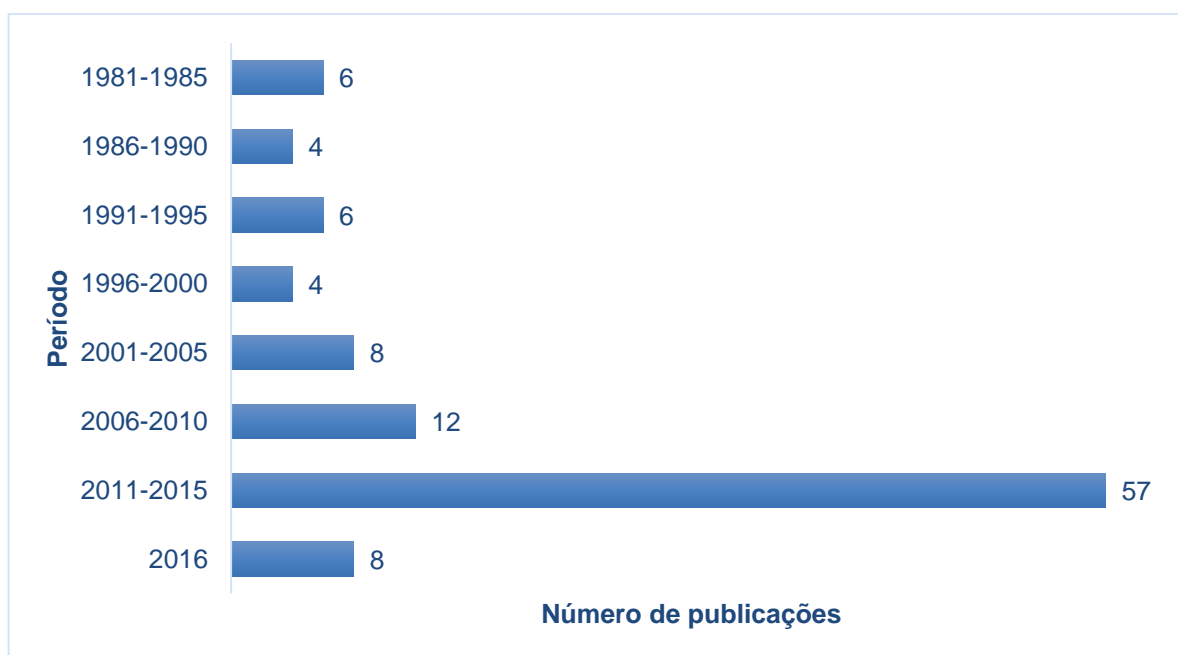


Figura 5. Número de artigos publicados sobre a utilização da RMN na caracterização de petróleo e derivados, segundo portal da CAPES, no período de 1981 a 2016.

Os espectros de RMN do petróleo bruto e seus derivados são muito complexos e contêm um grande número de sinais refletindo sua grande complexidade química. Portanto, para explorar o conteúdo de todas as informações dos dados de RMN, vários métodos de análise multivariada têm sido desenvolvidos e trabalhados em conjunto com essa técnica.³

1.3 Quimiometria

1.3.1 Histórico

O surgimento da quimiometria se deu com a evolução dos equipamentos de análise e com a maior facilidade de aquisição dos computadores, o que possibilitou o acesso de cientistas a uma grande quantidade de dados, tornando necessário o desenvolvimento de um método que possibilitasse a análise desses dados.¹⁷

Os principais fatos que marcaram o surgimento da quimiometria podem ser observados na Tabela 5:^{18, 19}

Tabela 5. Histórico da Quimiometria.^{18, 19}

1971	Introdução do termo quimiometria por Svant Wold
1974	Fundação da Sociedade Internacional de Quimiometria (<i>International Chemometrics Society</i>), por Wold e Kowalski
1975	Primeiro artigo com o termo Quimiometria no título, em que Kowalski apresenta a nova sociedade e a nova disciplina à comunidade científica como uma área de pesquisa
1977	Introdução da seção "Informática e Técnicas de Otimização" publicada na revista <i>Analytica Chimica Acta</i> , dedicada a Quimiometria
1987	Criação de dois periódicos dedicados à Quimiometria: <i>Chemometrics and Intelligent Laboratory Systems</i> e <i>Journal of Chemometrics</i> (fundados por Massart e Kowalsky, respectivamente)

No Brasil, a quimiometria foi introduzida no final da década de 1970, quando Kowalsky foi convidado por Roy E. Bruns para dar um curso de quimiometria em Campinas, iniciando-se o desenvolvimento da disciplina no país. O Brasil foi sede da “10th International Conference on Chemometrics in Analytical Chemistry”, em setembro do ano de 2006, evento que reuniu os maiores pesquisadores da área e permitiu que os pesquisadores brasileiros se inserissem no contexto mundial.²⁰

1.3.2 Aspectos Gerais

A quimiometria é uma área especificamente destinada à análise de dados químicos de natureza multivariada. Ao contrário do que muitos pensam não é uma disciplina da

matemática, mas sim da química, isto é, os problemas que se propõe a resolver são de interesse e originados na química, ainda que as ferramentas de trabalho provenham principalmente da matemática, estatística e computação. A quimiometria emprega métodos matemáticos e estatísticos para planejar ou selecionar experimentos de forma otimizada, além de fornecer dados quantitativos de uma propriedade de interesse e assim fornecer o máximo de informação química com a análise dos dados obtidos.²¹ A quimiometria aplica-se a inúmeros campos da química (Figura 6).²²

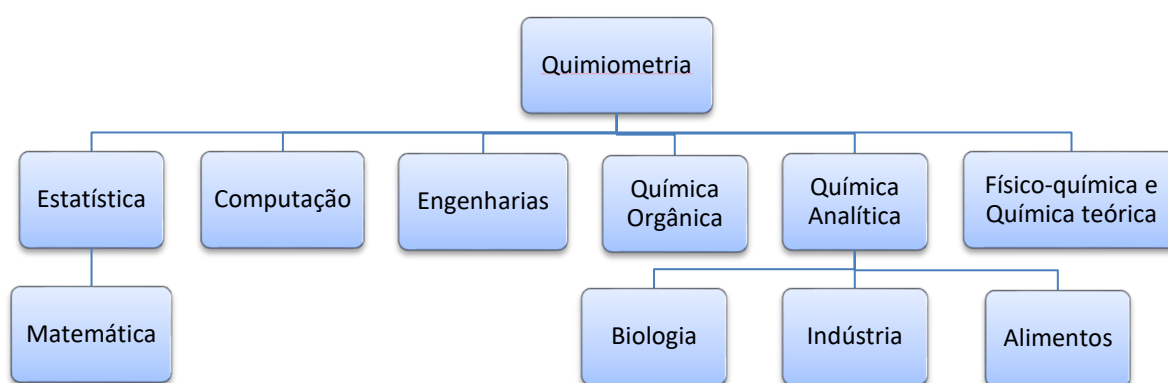


Figura 6. Relação da Quimiometria com outras disciplinas (adaptada).²²

A química analítica quantitativa teve grande impacto com o desenvolvimento das técnicas de calibração multivariada. Nesta, é possível estimar uma propriedade de interesse a partir de outras medições, que normalmente são espectros obtidos por procedimentos analíticos simples, rápidos, pouco onerosos e que dependem de pequena quantidade de amostra.²¹

A quimiometria é, portanto, muitas vezes utilizada como uma ferramenta de automação laboratorial e está relacionada com a análise multivariada. 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.²³ Ao invés de desenvolver métodos analíticos seletivos, como ocorre na análise univariada, passou-se a investir recursos no desenvolvimento de técnicas matemáticas capazes de transformar sinais oriundos de misturas de componentes, em informação útil sobre uma ou várias propriedades de interesse.

Outra característica dos métodos multivariados na determinação de uma propriedade de interesse é a chamada vantagem multicanal, quando apenas um canal ou variável é utilizado, a calibração torna-se extremamente sensível a ruídos nessa variável. A utilização de diversos canais minimiza ou mesmo elimina esta interferência, tornando a calibração mais robusta.²¹

Os métodos quimiométricos podem ser divididos em três grandes áreas: planejamento e otimização de experimentos, métodos qualitativos e de reconhecimento de padrões e calibração multivariada.²⁴ A aplicação de um ou outro método, ou até mesmo a combinação deles depende da natureza do problema que se deseja resolver, ou do tipo de informação que se deseja obter.²³

Ao se deparar com um conjunto de dados é necessária uma análise prévia inicial para se avaliar a qualidade das informações disponíveis. Os métodos de análise exploratória têm por objetivo fazer a avaliação inicial dos dados para descobrir que tipo de informação pode-se extrair deles, e assim definir as diretrizes para um tratamento mais aprofundado. Isto é feito através da utilização de algoritmos que permitem reduzir a dimensão dos dados, ou organizá-los numa estrutura que facilite a visualização de todos o conjunto, de forma global. 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, pode-se citar a análise por componentes principais (PCA – *Principal Components Analysis*).²⁵

Por outro lado, os métodos de calibração multivariada, relacionam-se ao desenvolvimento de modelos matemáticos que permitem estimar alguma propriedade de interesse com os sinais instrumentais. Entre os métodos de calibração, pode-se citar o método de mínimos quadrados parciais (PLS – *Partial Least Square*) e regressão linear múltipla (MLR – *Multiple Linear Regression*). Este pode ainda ser combinado a técnica de reconhecimento de padrão, isto é, um método de regressão combinado com uma técnica de classificação, sendo o caso do método dos mínimos quadrados parciais com análise de discriminante (PLS-DA – *Partial Least Squares Discriminant Analysis*).²²

Tendo em vista que os métodos de análise por componentes principais (PCA), mínimos quadrados parciais (PLS) regressão linear múltipla (RLM) e mínimos

quadrados parciais com análise de discriminante (PLS-DA) foram utilizados no presente trabalho, serão abordados de forma preferencial.

1.3.3 Análise por Componentes Principais (PCA)

A PCA encontra-se certamente entre as mais importantes ferramentas da análise multivariada, inclusive por constituir a base onde se fundamentam a maioria dos outros métodos multivariados de análise de dados. Como uma ferramenta de análise exploratória a PCA permite revelar a existência ou não de amostras anômalas, de relações entre as variáveis medidas, visualizar a estrutura dos dados, encontrar similaridades entre amostras e reduzir a dimensionalidade do conjunto de dados.²⁶

A análise por componentes principais é uma técnica de reconhecimento de padrão não-supervisionado, pois se trata de um método de agrupamento de amostras a partir das similaridades que estas possuem.²² A PCA consiste numa manipulação da matriz de dados, com objetivo de representar as variações presentes em muitas variáveis, através de um número menor de “fatores”.²¹

Em termos matemáticos, a PCA corresponde à decomposição de uma matriz de dados X , de dimensão $n \times m$, no produto de duas matrizes: a matriz dos scores T e a transposta da matriz dos pesos P^T , como mostra a equação 6, que pode ser representada pelo esquema da Figura 7:^{22, 25}

$$X = S \cdot P^T, \quad (\text{Eq. 6})$$

Onde: X – matriz ($n \times m$) de n observações e m variáveis;

S – valores de *scores*, que representa as coordenadas das observações no novo sistema de coordenadas;

P – valores dos pesos (do inglês, *loadings*), que representa a contribuição de cada variável antiga, determinados pelo cosseno do ângulo entre o eixo da componente principal e o eixo da variável original.

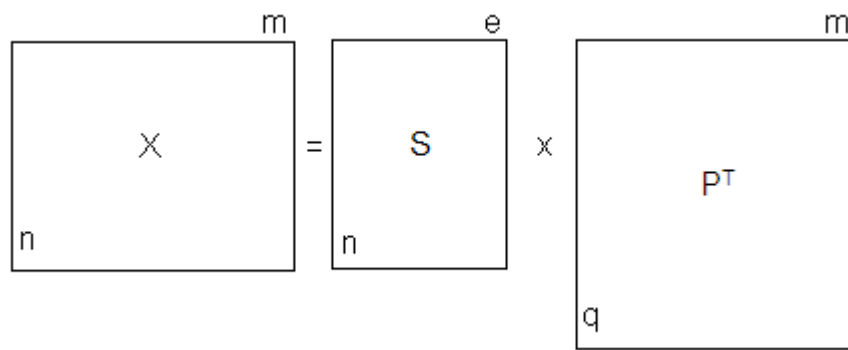


Figura 7. Representação da decomposição da matriz de dados X .

A PCA pode ser resumida como sendo uma transformação de um espaço dimensional com m variáveis (medidas), para um espaço de i novas variáveis, podendo i ser igual ou menor que m , de forma que o novo conjunto, forme um conjunto de vetores ortogonais. Como resultado desta transformação as novas variáveis obtidas, chamadas de componentes principais (PCs) ou autovetores, são combinações lineares das variáveis medidas originalmente e são ortogonais entre si, ou seja, são completamente não correlacionadas. As PCs estão ordenadas em sequência que vai daquela com maior explicação de variação dos dados (primeira componente principal) para aquela com menor explicação da variação dos dados (última componente principal). Isto cria a possibilidade de decomposição da matriz de dados em estrutura e ruído (Equação 7), ou em outras palavras, em variáveis significantes e não significantes, para a explicação da variação dos dados.²⁵

$$X = S \cdot P^T + R, \quad (\text{Eq. 7})$$

Onde: R – ruído ou matriz de resíduos, que é toda informação que permanece sem explicação

A projeção de cada amostra, de um universo de n amostras, neste novo sistema de eixos fornece os *scores*, e cada amostra terá então um valor de *score* para cada um dos novos eixos (as PCs). O gráfico de *scores* poderá revelar agrupamentos ou tendências das amostras analisadas, que poderiam ser de difícil visualização no caso das variáveis originais.

Como dito anteriormente, os pesos são capazes de mostrar quais variáveis originais têm maior importância na combinação linear de cada componente principal. Através

disto pode-se saber, por exemplo, que variáveis contribuem para a descrição deste ou daquele conjunto de amostras.^{22, 25}

A análise de componentes principais também é largamente utilizada como método de análise exploratória de dados, como já foi dito. Neste trabalho uma análise exploratória utilizando gráficos de PCA será apresentada.

1.3.4 Método de Mínimos Quadrados Parciais (PLS)

A calibração é uma das principais aplicações em quimiometria, e normalmente envolve o uso de um tipo de medida para prever o valor de um parâmetro de base ou propriedade.²² A vantagem mais significativa de uma calibração multivariada frente a uma calibração univariada é a redução de ruídos obtidos usando mais medidas para o mesmo fenômeno. Um modelo de calibração univariado pode fornecer resultados precisos, se não houver contribuição de sinal de outra fonte, pois se houver, os resultados estarão viciados, porém não haverá como notar que resultados incorretos estão sendo obtidos. Esse problema da não percepção não acontece com o tratamento multivariado e, pelo contrário, o uso de dados multivariados torna possível lidar com situações não seletivas. A seletividade química é suplementada pela seletividade do tratamento matemático.²⁷

Atualmente, a regressão PLS é o método de calibração multivariada mais utilizado em química analítica. Este método pode ser utilizado para a construção de modelos que permitem a previsão de propriedades físico-químicas em petróleo e derivados de uma forma mais econômica, mais rápida e com menor impacto ambiental, e os modelos construídos podem ser facilmente adaptados às rotinas de análise em laboratórios. O PLS utiliza as respostas analíticas, bem como as informações de interesse, para capturar a variância dos dados da matriz X e do vetor com a propriedade de interesse Y, através de suas decomposições sucessivas e simultâneas, correlacionando-as.²²

A regressão por PLS é uma extensão da PCA, uma vez que a construção do modelo PLS é feita de maneira semelhante a PCA. O que difere um método do outro é o fato da PLS além de projetar os dados independentes X num novo sistema de eixos (como

ocorre na PCA), a matriz dependente Y também é decomposta da mesma maneira, simultaneamente.^{22, 25}

O modelo PLS, portanto, ajusta uma função linear dos scores da matriz X aos scores da matriz Y de modo a minimizar os desvios. Porém, essa otimização simultânea ocasiona pequenas distorções dos pesos de maneira que, eles perdem sua ortogonalidade, e estas distorções entre os componentes principais na PLS fazem com que não sejam mais componentes principais (que são ortogonais) e sim variáveis latentes (VLs).

O número de variáveis latentes utilizadas na construção de um modelo é de extrema importância na obtenção dos resultados. Um modelo contendo um número de VLs menor que o ideal resultará em subajuste, pois não levará em consideração toda informação contida no conjunto de dados. Em contrapartida, quanto maior o número de VLs, maior será o erro, resultando em um sobreajuste. Assim, pode-se dizer que o número ótimo de VLs é aquele no qual a diminuição do erro é compensada pelo aumento de erro de superavaliação.²⁵

Portanto, a construção de um modelo PLS consiste de uma regressão entre os scores das matrizes X e Y em uma soma de “h” VL’s. O modelo PLS pode ser definido através de relações externas, que correlacionam, individualmente, as matrizes X e Y (conforme as equações 8 e 9), enquanto as internas correlacionam ambas as matrizes

$$X = S \cdot P^T + R_X = \sum s_h p_h^T + R_X, \quad (\text{Eq. 8})$$

$$Y = U \cdot Q^T + R_Y = \sum u_h q_h^T + R_Y, \quad (\text{Eq. 9})$$

onde X é a matriz de dados (medida instrumental), Y é a matriz de resposta (concentração, por exemplo), S e U são os scores de X e Y respectivamente; os elementos P e Q são os “pesos” e R_X e R_Y são os resíduos. A matriz de scores S é estimada pela combinação linear de X com coeficientes ponderados por W (pesos):

$$S = X \cdot W, \quad (\text{Eq. 10})$$

A partir de W, os coeficientes de regressão do modelo PLS podem ser estimados por:

$$b_{PLS} = W(P \cdot W)^{-1}q, \quad (\text{Eq. 11})$$

Finalmente, a partir do cálculo de b , o modelo linear PLS pode ser representado por:

$$Y = X \cdot b_{PLS}, \quad (\text{Eq. 12})$$

onde valores da propriedade de interesse para amostras futuras podem ser estimados pelo modelo utilizando a combinação linear do vetor x amostral pelos coeficientes do modelo.^{28, 25}

1.3.5 Regressão Linear Múltipla (MLR)

A regressão linear múltipla é o método de calibração multivariada mais simples. Ele pode ser usado quando não se conhece todas as espécies que geram resultados para a propriedade de interesse, desde que se tenha um número pequeno e seletivo de variáveis. O principal inconveniente em se usar a MLR está relacionado à colinearidade dos dados, visto que para calcular o vetor de coeficientes de regressão b é necessário que a matriz $X^T \cdot X$ seja inversível, ou seja, que forneça uma solução única. Para isso, o número de amostras deve ser superior ao número de variáveis. Considerando que na maioria dos casos são usados equipamentos que permitem medir um elevado número de variáveis (como o espectrômetro de infravermelho), é necessário selecionar, antes de aplicar a MLR, um número de variáveis que seja menor que o número de amostras, e que produzam uma informação “única”.²¹

O modelo MLR pode ser obtido a partir de uma matriz X de respostas instrumentais com dimensão $(m \times n)$, para m amostras e n variáveis. Além disso, se utiliza uma matriz/vetor de resposta Y com dimensão $(m \times k)$ para k fatores de resposta das m amostras. A relação matemática entre as duas matrizes é estabelecida pela equação 13:

$$Y_{(m \times k)} = X_{(m \times n)} \cdot b_{MLR (n \times k)}, \quad (\text{Eq. 13})$$

Onde b_{MLR} é a matriz que contém os coeficientes de regressão, calculados por mínimos quadrados a partir da pseudo-inversa de X , expressa a relação entre as variáveis dependentes e independentes.

$$b_{MLR (n \times k)} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y, \quad (\text{Eq. 14})$$

O sobrescrito -1 indica inverso da matriz e T a transposta da matriz.

A partir dos coeficientes determinados, os novos valores de y_{prev} podem ser estimados a partir da equação:

$$y_{prev} = X^* \cdot b_{MLR}, \quad (\text{Eq. 15})$$

Onde X^* representa a matriz de dados para novas amostras.

Como dito anteriormente, o método de MLR apresenta alguns problemas que limitam sua aplicação. Um deles é que o número de amostras deve ser igual ou superior que o número de variáveis. Uma vez que o modelo consiste em resolver um sistema de equações lineares simultâneas e caso se tenha menos equações que variáveis o sistema torna-se possível indeterminado. Em softwares estatísticos, normalmente o programa utiliza uma resposta possível para o sistema de equações e continua a rotina de cálculos da regressão, mas isso não impede que sejam utilizadas outras possíveis respostas.^{22, 39}

1.3.6 Método de Mínimos Quadrados Parciais com Análise de Discriminante (PLS-DA)

É uma técnica de reconhecimento de padrão supervisionada, no qual deve-se ter um conhecimento prévio das classes das amostras do conjunto de treinamento, e este é usado para prever a classificação de amostras desconhecidas.^{22, 29, 30}

O método de mínimos quadrados parciais com análise de discriminante (PLS-DA) é uma técnica de classificação linear que combina a regressão por mínimos quadrados parciais com o poder de discriminação de uma técnica de classificação.³⁰

O modelo PLS-DA é construído a partir dos algoritmos da regressão PLS. A PLS é um método multivariado de calibração inversa que busca uma relação direta entre a resposta instrumental e a propriedade de interesse (qualitativa ou quantitativa), como já foi mostrado. Como a PLS-DA é um método de classificação, a matriz ou vetor Y (propriedade de interesse) é codificada em 0 e 1, quando existem apenas duas

classes ($c = 2$). Para mais de duas classes, pode-se construir vários modelos com codificação 0 e 1, ou usar o algoritmo PLS2 através da construção de uma matriz ($n \times c$), onde cada coluna representa uma classe.

Uma etapa fundamental na construção do modelo PLS-DA é a determinação do número correto de variáveis latentes, uma vez que estas modelam as fontes de variabilidade dos dados relevantes. Essa escolha é comumente realizada utilizando a validação cruzada das amostras de calibração, onde algumas amostras são separadas em um conjunto de predição e os modelos são construídos com as demais, o erro obtido por esses modelos é calculado, e escolhe-se o número de VL com menor erro.

Na PLS-DA o vetor de cada classe é composto por linhas que são as amostras e colunas que são as informações de cada classe. A atribuição de cada classe obedece a probabilidade da amostra pertencer a uma ou outra classe. O cálculo por amostra é feito para todas as classes, aquele que apresentar maior probabilidade numérica será a classificação da amostra.^{29, 30}

1.3.7 Pré-processamento dos dados

Anteriormente à aplicação das ferramentas quimiométricas ao conjunto de dados a ser investigado, podem ser necessárias transformações dos dados originais, pois estes podem não ter uma distribuição adequada para a análise, dificultando a extração de informações úteis e interpretação dos mesmos. Basicamente, essas transformações consistem de pré-processamentos. O pré-processamento dos dados tem por objetivo remover variações sistemáticas não desejadas dos espectros, como mudanças na linha de base, efeitos de espalhamento e fatores externos, não controláveis, aumentando assim a qualidade das medições instrumentais.

Deve-se tomar cuidado com esse tipo de processamento, uma vez que os métodos conhecidos de pré-processamento têm vantagens e desvantagens. Primeiramente, muitos desses métodos, apesar de melhorar a qualidade de certos espectros, podem também degradar os dados em outros aspectos, por exemplo, o método de suavização é comumente utilizado para remover o ruído presente e melhorar a razão

sinal/ruído, mas quando usado de forma inadequada, pode distorcer o sinal original e comprometer a resolução do espectro.³¹

A seguir são apresentados exemplos de alguns métodos que podem ser utilizados:

- Correção do espalhamento multiplicativo (MSC, do inglês *multiplicative scatter correction*): tem por finalidade remover a variabilidade que pode ser causada por efeitos de espalhamento multiplicativo. Através desse pré-processamento são eliminados desvios de linha de base não lineares (*drift*) que ocorrem em medidas de reflectância em sólidos ou suspensões, devido ao espalhamento multiplicativo da luz. Este último se deve à falta de homogeneidade no tamanho das partículas.
- Variação normal padrão (SNV, do inglês *standard normal variate*): trata-se de outro método para correção do espalhamento multiplicativo da luz. Em geral as equações para a aplicação de SNV e MSC têm a mesma forma e, em muitos casos as duas abordagens produzem resultados semelhantes, de modo que eles são geralmente considerados como permutáveis. A diferença no cálculo ocorre basicamente no fato de que a SNV utiliza a média e o desvio padrão do espectro individual para sua execução, enquanto que o MSC depende de uma relação entre o espectro individual e um espectro de referência (coeficientes linear e angular) para a sua aplicação.³²
- Autoescalamento (AUTO): aplicado quando se deseja comparar variáveis com diferentes dimensões. Consiste, portanto, em se utilizar os dados adquiridos em uma mesma faixa de amplitude de sinal. O processo de autoescalamento consiste em centrar os dados na média e dividi-los pelo respectivo desvio padrão, sendo um para cada variável.
- Normalização (NORM): aplicado quando se deseja corrigir as variáveis. Consiste em dividir cada elemento da linha pela raiz quadrada da soma dos quadrados dessa linha.³³

Quando da utilização da técnica de regressão PLS, como recurso de pré-processamento todos os dados devem, ainda, ser centrados na média com o intuito de permitir que o sistema de eixo das variáveis latentes passe pela média, que conseqüentemente será a origem dos eixos.²³

1.3.8 Otimização dos modelos

A utilização dos modelos de calibração requer a otimização de alguns parâmetros. Quanto maior o número de parâmetros otimizados, maior o custo computacional e a possibilidade de superajuste aos dados. Para evitar este último problema, dois métodos de validação costumam ser aplicados.²²

- Validação externa: o modelo de calibração é testado por um grupo de amostras independentes, denominado conjunto de validação. Esse conjunto contém amostras conhecidas que não foram utilizadas na construção do modelo de calibração.^{20, 22}
- Validação interna: quando as próprias amostras do conjunto de calibração são utilizadas para validação do modelo; este procedimento é conhecido como validação cruzada ou *cross-validation*. Os métodos mais comuns de validação interna são:
 - *leave-one-out*: neste procedimento uma amostra é removida do conjunto de calibração para validação enquanto constrói-se o modelo com as $n-1$ amostras restantes. Neste caso, n modelos são construídos até que todas as amostras de calibração sejam utilizadas para validação;
 - *k-fold*: este procedimento consiste em dividir as n amostras de calibração em k subconjuntos mutuamente exclusivos de mesmo tamanho. Um subconjunto é utilizado para validação enquanto os $k-1$ restantes são utilizados para construção do modelo. O procedimento é repetido k vezes até que todos os subgrupos tenham sido utilizados como validação. Se os subconjuntos são escolhidos em blocos, o método é denominado *contiguos block*; caso sejam escolhidos aleatoriamente, *random block*; ou sejam retiradas de forma ordenada *venetian blinds*.

No procedimento de validação cruzada as amostras do conjunto de calibração são previstas por diferentes modelos construídos com as próprias amostras de calibração, mas nunca uma mesma amostra participando da calibração e validação simultaneamente.^{22, 23, 34}

1.3.8.1 Validação dos modelos de calibração

Para avaliação dos modelos de calibração, PLS e MLR, deve-se calcular os erros e estimar suas figuras de mérito.

- Cálculo dos erros

Para avaliação dos modelos PLS e MLR, primeiramente é feito o cálculo dos erros. A partir dos valores previstos das amostras de calibração durante a validação cruzada, a raiz quadrada do erro quadrático médio de validação cruzada (RMSECV) é calculada.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^N (y_{prev,i} - y_{ref,i})^2}{n}} \quad (\text{Eq. 16})$$

onde y_{ref} e y_{prev} são os valores de referência de calibração e estimado pelo procedimento de validação cruzada para n amostras do conjunto de calibração. No modelo PLS, o número ótimo de variáveis latentes é definido pelo gráfico do RMSECV em função do número de variáveis latentes. O número ótimo é definido pelo valor mínimo ou menor número de variáveis latentes no qual não se tem mais mudança significativa no valor de RMSECV.

Após otimização do modelo, este está pronto para ser aplicado em um conjunto de dados que não fez parte de sua construção ou otimização, conjunto este denominado de previsão ou teste. A qualidade dos modelos de calibração multivariada gerados normalmente é avaliada pela raiz quadrada do erro quadrático médio de previsão, RMSEP.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_{prev}} (y_{prev,i} - y_{ref,i})^2}{n_{prev}}} \quad (\text{Eq. 17})$$

onde n_{prev} é o número de amostras de previsão (ou teste), y_{ref} e y_{prev} são respectivamente, os valores de referência e previsto pelo modelo para as amostras do conjunto de previsão

O ajuste linear entre os valores de referência e previstos é verificado pelo coeficiente de determinação.

$$R^2 = 1 - \frac{\sum_i (y_{ref,i} - y_{prev,i})^2}{\sum_i (y_{ref,i} - y_{m\u00e9dio})^2} \quad (\text{Eq. 18})$$

Onde $y_{m\u00e9dio}$ s\u00e3o os valores m\u00e9dios, respectivamente, podendo ser aplicado tanto \u00e0s amostras do conjunto de calibra\u00e7\u00e3o ou previs\u00e3o. O valor de R^2 varia de 0 a 1, sendo que quanto mais pr\u00f3ximo de 1 melhor o ajuste do modelo.^{6, 23, 35, 36}

- An\u00e1lise dos res\u00edduos

Uma importante etapa de avalia\u00e7\u00e3o dos modelos est\u00e1 na an\u00e1lise dos res\u00edduos (r_i) do modelo, calculados por:

$$r_i = y_{ref,i} - y_{prev,i}, \quad (\text{Eq. 19})$$

Os res\u00edduos devem ser calculados tanto para o conjunto de calibra\u00e7\u00e3o quanto para o de previs\u00e3o e ambos devem ser avaliados quanto \u00e0 presen\u00e7a de erros sistem\u00e1ticos e tend\u00eancia. Havendo a presen\u00e7a destes tipos de erros nos res\u00edduos, o modelo gerado pode ser considerado insatisfat\u00f3rio.

- *Avalia\u00e7\u00e3o dos erros sistem\u00e1ticos*

Os erros sistem\u00e1ticos afetam a estimativa sempre no mesmo sentido, gerando resultados abaixo ou acima do valor esperado.³⁶ A presen\u00e7a desses erros \u00e9 calculada atrav\u00e9s do teste para vi\u00e9s, que representam o desvio m\u00e9dio de n medi\u00e7\u00f5es.

$$vi\u00e9s = \frac{\sum (y_i - \hat{y}_i)}{n}, \quad (\text{Eq. 20})$$

Esse teste \u00e9 executado atrav\u00e9s de um teste-t bicaudal no qual as hip\u00f3teses testadas s\u00e3o:

H_0 : vi\u00e9s = 0;

H_1 : vi\u00e9s \neq 0;

A partir do valor calculado de viés, o desvio padrão dos erros (SVD, do inglês *Standard Deviation of the Validation Errors*) e a estatística de teste (t_{calc}) são determinados por:

$$SVD = \sqrt{\frac{\sum_{i=1}^n \{[(y_{ref,i} - y_{prev,i}) - viés]^2\}}{n-1}} \quad (\text{Eq. 21})$$

$$t_{calc} = \frac{|viés|\sqrt{n}}{SVD}, \quad (\text{Eq. 22})$$

A partir daí o valor de t_{calc} é comparado com o valor de t tabelado (t_{tab}) da distribuição *t-student* com $n-1$ graus de liberdade. Se $t_{calc} < t_{tab}$, aceita-se a hipótese nula e considera-se que não há evidências da presença de erros sistemáticos nos resíduos. Caso contrário, a hipótese nula é rejeitada e admite-se a presença de erros sistemáticos.^{23, 34, 36}

- *Avaliação dos erros de tendência*

A presença de tendência não está diretamente associada à distribuição dos resíduos. Sendo assim, o procedimento normalmente adotado para avaliar tendência em resíduos de calibração multivariada é subjetivo, através da observação visual do gráfico dos resíduos em função dos valores de referência. Entretanto, neste procedimento diferentes analistas podem tomar decisões conflitantes quanto à existência de tendência ou não. Para evitar conclusões subjetivas foi implementado um teste de permutação não paramétrico para avaliar erros de tendência em resíduos de calibração multivariada. Esse teste baseia-se na probabilidade da tendência observada nos dados ser devida ao acaso portanto, em sua execução são testadas duas hipóteses:

H_0 : os resíduos r_i são independentes de y_i ;

H_1 : os resíduos r_i não são independentes de y_i conforme a Equação:

$$r_i = g(y_i) + \varepsilon_i, \quad (\text{Eq. 23})$$

onde ε_i é um erro aleatório independente e $g(y_i)$ alguma função polinomial que pode modelar a relação entre os resíduos e os valores de referência.

Na hipótese alternativa, a dependência dos resíduos com os valores de referência é proposta, e presume-se que todo o efeito aleatoriamente presente em y_i é devido somente a variável ε_i . Haverá evidências de tendência nos resíduos se o coeficiente polinomial de maior ordem (b_n) da equação polinomial $g(y_i)$ for estatisticamente significativo, ao nível de significância adotado. Assim, o grau do polinômio deve ser definido previamente à aplicação do teste. O teste pode ser resumido nos seguintes passos:

- I. Calcular o coeficiente polinomial b_n ajustado para os dados originais dos resíduos em função dos valores de referência. Assim, o coeficiente será denominado b_n^* ;
- II. Permutar randomicamente somente o vetor y contendo a propriedade de interesse;
- III. Calcular o coeficiente b_n^i para o i -ésimo vetor y permutado;
- IV. Comparar b_n^* com b_n^i
- V. Repetir as etapas (II) a (IV) k -vezes.

Como os coeficientes b_n^i dos ajustes permutados não tem sentido físico, por serem aleatórios, a distribuição destes coeficientes constituirá a distribuição para o teste. Sendo assim, o p -valor do teste é determinado pela proporção do número de vezes em que $b_n^* > b_n^i$. Se o p -valor do teste for menor que o nível de significância adotado, rejeita-se H_0 , e a tendência observada é significativa. Caso contrário, aceita-se H_0 e a hipótese dos resíduos serem aleatórios.^{23, 34, 36}

1.3.8.2 Validação dos modelos de classificação

Partindo de duas classes, positiva e negativa, os resultados de previsão dos modelos de classificação são avaliados pela sensibilidade, especificidade, exatidão e prevalência. A sensibilidade corresponde a habilidade do modelo em detectar amostras positivas quando elas são positivas, estimada por:

$$\text{sensibilidade} = \frac{TP}{TP+FN}, \quad (\text{Eq. 24})$$

Onde TP é o número de amostras positivas corretamente identificadas e FN é o número de amostras positivas que foram identificadas como negativas, falso

negativas. Um modelo com alta sensibilidade não falhará na identificação de uma amostra positiva, ou seja, FN será o menor possível, pois esse modelo não poderá correr o risco de não identificar uma amostra positiva.

A especificidade reflete a habilidade do modelo identificar corretamente amostras negativas:

$$especificidade = \frac{TN}{TN+FP}, \quad (\text{Eq. 25})$$

Onde TN é o número de amostras negativas corretamente identificadas e FP é o número de amostras negativas identificadas como positivas, falso positivas. Um teste de especificidade raramente identificará uma amostra negativa como positiva, ou seja, FP tenderá a zero. A exatidão é expressa pela relação entre os resultados obtidos pelo modelo e suas referências, é medida pela proporção de todos resultados corretos do modelo para todas as amostras testadas.

$$exatidão = \frac{TP+TN}{TP+TN+FP+FN}, \quad (\text{Eq. 26})$$

Embora na prática a questão importante seja determinar a probabilidade de uma amostra ser realmente positiva, quando o modelo de classificação identificar ela como positiva. Essa probabilidade é chamada de prevalência de valor positivo (PPV), e no caso do interesse em determinar a probabilidade de uma amostra ser negativa quando o modelo identificar como negativa, tem-se a prevalência de valor negativo (PNV).³⁷ Essas probabilidades, chamadas de prevalências, são estimadas por:

$$PPV = \frac{TP}{TP + FP}, \quad (\text{Eq. 27})$$

$$PNV = \frac{TN}{TN + FN}, \quad (\text{Eq. 28})$$

2 Objetivos

2.1 Objetivos Gerais

Este trabalho tem como objetivo principal a determinação de propriedades físico-químicas de amostras de petróleo e derivados a partir da espectroscopia de Ressonância Magnética Nuclear associada a ferramentas quimiométricas de calibração, reconhecimento de padrão e análise exploratória.

2.2 Objetivos Específicos

Para alcançar o objetivo geral, foram delineados os seguintes passos específicos

- Realizar medidas espectroscópicas de RMN de ^1H e ^{13}C em amostras de petróleo e derivados de petróleos com diferentes faixas temperaturas de destilação, e a partir dos espectros determinar os parâmetros moleculares;
- Determinar qual o melhor conjunto de variáveis obtidas dos espectros na construção dos modelos;
- Desenvolver metodologias analíticas utilizando a RMN e métodos quimiométricos de regressão para determinação de parâmetros físico-químicos de amostras de petróleo e derivados;
- Desenvolver metodologias analíticas utilizando a RMN e métodos quimiométricos de classificação para determinação de parâmetros físico-químicos de amostras de petróleo e derivados
- Testar e comparar o desempenho dos métodos de pré-processamento dos dados.

3 Aplicação 1: Determinação da gravidade API, viscosidade cinemática, ponto de fluidez e SARA em petróleo por espectroscopia RMN e métodos multivariados de calibração e classificação.

3.1 Introdução

O Petróleo apresenta grandes variações em sua composição e propriedades, e estes ocorrem não só em petróleos de diferentes áreas, mas também naqueles retirados de diferentes profundidades no mesmo poço de produção. Comercialmente sua composição, bem como de seus derivados, variam muito devido a suas diferentes origens e aos diferentes processos de refino pelos quais pode ser submetido. Além disso, sua qualidade tem mudado constantemente desde sua introdução no mercado como combustível.²

Historicamente, as propriedades físico-químicas tais como ponto de ebulição, densidade, viscosidade, odor e ponto de fluidez vêm sendo usadas para descrever os óleos. O petróleo pode ser classificado como leve ou pesado de acordo com a presença de compostos de baixo ponto de ebulição e a densidade relativa. Da mesma forma, o odor é usado para distinguir entre petróleos doces (com baixo teor de enxofre) ou azedos (com alto teor de enxofre). A viscosidade indica a facilidade de um óleo escoar (ou mais corretamente a resistência para escoar). Apesar de não estar diretamente relacionado a composição, os termos leve e pesado ou doce e azedo fornecem termos convenientes para descrição dos óleos.³⁸

Nesse contexto, o valor do petróleo depende da sua qualidade para o refino e de sua capacidade de ajuste a fim de gerar um determinado produto para atender a demanda do mercado. A partir de informações da qualidade do petróleo e derivados e do mercado a ser atendido, o refino moderno utiliza uma sofisticada combinação de calor, catalisador, e hidrogênio para reorganizar as moléculas de petróleo, de forma a aumentar o rendimento da formação de produtos ambientalmente saudáveis e com alto valor agregado. Processos de conversão incluem coque, hidrocrackeamento, e crackeamento catalítico para quebrar moléculas grandes em frações menores; hidrotratamento para reduzir teores de heteroátomos e aromáticos, criando produtos

ambientalmente aceitáveis; e isomerização e rearranjo para reorganizar moléculas e formar compostos com alto valor agregado como, por exemplo, a gasolina com um elevado índice de octano e diesel com elevado nível de cetano. Além disso, o conhecimento da composição molecular de petróleo permite o ambientalista considerar o impacto biológico da exposição ambiental.²

O valor do petróleo depende, portanto, da sua qualidade para o refino e de sua capacidade de ajuste a fim de gerar um determinado produto para atender a demanda do mercado. Assim, unidades de processamento de uma refinaria exigem métodos de testes analíticos que possam avaliar adequadamente matérias-primas e monitorar a qualidade do produto. Uma vez que as propriedades exigidas são determinadas, elas são controladas por testes e por análises apropriadas e auxiliam na elaboração de estratégias de transporte e refino, além de informar sobre potenciais derivados esperados.^{2, 39}

Métodos de caracterização são necessários para o planejamento do processo, avaliação de petróleo bruto, e controle operacional. A avaliação do petróleo produzido envolve cerca de 700 ensaios físico-químicos, consumindo de 10 a 70 litros de petróleo, em não menos de 4 meses, ao custo estimado de mais de 80 mil dólares. Dos 700 ensaios, 26 são diretamente no petróleo bruto.^{23, 34, 40} Devido à grande perspectiva de aumento da produção aliado ao longo tempo para análises e grandes quantidades de amostras, o desenvolvimento de metodologias analíticas rápidas e em menores quantidades de amostras são estratégicas para o controle do processo de produção e refino.³⁴

A avaliação dessas propriedades é, geralmente baseada nas normas da Sociedade americana de testes e materiais (ASTM – *American Society for Testing and Materials*).

A partir da grande variedade de petróleos existentes, a forma mais comum de expressar a densidade do óleo é a partir da escala hidrométrica idealizada pelo *American Petroleum Institute - API*, juntamente com a *National Bureau of Standards*. A densidade tem uma importância especial, uma vez que petróleo e principalmente produtos de petróleo são usualmente comercializados com base nos valores de densidade. A densidade pode ser determinada por vários métodos, tais como ASTM D1298 que descreve o teste usando hidrômetros de vidro, enquanto a D4052 usa

densímetro automático, no qual pequenos volumes de amostra são introduzidos num tubo de oscilação, onde a alteração da frequência de oscilação provocada pela variação da massa é usada em conjunto com os dados de calibração do equipamento fornecem a densidade da amostra.

A escala API em relação à densidade é dada pela equação:

$$API = \frac{141,5}{\rho(60^\circ/60^\circ F)} - 131,5, \quad (\text{Eq. 29})$$

Onde $\rho(60^\circ/60^\circ F)$ é a densidade específica do óleo a 60°F em relação à densidade da água na mesma temperatura. A gravidade API e a densidade são medidas inversamente proporcionais, portanto quanto maior a densidade, mais pesado será o petróleo e menor a sua gravidade API e vice-versa.^{2, 38} A gravidade API pode definir tipos de petróleos e direcioná-los para a correta unidade de destilação. Variação superior a 1 unidade na gravidade API pode modificar toda a logística de refino de um petróleo, uma vez que as unidades de destilação são construídas para destilar tipos específicos de petróleos.^{2, 34}

Utilizando-se da escala API, pode-se ainda classificar os óleos em leve, médio, pesado e extrapesado, de acordo com a Agência Nacional de Petróleo Gás Natural e Biocombustíveis (ANP).⁴⁰

A gravidade API e a viscosidade cinemática afetam a viabilidade econômica dos campos produtores, pois estas propriedades auxiliam no dimensionamento dos equipamentos utilizados na exploração e na estimativa do preço do barril do petróleo. Após iniciada a exploração, estas propriedades influenciam o processo de decisão da produção, pois controlam o intervalo de produção do reservatório e auxiliam na perfuração de novos poços.^{2, 34}

A viscosidade é uma importante propriedade do fluido pois é a medida da resistência de um fluido ao escoamento por razão das forças intermoleculares entre as moléculas ou grupos moleculares. Assim as medidas de viscosidade têm sido consideradas de particular importância pela indústria do petróleo, devido ao papel que desempenha na avaliação das propriedades do petróleo e, conseqüentemente, a produtividade de um reservatório.³⁸

As medidas de viscosidade e ponto de fluidez são realizadas principalmente para determinar comportamento fluido do óleo a baixas temperaturas, uma vez que existem algumas relações gerais de composição do óleo bruto que podem ser derivadas dessas propriedades.

A viscosidade é usualmente determinada em diferentes temperaturas por meio da medida do tempo necessário para um volume de líquido fluir pela ação da gravidade através de um viscosímetro capilar de vidro calibrado a uma temperatura rigorosamente controlada (ASTM D445). No teste esse tempo é medido em segundos, e a viscosidade cinemática será o produto entre o tempo de fluxo medido e a constante de calibração do viscosímetro² (Equação 30).

$$v = C \cdot t, \quad (\text{Eq. 30})$$

Onde v é a viscosidade cinemática, C é a constante de calibração do viscosímetro e t o tempo de escoamento em segundos.³⁸

O ponto de fluidez do petróleo corresponde a menor temperatura na qual o petróleo ainda flui em condições específicas. As temperaturas dos pontos de fluidez máximo e mínimo fornecem uma janela de temperaturas onde, dependendo da história térmica do petróleo, pode aparecer no estado líquido bem como no estado sólido. O ponto de fluidez pode ser utilizado para complementar outras medidas de comportamento de fluxo do óleo em baixas temperaturas, além desses dados serem particularmente úteis para o rastreamento do efeito da presença das parafinas na fluidez do petróleo.

No ensaio original (e ainda amplamente usado) para a determinação do ponto de fluidez (ASTM D97), uma amostra é aquecida a uma velocidade específica, tendo suas características de fluxo examinadas em intervalos de 3°C. A temperatura mais baixa na qual se observa movimento do óleo é registrada como o ponto de fluidez.²

Como dito anteriormente, o petróleo é constituído predominantemente de hidrocarbonetos,² os quais, de acordo com a estrutura, podem ser classificados em saturados ou parafinas, insaturados ou olefinas, aromáticos ou arenos. Os componentes presentes no petróleo são agrupados em quatro classes principais os hidrocarbonetos saturados, hidrocarbonetos aromáticos, resinas e asfaltenos.¹

Uma das formas de se fazer a separação dessas classes de compostos é baseando-se na diferença de solubilidade entre elas, esse método de fracionamento do petróleo é conhecido como análise SARA (saturados, aromáticos, resinas e asfaltenos)² de acordo com a norma ASTM D2007, o qual se baseia na diferença de solubilidade de cada uma das frações, por meio de cromatografia de coluna. Devido aos baixos teores de resinas e asfaltenos, e a semelhança entre essas duas frações alguns artigos trazem a separação em apenas três frações: saturados, aromáticos e polares (SAP).

Os hidrocarbonetos saturados são caracterizados por compostos cujos átomos de carbono são unidos somente por ligações simples e ao maior número possível de átomos de hidrogênio, no petróleo os compostos saturados são subdivididos em dois grupos: as parafinas que são constituídas por hidrocarbonetos de cadeia aberta lineares ou ramificadas; e os compostos naftênicos constituídos pelos hidrocarbonetos de cadeia fechada. Os hidrocarbonetos aromáticos são aqueles que apresentam pelo menos um anel benzênico em sua estrutura e, são encontrados na forma de naftenoaromáticos e benzotiofenos e seus derivados. Já as resinas e asfaltenos são moléculas grandes, com alta relação carbono/hidrogênio e presença de enxofre, oxigênio e nitrogênio (6,9 – 7,3%), o que confere a elas uma certa polaridade. Os teores de cada um desses compostos são os responsáveis pela classificação dos diversos tipos de petróleos, uma vez que essa divisão é feita de acordo seus constituintes em relação aos demais constituintes existentes. A classificação do petróleo de acordo com seus constituintes interessa desde os geoquímicos até os refinadores. Os primeiros visam caracterizar o óleo para relacioná-lo à rocha-mãe e medir o seu grau de degradação. Os refinadores querem saber a quantidade das diversas frações que podem ser obtidas, assim como sua composição e propriedades físicas.¹

Além das propriedades citadas acima, existem outras dezenas de ensaios que são realizados para que se haja uma caracterização e avaliação completa do petróleo. Às vezes, tem-se urgência no resultado de determinadas análises para uma tomada de decisão e isto fica prejudicado pela forma como as análises são feitas. Diante disto, surge a necessidade de se desenvolver técnicas mais rápidas, simples, econômicas, confiáveis e que provoquem menos impacto ambiental. Por isso, existe atualmente grande interesse da comunidade científica em tentar otimizar esses procedimentos

nos quesitos tempo, dinheiro e quantidade de amostra necessários para sua realização.^{3, 7, 23}

3.2 Objetivos

Pretende-se nesta etapa determinar propriedades físico químicas de amostras de petróleo utilizando a RMN associada a técnicas quimiométricas. Para isso, foram delineados os seguintes passos:

- Utilizar RMN de ^1H e ^{13}C com a finalidade de determinar parâmetros estruturais das amostras de petróleo, de acordo com os artigos Hasan (1983) e Poveda (2012);
- Construir modelos PLS a fim de determinar a gravidade API, viscosidade cinemática, ponto de fluidez e SARA em petróleos utilizando PLS e MLR;
- Construir modelos PLS-DA a fim de classificar as amostras de petróleo em leve, médio, pesado e extrapesado, a partir da gravidade API;
- Testar e comparar o desempenho dos métodos de pré-processamento dos dados: MSC, autoescalonamento, SNV e normalização, além de testes sem nenhum tipo de pré-processamento.

3.3 Metodologia

3.3.1 Obtenção das amostras

Neste trabalho foram utilizados 158 petróleos diferentes, com grau API variando de 11,4 a 54,0, ou seja, foram incluídas amostras de petróleos, desde leves até pesados. As amostras de petróleo foram cedidas pelo CENPES/Petrobras, armazenados em frascos apropriados e em temperatura adequada para posteriores análises.

3.3.2 Determinação dos parâmetros físico-químicos

As amostras de petróleo utilizadas neste trabalho, foram caracterizadas pelo LabPetro/UFES e/ou CENPES. As propriedades físico-químicas foram determinadas conforme os respectivos métodos padrão descritos na Tabela 6, onde são reportados os valores de reprodutibilidade dos respectivos métodos laboratoriais de cada propriedade trabalhada, exceto para viscosidade cinemática, uma vez que a norma

ASTM desta propriedade traz uma reprodutibilidade relativa em temperaturas diferentes das testadas neste trabalho.

Tabela 6. Métodos ASTM referentes às propriedades

Propriedade	Método	Reprodutibilidade	
Gravidade API	ISO 12185	1,1	
Ponto de Fluidez (Máx.)	ASTM D5853	18°C	
Viscosidade Cinemática	ASTM D7042	-	
SARA	ASTM D2007	Saturados	4,0%
		Aromáticos	3,3%
		Polares (resinas +	1,8%
		Asfaltenos	

3.3.3 Análises de RMN

As análises de Ressonância Magnética Nuclear de ^1H e ^{13}C foram realizadas no LabPetro no espectrômetro da marca Varian, modelo VNMRS 400, operando com um campo magnético de 9,4 T, utilizando uma sonda 5 mm BroadBand $^1\text{H}/^{19}\text{F}/\text{X}$ para as análises de ^1H e uma sonda 10 mm Broadband $^{15}\text{N}-^{31}\text{P}\{^1\text{H}-^{19}\text{F}\}$ para as análises de ^{13}C , ambas realizadas à temperatura de 25 °C.

Para as análises de ^1H , as amostras foram preparadas dissolvendo-se uma massa correspondente a 5% em diclorometano- d_2 , totalizando um volume de 600 μL de solução. As análises foram feitas conforme as condições descritas abaixo:

- Frequência: 399,73 MHz;
- Janela espectral: 6410,3 Hz;
- Tempo de Aquisição: 2,556 s;
- Tempo espera: 1 s;
- Pulso: 90°;
- Número de transientes: 64;
- Processamento: linebroadening de 0,5 Hz.

Para as análises de ^{13}C , as amostras foram preparadas dissolvendo-se uma massa correspondente a 40% do volume final, em uma solução 0,05 mol/L de $\text{Cr}(\text{acac})_3$ (acetilacetato de cromo III) com clorofórmio-*d*, totalizando um volume de 3 mL de solução. As análises foram feitas conforme as condições descritas abaixo:

- Frequência: 100,51 MHz;
- Janela espectral: 25510,2 Hz;
- Tempo de Aquisição: 1,285 s;
- Tempo espera: 7 s;
- Pulso: 90° ;
- Núcleo desacoplado: ^1H ;
- Modo do desacoplador: nny, ou seja, o desacoplador ficou desligado durante o pulso e o tempo de espera, e ligado somente durante a aquisição dos dados, com isso o aumento de intensidade do sinal ocasionado pelo Efeito Nuclear Overhauser (NOE) é eliminado, garantido assim um experimento quantitativo;
- Número de transientes: 1000;
- Processamento: alargamento exponencial de 5,0 Hz.

Os deslocamentos químicos foram obtidos tendo os solventes ou o tetrametilsilano (TMS) como sinal de referência. A fase e a linha de base foram ajustadas e corrigidas manualmente. Os espectros de RMN foram tratados utilizando-se os parâmetros moleculares, determinados a partir das áreas dos sinais específicos sugeridos pela literatura, segundo Hasan (1983) e Poveda (2012).^{15, 16}

As análises de RMN de ^1H , foram feitas em triplicata para a maioria das amostras, e as análises de RMN de ^{13}C foram feitas em triplicata apenas para três amostras.

3.3.4 Tratamento Quimiométrico

Os espectros de RMN das amostras de petróleo foram tratados utilizando-se os parâmetros moleculares, determinados a partir das áreas dos sinais específicos sugeridos segundo Hasan e Poveda. Os conjuntos de variáveis utilizados foram as integrais dos espectros de hidrogênio segundo os dois autores, as integrais dos espectros de carbono segundo os dois artigos, e o conjunto das integrais referentes

aos espectros de hidrogênio e carbono segundo ambos os autores. Como pode ser melhor explicado na Tabela 7.

Tabela 7. Conjuntos de variáveis utilizados na construção dos modelos.

Artigo Base	Conjunto de parâmetros moleculares		
	(Conjunto de variáveis)		
Hasan (1983) ¹⁶	RMN de ¹ H	RMN de ¹³ C	RMN de ¹ H e ¹³ C
Poveda (2012) ¹⁵	RMN de ¹ H	RMN de ¹³ C	RMN de ¹ H e ¹³ C

Vale ressaltar que, embora tenham sido analisadas 158 amostras no total, alguns parâmetros não foram avaliados em todas elas. Logo, o número de amostras utilizadas na construção dos modelos foi de acordo com os parâmetros físico-químicos fornecidos.

Os métodos quimiométricos aplicados aos dados foram o PLS e o MLR, utilizando-se ainda o método PLS-DA a fim de classificar as amostras em relação a gravidade API. A construção dos modelos foi realizada em 3 etapas principais:

- 1 – Pré-processamento dos dados;
- 2 – Construção dos modelos propriamente dita;
- 3 – Avaliação e validação dos modelos.

Todos os cálculos foram executados no software Matlab versão 7.

3.3.4.1 Modelos de calibração multivariada PLS e MLR

- **Tratamento e pré-processamento dos dados**

Anteriormente à aplicação das ferramentas quimiométricas ao conjunto de dados a ser investigado, foram aplicados e testados como recurso de tratamento dos dados a MSC, o autoescalamento, a SNV e a normalização, os modelos também foram construídos sem nenhum tipo de pré-processamento. Todos os dados foram, ainda, centrados na média antes da construção dos modelos PLS.

- **Construção dos modelos**

Anteriormente à construção dos modelos propriamente dita, foi feita a divisão das amostras, selecionando-se aquelas que constituíram os conjuntos de calibração e previsão. Para a seleção, aproximadamente 70% das amostras foram previamente selecionadas para o conjunto de calibração, enquanto que os 30% restantes foram para o conjunto de previsão.

Então, feita essa separação dos conjuntos de calibração e validação, foram construídos os modelos globais utilizando PLS e MLR, para cada parâmetro analisado, para cada conjunto de variáveis utilizados, e para cada tipo de pré-processamento e tratamento testados.

O número de variáveis latentes para cada modelo de calibração PLS foi otimizado pelo procedimento de validação cruzada utilizando o procedimento *k-fold* que consiste em dividir as amostras de calibração em *k* subconjuntos de mesmo tamanho, onde um subconjunto é utilizado para validação enquanto os restantes (*k-1*) são utilizados para construção do modelo. Os subconjuntos foram retirados de forma ordenada utilizando o método *venetian blind 5-fold*. Dessa forma, o número de VL's a ser utilizado na construção do modelo foi determinado pelo gráfico de RMSECV em função do número de variáveis latentes, ou seja, foi selecionado aquele com o número de variáveis latentes correspondente ao menor valor de RMSECV e R^2 mais próximo de 1.

- **Avaliação dos modelos**

A avaliação dos modelos se deu pela determinação e análise dos seguintes requisitos: coeficiente de determinação (R^2), curva obtida entre os valores previstos versus medidos, erros de validação cruzada (para PLS) e previsão.

A eficiência dos modelos de calibração multivariada foi avaliada pelo cálculo de três tipos de raiz quadrada do erro (RMSE – do inglês, *root mean square error*): a raiz quadrada do erro quadrático médio de calibração (RMSEC), definido como na equação 31, onde *n* corresponde ao número de amostras e VL corresponde ao número de variáveis latentes selecionadas:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^N (y_{prev,i} - y_{ref,i})^2}{n - VL - 1}} \quad (\text{Eq. 31})$$

Onde y_{ref} e y_{prev} são respectivamente, os valores de referência e previsto pelo modelo para as amostras de calibração.

Já o RMSE, determinado no procedimento de validação cruzada (RMSECV, do inglês *root mean square error of cross validation*) e de previsão (RMSEP, do inglês *root mean square error of prediction*), foi calculado conforme a Equação 32.²²

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_{prev} - y_{ref})^2}{N}} \quad (\text{Eq. 32})$$

Onde y_{ref} e y_{prev} são respectivamente, os valores de referência e previsto pelo modelo para as amostras de calibração ou validação e N corresponde ao número de amostras em cada conjunto (calibração ou previsão).

Esses valores de RMSE indicam a grandeza dos erros associados aos resultados obtidos e, como condição de aceitação dos modelos criados, não devem ser estatisticamente diferentes, ou seja, apresentar uma diferença grande em questão de unidades demonstrando, dessa forma, que os conjuntos de calibração e previsão são semelhantes. Como esses valores de RMSE são determinados na dimensão das propriedades que está se tentando determinar, sua avaliação depende da amplitude e da dimensão dos dados trabalhados e, nesse sentido, o cálculo do erro percentual de previsão facilita a avaliação dos modelos. Nesse contexto, também foi determinado o erro percentual de previsão a partir da Equação 33.^{23,36}

$$RMSE \% = 100 \cdot \frac{RMSEP}{\bar{y}_{previsão}}, \quad (\text{Eq. 33})$$

Onde $\bar{y}_{previsão}$ é a média dos valores de previsão.

Os modelos escolhidos foram submetidos a testes estatísticos, e tiveram suas figuras de mérito calculadas.^{23, 35, 36}

O teste de viés foi aplicado para verificação da presença de erros sistemáticos e, para a avaliação de tendência nos resíduos do modelo, foi aplicado o teste de permutação.^{23, 36}

Para avaliar o desempenho do modelo, considera-se que o RMSEP deve ser comparável à ASTM (método padrão) e a incerteza final de previsão é expressa pela Equação 34:

$$\sigma_{previs\tilde{a}o} = \sqrt{RMSEP^2 + \sigma_{ASTM}^2} \quad (\text{Eq. 34})$$

Onde $\sigma_{previs\tilde{a}o}$ é o erro total de medição do modelo de previsão e σ_{ASTM} é o valor de variabilidade em medidas feitas em forma de repetitividade ou reprodutibilidade. Neste trabalho esse cálculo foi feito apenas a partir dos valores de reprodutibilidade.

3.3.4.2 Modelos PLS-DA

Os modelos de PLS-DA foram aplicados somente aos conjuntos de variáveis que apresentaram melhores resultados para gravidade API nos métodos de calibração.

- **Tratamento e pré-processamento dos dados**

Anteriormente à aplicação das ferramentas quimiométricas PLS-DA aos conjuntos de dados que apresentaram melhores resultados na PLS, foram aplicados e testados como recurso de tratamento dos dados a MSC, o autoescalamento, a SNV e a normalização, os modelos também foram construídos sem nenhum tipo de pré-processamento.

- **Construção dos modelos**

Anteriormente à construção dos modelos propriamente dita, foi feita a divisão das amostras, selecionando-se aquelas que constituíram os conjuntos de treinamento e previsão. Para a seleção, aproximadamente 70% das amostras foram previamente selecionadas para o conjunto de treinamento, enquanto que os 30% restantes foram para o conjunto de previsão. Essa separação foi feita de maneira que 70% das amostras de cada classe fossem usadas na construção dos modelos e 30% de cada classe para a previsão, garantindo assim a representatividade de cada classe. As

amostras de treinamento foram utilizadas para calibrar os modelos de classificação PLS-DA, enquanto as amostras de previsão foram usadas somente no final para avaliar a capacidade preditiva do modelo construído.

Então, feita essa separação dos conjuntos de treinamento e validação, foram construídos os modelos globais utilizando PLS-DA, para o parâmetro gravidade API, para os conjuntos de variáveis que apresentaram melhores resultados na PLS e MLR, e para cada tipo de pré-processamento e tratamento testados.

O número de variáveis latentes para cada modelo de classificação foi otimizado pelo procedimento de validação cruzada, assim como na construção dos modelos PLS. Neste procedimento, o número ótimo de variáveis latentes foi avaliado pela taxa de erro de classificação na análise de discriminante.

Dessa forma, o número de VL's a ser utilizado na construção do modelo foi determinado pelo gráfico do erro em função do número de variáveis latentes, ou seja, foi selecionado aquele com o número de variáveis latentes correspondente ao menor erro.

- **Avaliação dos modelos**

A avaliação dos modelos PLS-DA se deu pela determinação e análise dos seguintes requisitos: sensibilidade, especificidade de cada classe, exatidão e prevalência.

Foram determinadas as taxas de erro e acerto dos modelos na classificação das amostras. A taxa de acerto é calculada como a média aritmética das taxas de acerto em cada classe, e o acerto em cada classe é dado pela média da especificidade e da sensibilidade dela. E finalmente, a taxa de erro é o complemento da taxa de acerto³⁰.

$$Taxa\ de\ acerto = \frac{\sum_{i=1}^c Taxa\ de\ acerto_{classe,i}}{c} \quad (Eq. 35)$$

Onde c corresponde ao número de classes do modelo.

$$Taxa\ de\ acerto_{classe} = \frac{especificidade_{classe} + sensibilidade_{classe}}{2} \quad (Eq. 36)$$

$$Taxa\ de\ erro = 1 - Taxa\ de\ acerto \quad (Eq. 37)$$

3.4 Resultados e Discussões

As 158 amostras de óleo estudadas apresentam gravidade API entre 11,4 e 54,0, que corresponde a óleos pesados, médios e leves de acordo com a classificação de petróleos por densidade proposta pela ANP.⁴⁰ A viscosidade cinemática a 50°C encontra-se na faixa de 0,6873 a 8200 mm².s⁻¹. O ponto de fluidez (PF) máximo dos óleos variou entre -60 e 39°C. Devido à grande faixa de óleos diferentes estudados, as concentrações de saturados, aromáticos, resinas e asfaltenos também apresentam grande variação. As propriedades físico-químicas foram medidas utilizando métodos analíticos padrões, e os dados estão reportados na Tabela 8.

Tabela 8. Propriedades físico-químicas selecionadas para os óleos estudados (continua).

Amostra	API	PF máximo (°C)	Viscosidade a 50°C (mm ² /s)	SARA (%)				
				Saturados	Aromáticos	Resinas	Asfaltenos	Polares
1	54,0	-60	0,7028	88,4	11,6	0,0	0,0	0,0
2	41,9	6	2,3507	71,9	16,2	11,8	0,1	11,9
3	17,0	-9	247,2000	41,6	29,8	27,2	1,4	28,6
4	32,5	30	20,5200	59,7	10,6	29,7	0,0	29,7
5	25,5	-54	15,2000	58,6	23,4	18,0	0,0	18,0
6	23,6	-46	43,2200	62,2	19,4	18,5	0,0	18,4
7	29,9	-27	7,3240	52,7	33,6	12,6	1,0	13,7
8	53,3	-60	0,6873	89,2	10,8	0,0	0,0	0,0
9	11,4	18	8200	37,0	28,7	23,4	9,6	34,3
10	26,8	9	21,2000	53,8	24,1	22,1	0,0	22,1
11	26,7	-12	14,7700	56,0	26,9	16,0	1,1	17,1
12	39,5	-	2,6450	84,0	8,4	7,6	0,0	7,6
13	29,0	-30	8,2140	52,1	27,2	18,7	2,0	20,7
14	28,4	9	12,7000	51,8	21,9	25,6	0,8	26,3
15	29,4	6	10,6200	58,3	21,9	19,8	0,0	19,8
16	24,4	-30	24,7800	44,8	22,8	31,0	1,4	32,4
17	29,6	12	10,4046	60,3	22,7	17,0	0,0	17,0
18	22,0	-30	38,5000	44,2	26,5	27,2	2,1	29,3
19	30,8	6	8,4070	57,7	24,2	17,5	0,5	18,1
20	30,6	3	9,1450	61,5	23,1	15,4	0,0	15,4
21	23,4	-27	25,6500	48,0	28,6	19,1	4,1	23,4
22	28,7	6	12,8200	58,4	19,4	22,2	0,0	22,2
23	29,2	12	9,0462	54,0	24,0	22,0	0,0	22,0
24	29,8	6	9,8340	38,0	27,3	33,4	1,3	34,7
25	32,1	3	6,7470	56,6	24,4	19,0	0,0	19,0
26	25,5	-45	19,8200	49,7	24,9	24,0	1,3	25,4
27	37,8	-18	3,2680	74,1	13,9	11,2	0,8	12,0
28	30,2	6	8,9340	63,1	18,2	18,6	0,1	18,7

Tabela 8. Propriedades físico-químicas selecionadas para os óleos estudados (continua).

Amostra	API	PF máximo (°C)	Viscosidade a 50°C (mm ² /s)	SARA (%)				
				Saturados	Aromáticos	Resinas	Asfaltenos	Polares
29	41,0	-6	2,2380	80,8	12,8	6,4	0,0	6,4
30	25,9	-15	16,6700	48,6	22,2	26,6	2,6	29,2
31	35,9	12	4,3920	46,0	35,1	18,9	0,0	18,9
32	28,0	-39	9,9990	53,5	29,4	16,2	0,9	17,1
33	27,5	-27	11,8300	56,2	30,7	11,9	1,2	13,1
34	29,5	-3	8,7992	61,5	22,6	14,7	0,8	15,9
35	40,5	18	3,3690	74,7	14,8	10,6	0,0	10,5
36	21,0	-35	86,5400	38,2	27,4	0,0	4,7	34,4
37	23,6	21	62,5800	40,4	16,2	42,8	0,7	43,4
38	52,0	-60	0,7003	90,6	9,4	0,0	0,0	0,0
39	36,7	15	4,6460	76,6	13,5	9,3	0,6	9,9
40	32,9	6	6,8820	70,8	14,7	14,0	0,8	14,5
41	30,8	8	7,0750	57,4	28,0	14,5	0,4	14,6
42	34,0	-6	4,6790	62,4	27,5	9,5	0,5	10,1
43	29,3	2,0	9,0920	58,2	26,5	14,8	0,5	15,3
44	22,7	-51	29,0900	40,2	37,2	20,8	1,9	22,6
45	28,4	8	12,8700	52,6	27,2	19,9	0,3	20,2
46	29,8	6	9,8340	38,0	27,3	33,4	1,3	34,7
47	27,8	12	14,4900	55,1	27,3	17,1	0,6	17,6
48	44,6	-9	1,9270	83,0	12,2	4,8	0,0	4,8
49	33,0	-3	5,7210	62,5	26,6	10,2	0,7	10,9
50	30,6	3	9,1450	61,5	23,1	15,4	0,1	15,4
51	52,4	-48	1,0708	88,0	12,0	0,0	0,0	0,0
52	28,4	15	12,31	50,5	25,7	23,8	0,0	23,8
53	29,6	12	9,0462	54,0	24,0	22,0	0,0	22,0
54	25,9	-15	16,6700	48,6	22,2	26,6	2,6	29,2
55	25,6	-54	15,2000	58,6	23,4	18,0	0,0	18,0
56	30,1	-27	7,3240	52,7	33,6	12,6	1,1	13,7
57	29,6	12	10,3757	60,3	22,7	17,0	0,0	17,0
58	29,4	6	10,6200	58,3	21,9	19,8	0,0	19,8
59	29,0	-30	8,2140	52,1	27,2	18,7	2,0	20,7
60	28,7	-42	9,6290	57,0	25,5	15,3	2,2	17,5
61	31,7	15	6,8850	57,1	26,2	14,7	2,0	16,7
62	29,3	12	10,8100	53,1	25,6	21,3	0,0	21,3
63	26,8	9	21,2	53,8	24,1	22,1	0,0	22,1
64	23,6	21	62,5800	40,4	16,2	42,8	0,7	43,4
65	28,4	9	12,7000	51,8	21,9	25,6	0,8	26,3
66	41,0	-6	2,2380	80,8	12,8	6,4	0,0	6,4
67	17,6	-15	146	38,3	31,4	26,0	4,2	30,3
68	29,0	6	11,03	52,1	30,1	17,8	0,0	17,8
69	30,8	6	8,0500	62,6	18,7	18,8	0,0	18,7
70	28,8	12	14,32	-	-	-	-	-

Tabela 8. Propriedades físico-químicas selecionadas para os óleos estudados (continua).

Amostra	API	PF máximo (°C)	Viscosidade a 50°C (mm ² /s)	SARA (%)				
				Saturados	Aromáticos	Resinas	Asfaltenos	Polares
71	29,6	-15	12,46	55,6	19,2	24,6	0,5	25,2
72	28,3	9	12,35	58,0	25,9	15,4	0,7	16,1
73	28,1	-24	14,2800	59,7	23,4	16,1	0,7	16,9
74	26,2	9	17,98	52,6	26,4	19,9	1,1	21,0
75	17,7	-21	123,4000	41,3	34,5	22,6	1,6	24,2
76	26,9	-6	14,5900	53,2	27,3	16,9	2,6	19,5
77	47,6	-18	1,3116	82,7	17,3	0,0	0,0	0,0
78	22,1	-33	34,5300	41,3	31,3	26,5	1,0	27,4
79	30,4	-33	5,7620	54,5	26,3	18,4	0,8	19,2
80	28,1	-33	13,8800	59,4	21,0	19,6	0,0	19,6
81	21,9	-33	26,5300	45,2	33,5	19,5	1,2	21,3
82	28,8	12	12,0000	55,3	27,4	-	-	17,3
83	29,7	3	8,8760	57,7	26,1	15,4	0,8	16,2
84	36,1	39	14,3900	77,5	7,0	15,4	0,1	15,5
85	30,2	9	7,9310	57,5	24,1	17,9	0,5	18,4
86	32,3	-33	5,5189	64,9	21,9	13,2	0,0	13,2
87	22,2	-39	24,9200	49,8	29,2	19,6	1,4	21,0
88	31,4	-12	7,0940	66,2	20,2	13,6	0,0	13,6
89	30,5	-24	9,4420	57,7	27,3	15,0	0,0	15,0
90	30,9	-3	8,1660	59,0	24,4	16,5	0,1	16,6
91	28,1	12	13,2600	59,2	23,5	16,4	1,0	17,3
92	22,6	-36	22,5900	48,1	33,3	17,8	0,7	18,6
93	30,2	9	8,5320	58,5	23,6	17,4	0,5	17,9
94	17,7	-24	162,4000	36,7	27,6	33,4	2,2	35,7
96	29,1	-27	8,7040	53,9	29,2	16,0	0,9	16,9
98	22,2	-42	23,1900	45,1	32,6	21,0	1,0	22,3
99	13,1	9	1379	36,6	26,3	30,7	6,4	37,1
100	37,8	-18	3,2680	74,1	13,9	11,2	0,8	12,0
101	26,8	9	21,2000	53,8	24,1	22,1	0,0	22,1
102	30,2	6	8,9340	63,1	18,2	18,6	0,1	18,7
103	27,9	-3	9,1820	56,3	25,2	16,9	1,6	18,5
104	49,4	-	1,1307	87,0	13,0	0,0	0,0	0,0
105	49,6	-	1,1045	86,4	13,4	-	-	0,2
106	52,0	-60	0,7003	90,6	9,4	0,0	0,0	0,0
107	53,3	-60	0,6873	89,2	10,8	0,0	0,0	0,0
108	54,0	-60	0,7028	88,4	11,6	0,0	0,0	0,0
109	46,4	-21	1,6018	80,8	8,5	10,7	0,0	10,7
110	28,0	3	13,1500	59,8	24,2	16,0	0,0	16,0
112	19,4	-35	81,6700	40,2	33,3	23,4	3,1	26,5
115	36,7	15	4,6460	76,6	13,5	9,3	0,6	9,9
116	48,5	-	1,1860	83,0	17,0	0,0	0,0	0,0
117	29,8	6	9,8340	38,0	27,3	33,4	1,3	34,7

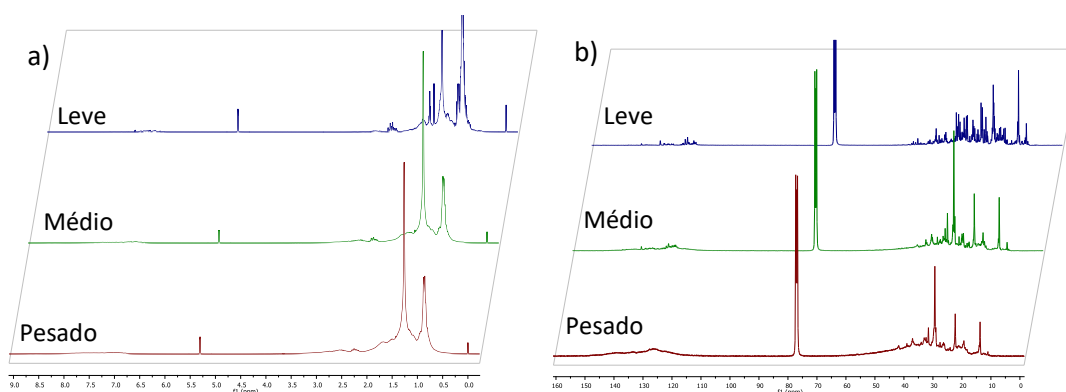
Tabela 8. Propriedades físico-químicas selecionadas para os óleos estudados (continua).

Amostra	API	PF máximo (°C)	Viscosidade a 50°C (mm ² /s)	SARA (%)				
				Saturados	Aromáticos	Resinas	Asfaltenos	Polares
118	41,0	-6	2,2380	80,8	12,8	6,4	0,0	6,4
120	53,7	-	0,9171	87,9	12,1	0,0	0,0	0,0
121	28,5	12	12,1890	51,3	29,1	19,7	0,0	19,6
122	29,0	6	11,0300	52,1	30,1	17,8	0,0	17,8
123	32,6	-1	7,3542	72,1	11,2	15,7	1,0	16,7
124	26,7	-12	14,7700	56,0	26,9	16,0	1,1	17,1
125	28,4	9	12,7000	51,8	21,9	25,6	0,8	26,3
126	29,7	3	8,8760	57,7	26,1	15,4	0,8	16,2
127	17,1	-18	250,4000	-	-	-	1,6	-
128	30,6	-33	6,0800	59,7	26,2	13,5	0,6	14,1
129	45,6	-39	1,4182	-	-	-	-	-
130	34,4	-6	5,1420	66,4	21,0	11,3	1,3	12,6
131	23,6	12	61,6000	49,2	24,5	24,2	2,1	26,3
132	28,7	3	11,8900	62,3	20,6	17,1	0,0	17,1
133	19,8	-36	56,3500	-	-	-	0,8	-
134	38,5	9	3,4130	82,0	17,8	0,2	0,0	0,2
135	29,5	6	9,0770	64,4	19,8	-	-	15,8
136	19,0	-30	90,9300	65,9	33,6	-	-	0,5
137	28,3	9	13,9100	51,5	21,7	25,9	1,0	26,8
138	13,5	-15	271,5000	-	-	-	10,3	-
139	28,8	9	12,3600	50,8	22,9	25,6	0,7	26,3
140	31,9	3	6,9950	60,0	24,1	15,8	0,1	15,9
141	13,1	12	5849,0	29,2	23,6	41,2	6,0	47,2
142	29,8	12	10,6048	54,7	21,4	24,0	0,0	23,9
143	28,5	9	12,8600	53,4	24,7	21,3	0,6	21,9
144	28,9	3	13,9000	52,0	20,5	26,7	0,8	27,5
145	18,3	-30	90,3100	39,4	31,5	26,4	2,6	29,1
146	20,6	-42	37,8100	-	-	-	3,1	-
147	31,3	-9	8,2780	57,5	23,7	18,8	0,0	18,8
148	30,7	-3	8,3990	58,4	20,5	20,6	0,5	21,1
149	26,9	-9	13,9400	48,2	26,2	24,0	1,5	25,6
150	30,6	-9	7,4860	63,9	21,6	14,4	0,1	14,5
151	32,2	9	6,3130	60,9	21,8	17,3	0,0	17,3
152	29,3	-27	9,3120	56,6	22,7	19,8	0,8	20,7
153	28,4	-18	11,8300	51,0	21,4	26,7	0,9	27,6
154	29,0	6	19,3300	48,6	24,6	24,8	2,0	26,8
155	27,6	9	15,1500	50,5	25,8	22,8	0,9	23,7
156	18,4	-15	101,0000	44,9	32,2	-	-	22,9
157	30,3	-12	8,2100	64,1	17,8	17,2	0,9	18,1
158	29,5	-3	9,6280	57,1	22,6	20,3	0,0	20,3
159	27,6	6	18,2400	48,5	23,4	-	-	28,1
160	25,9	12	26,1600	46,8	24,0	-	-	29,2

Tabela 8. Propriedades físico-químicas selecionadas para os óleos estudados (conclusão).

Amostra	API	PF máximo (°C)	Viscosidade a 50°C (mm ² /s)	SARA (%)				
				Saturados	Aromáticos	Resinas	Asfaltenos	Polares
161	27,7	12	14,9400	52,3	24,5	22,5	0,7	23,2
162	28,5	6	13,2500	50,7	24,9	-	-	24,4
163	30,6	0	7,8760	-	-	-	-	-
164	15,2	-	1203,00	-	-	-	-	-

Os espectros de RMN de ¹H e ¹³C dos petróleos brutos foram registrados, e como exemplo, os espectros de três óleos típicos com densidades e viscosidades diferentes, são mostrados na Figura 8. E percebe-se que os espectros dos petróleos analisados diferem sutilmente, uma vez que apresenta inúmeros sinais.

**Figura 8.** Espectros de três amostras de petróleo - a) de ¹H e b) de ¹³C

Os deslocamentos químicos dos principais grupos funcionais das moléculas constituintes foram reportados, de acordo com os dois artigos utilizados como base para esse assinalamento.

É comum e necessário muitas vezes manipular a amostra, diluindo-a em um solvente por exemplo, a fim de garantir uma melhor resolução espectral, porém essa prática quando aplicada a amostras complexas, compromete a repetitividade dos resultados. Na verdade, qualquer variação observada nos dados de RMN devem estar relacionados com a composição intrínseca da amostra, isto é, alterações nos deslocamentos químicos ou linha de base não devem ser devido a fatores externos, tais como o preparo da amostra.³ Para solucionar este problema, neste trabalho as variáveis utilizadas são as áreas de integração dos constituintes do petróleo, para que alterações no deslocamento químico não interfiram nos resultados, uma vez que dessa forma leva-se em consideração intervalos de frequência e não valores

individuais. Com isso, garantimos que as variações ocorridas entre os espectros se deve a modificações no ambiente químico do hidrogênio e carbono.

Na Tabela 9, são especificados o número de amostras utilizadas na construção dos modelos PLS, MLR e PLS-DA, para cada conjunto de variáveis e para cada propriedade físico-química estudada, uma vez que, como dito anteriormente, alguns parâmetros não foram avaliados em todas as amostras e as análises de RMN de ^1H , foram feitas em triplicata para a maioria das amostras, o que não ocorreu nas análises de RMN de ^{13}C que apenas três amostras foram feitas em triplicata.

Tabela 9. Número de amostras utilizado na construção dos modelos para cada propriedade físico-química estudada.

Conjuntos de Variáveis	Número de amostras por propriedade físico-química				
	API	Ponto de Fluidez máximo	Viscosidade a 50°C	SARA	
Hasan (1983) ¹⁶	RMN de ^1H	327	317	315	291
	RMN de ^{13}C	155	149	148	140
	RMN de ^1H e ^{13}C	132	128	126	120
Poveda (2012) ¹⁵	RMN de ^1H	327	317	315	291
	RMN de ^{13}C	155	149	148	140
	RMN de ^1H e ^{13}C	132	128	126	120

Os modelos PLS, MLR e PLS-DA foram então construídos e, neste trabalho, foram testados inclusive os mesmos tipos de pré-processamento dos dados.

Exemplificando, para as técnicas avaliadas PLS e PLS-DA, a escolha do melhor número de VL's dos modelos construídos foi baseada na avaliação do RMSECV. O número apropriado de VL's 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. Como exemplo, na Figura 9 é mostrado o gráfico do número de RMSECV vs variáveis latentes.

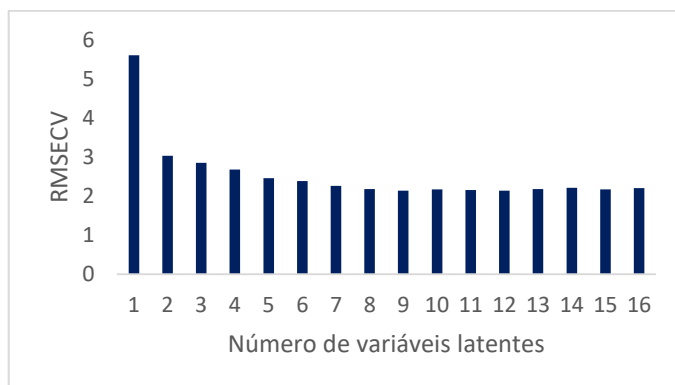


Figura 9. Gráfico utilizado para a seleção das VL's em função do RMSECV para a previsão da densidade por RMN de ^1H e ^{13}C segundo Poveda¹⁵ com pré-tratamento do tipo MSC.

Com base no gráfico da Figura 9, pode-se utilizar 5 VL's, pois não há uma diferença significativa no valor de RMSECV.

Após o pré-processamento dos dados, estas novas matrizes foram utilizadas na elaboração dos modelos de calibração. Os gráficos resultantes da utilização das ferramentas de regressão multivariada, não são mostrados para evitar excesso de informações, sendo mostrado somente os gráficos dos melhores modelos obtidos para cada propriedade físico-química. Os resultados dos parâmetros estatísticos dos modelos PLS e MLR construídos para as propriedades estudadas são apresentados nas Tabelas 10, 11 e 12, onde os melhores modelos foram destacados em cinza. Os modelos MLR apresentaram menores erros de previsão (RMSEP) e maiores coeficientes de determinação (R^2_p) que os modelos PLS.

A Tabela 10 mostra os parâmetros estatísticos calculados para todos os modelos construídos para as propriedades gravidade API e ponto de fluidez. Na modelagem da propriedade gravidade API, o melhor modelo PLS foi construído utilizando os parâmetros moleculares segundo Hasan dos espectros de ^1H e ^{13}C , com pré-processamento do tipo MSC e 6 variáveis latentes, enquanto o modelo MLR utilizou os parâmetros moleculares segundo Poveda dos espectros de ^1H e ^{13}C . Os modelos PLS e MLR apresentaram RMSEP's de 1,26 e 1,13 respectivamente. A boa relação entre os valores previstos pelo modelo MLR e medidos pelo método de referência é observado no coeficiente de determinação: 0,8779. Para a propriedade ponto de fluidez, os melhores modelos PLS e MLR foram construídos utilizando os parâmetros moleculares do artigo Hasan dos espectros de ^1H e ^{13}C , no caso do PLS com pré-

processamento do tipo SNV e 2 variáveis latentes. Os modelos PLS e MLR apresentaram RMSEP's de 14,5 e 14,2 respectivamente.

Tabela 10. Resultados dos modelos PLS e MLR na determinação das propriedades físico-químicas gravidade API e ponto de fluidez de petróleo.

Conjunto de variáveis	Técnica de regressão	Pré-processamento	Gravidade API			Ponto de Fluidez		
			VL	RMSEP	R ² _p	VL	RMSEP	R ² _p
H_Hasan	PLS	msc	3	2,06	0,8998	2	14,6	0,4946
		-	2	2,12	0,8939	2	14,7	0,4974
		auto	3	2,06	0,8996	2	14,7	0,4970
		snv	2	2,30	0,8767	3	15,0	0,4919
		norm	2	2,21	0,8858	2	15,1	0,4800
	MLR	-	-	2,02	0,8996	-	14,9	0,4789
H_Poveda	PLS	msc	4	1,85	0,906	3	15,3	0,4878
		-	4	1,81	0,9067	4	15,3	0,5079
		auto	4	1,68	0,9207	3	15,3	0,5177
		snv	2	2,40	0,8442	3	15,2	0,4960
		norm	2	2,23	0,8625	3	15,4	0,4892
	MLR	-	-	1,75	0,9093	-	14,6	0,5459
C_Hasan	PLS	msc	4	2,79	0,8025	4	21,3	0,0245
		-	4	2,78	0,8122	3	21,0	0,0199
		auto	4	2,79	0,8104	3	21,1	0,0128
		snv	4	2,97	0,8054	3	18,5	0,3412
		norm	5	2,88	0,8165	3	18,0	0,3911
	MLR	-	-	2,70	0,7975	-	20,1	0,0070
C_Poveda	PLS	msc	4	2,25	0,8759	2	16,2	0,2171
		-	1	4,02	0,6234	2	16,5	0,1912
		auto	1	2,76	0,7995	2	16,4	0,2124
		snv	4	2,24	0,8765	2	16,1	0,2185
		norm	4	2,17	0,8816	2	16,2	0,2067
	MLR	-	-	2,06	0,8804	-	15,5	0,2200
HeC_Hasan	PLS	msc	6	1,26	0,9449	2	14,5	0,5549
		-	7	1,47	0,9386	2	14,7	0,5555
		auto	7	1,51	0,9348	2	14,8	0,5518
		snv	7	1,33	0,9416	2	14,5	0,5554
		norm	7	1,44	0,9328	2	14,6	0,5564
	MLR	-	-	1,45	0,9216	-	14,2	0,5188
HeC_Poveda	PLS	msc	5	1,46	0,8113	2	15,6	0,4085
		-	1	2,14	0,6261	1	15,6	0,4215
		auto	1	2,05	0,7411	2	16,4	0,4003
		snv	5	1,43	0,818	2	15,5	0,4086
		norm	5	1,36	0,843	2	15,6	0,4153
	MLR	-	-	1,13	0,8779	-	15,3	0,2506

A viscosidade cinemática varia exponencialmente com a temperatura de medição do óleo. Para petróleos pesados e extrapesados, normalmente esta propriedade é medida em elevadas temperaturas (acima de 50°C) para posteriormente ser estimada a 40°C. Na modelagem da propriedade viscosidade cinemática a 50°C (Tabela 11), os modelos foram construídos utilizando a viscosidade depois de submetida a duas operações logarítmicas ($\log [\log (\text{viscosidade a } 50^\circ\text{C})]$). Os melhores modelos PLS e MLR foram construídos utilizando os parâmetros moleculares segundo Hasan dos espectros de ^1H e ^{13}C , no caso do PLS com pré-processamento do tipo SNV e 8 variáveis latentes. Os modelos PLS e MLR apresentaram RMSEP's de 13,93 e 13,61 $\text{mm}^2.\text{s}^{-1}$ respectivamente.

Tabela 11. Resultados dos modelos PLS e MLR na determinação da propriedade físico-química viscosidade cinemática a 50°C de petróleo (continua).

Conjunto de variáveis	Técnica de regressão	Pré-processamento	Densidade			
			VL	RMSEP ($\log(\log(\text{visc.}))$)	RMSEP (visc.)	R ² _p
H_Hasan	PLS	msc	3	0,1100	19,420	0,7205
		-	3	0,1083	19,196	0,7279
		auto	2	0,1076	19,105	0,7284
		snv	4	0,1077	19,118	0,7337
		norm	3	0,1091	19,301	0,7239
	MLR	-	-	2,0658	-	0,8978
H_Poveda	PLS	msc	5	0,1154	20,154	0,7331
		-	5	0,1105	19,486	0,7468
		auto	5	0,1050	18,772	0,7645
		snv	5	0,1062	18,925	0,7518
		norm	6	0,1120	19,688	0,7379
	MLR	-	-	0,1067	18,989	0,7484
C_Hasan	PLS	msc	4	0,0746	15,396	0,7552
		-	4	0,0770	15,631	0,7312
		auto	4	0,0762	15,552	0,7373
		snv	4	0,0834	16,282	0,6908
		norm	4	0,0804	15,972	0,7082
	MLR	-	-	2,3751	-	0,8172
C_Poveda	PLS	msc	4	0,0905	17,048	0,6737
		-	4	0,0898	16,971	0,6774
		auto	1	0,1178	20,493	0,676
		snv	4	0,0903	17,026	0,6746
		norm	3	0,0900	16,993	0,678
	MLR	-	-	0,0851	16,461	0,6754

Tabela 11. Resultados dos modelos PLS e MLR na determinação da propriedade físico-química viscosidade cinemática a 50°C de petróleo (conclusão).

Conjunto de variáveis	Técnica de regressão	Pré-processamento	Densidade			
			VL	RMSEP (log(log(visc.)))	RMSEP (visc.)	R ² _p
HeC_ Hasan	PLS	msc	8	0,0624	14,273	0,5573
		-	7	0,0604	14,100	0,5584
		auto	1	0,0881	16,784	0,4839
		snv	8	0,0584	13,929	0,5834
		norm	7	0,0602	14,083	0,5605
	MLR	-	-	0,0545	13,605	0,5715
HeC_ Poveda	PLS	msc	8	0,0853	16,483	0,6173
		-	1	0,0888	16,861	0,5584
		auto	1	0,0859	16,547	0,6392
		snv	8	0,0853	16,483	0,6196
		norm	6	0,0802	15,952	0,6193
	MLR	-	-	1,1251	-	0,9567

Para a modelagem do parâmetro SARA (Tabela 12), foram construídos modelos separadamente para cada classe de compostos, saturados, aromáticos, resinas e asfaltenos.

O modelo de calibração para o teor de saturados utilizando os parâmetros moleculares segundo Hasan dos espectros de ¹³C apresenta erros de 5,2% para PLS com 4 variáveis latentes e nenhum pré-tratamento, e 5,1% para MLR. Na modelagem do teor de aromáticos utilizando os parâmetros moleculares segundo Poveda dos espectros de ¹H, os erros foram de 2,9% para ambas as técnicas testadas, uma vez que para PLS utilizou-se 3 variáveis latentes e o autoescalamento para pré-processar os dados. Os melhores modelos para o teor de resinas, assim como ocorreu no teor de saturados utilizaram parâmetros moleculares segundo Hasan dos espectros de ¹³C, apresentando erros de 4,8% para PLS com 5 variáveis latentes e normalização como técnica de pré-tratamento, e 4,6% para MLR. Finalmente, para o teor de asfaltenos, o melhor modelo PLS foi construído utilizando os parâmetros moleculares segundo Hasan dos espectros de ¹³C, com normalização dos dados como forma de pré-processamento e 2 variáveis latentes, enquanto o modelo MLR utilizou os parâmetros moleculares segundo Poveda dos espectros de ¹H e ¹³C. Os modelos PLS e MLR apresentaram RMSEP's de 0,54 e 0,57% respectivamente.

Tabela 12. Resultados dos modelos PLS e MLR na determinação do SARA de petróleo.

Conjunto de variáveis	Técnica de regressão	Pré-processamento	SARA											
			Saturados			Aromáticos			Resinas			Asfaltenos		
			VL	RMSEP	R ² _P	VL	RMSEP	R ² _P	VL	RMSEP	R ² _P	VL	RMSEP	R ² _P
H_Hasan	PLS	msc	2	6,9	0,6006	2	3,1	0,5870	2	6,0	0,3306	3	1,11	0,3826
		-	2	6,5	0,6329	2	3,1	0,6004	2	5,9	0,3492	1	1,22	0,2097
		auto	2	6,4	0,6500	2	3,1	0,5906	1	5,8	0,3492	1	1,17	0,2963
		Snv	2	6,9	0,6013	3	3,2	0,5705	3	5,9	0,3548	3	1,11	0,3814
		norm	2	6,7	0,6163	3	3,2	0,5842	2	5,9	0,3397	3	1,11	0,3842
	MLR	-	-	6,3	0,6440	-	3,0	0,6005	-	5,7	0,3525	-	1,09	0,3805
H_Poveda	PLS	msc	4	6,1	0,6473	4	3,2	0,5376	5	5,8	0,3614	4	1,09	0,4074
		-	4	5,7	0,6675	3	3,1	0,5558	5	5,6	0,3750	5	1,11	0,3927
		auto	3	5,9	0,6480	3	2,9	0,5855	4	5,7	0,3632	3	1,03	0,4957
		Snv	4	6,0	0,6484	3	3,2	0,5167	5	5,7	0,3695	4	1,11	0,3706
		norm	4	5,9	0,6573	3	3,1	0,5433	4	5,8	0,3427	4	1,11	0,3668
	MLR	-	-	5,6	0,6717	-	2,9	0,5844	-	5,5	0,3677	-	1,00	0,4804
C_Hasan	PLS	msc	4	5,3	0,6932	3	3,4	0,4447	4	5,0	0,4693	3	0,65	0,4181
		-	4	5,2	0,7049	1	3,5	0,3904	4	4,9	0,4850	5	0,76	0,3681
		auto	4	5,2	0,7033	1	3,5	0,3878	4	4,9	0,4845	5	0,76	0,3681
		Snv	4	5,3	0,6934	3	3,3	0,4572	5	4,8	0,5004	4	0,57	0,5175
		norm	5	5,2	0,7107	3	3,4	0,4464	5	4,9	0,4963	2	0,54	0,5347
	MLR	-	-	5,1	0,6785	-	3,1	0,4791	-	4,6	0,4827	-	0,70	0,3681
C_Poveda	PLS	msc	4	6,6	0,5321	1	4,1	0,2430	2	6,4	0,2499	5	0,82	0,1402
		-	4	6,1	0,5919	2	4,1	0,2830	3	6,1	0,3225	3	0,81	0,1726
		auto	1	6,4	0,5247	1	3,9	0,3230	2	6,2	0,2868	1	0,85	0,1101
		Snv	4	6,5	0,5333	1	4,1	0,2430	2	6,4	0,2496	7	0,74	0,2295
		norm	1	6,2	0,5772	1	4,1	0,2583	3	6,2	0,3159	7	0,72	0,2429
	MLR	-	-	5,6	0,6098	-	3,6	0,3851	-	5,9	0,3137	-	0,80	0,1295
HeC_Hasan	PLS	msc	1	6,0	0,5723	3	3,1	0,5761	6	6,0	0,2827	2	0,69	0,3316
		-	3	6,5	0,5327	3	3,1	0,5760	7	6,1	0,2719	2	0,70	0,3521
		auto	3	6,4	0,5488	3	3,1	0,5432	5	5,8	0,2871	2	0,69	0,3742
		Snv	3	6,6	0,5173	4	3,2	0,5196	8	6,2	0,2822	2	0,69	0,3321
		norm	3	6,5	0,5264	3	3,1	0,5758	3	5,8	0,2627	4	0,66	0,4540
	RLM	-	-	5,8	0,5766	-	3,3	0,4232	-	5,4	0,2829	-	0,82	0,3574
HeC_Poveda	PLS	msc	3	6,7	0,3061	3	3,3	0,3659	3	6,0	0,0745	2	0,75	0,1861
		-	1	6,7	0,2811	4	3,4	0,3475	3	6,0	0,0904	2	0,75	0,2302
		auto	1	6,6	0,3133	5	3,7	0,2354	4	5,9	0,1357	1	0,71	0,3223
		Snv	3	6,7	0,3063	5	3,5	0,3111	3	6,0	0,0751	2	0,75	0,1859
		norm	3	6,6	0,3234	5	3,5	0,3256	3	6,0	0,0824	2	0,73	0,2169
	MLR	-	-	6,2	0,3349	-	3,5	0,2139	-	5,8	0,0838	-	0,57	0,5486

O teste-t para erros sistemáticos foi aplicado aos resíduos dos modelos construídos e não apresentaram resultados significativos ao nível de significância de 5%. Nenhum modelo apresentado nas Tabelas 10, 11 e 12 apresentou erros sistemáticos. Os erros de tendência nos resíduos de calibração ou de previsão, foram analisados e alguns modelos apresentaram valores de p menores do que o nível de significância adotado de 0,05. Nas propriedades de ponto de fluidez e viscosidade cinemática, os modelos escolhidos apresentaram erros de tendência, para isso foram selecionados os melhores modelos dessas propriedades que não eram tendenciosos, os quais foram destacados nas Tabelas 10, 11 e 12 na cor amarela. Na modelagem do ponto de fluidez houve problema de tendência no modelo PLS, logo o modelo escolhido foi substituído por aquele que utilizou os parâmetros moleculares do artigo Hasan dos espectros de ^1H , com um erro RMSEP de $14,6^\circ\text{C}$. para a viscosidade cinemática a 50°C ambos os modelos, PLS e MLR, apresentaram erros de tendência, afim de solucionar esse problema foram selecionados novos modelos, para a técnica PLS utilizando os parâmetros moleculares do artigo Hasan dos espectros de ^{13}C , com pré-processamento do tipo MSC e 4 variáveis latentes, enquanto o modelo MLR utilizou os parâmetros moleculares do artigo Poveda dos espectros de ^{13}C . Os modelos PLS e MLR apresentaram RMSEP's de 15,40 e 16,46 respectivamente.

Vale destacar que alguns modelos selecionados geraram resultados com altos valores de RMSEP e baixos valores de R^2 . Essas peculiaridades poderão também ser facilmente visualizadas adiante, através da observação dessa relação entre os valores medidos e previstos, na discussão dos melhores modelos para cada propriedade estudada. A partir de agora cada uma das propriedades serão discutidas separadamente, utilizando os melhores modelos obtidos para cada uma delas e validando-os.

- *Gravidade API*

Para os modelos criados, a boa relação entre os valores previstos por métodos quimiométricos de calibração e medidos pelo método de referência, para as amostras de calibração e as amostras de teste, podem ser observados na Figura 10.

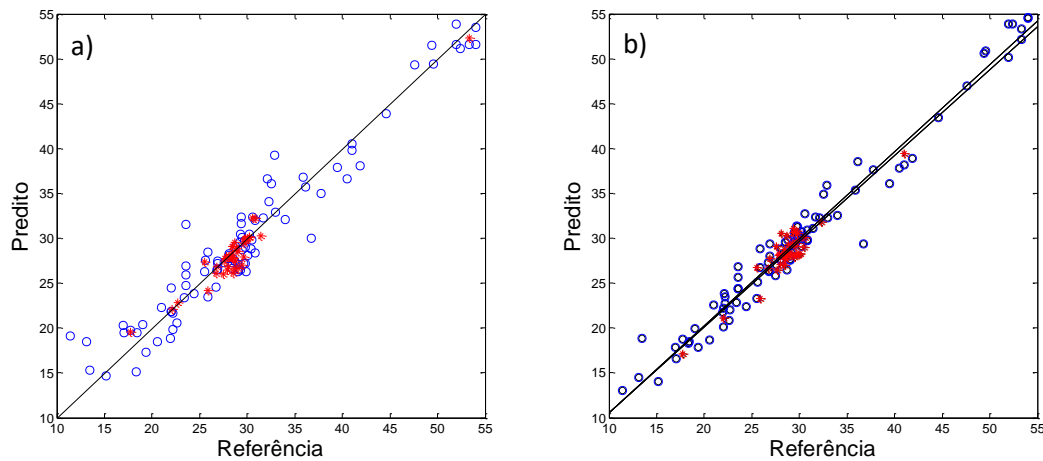


Figura 10. Relação entre valores de referência (o) e previstos (*) da propriedade gravidade API dos modelos criados por (a) PLS e (b) MLR.

Os parâmetros estatísticos para os modelos PLS e MLR selecionados para a gravidade API são mostrados na Tabela 13.

Tabela 13. Valores calculados para a validação dos modelos selecionados para gravidade API, por PLS e MLR.

Parâmetros	PLS	MLR
VL's	6	-
RMSEC	2,54	1,69
$R^2_{\text{Calibração}}$	0,9346	0,9704
RMSEP	1,26	1,13
RMSEP%	4,44	3,98
$R^2_{\text{Previsão}}$	0,9449	0,8779
$\sigma_{\text{previsão}}$	1,67	1,58

Na indústria do petróleo é aceito variação máxima de 1,1 na gravidade API para óleos de mesma procedência. Neste trabalho, foram utilizados óleos da bacia sedimentar da costa brasileira. Devido ao uso de amostras muito diferentes, uma melhor comparação da exatidão dos modelos pode ser avaliada pelo RMSE %, mostrado na equação 33. O modelo PLS, o RMSEP de 1,26 representa 4,44% do erro médio de

previsão, enquanto que o RMSEP de 1,13 obtido pelo modelo MLR representa apenas 3,98%.

Comparando-se, ainda, os valores de reprodutibilidade dos respectivos métodos laboratoriais (Tabela 6) para determinação da gravidade API e os valores de RMSEP e desvio padrão de previsão calculados (Tabela 13), os valores de desvio padrão de previsão calculados para o modelo PLS indicam que parte do erro encontrado é devido ao modelo estatístico. Essa comparação entre os valores de RMSEP, reprodutibilidade e desvio padrão de previsão também pode ser visualizada na Figura 11.

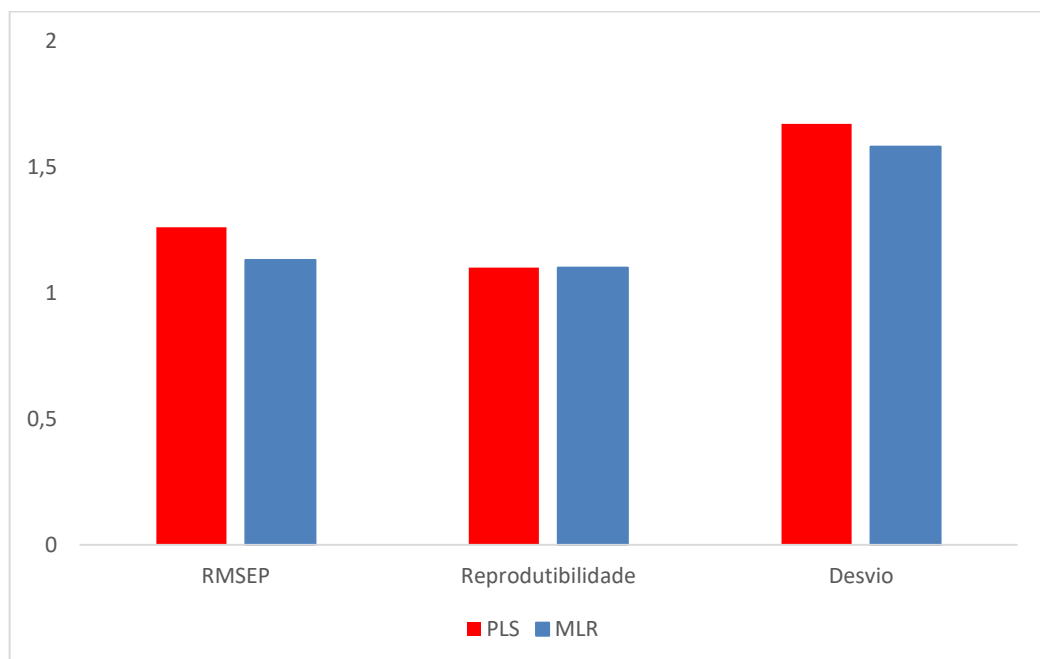


Figura 11. Relação entre os erros experimentais e calculados para a gravidade API.

Os resultados mostram que há boa exatidão para a previsão da gravidade API, uma vez que os valores de referência e previstos não apresentaram diferença significativa.

As variáveis mais significativas para a construção dos modelos PLS e MLR escolhidos podem ser identificadas pelos pesos ou pelos coeficientes do modelo de regressão, os quais podem ser observados na Tabela 14, para ambos modelos:

Tabela 14. Coeficientes de regressão do modelo PLS, com 6 variáveis latentes e pré-processamento MSC, e coeficientes do modelo MLR, na determinação da gravidade API.

	MLR		PLS	
	Parâmetros (região do espectro - ppm)	Coefficientes	Parâmetros (região do espectro - ppm)	Coefficientes
H_ _{Poveda}	0,10 - 1,00	6,92	6,00 - 9,00	-3,81
	1,00 - 1,50	6,25	0,50 - 4,00	1,57
	1,50 - 2,00	6,38	0,50 - 1,00	-1,94
	2,00 - 4,50	5,50	1,00 - 2,00	-2,51
	6,00 - 7,20	5,26	2,00 - 4,00	-3,57
	7,20 - 9,00	3,79	0,00 - 70,00	0,92
C_ _{Poveda}	9,00 - 12,00	-53,24	110,00 - 160,00	-4,82
	3,00 - 18,50	45,93	110,00 - 130,00	6,00
	18,50 - 21,50	44,10	128,00 - 136,00	13,86
	21,50 - 50,00	45,18	129,00 - 137,00	-9,02
	50,00 - 60,00	41,07	137,00 - 160,00	3,32
	60,00 - 70,00	34,14		
	85,00 - 129,20	46,00		
	129,20 - 137,00	45,35		
137,00 - 160,00	42,92			
190,00 - 220,00	47,26			

De acordo com a Tabela 14, as regiões do espectro mais importantes na construção dos modelos para a gravidade API, estão principalmente nos espectros de RMN de ¹³C para os dois casos. No modelo MLR todo o espectro de carbono tem grande contribuição, além da região de hidrogênios de aldeídos e ácidos carboxílicos. Já no modelo PLS a maior contribuição se deve a região de carbonos aromáticos em junção de anéis e protonados. Em ambos os casos, as variáveis mais importantes nos espectros de ¹³C, se devem aos compostos aromáticos, os quais têm suas quantidades aumentadas com o aumento da densidade dos óleos.

Segundo ANP, utilizando a escala API, o petróleo pode ser classificado em:⁴⁰

- Leve ($^{\circ}\text{API} \geq 31$);
- Médio ($22 \leq ^{\circ}\text{API} < 31$),
- Pesado ($10 \leq ^{\circ}\text{API} < 22$) e
- Extrapesado ($^{\circ}\text{API} < 10$).

A fim de classificar as amostras estudadas nas classes acima, aplicou-se o método de reconhecimento de padrão PLS-DA aos conjuntos de variáveis que apresentaram

melhores resultados para os métodos de calibração multivariada, sendo eles os parâmetros moleculares dos espectros de ^1H e ^{13}C segundo Hasan e Poveda. Testando-se os mesmos tipos de pré-processamentos utilizados na construção dos modelos de regressão. Foram utilizadas 132 amostras de petróleo na modelagem, as quais estão distribuídas nas classes de acordo com a Figura 12.

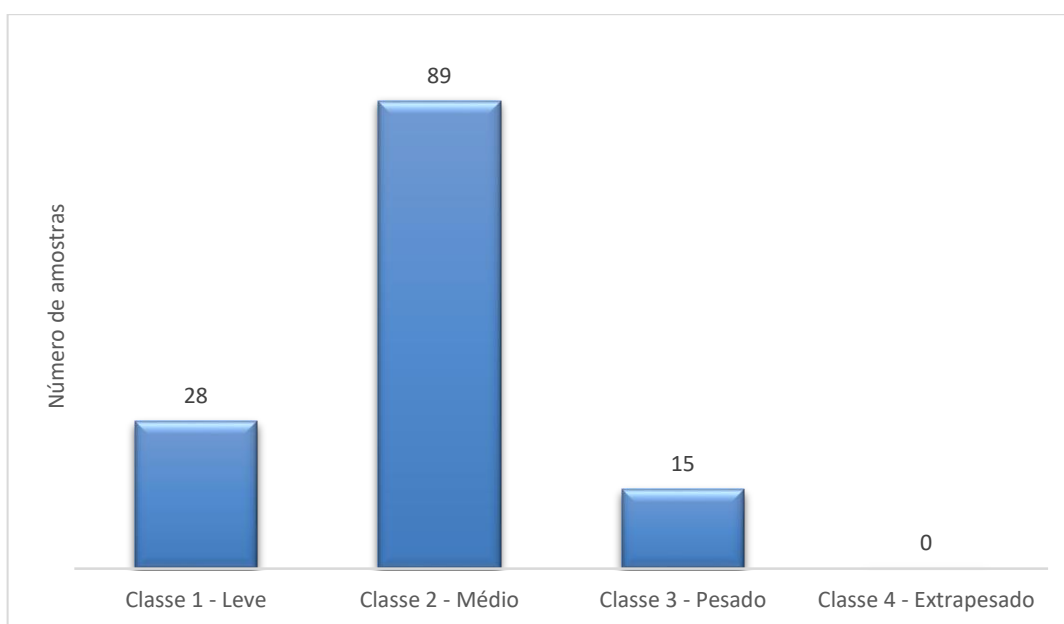


Figura 12. Distribuição das amostras nas classes

Assim como nos modelos PLS, foram selecionados os números ótimos de variáveis latentes a partir do erro de validação cruzada, para este procedimento as amostras foram divididas em 5 grupos pela técnica *venetian blinds*. Definido o número de variáveis latentes seguiu-se para a construção dos modelos PLS-DA.

Os resultados dos modelos construídos podem ser observados na Tabela 15, na qual o melhor modelo foi destacado. O melhor modelo PLS-DA foi construído utilizando os parâmetros moleculares do artigo Poveda dos espectros de ^1H e ^{13}C , com pré-processamento do tipo normalização e 4 variáveis latentes. Apresentando uma taxa de acerto e exatidão de 100%, com 8% de amostras não classificadas.

Tabela 15. Resultados dos modelos PLS-DA para a determinação da classe das amostras de petróleo baseadas na gravidade API

Conjunto de variáveis	Pré-processamento	Parâmetros Estatísticos				
		VL's	Taxa de acerto	Taxa de erro	Exatidão	Não classificadas
HeC_Hasan	msc	3*	0,83	0,17	1,00	0,15
	-	2*	0,83	0,17	1,00	0,15
	auto	3*	0,83	0,17	1,00	0,15
	snv	2	0,83	0,17	1,00	0,18
	norm	2	0,83	0,17	1,00	0,15
HeC_Poveda	msc	2	0,83	0,17	1,00	0,15
	-	2	0,77	0,23	0,94	0,21
	auto	2	0,83	0,17	1,00	0,15
	snv	3	0,79	0,21	0,97	0,13
	norm	4	1,00	0,00	1,00	0,08

As matrizes de confusão obtidas na construção do modelo escolhido e no procedimento de validação cruzada são mostradas na Tabela 16.

Tabela 16. Matrizes de confusão obtidas na construção do modelo (ajuste) e na validação cruzada.

Classes de referência	Classes calculadas			
	Leve	Médio	Pesado	Não classificado
Ajuste				
Leve	12	2	0	6
Médio	0	48	0	14
Pesado	0	1	2	8
Validação Cruzada				
Leve	11	2	0	7
Médio	1	47	4	10
Pesado	0	1	2	8

Os parâmetros de classificação, tais como taxa de acerto e erro, especificidade e sensibilidade das classe, exatidão e prevalências são obtidos a partir das matrizes de confusão, a fim de avaliar o desempenho do modelo construído (Tabela 17).

Tabela 17. Parâmetros estatísticos obtidos no ajuste, validação cruzada e teste para avaliação do modelo de classificação construído.

		Ajuste	Validação Cruzada	Teste	
Modelo	Exatidão	0,95	0,88	1,00	
	Taxa de acerto	0,89	0,86	1,00	
	Taxa de erro	0,11	0,14	0,00	
Classe	Leve	Especificidade	1,00	0,98	1,00
		Sensibilidade	0,86	0,85	1,00
		PPV – leve	1,00	0,92	0,88
		PNV – leve	0,91	0,90	0,97
Médio	Especificidade	0,82	0,81	1,00	
	Sensibilidade	1,00	0,90	1,00	
	PPV – médio	0,94	0,94	1,00	
	PNV – médio	0,87	0,75	1,00	
Pesado	Especificidade	1,00	0,94	1,00	
	Sensibilidade	0,67	0,67	1,00	
	PPV – pesado	1,00	0,33	1,00	
	PNV – pesado	0,91	0,91	0,95	
Não classificadas		0,30	0,27	0,08	

Outra ferramenta relevante para análise e compreensão da técnica de reconhecimento de padrão são os gráficos de *scores* e *loadings*. Nas Figuras 13 e 14, são mostrados os *scores* e *loadings* nas duas primeiras variáveis latentes, respectivamente.

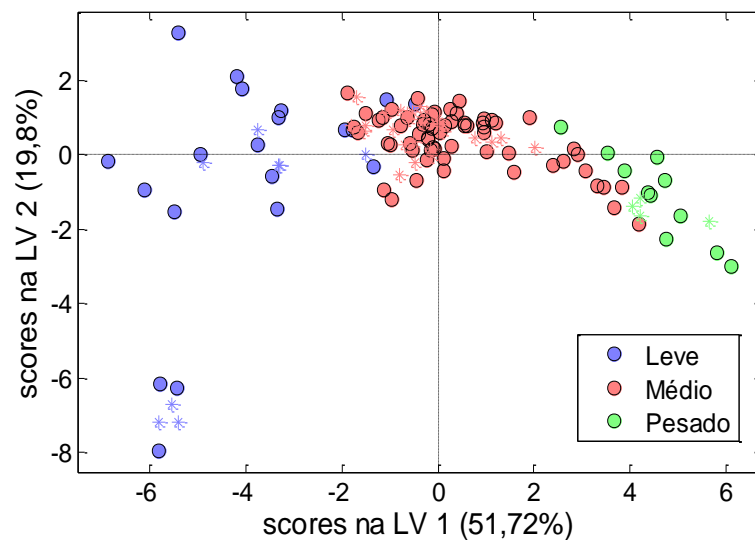


Figura 13. Gráfico de *scores* do modelo selecionado nas duas primeiras variáveis latentes.

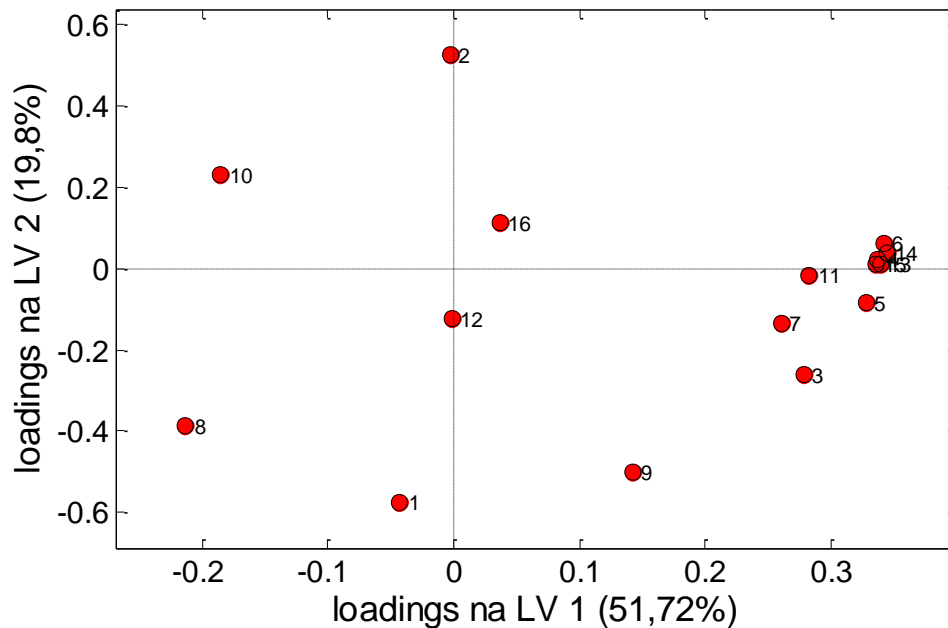


Figura 14. Gráfico de *loadings* do modelo selecionado nas duas primeiras variáveis latentes.

Percebe-se que a primeira variável latente detém 51,72% da variância explicada, enquanto a segunda variável latente explica 19,8% dela, logo a primeira variável latente contém a maior parte da informação. Porém, a variância explicada não é suficiente para afirmar a capacidade de discriminação em classes do modelo. Observando o gráfico de *scores* (Figura 13), pode-se perceber padrões interessantes de distribuição das amostras. Mesmo que haja um certo grau de sobreposição entre as amostras, como ocorre principalmente entre as classes de óleos médios e pesados, as amostras leves têm valores maiores na primeira variável latente e as amostras pesadas valores menores. Essa separação pode ser melhor entendida, quando se observa o gráfico de *loadings* (Figura 14).

Ambas variáveis latentes são responsáveis pela separação das amostras nas três classes, uma vez que todas as variáveis têm efeito na discriminação das classes, algumas com maior e outras com menor contribuição. É interessante destacar que, como a separação se dá de maneira muito acentuada ao longo da primeira variável latente, podemos perceber que as variáveis responsáveis por separar as amostras leves (variáveis que se encontram em valores negativos da primeira VL) são aquelas que correspondem a carbonos alifáticos protonados do tipo CH, CH₂ e CH₃, os quais correspondem aos compostos encontrados em maiores concentrações nos petróleos

leves. Enquanto as variáveis responsáveis pela separação das amostras pesadas (variáveis que se encontram em valores positivos da primeira VL) são aquelas que correspondem basicamente a hidrogênios e carbonos aromáticos ou ligados a anéis aromáticos, compostos estes que têm suas concentrações aumentadas nos óleos pesados em relação aos leves.

- *Ponto de Fluidez*

Os modelos criados para o ponto de fluidez máximo, resultaram em altos valores de RMSEP e baixos valores de R^2 , observando-se, na Figura 15, a relação entre valores medidos e previstos percebe-se que os modelos selecionados para a previsão do ponto de fluidez tiveram desempenho bastante ruim, como todos os valores previstos praticamente constantes, para todo o conjunto testado e com intervalos de confiança bastante extensos.

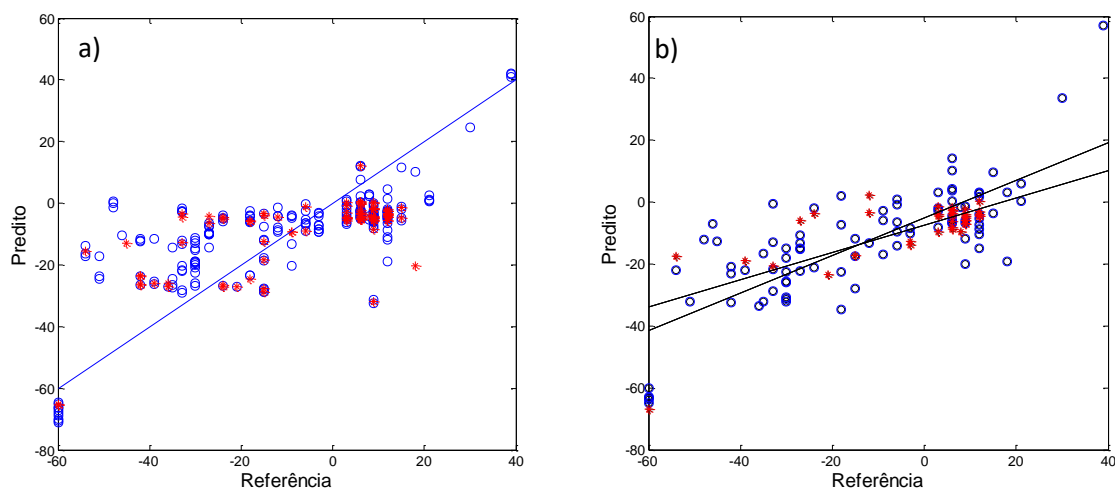


Figura 15. Relação entre valores de referência (o) e previstos (*) da propriedade ponto de fluidez máximo dos modelos criados por (a) PLS e (b) MLR

Os parâmetros estatísticos obtidos para os modelos PLS e MLR selecionados para o ponto de fluidez são mostrados na Tabela 18.

Tabela 18. Valores calculados para a validação dos modelos selecionados para ponto de fluidez máximo, por PLS e MLR.

Parâmetros	PLS	MLR
VL's	2	-
RMSEC	16,1	15,7
R ² Calibração	0,5146	0,5232
RMSEP	14,6	14,9
RMSEP%	170,8	151,1
R ² Previsão	0,4946	0,4789
σ previsão	23,2	23,4

O método de referência ASTM para ponto de fluidez máximo apresenta um intervalo de incerteza de 18°C. Devido ao uso de amostras muito diferentes, uma melhor comparação da exatidão dos modelos foi feita pelo RMSE %, mostrado na equação 33. O modelo PLS, o RMSEP de 14,6 representa 170,8% do erro médio de previsão, enquanto que o RMSEP de 14,9 obtido pelo modelo MLR representa 151,1%, valores esses muito altos, mas que não indicam que o modelo seja inadequado, uma vez que o intervalo de incerteza do método original é bastante alto e os modelos construídos apresentaram erros menores que a ASTM.

Observando-se os valores de reprodutibilidade dos respectivos métodos laboratoriais (Tabela 6) para determinação do ponto de fluidez máximo e os valores de RMSEP e desvio padrão de previsão calculados (Tabela 18), observa-se que o erro adicionado à previsão devido ao modelo estatístico é baixo quando comparado à reprodutibilidade do método padrão. Essa comparação entre os valores de RMSEP, reprodutibilidade e desvio padrão de previsão também pode ser visualizada na Figura 16.

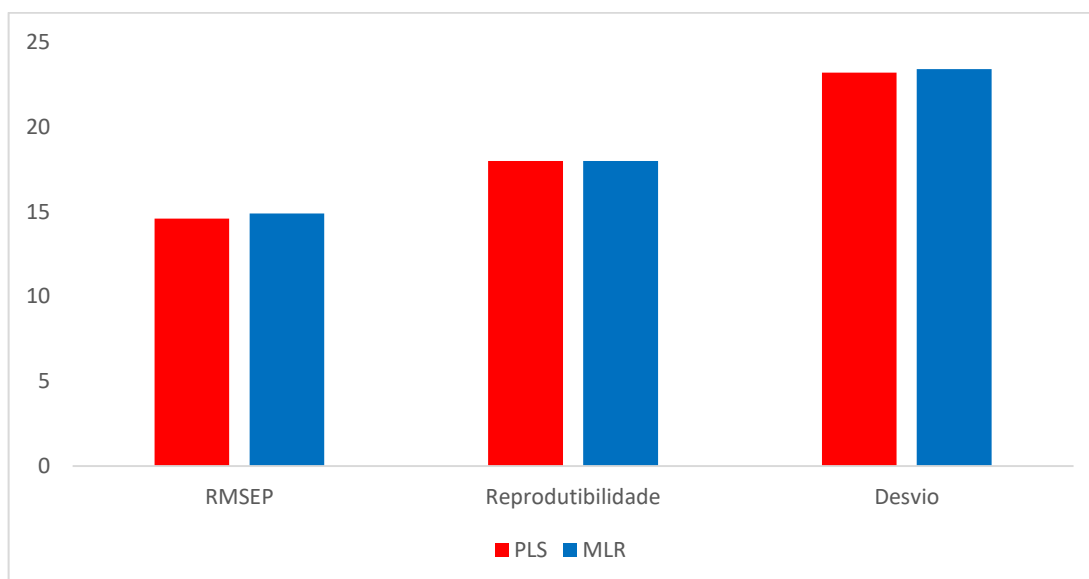


Figura 16. Relação entre os erros experimentais e calculados para o ponto de fluidez máximo.

As variáveis mais significativas para a construção dos modelos PLS e MLR escolhidos podem ser observados na Tabela 19, para ambos modelos:

Tabela 19. Coeficientes de regressão do modelo PLS, com 6 variáveis latentes e pré-processamento MSC, e coeficientes do modelo MLR, na determinação do ponto de fluidez máximo.

	MLR		PLS	
	Parâmetros (região do espectro - ppm)	Coeficientes	Parâmetros (região do espectro - ppm)	Coeficientes
H_Hasan	6,00 - 9,00	-23,10	6,00 - 9,00	0,58
	0,50 - 4,00	-35,31	0,50 - 4,00	-2,30
	0,50 - 1,00	6,09	0,50 - 1,00	-0,66
	1,00 - 2,00	10,62	1,00 - 2,00	4,09
	2,00 - 4,00	4,61	2,00 - 4,00	-1,70
C_Hasan	0,00 - 70,00	26,51		
	110,00 - 160,00	-90,68		
	110,00 - 130,00	119,19		
	128,00 - 136,00	-76,79		
	129,00 - 137,00	176,85		
	137,00 - 160,00	112,69		

De acordo com a Tabela 19, as regiões do espectro mais importantes na construção dos modelos para o ponto de fluidez máximo, estão principalmente nos espectros de RMN de ^{13}C para o modelo MLR, com destaque para os sinais que se encontram nas regiões de carbonos aromáticos ligados a cadeias carbônicas ou carbonos aromáticos protonados. No modelo PLS os sinais nas regiões com sinais de hidrogênios alifáticos

têm maior contribuição, com destaque para hidrogênios parafínicos e naftênicos do tipo CH₂ ligados a sistemas aromáticos na posição β e alquílicos terminais.

Propriedades como densidade, ponto de fluidez e viscosidade estão diretamente relacionadas a composição química do petróleo, além de serem fundamentais na indústria petrolífera, uma vez que influenciam em todo o processamento do óleo.

O ponto de fluidez fornece informações sobre as características de fluxo do óleo em baixas temperaturas, e isso está relacionado a densidade do petróleo, pois petróleos mais pesados terão maior dificuldade de fluir quando comparados a óleos mais leves. Porém essa relação não pode ser visualizada ao se comparar os resultados dessas propriedades para um conjunto de amostras, conforme a Figura 17. Embora, algumas relações gerais da composição do petróleo, possam influenciar no ponto de fluidez, assim como a gravidade API, porém não existe uma relação direta entre elas. Esperava-se que amostras com menores valores de gravidade API (óleos mais pesados) apresentassem maiores valores de ponto de fluidez, mas essa relação não foi observada no gráfico. Comumente, óleos com baixos pontos de fluidez são mais aromáticos e maiores pontos de fluidez caracterizam petróleos parafínicos, regra essa que não é seguida pela densidade. Para facilitar essa discussão foram colocados no gráfico da Figura 16, linhas nos limites de classe para as duas propriedades, gravidade API (já citado anteriormente) e ponto de fluidez, sendo eles:

- Leve ($^{\circ}\text{API} \geq 31$ e $\text{PF} < -9^{\circ}\text{C}$);
- Médio ($22 \leq ^{\circ}\text{API} < 31$ e $-9 \leq \text{PF} \leq 9$);
- Pesado ($10 \leq ^{\circ}\text{API} < 22$ e $\text{PF} > 9^{\circ}\text{C}$);
- Extrapesado ($^{\circ}\text{API} < 10$).

Com isso é possível destacar que a maioria das amostras com gravidade API na faixa de óleos médios também são classificados como óleos médios baseando-se no ponto de fluidez, porém é possível observar também que algumas amostras classificadas como óleo médio pela densidade são classificadas como leves de acordo com o ponto de fluidez. Amostras classificadas como leves pela gravidade API podem ser encontradas nas três classes quando observamos o ponto de fluidez, e várias amostras pesadas pela densidade são classificadas como leves pelo ponto de fluidez.

Portanto percebe-se que não há relação direta entre essas duas propriedades, apenas para algumas amostras.

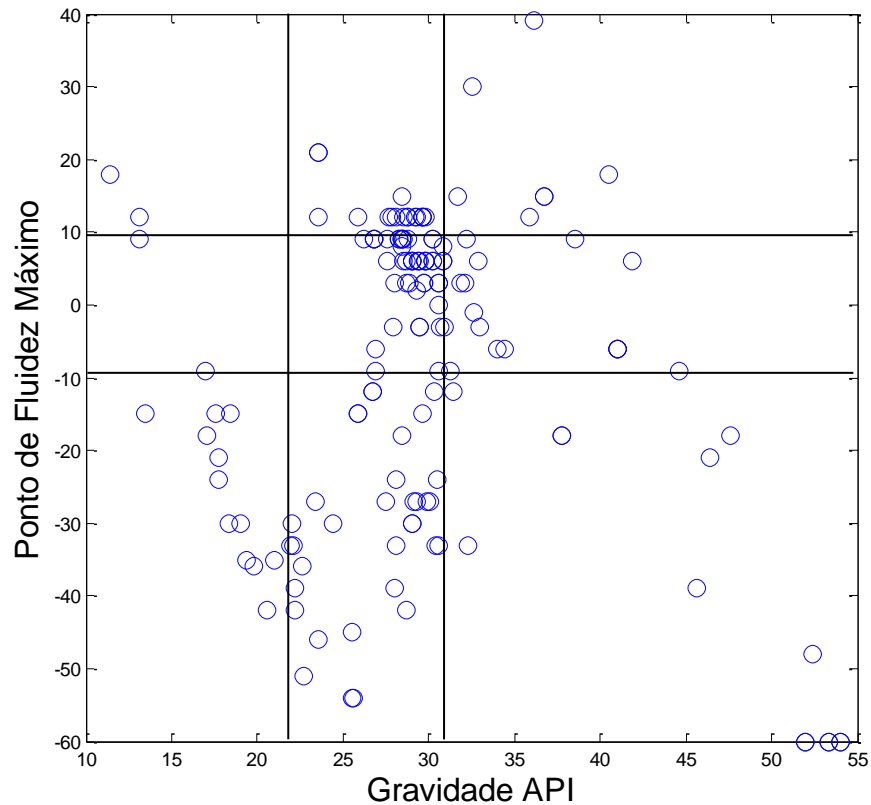


Figura 17. Correlação entre a gravidade API e o ponto de fluidez de um conjunto de amostras de petróleo.

- *Viscosidade cinemática*

Os modelos criados para viscosidade cinemática, foram construídos utilizando a viscosidade depois de submetida a duas operações logarítmicas, como relatado no tópico da metodologia. Os modelos resultaram em bons valores de RMSEP, e altos valores de RMSEC (Tabela 20). A boa relação entre os valores previstos por métodos quimiométricos de calibração e medidos pelo método de referência, para as amostras de calibração e as amostras de teste são mostrados na Figura 18.

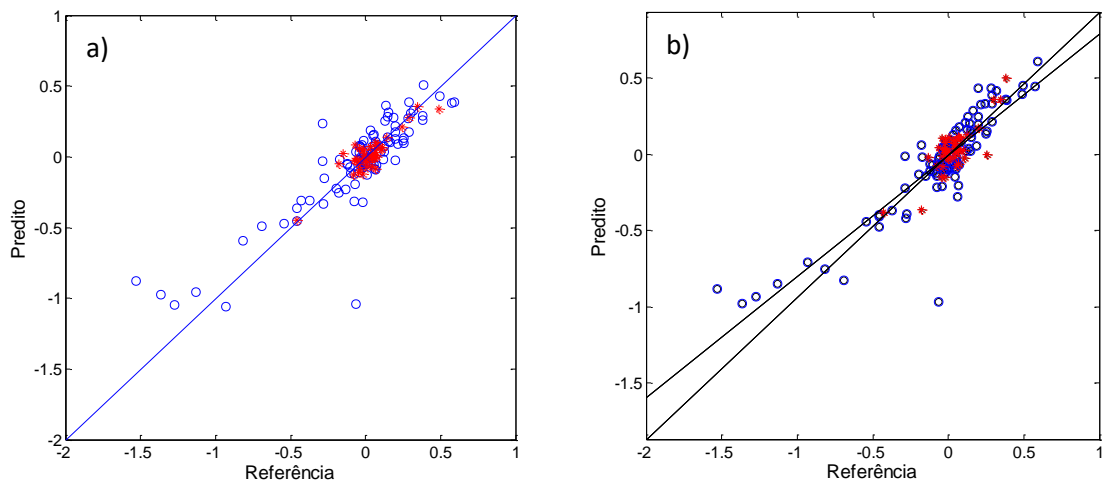


Figura 18. Relação entre valores de referência (o) e previstos (*) para $\log [\log (\text{viscosidade cinemática a } 50^{\circ}\text{C})]$ dos modelos criados por (a) PLS e (b) MLR

Os parâmetros estatísticos para os modelos PLS e MLR selecionados para a viscosidade cinemática a 50°C são mostrados na Tabela 20.

Tabela 20. Valores calculados para a validação dos modelos selecionados para viscosidade cinemática a 50°C , por PLS e MLR.

Parâmetros	PLS	MLR
VL's	4	-
RMSEC	31,4	28,0
$R^2_{\text{Calibração}}$	0,7661	0,7947
RMSEP	15,4	16,5
RMSEP%	11,44	12,23
$R^2_{\text{Previsão}}$	0,7552	0,6754

O método de referência ASTM para a viscosidade traz um valor de reprodutibilidade relativa, não nos permitindo comparar com os valores obtidos pelo método quimiométrico construído.

As variáveis mais significativas para a construção dos modelos PLS e MLR escolhidos podem ser observados na Tabela 21, para ambos modelos:

Tabela 21. Coeficientes de regressão do modelo PLS, com 6 variáveis latentes e pré-processamento MSC, e coeficientes do modelo MLR, na determinação da viscosidade cinemática a 50°C.

MLR		PLS	
Parâmetros (região do espectro - ppm)	Coeficientes	Parâmetros (região do espectro - ppm)	Coeficientes
C_Poveda	3,00 - 18,50	C_Hasan	0,00 - 70,00
	18,50 - 21,50		110,00 - 160,00
	21,50 - 50,00		110,00 - 130,00
	50,00 - 60,00		128,00 - 136,00
	60,00 - 70,00		129,00 - 137,00
	85,00 - 129,20		137,00 - 160,00
	129,20 - 137,00		
	137,00 - 160,00		
	190,00 - 220,00		

De acordo com a Tabela 21, as regiões do espectro mais importantes na construção dos modelos para a viscosidade cinemática, estão nos espectros de RMN de ^{13}C para ambos os modelos, sendo que para o modelo MLR a importância das variáveis está igualmente distribuída por todo o espectro, enquanto no modelo PLS as variáveis de maior contribuição são regiões com sinais de carbonos aromáticos em junção de anéis e carbonos aromáticos ligados a grupos metílicos.

- SARA

Na modelagem dos teores de saturados, aromáticos, resinas e asfaltenos, a relação entre os valores previstos por métodos quimiométricos de calibração e medidos pelo método de referência SARA, para as amostras de calibração e as amostras de teste são mostrados nas Figura 19, 20, 21 e 22. Os modelos criados resultaram em altos valores de RMSEP e baixos valores de R^2 , observando-se, nas Figuras 21 e 22, a relação entre valores medidos e previstos percebe-se que os modelos selecionados para a previsão tiveram desempenho bastante ruim, principalmente para resinas e asfaltenos, o que pode ter sido influenciado pelo grande número de amostras com teores iguais a zero.

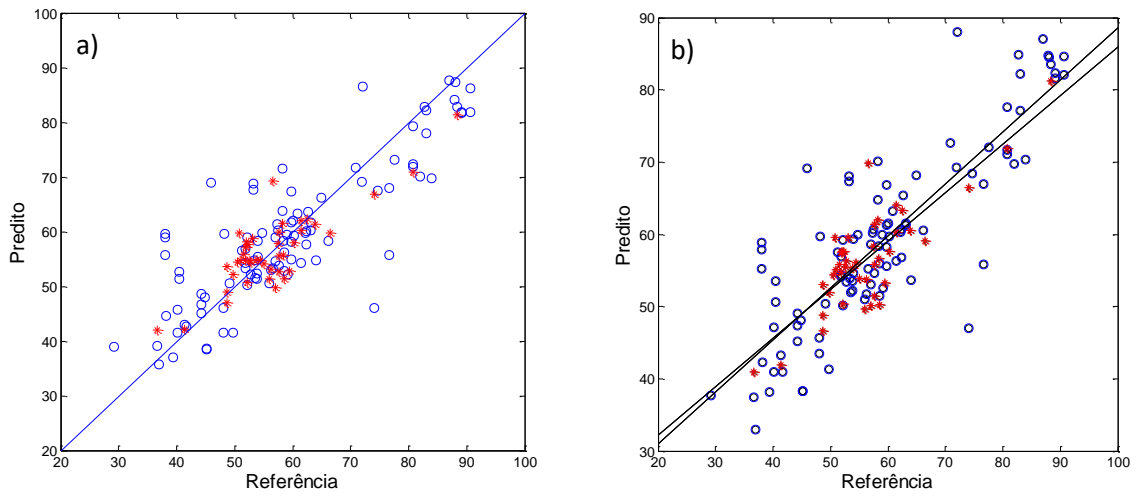


Figura 19. Relação entre valores de referência (o) e previstos (*) para os teores de saturados dos modelos criados por (a) PLS e (b) MLR

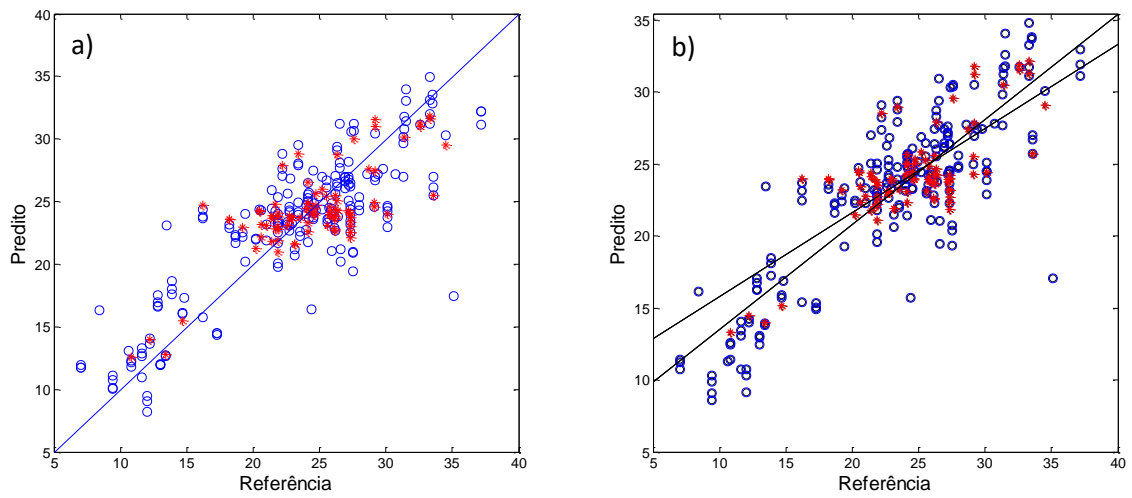


Figura 20. Relação entre valores de referência (o) e previstos (*) para os teores de aromáticos dos modelos criados por (a) PLS e (b) MLR

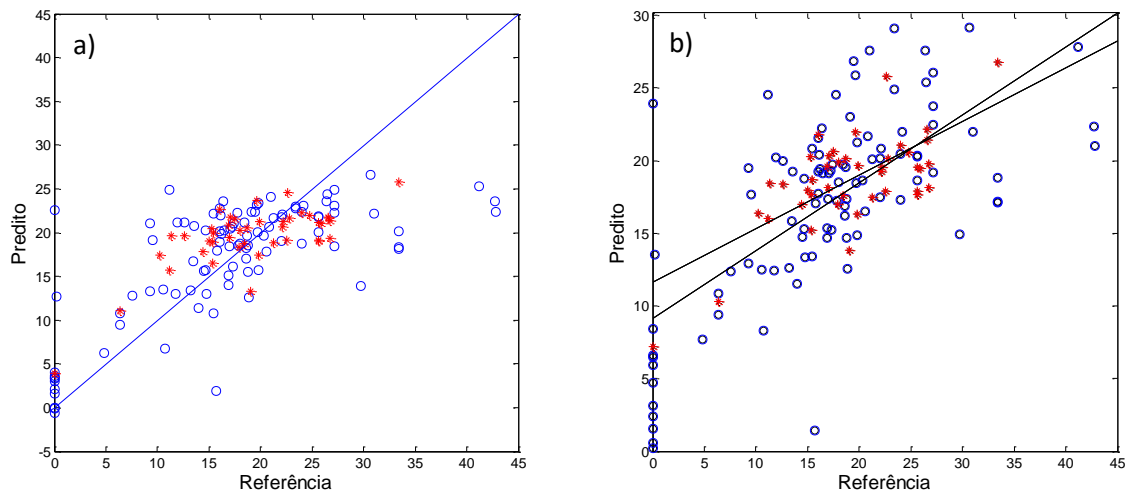


Figura 21. Relação entre valores de referência (o) e previstos (*) para os teores de resinas dos modelos criados por (a) PLS e (b) MLR

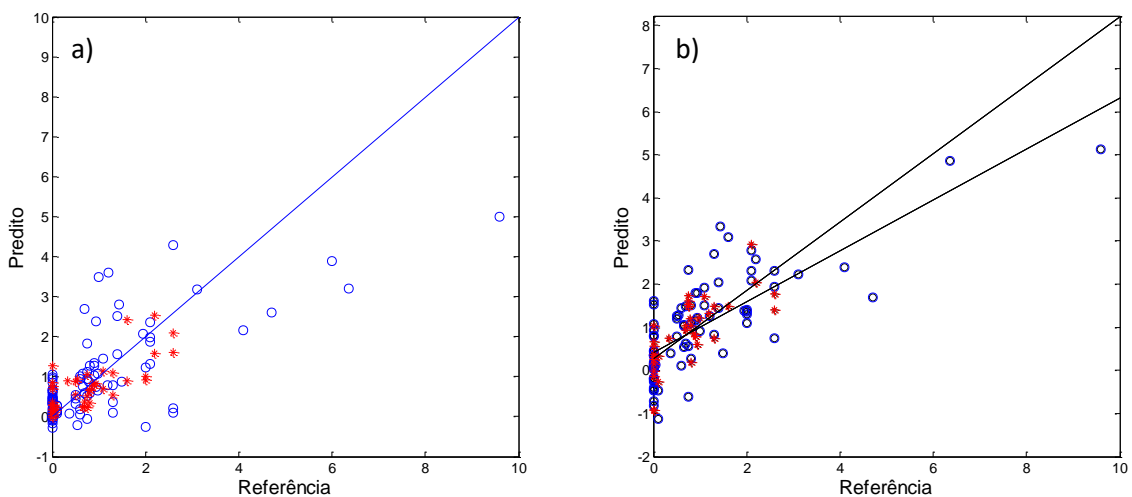


Figura 22. Relação entre valores de referência (o) e previstos (*) para os teores de asfaltenos dos modelos criados por (a) PLS e (b) MLR

Os parâmetros estatísticos obtidos para os modelos PLS e MLR selecionados são mostrados na Tabela 22. A reprodutibilidade sugerida na ASTM, os valores de resinas e asfaltenos são dados como teor de polares, logo para o cálculo dos erros e do desvio os parâmetros da tabela são mostrados em função dos polares, dado pela combinação dos erros individuais das classes de resinas e asfaltenos.

Tabela 22. Valores calculados para a validação dos modelos selecionados para saturados, aromáticos, resinas e asfaltenos, por PLS e MLR.

Parâmetros	Saturados		Aromáticos		Resinas		Asfaltenos	
	PLS	MLR	PLS	MLR	PLS	MLR	PLS	MLR
VL's	4	-	3	-	5	-	2	-
R ² Calibração	0,7156	0,7205	0,7245	0,7311	0,5016	0,4674	0,5265	0,5910
RMSEC	8,23	7,95	3,57	3,49	Polares - PLS = 7,02 MLR = 7,01			
RMSEP	5,18	5,06	2,95	2,88	Polares - PLS = 4,87 MLR = 4,64			
RMSEP%	8,9	8,6	12,5	12,3	Polares - PLS = 25,2 MLR = 24,0			
$\sigma_{\text{previsão}}$	6,55	6,45	4,43	4,38	Polares - PLS = 5,19 MLR = 4,98			
R ² Previsão	0,7049	0,6785	0,5855	0,5844	0,5004	0,4827	0,5347	0,5486

Apesar dos modelos terem gerado valores maiores que os especificados na ASTM para saturados e resinas, os resultados são aceitáveis para a indústria do petróleo, tendo em vista que os RMSEP's obtidos são menores de 5% m/m, que é o valor máximo de variação permitido na análise de SARA na indústria do petróleo,³⁴ exceto quando observamos o teor de saturados que ultrapassam um pouco esse teor.

Observando-se os valores de reprodutibilidade dos respectivos métodos laboratoriais (Tabela 6) para determinação de SARA e os valores de RMSEP e desvio padrão de previsão calculados (Tabela 22), observa-se que, para a previsão do teor de aromáticos o erro adicionado à previsão devido ao modelo estatístico é baixo quando comparado à reprodutibilidade do método padrão. Já para a previsão dos teores de saturados e polares há uma contribuição proporcional do modelo estatístico e do método laboratorial ao erro de previsão calculado. Essa comparação entre os valores de RMSEP, reprodutibilidade e desvio padrão de previsão também pode ser visualizada na Figura 23.

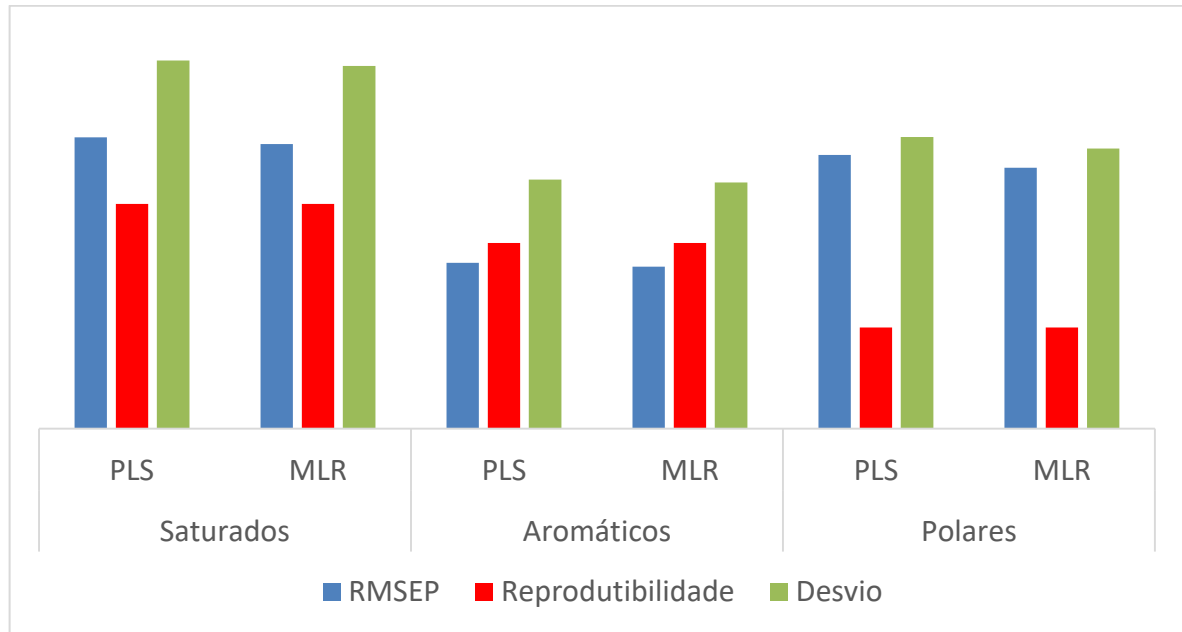


Figura 23. Relação entre os erros experimentais e calculados para os teores de saturados, aromáticos, resinas e asfaltenos.

As variáveis mais significativas para a construção dos modelos PLS e MLR escolhidos podem ser observados nas Tabela 23, para os teores de saturados, aromáticos, resinas e asfaltenos:

Tabela 23. Coeficientes de regressão do modelo PLS e modelo MLR, na determinação do teor de saturados, aromáticos, resinas e asfaltenos (continua).

Saturados			
	Parâmetros	Coefficientes	Coefficientes
	(região do espectro - ppm)	MLR	PLS
C_Hasan	0,00 - 70,00	0,86	4,32
	110,00 - 160,00	44,72	-4,32
	110,00 - 130,00	-45,71	6,27
	128,00 - 136,00	29,16	20,08
	129,00 - 137,00	-72,39	-16,00
	137,00 - 160,00	-45,25	7,00
Aromáticos			
	Parâmetros	Coefficientes	Coefficientes
	(região do espectro - ppm)	MLR	PLS
H_Poveda	0,10 - 1,00	-0,39	-1,36
	1,00 - 1,50	0,07	-0,27
	1,50 - 2,00	0,84	0,96
	2,00 - 4,50	0,58	1,97
	6,00 - 7,20	14,23	1,66
	7,20 - 9,00	0,20	1,54
	9,00 - 12,00	-88,64	-1,15

Tabela 23. Coeficientes de regressão do modelo PLS e modelo MLR, na determinação do teor de saturados, aromáticos, resinas e asfaltenos (conclusão).

Resinas				
	Parâmetros	Coeficientes	Coeficientes	
	(região do espectro - ppm)	MLR	PLS	
C_Hasan	0,00 - 70,00	0,12	592,11	
	110,00 - 160,00	-4,93	-19,42	
	110,00 - 130,00	7,49	-43,21	
	128,00 - 136,00	-24,99	-956,38	
	129,00 - 137,00	27,63	546,59	
	137,00 - 160,00	6,60	-119,70	
Asfaltenos				
	Parâmetros	Coeficientes	Parâmetros	Coeficientes
	(região do espectro - ppm)	MLR	(região do espectro - ppm)	PLS
H_Poveda	0,10 - 1,00	-1,26	0,00 - 70,00	-183,37
	1,00 - 1,50	-1,69	110,00 - 160,00	-12,37
	1,50 - 2,00	-1,46	110,00 - 130,00	1,18
	2,00 - 4,50	-1,61	128,00 - 136,00	-47,00
	6,00 - 7,20	-6,21	129,00 - 137,00	-52,41
	7,20 - 9,00	1,57	137,00 - 160,00	47,35
	9,00 - 12,00	18,23		
C_Poveda	3,00 - 18,50	32,25		
	18,50 - 21,50	30,33		
	21,50 - 50,00	31,68		
	50,00 - 60,00	32,31		
	60,00 - 70,00	31,80		
	85,00 - 129,20	31,91		
	129,20 - 137,00	31,91		
137,00 - 160,00	31,75			
190,00 - 220,00	32,24			

A região de 128 a 137 ppm no espectro de RMN de ^{13}C , apesar de apresentar baixa intensidade, tem grande importância para a modelagem do teor de saturados, para ambas as técnicas utilizadas, essa região confere deslocamentos químicos atribuídos a carbonos aromáticos substituídos por metila, carbonos aromáticos em junção de 2 ou 3 anéis aromáticos. Esperava-se que o início do espectro tivesse maior contribuição na modelagem dos saturados por estar relacionado a carbonos de grupos alifáticos, que constituem a fração de saturados. Entretanto, é observada uma maior contribuição da região com predominância de compostos aromáticos.

As variáveis selecionadas para a modelagem de aromáticos apresentam predominância do final do espectro. No modelo MLR a região de maior importância no espectro de RMN de ^1H corresponde a hidrogênios de aldeído ou ácidos carboxílicos, enquanto no modelo PLS a maior contribuição está na região de hidrogênios parafínicos e naftênicos do tipo CH , CH_2 e CH_3 , ligados a sistemas aromáticos na posição α e hidrogênios ligados a heteroátomos, como oxigênio e enxofre.

Na modelagem do teor de resinas a região de maior contribuição no modelo MLR corresponde a carbonos aromáticos substituídos por metila, carbonos aromáticos em junção de 2 ou 3 anéis aromáticos, assim como nos saturados, enquanto no modelo PLS as regiões iniciais e finais do espectro apresentam grande importância. Em contrapartida, o teor de asfaltenos foi determinado com maior contribuição da região de alifáticos no caso do modelo PLS, enquanto no modelo MLR todo o espectro de carbono tem grande importância e contribui da mesma forma.

3.5 Conclusões

A técnica de RMN associada a métodos quimiométricos multivariados é eficiente para determinar propriedades em petróleo, uma vez que a maior parte dos modelos desenvolvidos obteve um resultado satisfatório. Vale destacar que as propriedades químicas com influência direta no perfil espectral, tais como gravidade API, viscosidade cinemática e teores de saturados e aromáticos nos modelos SARA geraram modelos mais robustos e com melhor capacidade preditora, quando comparados aos modelos selecionados para a predição das propriedades restantes, ponto de fluidez máximo, resinas e asfaltenos.

De modo geral, comparando-se as duas técnicas quimiométricas testadas, houve uma melhor performance daqueles que utilizaram a regressão linear múltipla, entretanto os modelos desenvolvidos por PLS não podem ser descartados, pois forneceram resultados aceitáveis.

Vale destacar que para a previsão do ponto de fluidez, viscosidade e teor de aromáticos, o erro adicionado à previsão devido ao modelo estatístico é baixo quando comparado à reprodutibilidade do método padrão, tanto por MLR quanto por PLS, o erro final dos melhores modelos selecionados para previsão para as propriedades

restantes (gravidade API, teor de saturados e polares) pode ser viável para aplicação prática. Isso ocorre pois métodos muito automatizados em geral possuem alta precisão e exatidão, mas para a indústria do petróleo uma rápida estimativa do parâmetro com menor exatidão, na maioria das vezes, é aceitável.

A espectroscopia RMN associada a técnica de reconhecimento de padrão PLS-DA se mostrou eficiente na classificação das amostras de petróleo, de acordo com a ANP, em amostras leves, médias, pesadas e extrapesadas com taxas de erro baixos e altos graus de exatidão.

4 Aplicação 2: Análise exploratória de frações de petróleo por RMN e PCA na investigação da acidez dos óleos originais

4.1 Introdução

Em estado bruto o petróleo não tem aplicabilidade prática, mas quando refinado ele fornece combustíveis líquidos valiosos, solventes, lubrificantes, e muitos outros produtos. Os combustíveis derivados do petróleo contribuem com aproximadamente um terço a um meio do suprimento total de energia no mundo.²

O processo físico básico para a separação dos derivados do petróleo em uma refinaria é a destilação. O petróleo contém milhares de compostos diferentes que variam em massa molar de $16 \text{ g} \cdot \text{mol}^{-1}$ (metano, CH_4) a mais de $2000 \text{ g} \cdot \text{mol}^{-1}$.^{2, 38} Esta larga escala nas massas molares conduz aos pontos de ebulição que variam de -160°C (ponto de ebulição do metano) a mais do que 600°C , que é o ponto de ebulição de compostos pesados no óleo cru. Um grupo de hidrocarbonetos, portanto, pode ser separado com a destilação de acordo com o ponto de ebulição dos compostos mais leves e mais pesados nas misturas.

De fato, durante a destilação um petróleo bruto é convertido em uma série de frações do petróleo, na qual cada uma é uma mistura de um número limitado de hidrocarbonetos com uma escala específica do ponto de ebulição. As frações com uma escala mais larga de pontos de ebulição contêm o maior número de hidrocarbonetos. Todas as frações de uma coluna de destilação têm uma escala de ebulição conhecida, exceto o resíduo, para o qual o ponto de ebulição superior geralmente não é conhecido. Porém, a natureza infinitamente variável de fatores composicionais faz com que todos os óleos crus e produtos de petróleo processados numa refinaria sejam diferentes entre si. Essa variabilidade representa uma característica química do tipo *fingerprint* ou impressão digital para cada petróleo e fornece uma base para caracterizá-lo.²

A partir do conhecimento dos dados de distribuição dos intervalos de ebulição característicos de cada corte que vão compor e definir a curva de destilação de um determinado petróleo é possível prever o percentual (rendimento) dos diferentes cortes ou frações, tais como gasolina e querosene, óleo diesel e outros, o que é

fundamental para o controle do processo de refino e para a garantia da qualidade dos produtos.

A diferença de volatilidade entre os compostos químicos que constituem o petróleo (compostos leves, médios e pesados), é a base fundamental para a separação das várias faixas de hidrocarbonetos que o compõem. Define-se, a partir desse conceito, que petróleos mais pesados são constituídos por hidrocarbonetos mais pesados (aqueles que possuem maiores pontos de ebulição).⁴¹

Frações (ou “Cortes”) são definidas como as faixas de hidrocarbonetos cujo ponto de ebulição encontra-se dentro de uma determinada faixa de temperatura (“pontos de corte”). Quando se compara diferentes petróleos para um mesmo perfil de destilação, o que muda em si não é o intervalo de temperatura de corte, mas sim o rendimento de cada produto que é obtido dentro daquelas faixas pré-determinadas.¹

Assim, é possível separar, por meio da destilação, os compostos leves, médios e pesados que compõem o petróleo, conforme Tabela 24.

Tabela 24. Frações típicas do petróleo¹

Fração	TEB (°C)	Composição (aprox.)
Gás residual	-	C ₁ – C ₂
GLP	< 40	C ₃ – C ₄
Gasolina	40 – 175	C ₅ – C ₁₀
Querosene	175 – 235	C ₁₁ – C ₁₂
Gasóleo Leve	235 – 305	C ₁₃ – C ₁₇
Gasóleo Pesado	305 – 400	C ₁₈ – C ₂₅
Lubrificantes	400 – 510	C ₂₆ – C ₃₈
Resíduos	> 510	C ₃₈ ⁺

Com base nos valores de massa e volume obtidos a cada faixa de corte do petróleo nos processos de destilação atmosférica e a vácuo, seguindo as normas ASTM D2892⁴² e ASTM D5236⁴³, respectivamente, é possível determinar a curva de destilação do óleo cru. Esta é conhecida como curva de Pontos de Ebulição Verdadeiros (PEV) e é representada por um gráfico de percentual acumulado *versus*

temperatura de ebulição. Na destilação de petróleo em escala laboratorial é possível analisar a quantidade dos cortes que estão sendo destilados, seja por métodos padronizados ou não, e com isso pode-se determinar os processos de tratamento necessários para enquadrar estes derivados às especificações do mercado.

No refino de petróleo, ácidos naftênicos são compostos indesejáveis associados a corrosão e a formação de incrustações.⁴¹

O número de acidez total (NAT) é definido pela ASTM D664 como a quantidade de base, expresso em miligramas de KOH, necessária para neutralizar os componentes ácidos em 1 grama de óleo.⁴⁴

A corrosão provocada por ácidos naftênicos, a qual gera graves problemas na cadeia do refino do petróleo, está relacionada ao número de acidez total (NAT) presente no óleo cru bem como pela forma com que estes se distribuem nas frações destiladas, suas estruturas e a temperatura de operação, entre outros. Maiores concentrações de ácidos naftênicos são encontrados normalmente em frações do petróleo com pontos de ebulição no intervalo de 371°C a 426°C, desta maneira as frações mais voláteis tendem a serem relativamente menos ácidas comparadas às pesadas.⁴⁵

4.2 Objetivos

Pretende-se nesta etapa fazer uma análise exploratória da acidez de amostras de petróleo a partir de análises de RMN das frações desses petróleos obtidas por destilação fracionada dos petróleos e PCA. Para isso, foram delineados os passos:

- Utilizar RMN de ^1H e ^{13}C com a finalidade de determinar parâmetros estruturais das frações de petróleo, de acordo com os artigos Hasan (1983) e Poveda (2012);
- Construir modelos PCA a fim de fazer uma análise exploratória dos dados para a determinação da acidez do óleo original a partir de parâmetros estruturais das frações;
- Testar e comparar o desempenho dos métodos de pré-processamento dos dados: centrar na média e autoescalonamento, além de testes sem nenhum tipo de pré-processamento.

4.3 Metodologia

4.3.1 Obtenção das amostras

Nesta etapa do trabalho foram utilizadas 138 frações de petróleo obtidas por destilação de oito diferentes petróleos, com grau API variando de 20,7 a 46,4, ou seja, foram incluídas amostras derivadas de diferentes tipos de petróleos, desde leves até pesados. As amostras que deram origem às frações apresentam valores de NAT distintos, que variavam de 0,04 a 1,42 mg de KOH/g de óleo. Os petróleos utilizados foram destilados conforme ASTM D2892 e ASTM D5236. Os cortes obtidos (amostras) foram cedidos pelo laboratório de destilação do LabPetro, armazenados em frascos apropriados e em temperatura adequada para posteriores análises. As propriedades físico-químicas foram determinadas conforme os respectivos métodos padrão descritos na ASTM.

4.3.2 Análises de RMN

As análises de Ressonância Magnética Nuclear de ^1H e ^{13}C foram realizadas no LabPetro no espectrômetro da marca Varian, modelo VNMRS 400, operando com um campo magnético de 9,4 T, utilizando uma sonda 5 mm BroadBand $^1\text{H}/^{19}\text{F}/\text{X}$ para as análises de ^1H e uma sonda 10 mm Broadband $^{15}\text{N}-^{31}\text{P}\{^1\text{H}-^{19}\text{F}\}$ para as análises de ^{13}C , ambas realizadas à temperatura de 25 °C.

Para as análises de ^1H , as amostras foram preparadas dissolvendo-se uma massa correspondente a 5% em clorofórmio-*d*, totalizando um volume de 600 μL de solução. As análises foram feitas conforme as condições descritas abaixo:

- Frequência: 399,73 MHz;
- Janela espectral: 6410,3 Hz;
- Tempo de Aquisição: 2,556 s;
- Tempo espera: 1,0 s;
- Pulso: 90°;
- Número de transientes: 32;
- Processamento: linebroadening de 0,5 Hz.

Para as análises de ^{13}C , as amostras foram preparadas dissolvendo-se uma massa correspondente a 40% do volume final, em uma solução 0,05 mol/L de $\text{Cr}(\text{acac})_3$ (acetilacetato de cromo III) com clorofórmio-*d*, totalizando um volume de 3 mL de solução. As análises foram feitas conforme as condições descritas abaixo:

- Frequência: 100,51 MHz;
- Janela espectral: 25510,2 Hz;
- Tempo de Aquisição: 1,285 s;
- Tempo espera: 7,0 s
- Pulso: 90° ;
- Núcleo desacoplado: ^1H ;
- Modo do desacoplador: nny, ou seja, o desacoplador ficou desligado durante o pulso e o tempo de espera, e ligado somente durante a aquisição dos dados, com isso o aumento de intensidade do sinal ocasionado pelo Efeito Nuclear Overhauser (NOE) é eliminado, garantido assim um experimento quantitativo;
- Número de transientes: 1000;
- Processamento: alargamento exponencial de 5,0 Hz.

Os deslocamentos químicos foram obtidos tendo os solventes ou o tetrametilsilano (TMS) como sinal de referência. A fase e a linha de base foram ajustadas e corrigidas manualmente. Os espectros de RMN foram tratados utilizando-se os parâmetros moleculares, determinados a partir das áreas dos sinais específicos sugeridos pela literatura, nos artigos Hasan (1983) e Poveda (2012).^{15, 16}

4.3.3 Tratamento Quimiométrico

Os espectros de RMN das frações foram tratados utilizando-se os parâmetros moleculares, determinados a partir das áreas dos sinais específicos sugeridos nos os dois artigos utilizados como base. Os conjuntos de variáveis utilizados foram de acordo com:

Tabela 25. Conjuntos de variáveis utilizados na construção dos modelos.

Artigo Base	Conjunto de parâmetros moleculares		
	(Conjunto de variáveis)		
Hasan (1983) ¹⁶	RMN de ¹ H	RMN de ¹³ C	RMN de ¹ H e ¹³ C
Poveda (2012) ¹⁵	RMN de ¹ H	RMN de ¹³ C	RMN de ¹ H e ¹³ C

Inicialmente, foram aplicados e testados como recurso de tratamento dos dados o autoescalonamento e centrar na média, os modelos também foram construídos sem nenhum tipo de pré-processamento.

O método quimiométrico principal aplicado aos dados foi a PCA, a fim de investigar alguma tendência entre as frações de um mesmo petróleo e entre as frações de petróleos diferentes, bem como relacionar essas tendências com propriedades físico-químicas dos petróleos originais. Neste trabalho observou-se o comportamento da acidez dos petróleos originais a partir frações.

Todos os cálculos foram executados no software Matlab versão 7.

4.4 Resultados e Discussões

Neste trabalho foram utilizados oito petróleos diferentes, sendo um óleo pesado, cinco óleos médios e dois óleos leves, classificados de acordo com a Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – ANP⁴⁰, Tabela 26. Não foi observada uma relação entre a gravidade API e a acidez dos óleos.

O Óleo 1 apresentou número de acidez total (NAT) muito elevada (1,4214 mg de KOH/g), seguido pelo Óleo 4 com número de acidez de 0,5014 mg de KOH/g. Os demais óleos podem ser avaliados em dois grupos: Óleos 2, 3 e 6 com número de acidez muito baixa, menor que 0,1 mg de KOH/g e Óleos 5, 7 e 8 com número de acidez de aproximadamente 0,2 mg de KOH/g.

Tabela 26. Propriedades dos óleos originais utilizados nas destilações.

Amostra	Gravidade API	Classe	NAT (mg de KOH/g)
1	22,3	Médio	1,4214
2	42,2	Leve	0,0500
3	46,4	Leve	0,0400
4	20,7	Pesado	0,5014
5	27,0	Médio	0,2824
6	28,4	Médio	0,0434
7	27,0	Médio	0,2357
8	26,9	Médio	0,2475

Os parâmetros moleculares do espectro de RMN de ^1H , de acordo os trabalhos de Hasan (1983)¹⁶ e Poveda (2012)¹⁵, podem ser relacionados a hidrogênios que podem contribuir para o aumento da gravidade API do óleo ou para sua acidez. O aumento das áreas dos sinais dos hidrogênios ao H alifático (região de 0,5 a 4,0 ppm) pode indicar um óleo mais parafínico, com menor densidade API.⁴⁶ Utilizando os espectros de RMN de ^{13}C podemos obter relações de carbonos ligados a grupos poliaromáticos relacionados às resinas e asfaltenos presentes no petróleo, as quais se acumulam em frações mais pesadas do óleo. Espera-se que um óleo com maior teor de componentes aromáticos reflita em suas frações, após a destilação, um maior teor de componentes aromáticos, o que podem ser discriminados pelos espectros de RMN associados a análise de dados multivariados.

Na aplicação da análise de componentes principais foram utilizados como forma de pré-processamento dos dados o centrar na média e o autoescalamento, além da aplicação sem nenhum tipo de pré-processamento. Serão apresentados os melhores resultados referentes a cada conjunto de variáveis, com o melhor pré-processamento em cada um deles.

Para os parâmetros moleculares obtidos dos espectros de RMN de ^1H de acordo com Hasan (1983)¹⁶ com os dados centrados na média, o gráfico dos *scores* da PCA indica padrões de distribuição das amostras de frações de petróleo relacionada a acidez dos óleos originais (Figura 24). As duas primeiras componentes principais são responsáveis pela separação das amostras, porém é interessante destacar que a separação se dá de maneira acentuada ao longo da segunda componente (variância

explicada de 41,15%). As variáveis responsáveis por essa separação são principalmente os hidrogênios aromáticos com alifáticos e hidrogênios ligados a carbonos saturados na posição α em relação ao sistema aromático, como pode ser observado pelos *loadings*.

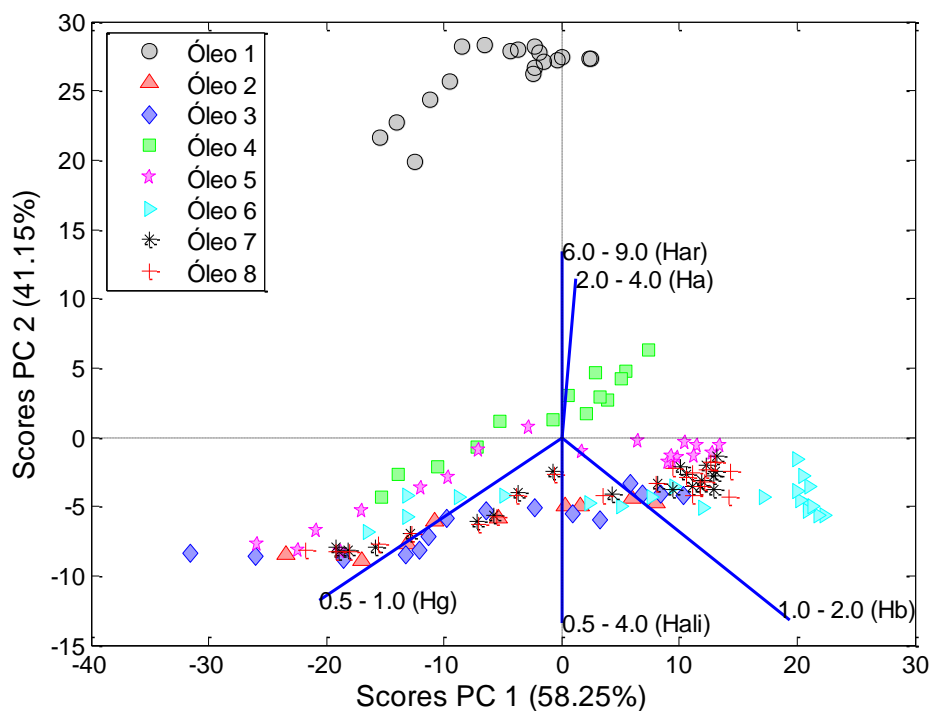


Figura 24. Biplot da PCA utilizando parâmetros moleculares médios do espectro de RMN de ^1H de acordo com Hasan (1983)¹⁶. Os vetores em azul representam os *loadings* da PCA.

Utilizando informações dos espectros de RMN de ^{13}C obtidos do artigo Hasan (1983)¹⁶ também com os dados centrados na média, observa-se apenas uma discriminação das amostras (frações) dos petróleos 1 e 4, os mais ácidos entre os estudados (Figura 25), mostrando que as informações geradas nos espectros de RMN de ^{13}C não são capazes de relacionar as amostras das frações com a acidez dos óleos originais.

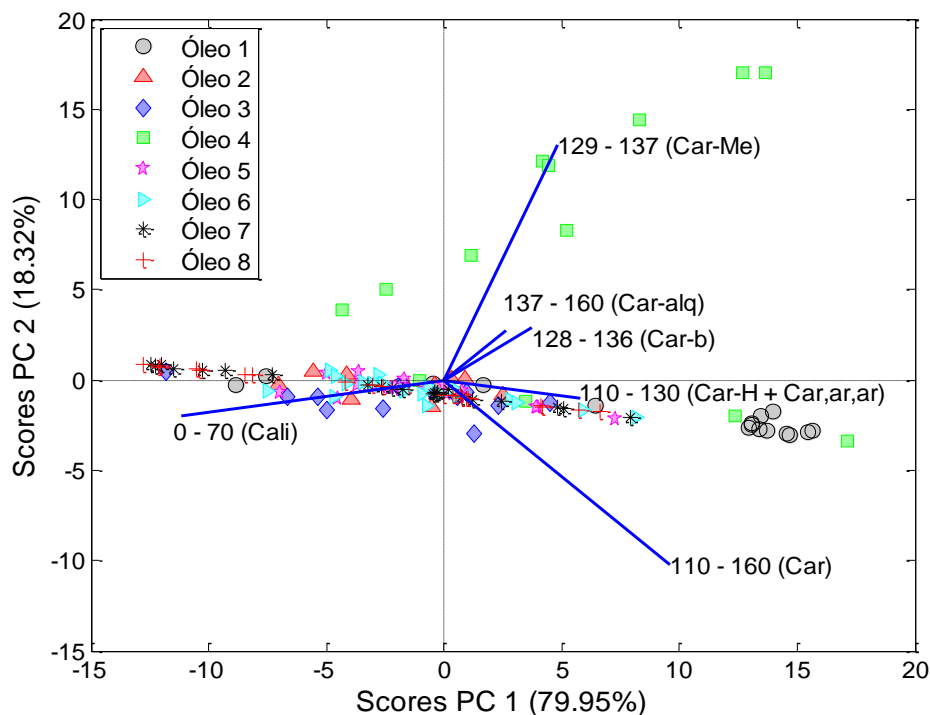


Figura 25. Biplot da PCA utilizando parâmetros moleculares médios do espectro de RMN de ^{13}C de acordo com Hasan (1983)¹⁶. Os vetores em azul representam os *loadings* da PCA.

Quando são utilizadas áreas de integração de ambos os espectros de RMN de ^1H e ^{13}C , de acordo com Hasan (1983)¹⁶, com os dados centrados na média, o gráfico de *scores* apresenta um padrão de distribuição muito semelhante ao mostrado na Figura 24, para espectros de RMN de ^1H , permitindo também relacionar as amostras de frações de petróleo com a acidez dos óleos originais (Figura 26). As variáveis responsáveis por essa separação são principalmente os Hidrogênios aromáticos com alifáticos, hidrogênios ligados a carbonos saturados na posição α em relação ao sistema aromático além das regiões que apresentam os carbonos aromáticos, como pode ser observado pelos *loadings*.

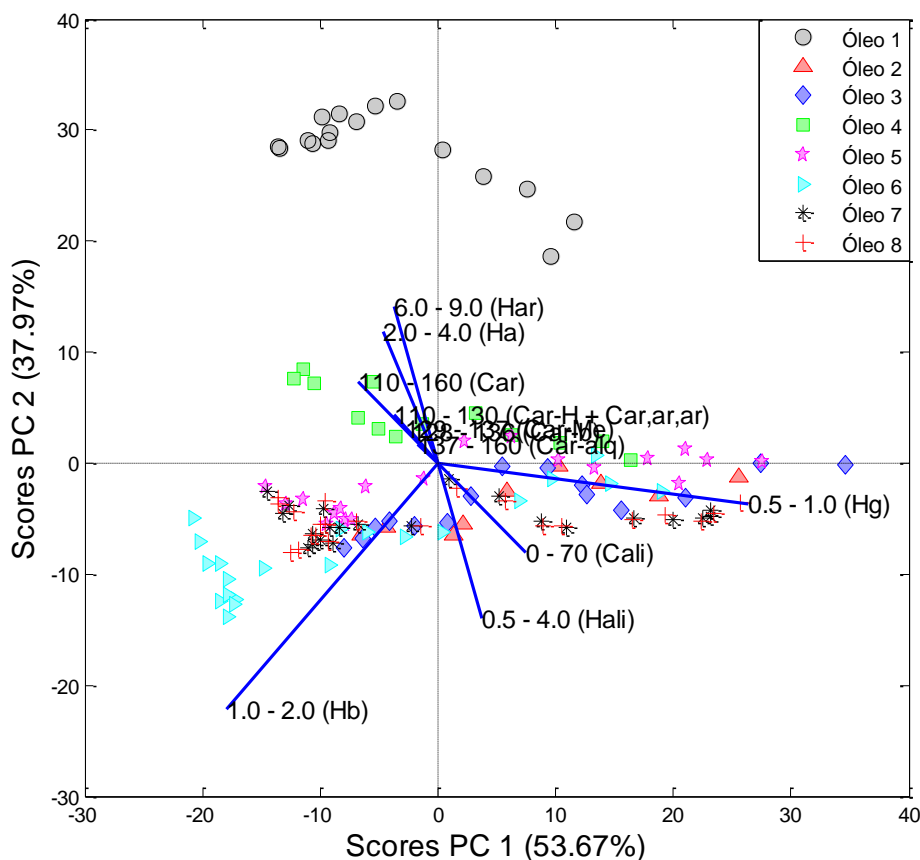


Figura 26. Biplot da PCA utilizando parâmetros moleculares médios do espectro de RMN de ^1H e ^{13}C de acordo com Hasan (1983)¹⁶. Os vetores em azul representam os *loadings* da PCA

Utilizando os parâmetros moleculares médios do espectro de RMN de ^1H de acordo com Poveda (2012)¹⁵ foi possível obter uma grande distinção das frações do Petróleo 1, Figura 27. As frações do óleo 4, pouco se distinguem das demais, talvez pela pequena diferença de acidez entre eles. Assim como ocorreu para os espectros de RMN de ^1H de acordo com Hasan (1983)¹⁶ a separação se dá de maneira mais acentuada ao longo da segunda componente. Onde as variáveis responsáveis por essa separação estão principalmente na região de aromáticos, porém a região de alifáticos também tem grande contribuição, como pode ser observado pelos *loadings*.

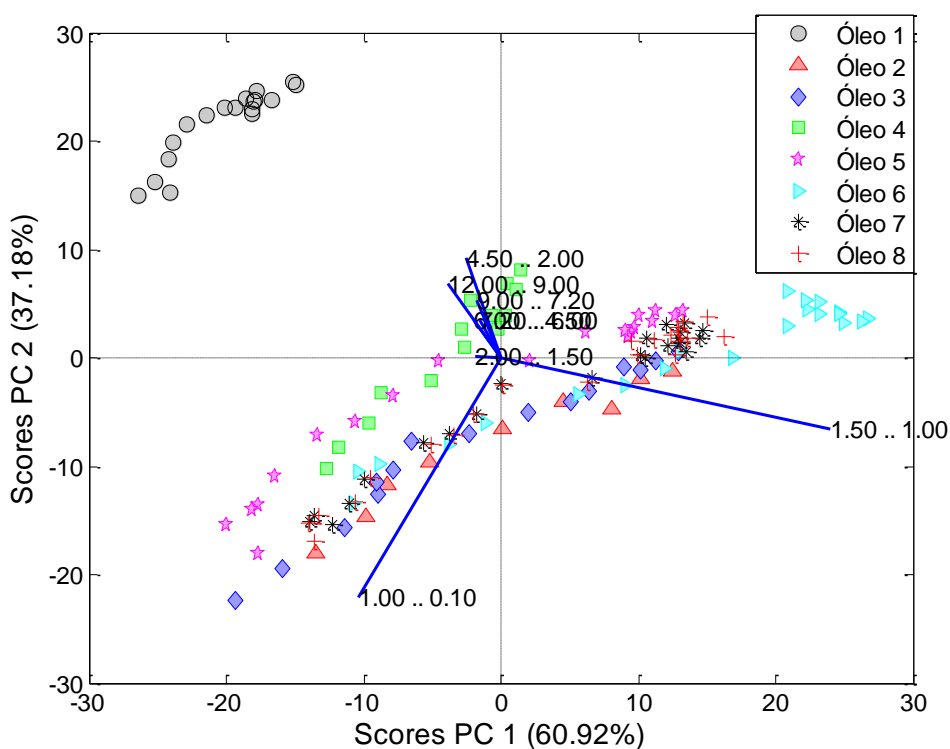


Figura 27. Biplot da PCA utilizando parâmetros moleculares médios do espectro de RMN de ^1H de acordo com Poveda (2012)¹⁵. Os vetores em azul representam os *loadings* da PCA.

Utilizando informações dos espectros de RMN de ^{13}C , segundo Poveda (2012)¹⁵ para os dados centrados na média, pouco se diferenciou as amostras dos Petróleos estudados, Figura 28, mostrando que as informações geradas nos espectros de RMN de ^{13}C não são capazes de relacionar as amostras das frações com a acidez dos óleos originais.

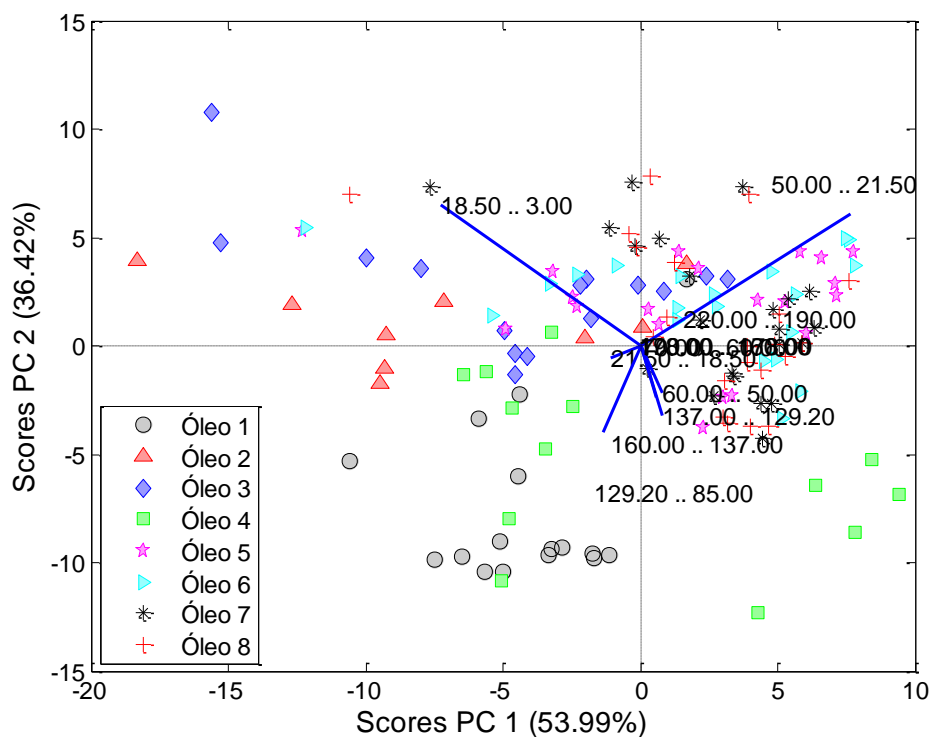


Figura 28. Biplot da PCA utilizando parâmetros moleculares médios do espectro de RMN de ^{13}C de acordo com Poveda (2012)¹⁵. Os vetores em azul representam os *loadings* da PCA.

Quando são utilizadas áreas de integração de ambos os espectros de RMN de ^1H e ^{13}C , de acordo com Poveda (2012)¹⁵, com os dados centrados na média, o gráfico de *scores* apresenta um padrão de distribuição muito semelhante ao mostrado na Figura 27, para espectros de RMN de ^1H , permitindo também relacionar as amostras de frações de petróleo com a acidez dos óleos originais (Figura 29). Onde as variáveis responsáveis por essa separação estão principalmente na região de aromáticos, tanto dos espectros de RMN de ^1H como ^{13}C , porém a região de alifáticos também tem grande contribuição, como pode ser observado pelos *loadings*.

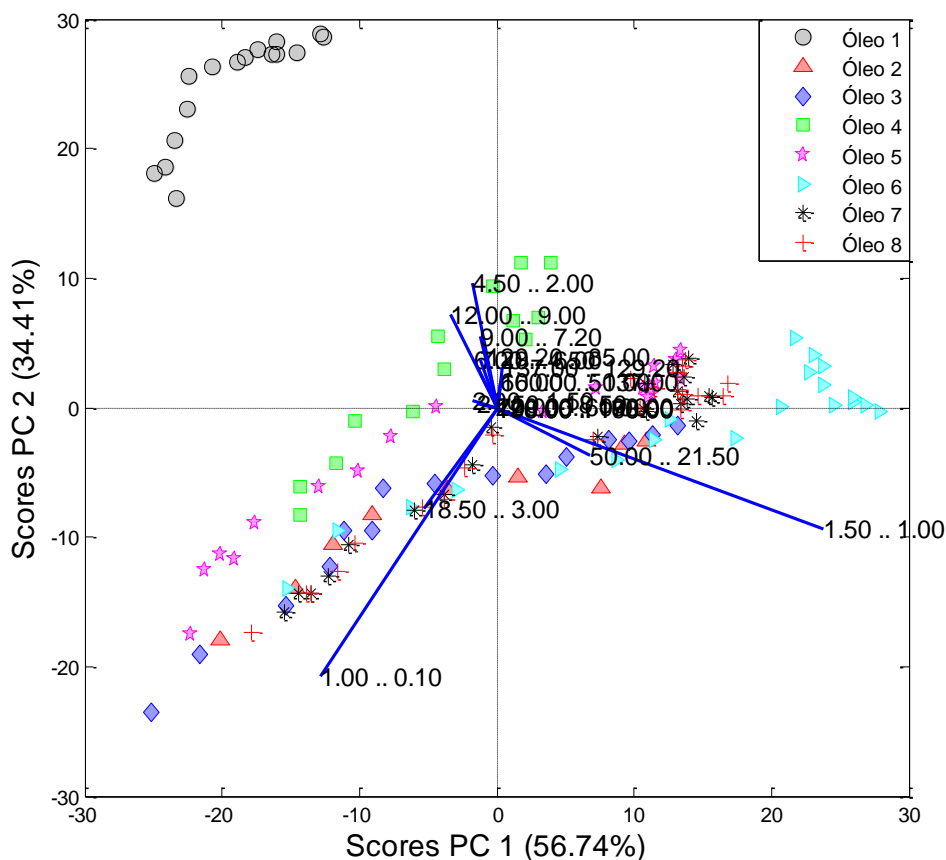


Figura 29. Biplot da PCA utilizando parâmetros moleculares médios do espectro de RMN de ^1H e ^{13}C de acordo com Poveda (2012)¹⁵. Os vetores em azul representam os *loadings* da PCA

4.5 Conclusões

A análise exploratória mostrou que existe uma relação entre a distribuição das frações e a acidez das amostras de petróleo originais, a partir dos parâmetros estruturais determinados por ressonância magnética nuclear. Utilizando informações dos espectros de RMN de ^1H e ^{13}C é possível discriminar frações de petróleo de acordo a acidez do óleo que as deram origem. Os espectros de RMN de ^1H apresenta melhor distinção que os espectros de RMN de ^{13}C .

5 Conclusões Gerais

A maior parte dos modelos desenvolvidos obteve um resultado satisfatório, demonstrando que os espectros de ressonância magnética nuclear, combinados com métodos quimiométricos, podem ser usados para determinar propriedades em petróleo.

Vale ressaltar que para os modelos de calibração, em sua maioria, obtiveram melhores resultados utilizando-se os parâmetros estruturais sugeridos pelo autor Hasan¹⁶. E o melhor conjunto de variáveis se baseiam nas áreas de integração fornecidas pelos espectros de ¹³C.

A maior parte dos modelos tiveram sua habilidade previsora melhorada quando os diversos métodos de pré-processamento foram aplicados, havendo uma superioridade preditiva quanto à aplicação do método MSC, em relação aos demais métodos de pré-processamento. Para os resultados de análise exploratória os melhores resultados foram obtidos com os dados centrados na média.

6 Referências

- ¹ THOMAS, J. E.. **Fundamentos de Engenharia de Petróleo**. 2. Ed. Rio de Janeiro: Interciência: PETROBRÁS, 2004. 271 p.
- ² SPEIGHT, J. G.. **Handbook of Petroleum Product Analysis**. New Jersey: John Wiley & Sons, 2002. 454 p.
- ³ MASILI, A.; PULIGHEDDU, S.; SASSU, L.; SCANO, P.; LAI, A.. Prediction of physical-chemical properties of crude oils by ¹H NMR analysis of neat samples and chemometrics. *Magnetic Resonance in Chemistry*. **50**, 729-738, 2012.
- ⁴ THE INSTITUTE OF PETROLEUM (Org.). **Modern Petroleum Technology**. v. 2. 6. Ed. London: John Wiley & Sons, 2000. 478 p.
- ⁵ NIELSEN, K. E., DITTMER, J., MALMENDAL, A., NIELSEN, N. C.. Quantitative analysis of constituents in heavy fuel oil by ¹H nuclear magnetic resonance (NMR) spectroscopy and multivariate data analysis. *Energy & Fuels*. **22**, 4070-4076, 2008.
- ⁶ PORTELA, N. A.; OLIVEIRA, E. C. S; NETO, A. C.; RODRIGUES, R. R. T.; SILVA, S. R. C.; CASTRO, E. V. R.; FILGUEIRAS, P. R.. Quantification of biodiesel in petroleum diesel by ¹H NMR: Evaluation of univariate and multivariate approaches. *Fuel*. **166**, 12-18, 2016.
- ⁷ MOLINA V., D.; URIBE, U. N.; MURGICH, J.. Correlations between SARA fractions and physicochemical properties with ¹H NMR spectra of vacuum residues from Colombian crude oils. *Fuel*, **89**, 185-192, 2010.
- ⁸ LINDON, J. C.; TRANTER, G. E.; HOLMES, J. L.. Encyclopedia of Spectroscopy and Spectrometry Part 2 (M-Z). London: Elsevier, 2000. 1405 p.
- ⁹ CLARIDGE, T. D. W.. High-Resolution NMR Techniques in Organic Chemistry. 2. Ed. Oxford: Elsevier, 2009. 383 p.
- ¹⁰ GIL, V. M. S.; GERALDES, C. F. G. C.. Ressonância Magnética Nuclear: Fundamentos, Métodos e Aplicações. 2. Ed. Lisboa: Fundação Calouste Gulbenkian, 2002. 1004 p.

- ¹¹ LEVITT, M. H.. Spin Dynamics: Basics of Nuclear Magnetic Resonance. 2. Ed. Southampton: John Wiley & Sons Ltd, 2008. 713 p.
- ¹² MACOMBER, R. S.. A Complete Introduction to Modern NMR Spectroscopy. New York: wiley-Interscience, 1998. 378 p.
- ¹³ OLIVEIRA, S. C.. **Revisão bibliográfica acerca dos princípios de ressonância magnética nuclear**. Monografia. Departamento de Engenharia Elétrica, Faculdade de Tecnologia, Universidade de Brasília, Brasília, 2008. 97 p.
- ¹⁴ BARBOSA, L. R.. **Estudo Teórico-Experimental de Compostos Cage-Like com Potencial Atividade Biológica: Um estudo por Ressonância Magnética Nuclear**. Dissertação de Mestrado. Programa de pós-graduação em Química, Centro de Ciências Exatas, Universidade Federal do Espírito Santo, Vitória, 2013. 294p.
- ¹⁵ POVEDA, J. C.; MOLINA, D. R.. Average molecular parameters of heavy crude oils and their fractions using NMR spectroscopy. *Journal of Petroleum Science and Engineering*, **84-85**, 1-7, 2012.
- ¹⁶ HASAN, M. U.; ALI, M. F.; BUKHARI, A.. Structural characterization of Saudi Arabian heavy crude oil by n.m.r. spectroscopy. *Fuel*, **62**, 518-523, 1983.
- ¹⁷ BROWN, S. D.. Has the chemometrics revolution ended? Some views on the past, present and future of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **30**, 49-58, 1995.
- ¹⁸ GELADI, P.. Chemometrics in spectroscopy. Part 1. Classical Chemometrics. *Spectrochimica Acta Part B*, **58**, 767-782, 2003.
- ¹⁹ HOPKE, P. K. The evolution of chemometrics. *Analytica Chimica Acta*, **500**, 365-377, 2003.
- ²⁰ OLIVEIRA, E. C. S.. **Determinação de adulteração de etanol hidratado com metanol por espectrometria de infravermelho médio associada a técnicas quimiométricas**. Monografia. Departamento de Química, Centro de Ciências Exatas, Universidade Federal do Espírito Santo, Vitória, 2010, 63p.

- ²¹ FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O.. Quimiometria I: Calibração Multivariada, um tutorial. *Química Nova*, **22-5**, 724-731, 1999.
- ²² BRERETON, R. G.. Chemometrics: Data Analysis for the Laboratory and Chemical Plant. Grã-Bretanha: John Wiley & Sons Ltd, 2003. 498 p.
- ²³ ROCHA, J. T. C.. **Aplicação de espectroscopia no infravermelho próximo (NIR) e médio (MIR) associada a métodos quimiométricos, para avaliação de parâmetros físico-químicos em frações de petróleo.** Tese de Doutorado. Pós-graduação em Química, Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora, Juiz de Fora, 2016, 108 p.
- ²⁴ NETO, B. B.; SCARMÍNIO, I. S.; BRUNS, R. E. 25 anos de Quimiometria no Brasil. *Química Nova*, **29-6**, 1401-1406, 2006.
- ²⁵ ROCHA, J. T. C.. **Emprego de espectrometria no infravermelho e métodos quimiométricos para a identificação e quantificação de petróleos a partir de misturas de frações de diesel.** Dissertação de Mestrado. Programa de Pós-graduação em Química, Centro de Ciências Exatas, Universidade Federal do Espírito Santo, Vitória, 2009, 104 p.
- ²⁶ LYRA, W. S.; SILVA, E. C.; ARAÚJO, M. C. U.; FRAGOSO, W. D.. Classificação periódica: um exemplo didático para ensinar análise de componentes principais. *Química Nova*, **33-7**, 1594-1597, 2010.
- ²⁷ FORINA, M.; LANTERI, S.; CASALE, M.. Multivariate calibration. *Journal of chromatography A.*, **1158**, 61-93, 2007.
- ²⁸ GELADI P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. *Analytical Chemical Acta*, **185**, 1-17, 1986.
- ²⁹ ALMEIDA, M. R.; CORREA, D. N.; ROCHA, W. F. C.; SCAFI, F. J. O.; POPPI, R. J.. Discrimination between authentic and counterfeit banknotes using Raman spectroscopy and PLS-DA with uncertainty estimation. *Microchemical Journal*, **109**, 170-177, 2013.

- ³⁰ BALLABIO, D.; CONSONNI, V.. Classification tools in chemistry. Part 1: Linear models. PLSDA. *Analytical Methods*, **5**, 3790-3798, 2013.
- ³¹ XU, L.; ZHOU, Y. P.; TANG, L. J.; WU, H. L.; JIANG, J. H.; SHEN, G. L.; YU, R. Q.. Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration. *Analytica Chimica Acta*, **616**, 138-143, 2008.
- ³² FEARN, T.; RICCIOLI, C.; GARRIDO-VARO, A.; GUERRERO-GINEL, J. E.. On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems*, **96**, 22-26, 2009.
- ³³ ELSHEREEF, R.; BUDMAN, H.; MORESOLI, C.; LEGGE, R. L.. Fluorescence spectroscopy as a tool for monitoring solubility and aggregation behavior of β -lactoglobulin after heat treatment. *Biotechnology and Bioengineering*, **95-5**, 863-874, 2006.
- ³⁴ FILGUEIRAS, P. R.. **Regressão por vetores de suporte aplicado na determinação de propriedades físico-químicas de petróleo e biocombustíveis.** Tese de Doutorado. Instituto de Química, Universidade Estadual de Campinas, Campinas, 2014, 140 p.
- ³⁵ VALDERRAMA P.; BRAGA J. W.; POPPI R. J.. Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. *J. Agric. Food Chem*, **55-21**, 8331-8338, 2007.
- ³⁶ FILGUEIRAS, P. R.; ALVES, J. C. L.; SAD, C. M. S.; CASTRO, E. V. R.; DIAS, J. C. M.; POPPI, R. J.. Evaluation of trends in residuals of multivariate calibration models by permutation test. *Chemometrics and Intelligent Laboratory Systems*, **133**, 33-41, 2014.
- ³⁷ CUNHA-NETO, A.; OLIVEIRA, E. C. S.; LACERDA-JUNIOR, V.; CASTRO, E. V. R.; ROMÃO, W.; SILVA, S. R. C.; PEREIRA, R. G.; STEN, T.; FILGUEIRAS, P. R.; POPPI, R. J.. Quality control of ethanol fuel: Assessment of adulteration with methanol using ¹H NMR. *Fuel*, **135**, 387-392, 2014.

- ³⁸ SPEIGHT, J. G.. **Handbook of Petroleum Analysis**. New York: John Wiley & Sons, 2001. 489 p.
- ³⁹ FILGUEIRAS, P. R. **Determinação da composição de blends de petróleo utilizando FTIR-ATR e calibração multivariada**. Dissertação de Mestrado. Programa de Pós-graduação em Química, Centro de Ciências Exatas, Universidade Federal do Espírito Santo, Vitória, 2011, 113 p.
- ⁴⁰ ANP. Agência Nacional de Petróleo, Gás Natural e Biocombustíveis. Disponível em: <<http://www.anp.gov.br>> Acesso em: 06 junho 2016
- ⁴¹ CHIMIN, R. Q. F.. **Estudo da Evolução de Cloretos no Processo de Destilação de Petróleo**. Dissertação de Mestrado. Programa de Pós-graduação em Química, Centro de Ciências Exatas, Universidade Federal do Espírito Santo, Vitória, 2013, 126 p.
- ⁴² Annual Book of ASTM Standards, Method D 2892-15. West Conshohocken, PA, USA: American Society for Testing and Materials, 2015.
- ⁴³ Annual Book of ASTM Standards, Method D 5236-13. West Conshohocken, PA, USA: American Society for Testing and Materials, 2013.
- ⁴⁴ Annual Book of ASTM Standards, Method D 664-11a. West Conshohocken, PA, USA: American Society for Testing and Materials, 2011.
- ⁴⁵ PARISOTTO, G.. **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**. Dissertação de Mestrado. Programa de Pós-graduação em Química, Universidade Federal de Santa Maria, Santa Maria, 2007, 105 p.
- ⁴⁶ FILGUEIRAS, P. R.; PORTELA, N. A.; SILVA, S. R. C.; CASTRO, E. V. R.; OLIVEIRA, L. M. S. L.; DIAS, J. C. M.; CUNHA-NETO, A.; ROMÃO, W.; POPPI, R. J.. Determination of Saturates, Aromatics, and Polars in Crude Oil by ¹³C NMR and Support Vector Regression with Variable Selection by Genetic Algorithm. *Energy&Fuels*, **30**, 1972-1978, 2016.

ANEXO 1

Trabalhos Publicados

Fuel 166 (2016) 12–18



Contents lists available at ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel

Quantification of biodiesel in petroleum diesel by ^1H NMR: Evaluation of univariate and multivariate approaches



Natália A. Portela, Emanuele C.S. Oliveira, Alvaro C. Neto, Rayza R.T. Rodrigues, Samanta R.C. Silva, Eustáquio V.R. Castro, Paulo R. Filgueiras*

Center of Competence in Petroleum Chemistry – NCQP, Laboratory of Research and Development of Methodologies for Analysis of Oils – LabPetro, Chemistry Department, Federal University of Espírito Santo, Av. Fernando Ferrari, 514, Vitória, Espírito Santo CEP: 29075-910, Brazil

ARTICLE INFO

Article history:

Received 7 July 2015

Received in revised form 19 October 2015

Accepted 20 October 2015

Available online 27 October 2015

Keywords:

Biodiesel

 ^1H NMR

Partial least squares

Figures of merit

ABSTRACT

The commercially available biodiesel in Brazil is a mixture formed by two different raw materials: oleaginous plants and animal fats. Methanol is the alcohol commonly used in the transesterification of triglycerides, however, due to high ethanol production in Brazil, it can be used by manufacturers, producing a complex commercial blend. In this paper, the biodiesel content in diesel was evaluated using the hydrogen nuclear magnetic resonance (^1H NMR) technique associated with univariate and multivariate calibrations. Two univariate models were evaluated. The first one used the ratio of area of the signals of olefins to aliphatics, and the second used the ratio of the area of the signals of the methyl group from the ester to the aliphatics. The multivariate model was constructed using partial least squares (PLS) regression. The results show that using the area of the signals related to the olefins method presented an accuracy similar to the PLS model. However, the multivariate method has a better analytical sensitivity and lower limits of detection and quantification. Using only part of the ^1H NMR spectrum to construct the model can be less efficient than using a broader set of information encompassing the full spectrum.

© 2015 Elsevier Ltd. All rights reserved.



Contents lists available at ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel

Quality control of ethanol fuel: Assessment of adulteration with methanol using ^1H NMR



Álvaro C. Neto^a, Emanuele C.S. Oliveira^a, Valdemar Lacerda Jr.^{a,*}, Eustáquio V.R. Castro^a, Wanderson Romão^{a,b}, Renzo C. Silva^a, Roni G. Pereira^c, Tiago Sten^d, Paulo R. Filgueiras^c, Ronei J. Poppi^e

^a NMR Laboratory, Center of Competence in Petroleum Chemistry-MCQP, Laboratory of Research and Development of Methodologies for Analysis of Oils-LabPetro, Chemistry Department, Federal University of Espírito Santo, Av. Fernando Ferrari, 514, Vitória, Espírito Santo CEP: 29075-910, Brazil

^b Federal Institute of Education, Science and Technology of Espírito Santo, 29106-010 Vila Velha, ES, Brazil

^c Centro Leste Faculty, Rodovia ES 010 – Km 6 – Manguihuas, Serra, Espírito Santo CEP: 29173-087, Brazil

^d Vila Velha University, Avenida Comendador José Dantas de Melo, n°21, Boa Vista, Vila Velha, Espírito Santo CEP: 29162-920, Brazil

^e Chemistry Institute, State University of Campinas, POB: 6154, 13084-971 Campinas, Brazil

HIGHLIGHTS

- Development of the new analytical methodology to control the quality of HEAF.
- ^1H NMR is used to detect and quantify the presence of methanol as adulterant of HEAF.
- Adulterated and unadulterated samples were classified in high precision and accuracy.
- Discriminant Analysis coupled with PCA-DA is applied to ^1H NMR data.

ARTICLE INFO

Article history:

Received 13 May 2014

Received in revised form 6 July 2014

Accepted 8 July 2014

Available online 21 July 2014

Keywords:

NMR

Ethanol fuel

Adulteration with methanol

ABSTRACT

Recently in Brazil, a large number of severe cases of adulteration of hydrated ethyl alcohol fuel (HEAF) with methanol have been observed, causing losses to the tax authorities and consumers. A rapid and reliable analysis is demanded to assess the presence of this contaminant, whose physical and chemical properties are very similar to HEAF. The ^1H Nuclear Magnetic Resonance (^1H NMR) is evaluated to detect and quantify the presence of methanol as adulterant of HEAF using univariate and multivariate calibration. For univariate calibration, the ratio of the methyl area of methanol by the methylene group area was used and methanol content was easily quantified. Additionally, adulterated and unadulterated samples were successfully classified with high precision and accuracy when Discriminant Analysis coupled with Principal Component Analysis (PCA-DA) model are applied to ^1H NMR data.

© 2014 Elsevier Ltd. All rights reserved.