

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

Calibração multivariada associada à ressonância magnética nuclear na caracterização de petróleos

Multivariate calibration and nuclear magnetic resonance for petroleum characterization

Andressa Pagotto Vieira

Dissertação de Mestrado em Química

**Vitória
2019**

Andressa Pagotto Vieira

Calibração multivariada associada à ressonância magnética nuclear na caracterização de petróleos

Dissertação 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 Mestre em Química

Área de Concentração: Química

Linha de Pesquisa: Química do Petróleo e Biocombustíveis

Orientador: Prof. Dr. Paulo Roberto Filgueiras

Co-orientador(a): Prof. Dr. Álvaro Cunha Neto

**VITÓRIA
2019**

V657c Vieira, Andressa Pagotto, 1987-
Calibração multivariada associada à ressonância magnética
nuclear na caracterização de petróleos / Andressa Pagotto Vieira. -
2019.
68 f. : il.

Orientador: Paulo Roberto Filgueira.
Coorientador: Alvaro Cunha Neto.
Dissertação (Mestrado em Química) - Universidade Federal do
Espírito Santo, Centro de Ciências Exatas.

1. Petróleo. 2. Ressonância magnética nuclear. 3.
Quimiometria. 4. Calibração multivariada. I. Filgueira, Paulo
Roberto. II. Cunha Neto, Alvaro. III. Universidade Federal do
Espírito Santo. Centro de Ciências Exatas. IV. Título.

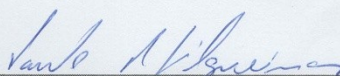
CDU: 54

“Calibração multivariada associada à Ressonância Magnética Nuclear na caracterização de petróleos”

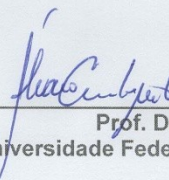
Andressa Pagotto Vieira

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

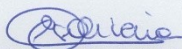
Aprovada em 22/02/2019 por:



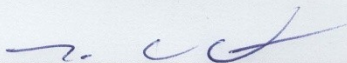
Prof. Dr. Paulo Roberto Filgueiras
Universidade Federal do Espírito Santo
Orientador



Prof. Dr. Álvaro Cunha Neto
Universidade Federal do Espírito Santo
Coorientador



Profª. Drª. Emanuele Catarina da Silva Oliveira
Instituto Federal do Espírito Santo



Prof. Dr. Eustáquio Vinícius Ribeiro de Castro
Universidade Federal do Espírito Santo

Dedico à minha família, que sempre acreditou em mim

AGRADECIMENTOS

A Deus, que sempre me sustentou em todos os momentos.

Aos meus pais, Rosa e Élio, pelos ensinamentos e amor que me transmitiram.

Agradeço principalmente à minha mãe, pela dedicação, apoio e incentivo.

Ao meu irmão André pelo carinho e compreensão.

À minha amiga Ívina que me incentivou e esteve junto em todos os momentos.

Obrigada pelo carinho, paciência e ensinamentos. Gratidão por acreditar em mim.

Ao Fred pelo amor e companheirismo.

Às minhas amigas Natália Portela e Arielle pelo companheirismo e ajuda. Estar com vocês tornou essa jornada mais leve.

Aos amigos do Laboratório de Quimiometria do LabPetro/UFES pela colaboração e principalmente pela amizade: Érica, Márcia, Betina, Pedro, Éllisson, Gabriela, Francine, Layla e Mariana.

Ao meu orientador, Prof. Dr. Paulo Roberto Filgueiras, pelos ensinamentos, paciência e confiança.

À banca de qualificação, os professores Prof. Dr. Valdemar e Prof. Dr. Eustáquio pelas correções e contribuições.

À banca de defesa, Prof. Dr. Eustáquio e Profa Dra. Emanuele como membros titulares por aceitarem o convite.

Ao Cenpes/Petrobras, pelas análises que possibilitaram realizar este trabalho.

Ao Laboratório de RMN, Laboratório de Caracterização do Labpetro/UFES, pelos recursos que permitiram a execução deste trabalho.

À CNPQ, CAPES e FAPES, pelo apoio financeiro.

Há momentos em que as tribulações acontecem em nossas vidas e não podemos evitá-las. Mas estão ali por algum motivo. Só quando às ultrapassamos, entendemos por que estavam ali.
(Paulo Coelho)

LISTA DE FIGURAS

Figura 1. Faixa dos parâmetros de Solubilidade para solventes e não-solventes em 0,1g de petróleo/25mL do líquido. Fonte: (WIEHE; KENNEDY, 2000).....	24
Figura 2: Níveis de energia dos núcleos de hidrogênio quando aplicado um campo magnético de magnitude B_0	28
Figura 3. Esboço do algoritmo do MPA. Adaptado pela Autora.....	37
Figura 4. Esquema da obtenção e tratamento dos dados das amostras de petróleos.....	42
Figura 5. Espectro de RMN de ^1H : (a) sem alinhamento (b) depois do alinhamento pelo algoritmo <i>icoshift</i>	46
Figura 6. Espectro de RMN de ^1H de três petróleos: (a) leve, (b) médio e (c) pesado.....	47
Figura 7. Histograma das amostras de petróleos: (a) fator de caracterização; (b) nitrogênio total; (c) parâmetro de solubilidade; (d) enxofre total; (e) ponto de fluidez mínimo; (f) ponto de fluidez máximo; viscosidade a (g) 20°C; (h) 30°C; (i) 40°C; (j) 50°C.....	49
Figura 8. Valores de R^2 para as propriedades de interesse dos modelos PLS, iPLS, siPLS, OPLS e MPA: 1) fator de caracterização, 2) nitrogênio total, 3) enxofre total, 4) parâmetro de solubilidade, 5) ponto de fluidez mínimo, 6) ponto de fluidez máximo, viscosidade cinemática à 7) 20°C, 8) 30°C, 9) 40°C e 10) 50°C.....	52
Figura 9. Gráfico dos valores de referência e previstos para as seguintes propriedades: (a) fator de caracterização; (b) enxofre total; (c) log viscosidade cinemática à 40 °C; (d) log viscosidade cinemática à 50 °C; (e) ponto de fluidez mínimo; (f) ponto de fluidez máximo; (g) parâmetro de solubilidade e (h) nitrogênio total.....	58

LISTA DE TABELAS

Tabela 1. Parâmetros moleculares médios de RMN de ^1H	30
Tabela 2. Quantidade de amostras de calibração e teste para modelagem quimiométrica.....	44
Tabela 3. Valores de RMSEC e RMSEP dos modelos PLS, iPLS, siPLS e OPLS de predição das propriedades de interesse a partir de espectros de RMN de ^1H	51
Tabela 4: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Fator de caracterização.....	53
Tabela 5: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Viscosidade cinemática à 20°C.....	54
Tabela 6: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Viscosidade cinemática à 30°C.....	55
Tabela 7: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Viscosidade cinemática à 40°C.....	55
Tabela 8: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Viscosidade cinemática à 50°C.....	55
Tabela 9: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Ponto de fluidez mínimo.....	56
Tabela 10: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Ponto de fluidez máximo.....	57
Tabela 11: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Teor de nitrogênio total.....	58

LISTA DE ABREVIATURAS E SIGLAS

ANP	Agência Nacional do Petróleo, Gás e Biocombustíveis.
API	Instituto Americano de Petróleo, do inglês <i>American Petroleum Institute</i> .
ASTM	Sociedade americana de testes de materiais, do inglês <i>American Society for Testing and Materials</i> .
ATR-FTIR	Espectroscopia no infravermelho por transformada de Fourier com reflectância total atenuada, do inglês <i>Fourier Transform Infrared – Attenuated Total Reflectance Spectroscopy</i> .
CENPES	Centro de Pesquisa Leopoldo Americo Miguez de Mello
CV	Validação cruzada, do inglês <i>cross-validation</i> .
DS	Destilação simulada
ER	Erro relativo
FT-IR	Espectroscopia no infravermelho por Transformada de Fourier, do inglês <i>Fourier Transform Infrared</i> .
FT-ICR MS	Espectrometria de massas de ressonância ciclôtrônica de íons por transformada de Fourier
GC	Cromatografia gasosa.
HCA	Análise hierárquica de agrupamentos, do inglês <i>Hierarchical Cluster Analysis</i>
HTGC	Cromatografia gasosa de alta temperatura, do inglês <i>High Temperature Gas Chromatography</i>
IF	Infravermelho
ISO	Organização Internacional de Padronização, do inglês <i>International Organization for Standardization</i>
iPLS	Mínimos quadrados parciais por intervalos
kPa	Unidade de pressão Pascal.
KUOP	Fator de caracterização <i>K</i>
LabPetro	Laboratório de Pesquisa e Desenvolvimento de Metodologias para Análise de óleos Pesados.
LDA	Análise de discriminante linear
MCS	Amostragem Monte Carlo, do inglês <i>Monte Carlo sampling</i>
MPA	Análise de modelo populacional, do inglês <i>Model population analysis</i>
MSC	Correção multiplicativa de sinal, do inglês <i>Multiplicative Scatter</i>

Correction.

°C	Graus Celsius.
OPLS	Projeções ortogonais a estruturas latentes, do inglês <i>Orthogonal projections to latent structures</i>
OPLS-DA	Projeções ortogonais a estruturas latentes com análise discriminante, do inglês <i>orthogonal partial least squares discriminant</i>
OSC	Correção de sinal ortogonal
PCA	Análise por componentes principais, do inglês <i>Principal Component Analysis</i>
PCR	Regressão por componentes principais, do inglês <i>Principal Component Regression</i>
PLS	Regressão por Mínimos Quadrados Parciais, do inglês <i>Partial Least Squares</i>
R ²	Coeficiente de determinação.
RLM	Regressão linear múltipla, do inglês <i>Multiple linear regression</i>
RMN	Ressonância magnética nuclear
RMN de ¹³ C	Ressonância magnética nuclear de carbono 13
RMN de ¹ H	Ressonância magnética nuclear de hidrogênio
RMSEC	Raiz quadrada do erro quadrático médio do conjunto de calibração, do inglês <i>Root Mean Square Error of Calibration.</i>
RMSECV	Raiz quadrada do erro quadrático médio de validação cruzada, do inglês <i>root mean square error of cross validation.</i>
RMSEP	Raiz quadrada do erro quadrático médio do conjunto de previsão, do inglês <i>Root Mean Square Error of Prediction.</i>
SARA	Método cromatográfico de separação em classes: Saturados, Aromáticos, Resinas e Asfaltenos.
SIMCA	Modelagem independente flexível por analogia de classe, do inglês <i>Soft independent modelling by class analogy</i>
siPLS	Mínimos quadrados parciais por sinergismo
SNV	variação normal padral, do inglês, <i>Standard Normal Variate.</i>
SVR	Regressão por vetor de suporte, do inglês <i>Support Vector Regression.</i>
TAN	número de acidez total

TIAC	Temperatura inicial de aparecimento de cristais.
VL	Variável latente, do inglês <i>latent variables</i>

LISTA DE SÍMBOLOS

R^2	coeficiente de determinação
R^2_p	coeficiente de determinação de predição
b_{PLS}	coeficientes de regressão do modelo PLS
W	<i>coeficientes ponderados (pesos)</i>
Y	matriz ou vetor com a propriedade de interesse
$\bar{y}_{previsão}$	média dos valores de previsão
n_{prev}	número de amostras de previsão (ou teste)
T	temperatura
S	valores de <i>scores</i> da matriz X
U	valores de <i>scores</i> da matriz Y
y_{ref}	valores de referência de calibração
P	valores dos pesos (do inglês, <i>loadings</i>) da matriz X
Q	valores dos pesos (do inglês, <i>loadings</i>) 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

O uso da espectroscopia de ressonância Magnética Nuclear (RMN) associada com ferramentas quimiométricas tem se tornado uma metodologia eficiente para estimar propriedades físico-químicas de petróleos brasileiros com sua elevada precisão, rapidez e confiabilidade. Neste trabalho, a técnica de RMN de hidrogênio (RMN de ^1H) associada à calibração multivariada PLS, OPLS e seleção de variáveis (iPLS, siPLS e MPA) foi utilizada para estimar o fator de caracterização, teor de nitrogênio total, parâmetro de solubilidade, teor de enxofre total, ponto de fluidez máximo e mínimo e viscosidade cinemática à 20°C, 30°C, 40°C e 50°C de 138 amostras de óleo bruto brasileiro. Os modelos foram avaliados pelo coeficiente de determinação (R^2), erro quadrático médio de previsão (RMSEP) e resíduos. O teor de enxofre total e o parâmetro de solubilidade tiveram suas estimativas melhoradas por PLS e OPLS, respectivamente, com um RMSEP de 0,09 wt% (enxofre total) e 1,06 (MPa)^{1/2} (parâmetro de solubilidade). O MPA apresentou melhores resultados na estimativa do fator de caracterização, teor de nitrogênio total, viscosidade cinemática à 40°C e ponto de fluidez mínimo e máximo com RMSEP de 0,06, 0,048 wt%, 0,036, 14,6°C e 19,2°C respectivamente. Os modelos mínimo e máximo para o ponto de fluidez tiveram os desempenhos menos satisfatórios, talvez devido a complexidade desta propriedade. Com resultados bem promissores, o MPA tem mostrado que é uma técnica quimiométrica inovadora.

Palavras-chave: petróleo, RMN, PLS, OPLS, iPLS, siPLS e MPA

ABSTRACT

The use of Nuclear Magnetic Resonance (NMR) spectroscopy with chemometrics tools has become an efficient methodology for estimating physicochemical properties of Brazilian oils with their high precision, speed and reliability. In this work, the hydrogen NMR (^1H NMR) technique associated to the multivariate calibration PLS, OPLS and variable selection (iPLS, siPLS, and MPA) was used to estimate the UOP characterization factor K (KUOP), nitrogen content, solubility parameter, sulfur content, maximum and minimum pour point and kinematic viscosity at 20 °C, 30 °C, 40 °C and 50 °C of 138 Brazilian crude oil samples. The models were evaluated by coefficient of determination (R^2), root mean square error of predicted (RMSEP) and the residuals. The sulfur content and the solubility parameter had their estimates improved by PLS and OPLS, respectively, with a RMSEP of 0.09 wt% (sulfur content) and 1.06 (MPa)^{1/2} (solubility parameter). Model population analysis presented better results in the estimation of KUOP, nitrogen content, kinematic viscosity at 40 °C and pour point with prediction mean error of 0.06, 0.048 wt%, 0.036 and (14.6 °C and 19.2 °C) respectively. Minimum and maximum pour point models had the weakest performances; suggesting the complexity of this property. With very promising results, MPA has shown that it is an innovative chemometrics technique.

Keywords: crude oil; NMR; PLS; OPLS; MPA.

SUMÁRIO

1.Introdução	16
2.Fundamentos Teóricos.....	23
2.1.Caracterização Petróleo.....	23
2.2.RMN e petróleo.....	27
2.2.1.Ressonância magnética nuclear.....	27
2.2.2.RMN na caracterização de petróleos.....	29
2.3.Quimiometria.....	31
2.3.1..PLS.....	32
2.3.2.OPLS.....	34
2.3.3.Seleção de Variáveis.....	35
2.3.3.1.iPLS e siPLS.....	35
2.3.3.2.MPA.....	36
2.3.4.Pré-tratamento de dados.....	38
2.3.5. Avaliação dos Modelos.....	39
3.Objetivos.....	41
3.1.Objetivos Gerais.....	41
3.2.Objetivos Específicos.....	41
4.Metodologia.....	42
4.1.Ensaio espectroscópicos de RMN de ^1H	43
4.2.Modelos de Calibração.....	43
5.Resultados e Discussão.....	45
5.1.Análise do espectro de RMN de ^1H	45
5.2.Calibração Multivariada.....	50
6.Conclusão.....	60
7.Referências Bibliográficas.....	61

1.INTRODUÇÃO

Petróleo é uma mistura complexa de hidrocarbonetos e entidades de composição oxigenada, nitrogenada, sulfurada e metálica.¹ Sua qualidade e valor de mercado são determinados através de ensaios de caracterização físico-química. Conhecer estas propriedades é importante para prever possíveis mudanças que possam ocorrer na composição dos óleos afim de auxiliar nas etapas de recuperação, processamento primário, transporte e avaliação dos petróleos.² Os métodos atuais padrão para caracterizar petróleos são embasados em uma série de testes laboriosos e envolvem cerca de 700 ensaios físico-químicos. Na realização de todos os ensaios usa-se uma grande quantidade de amostras de óleo (de 10 a 70 litros) durante um longo espaço de tempo (não menos de 4 meses), com isso, o desenvolvimento de metodologias analíticas rápidas e que carecem de menores quantidades de amostras se tornam atrativas visando agilizar decisões operacionais.³

O principal problema associado aos métodos padrões utilizados para a caracterização de petróleos é o longo tempo para realizar as análises. O uso de tecnologias conhecidas capazes de caracterizar eficientemente petróleos brutos de forma que contribua para a redução do tempo das análises já se encontram na literatura. Algumas dessas técnicas são: espectroscopia na região do infravermelho, ressonância magnética nuclear e cromatografia gasosa.⁴⁻⁷ A espectroscopia no infravermelho é a ferramenta alternativa mais utilizada para a análise de petróleos e seus derivados. É um procedimento rápido onde são utilizadas pequenas quantidades de amostra, demandando o mínimo de pré-tratamento.^{3,9-10} O infravermelho possui algumas limitações associadas à técnica, como a sobreposição que ocorre entre as faixas de frequência que pode ser superada pelo uso da espectroscopia de ressonância magnética nuclear (RMN). A RMN oferece uma melhor resolução espectral com a capacidade de fornecer informações sobre a natureza química de tipos individuais de átomos como o hidrogênio e o carbono em misturas diferentes e complexas de óleos brutos.^{4,11}

A aplicação da espectroscopia de RMN na indústria de petróleos está crescendo em todo mundo na identificação e quantificação de diferentes tipos

de hidrocarbonetos, fornecendo informações sobre a composição de alifáticos, olefínicos e aromáticos.¹¹⁻¹² Abrange uma série de amostras de óleos, a partir da caracterização de frações mais leves a frações mais pesadas (resinas, asfaltenos)¹³, permitindo análise rápida, não destrutiva e quantitativa.¹⁴

Uma das características mais atrativas dos modernos instrumentos analíticos, como a Ressonância Magnética Nuclear, é a quantidade de informações que podem ser obtidas em um único ensaio. O espectro de RMN de ^1H de petróleos é muito complexo por conter um grande número de sinais altamente sobrepostos e aglomerados devido à distribuição dos deslocamentos químicos.¹⁵ Com isso, torna-se necessária a utilização de ferramentas sofisticadas para tratar e extrair as informações químicas mais relevantes.¹⁶⁻¹⁷

Em 2007, Molina e colaboradores¹⁵ publicaram o primeiro artigo que associava espectros de RMN de diferentes óleos crus colombianos com calibração multivariada (foi utilizada a regressão por mínimos quadrados parciais - PLS). Foi desenvolvido um método no qual o espectro foi dividido em 12 regiões integrais dos sinais de ressonância que foram correlacionadas com as propriedades físico-químicas (densidade, gravidade API, número de acidez total, teor de enxofre, ponto de fluidez, dentre outras). Como as regiões integrais do espectro abrange grupos atômicos bem definidos, foi possível obter valores elevados de correlação para todos os modelos PLS utilizados. Os valores de R^2 dos modelos de predição para as propriedades físico-químicas foram maiores que 0,95.

Já em 2010, Molina e colaboradores¹⁸ correlacionaram os espectros de RMN e as propriedades de interesse com PLS e RLM para prever o ponto de ebulição, densidade, índice de octanagem da gasolina e composição SARA (Saturados, Aromáticos, Resinas e Asfaltenos). Os modelos de previsão para SARA obtidos por PLS tiveram $R^2 > 0,97$. Os modelos obtidos com RLM para SARA também mostraram bons resultados, porém para asfaltenos o R^2 foi razoável (0,83). O R^2 para as outras propriedades foram maiores que 0,98 para ambos os modelos. Neste estudo os autores mostraram que diferentes ferramentas quimiométricas podem ser eficientes, tornando os métodos de correlação rápidos e menos onerosos que os convencionais.

Masili e colaboradores (2012)¹⁹ utilizaram amostras de petróleos extra-leves a extra-pesados de 28 diferentes campos de extração. Foi adotada uma

estratégia de pré-tratamento de dados espectrais para caracterizar os petróleos brutos de maneira rápida, confiável e econômica. Antes da modelagem multivariada a área espectral de 0,22 a 9 ppm foi dividida em regiões fixas de 0,44 ppm, enquanto os dados entre 4,20 e 5,90 ppm foram omitidos para evitar a inclusão do sinal da água e a região do solvente. Tal procedimento foi preferido ao método alternativo pois usa pontos dos dados originais para reduzir a dimensão dos dados da matriz. A área total foi normalizada e a matriz de dados foi centrada na média, implicando que o espectro médio foi subtraído de cada espectro individual. As principais propriedades geralmente empregadas para avaliar a qualidade e comportamento dos petróleos durante o refino estimadas por PLS foram: fator de caracterização KUOP, densidade, número de acidez total, teor de enxofre e curva PEV, obtendo-se resultados úteis para previsões. A previsão destas propriedades resultou estar na faixa com os métodos ASTM correspondentes. Escolher o pré-tratamento correto dos dados, definir o número adequado de amostras para cada conjunto de dados e avaliar a dimensionalidade do modelo foram questões consideradas e melhoraram a robustez do mesmo, com resultados de R^2 acima de 0,85.

Ainda segundo Masili¹⁹, durante a análise de RMN as amostras devem ser coletadas e preparadas da forma mais homogênea possível, pois vários fatores físico-químicos como variação do pH podem ocasionar desalinhamento espectral, afetando a reprodutibilidade do modelo. Para a construção do modelo de calibração multivariada PLS se torna interessante alinhar os espectros de RMN de ^1H para corrigir pequenos desvios nos deslocamentos químicos causados pelo ambiente ao qual os hidrogênios são submetidos. Estes pequenos deslocamentos nas variáveis podem fazer com que o modelo PLS exija um número maior de componentes para explicar a mesma variação causada pelo deslocamento de um pico no espectro, entretanto, esse alinhamento não interfere na análise espectral.

Um marco importante na calibração multivariada com espectros de RMN foi o desenvolvimento do algoritmo *icoshift* (do inglês, *interval correlation shifting*), uma ferramenta computacionalmente eficiente e versátil para o alinhamento de sinais de RMN em grandes conjuntos de dados espectrais desenvolvida por Savorani e colaboradores (2010)²⁰. Nos estudos de Savorani e colaboradores, três conjuntos de dados experimentais de RMN com vários problemas

de alinhamento a serem resolvidos foram utilizados para avaliar o desempenho do algoritmo *icoshift*. No primeiro caso foram usados dados de vinhos para prever ácido láctico: o RMSECV baixou de 0,351 para 0,177 e o R^2 melhorou de 0,79 para 0,95. O segundo caso foi com dados de banho de níquel de galvanoplastia, que é particularmente desafiador devido as múltiplas mudanças de sinal presentes que são de magnitude muito diferentes, dependendo da suscetibilidade do pH das moléculas. Para superar essa limitação, o algoritmo *icoshift* foi implementado para encontrar a mudança permitida de cada intervalo, tendo todos os intervalos selecionados bem alinhados e o alinhamento corrigido de maneira ideal. O caso 3 são de dados metabonômicos de plasma de rato e foram adquiridos sob condições experimentais estritamente definidas por uma lista de protocolos, limitando o máximo risco possível de adquirir espectros desalinhados. Os espectros foram alinhados de acordo com os sinais que não foram afetados pelo pH. Na prática, o *icoshift* procurou a melhor correlação cruzada local para a região de referência fornecida e, em seguida, deslocou todos os espectros para a esquerda ou para a direita, de acordo com os índices de deslocamento encontrados. Embora a correção seja pequena, melhorou a homogeneidade do conjunto de dados e ajudou na interpretação dos resultados de uma análise de dados multivariada subsequente de forma rápida. Logo, o uso do *icoshift* no alinhamento do sinal do espectro melhora significativamente o modelo PLS, tendo uma melhor correlação entre os valores medidos e previstos, além de reduzir os valores do erro quadrático médio de validação cruzada (RMSECV) e melhorar a exatidão do modelo.

Estudos recentes como o de Filgueiras e colaboradores (2015)¹⁷ utilizaram o alinhamento *icoshift* em espectros de RMN de ^1H e Regressão por Vetores de Suporte (SVR) para estimar a temperatura equivalente a 10%, 50% e 90% do volume destilado em óleos crus. Intervalos de confiança para os valores previstos foram calculados e comparados para os diferentes métodos adotados (PLS, ePLS, SVR, eSVR). O procedimento eSVR melhorou a precisão das previsões da temperatura de destilação em relação ao padrão PLS, ePLS e SVR (exemplo, R^2 à 10%: PLS – 0,7972, ePLS – 0,8047, SVR – 0,6802 e SVR – 0,8949). O estudo de diferentes metodologias de calibração multivariada mostram a importância em se utilizar algoritmos distintos para estimar propriedades de matrizes complexas como os petróleos.

Duarte e colaboradores (2016)⁵ desenvolveram uma metodologia associando espectros de RMN de ^1H e calibração multivariada (PLS) para estimar as propriedades densidade API, resíduo de carbono, nitrogênio orgânico básico e TIAC (temperatura inicial de aparecimento de cristais) a fim de economizar tempo e quantidade de amostra. A RMN vem ganhando proeminência para análise de petróleos brutos a nível qualitativo e quantitativo na investigação de propriedades físico-químicas em misturas complexas. Ou seja, a partir do espectro de RMN de ^1H é possível estimar diferentes propriedades de petróleos brutos quando aplicada a modelagem quimiométrica. Neste trabalho, antes da modelagem quimiométrica, foi feito o alinhamento dos espectros com o algoritmo *icoshift*. Após o alinhamento foi realizada uma avaliação completa sobre o conjunto de amostras para verificar a ocorrência de *outliers* espectrais (a remoção da amostra anômala é muito importante para evitar problemas na modelagem, diminuindo a capacidade preditiva dos modelos). Nesse contexto, a análise de componentes principais (PCA) foi efetuada. Foram obtidos R^2 de 0,945 (densidade API), 0,802 (resíduo de carbono), 0,857 (TIAC) e 0,789 (nitrogênio orgânico básico).

Duarte e colaboradores (2017)²¹ apresentaram mais metodologias alternativas na caracterização de petróleos brutos. Nesse estudo estimou-se 18 temperaturas de destilação de 5 a 90% de evaporados de petróleos brasileiros destilados usando RMN de ^1H associado a regressão multivariada PLS confrontando com dados de destilação simulada (DS). Enfatizou-se novamente que o alinhamento com *icoshift* é um passo importante quando a RMN é utilizada para fins multivariados e que, além do alinhamento, o procedimento de pré-tratamento do espectro também se faz necessário. O pré-tratamento utilizado foi a variação normal padrão (SNV)²² em que se tem um melhor ajuste de linha de base, além de modelos mais precisos e parcimoniosos. O SNV atenua substancialmente as variações espectrais causadas principalmente por interferentes da matriz. O RMSEP foi o parâmetro de modelagem mais importante monitorado, mostrando um comportamento decrescente ao longo do intervalo estudado e apresentou valores inferiores a 5% para alguns modelos.

As últimas pesquisas de RMN associada a métodos quimiométricos mostram que a técnica é rápida e eficaz para estimar propriedades físico-químicas de petróleos e derivados. Porém, com o decorrer do desenvolvimento cien-

tífico, tem-se a necessidade de melhorar o desempenho das metodologias. Na quimiometria, um avanço nos estudos de calibração multivariada engloba a seleção de variáveis. Técnicas de seleção de variáveis como iPLS e siPLS fornecem uma visão geral da informação relevante em diferentes subdivisões espectrais focando nas mais importantes regiões e removendo as interferências das outras regiões.²³ No trabalho de Ferrão e colaboradores²⁴, os algoritmos iPLS e siPLS foram capazes de selecionar a região espectral mais adequada para cada propriedade estudada do biodiesel (gravidade específica, teor de enxofre e ponto de fulgor onde o algoritmo siPLS produziu modelos melhores do que o PLS).

Um outro método de seleção de variáveis usado foi o modelo de análise populacional (MPA). O MPA foi proposto pela primeira vez em 2010 pelo grupo de pesquisas do Prof. Liang²⁵ e é uma estrutura para o desenvolvimento de novos tipos de algoritmos para melhorar a modelagem de sistemas analíticos complexos utilizando ferramentas estatísticas (coeficiente de regressão, erros de previsão, viés) para extrair informações importantes do modelo.²⁶⁻²⁹ O MPA tem mostrado ser útil para o desenvolvimento de algoritmos para detecção de outliers³⁰, seleção de variáveis²⁵ e avaliação de modelo.³¹

Afim de se obter um método com melhor capacidade de previsão, Deng e colaboradores (2015)²⁷ propuseram uma estratégia para selecionar o melhor número de variáveis latentes nos modelos PLS com base no MPA. O conjunto de 22 dados de Ultra Violeta de uma mistura contendo 4-hidroxibenzidrazida, 8-hidroxiquinolina, ácido 3,4-di-hidroxibenzóico e 2-mercaptobenzimidazole, 128 dados de NIR da carragenina e 54 dados de NIR de soja foram usados para investigar o método. Um critério que combina a capacidade de previsão e a estabilidade do modelo é usado para selecionar o número ideal de variáveis latentes. A capacidade de previsão do modelo é medida pela validação cruzada e a estabilidade do modelo é medida pela reprodutibilidade de vetores de regressão em sub-modelos. Combinado com outros métodos baseados em vetores de regressão, esse método tem algumas vantagens: (1) a informação da capacidade de previsão e da estabilidade do modelo são obtidas através de múltiplos modelos em vez de um modelo único; (2) é aplicável aos dados com variáveis discretas; (3) é mais sensível ao super-ajuste. A estratégia também é apli-

cável para evitar o ajuste excessivo de outros modelos como PCR e outros tipos de dados como de relação quantitativa estrutura- atividade (QSAR).

Deng e colaboradores (2018)²⁹ resumiram algumas aplicações do MPA no campo da avaliação de modelos. A avaliação dos modelos é frequentemente realizada comparando diferentes modelos em um conjunto de teste independente. Um dos algoritmos de seleção de amostras usado com mais frequência é o método Kennard-Stone³² (geralmente 30% teste e 70% calibração) por garantir o máximo da variabilidade do conjunto de amostras. Métricas individuais como a raiz da média quadrática dos resíduos de previsão (RMSEP) frequentemente aplicada no conjunto de testes independente para comparação de modelos podem ser vulneráveis devido à sua dependência na divisão de amostra. Com a ajuda do MPA, obteve-se um valor de RMSEP para cada modelo usando sub-conjuntos de dados gerados aleatoriamente. Em comparação com métricas únicas, a aplicação de várias métricas reduziu a probabilidade de tirar conclusões equivocadas. Usou-se análises estatísticas como o teste t-Student para avaliar a significância da diferença do RMSEP entre os modelos²⁶ mostrando que a técnica é realmente promissora, sendo possível obter informações adicionais importantes durante a avaliação com uma melhora no desempenho do modelo.

2.FUNDAMENTOS TEÓRICOS

2.1.Caracterização do petróleo

A avaliação do petróleo é de fundamental importância para auxiliar na elaboração de estratégia de recuperação, transporte, avaliação e sobre os potenciais derivados.¹ A avaliação do óleo cru pode envolver mais 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.³ Nesse trabalho foram previstas as seguintes propriedades físico-químicas do petróleo:

a) Parâmetro de Solubilidade

No Brasil, com a descoberta de novos campos com grande potencial de exploração, as indústrias petrolíferas tem a necessidade em misturar óleos de diferentes origens para viabilizar o processamento da matéria prima além de aumentar o volume total de petróleo quando se deseja obter um maior rendimento dos poços com baixa produção.³³ Um dos trabalhos desenvolvidos a respeito da possibilidade das misturas de petróleos ou suas frações é o proposto por Wiehe e Kennedy (2007)³⁴, no qual Mostrou que misturas de petróleos de diferentes origens podem ocasionar a precipitação dos asfaltenos e com isso, o início de precipitação dos asfaltenos está bem relacionado com o parâmetro de solubilidade e o parâmetro de floculação dos petróleos.³⁵ O parâmetro de solubilidade de Hildebrand³⁴ é utilizado para expressar a compatibilidade de petróleos em misturas, partindo-se do princípio que, independentemente da composição do sistema, os asfaltenos precipitam sempre em um determinado parâmetro de solubilidade, conhecido como parâmetro de floculação, cujo valor referenciado é $16,35 \text{ (MPa)}^{1/2}$. O parâmetro de solubilidade do petróleo está situado em uma faixa entre $15,95 \text{ (MPa)}^{1/2}$ e $18,20 \text{ (MPa)}^{1/2}$.³⁴

A Figura 1 representa essa faixa para os petróleos em solventes e não solventes. Na parte da escala abaixo de $15,95 \text{ (MPa)}^{1/2}$ os asfaltenos permanecem insolúveis em *n*-alcanos e metilciclohexano. Acima de $18,2 \text{ (MPa)}^{1/2}$ são solúveis a ciclohexano entre outros da escala de forma crescente

de potencial de solubilidade.³³⁻³⁴ A faixa de parâmetro de solubilidade proposta por Wiehe e Kennedy tem sido bastante eficiente quando se pretende avaliar o parâmetro de solubilidade de uma amostra de petróleo em termos de determinação do seu início de precipitação. Este parâmetro é muito útil quando se pretende avaliar a possibilidade de mistura de dois ou mais petróleos, sendo que as frações mais leves do petróleo, como por exemplo, nafta, condensado, etc. são considerados no sistema como floculantes.³³

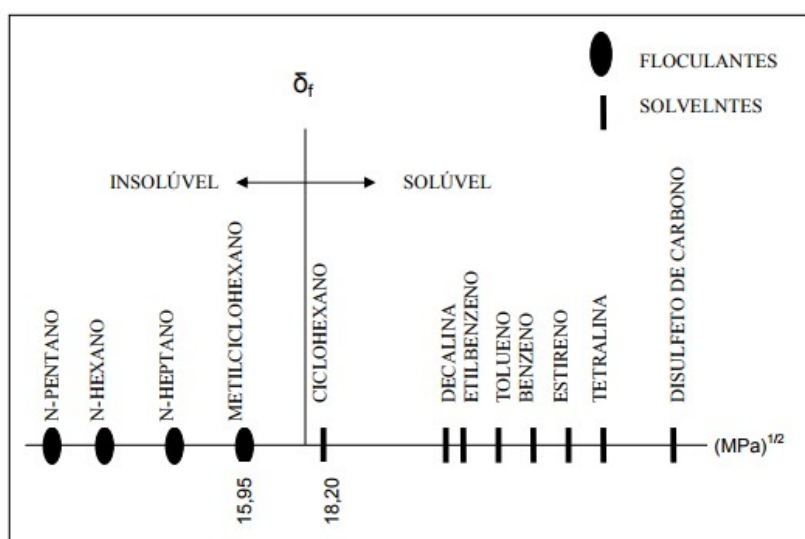


Figura 1. Faixa dos parâmetros de Solubilidade para solventes e não-solventes em 0,1g de petróleo/25mL do líquido.³⁹

b) Fator de Caracterização UOP

É uma classificação utilizada em especial pelos refinadores. Foi estabelecida pelo *U. S. Bureau of Mines* como fator de caracterização UOP (KUOP). O Método UOP 375-59³⁶ é um indicativo da origem e natureza de um petróleo, dando informação quanto ao grupo predominante. O fator de caracterização KUOP é obtido a partir da Equação 1, onde T_B é o ponto de ebulição médio (em °Rankine) e G a densidade específica a 60°F.

$$K_{UOP} = \frac{\sqrt[3]{T_B}}{G} \quad (1)$$

Onde se tem a classificação proposta pela *Universal Oil Products* (UOP)³⁶: $K_{UOP} \geq 12$: petróleos predominantemente parafínicos; $K_{UOP} \leq 10$:

petróleos predominantemente aromáticos; e, $10 < K_{UOP} \leq 11,8$: petróleos com caráter naftênico.

c) Teor de enxofre

A qualidade de muitos produtos petrolíferos está relacionada à quantidade de enxofre presente, sendo esta outra forma de caracterizar o petróleo. O conhecimento da concentração de enxofre é necessário para fins de processamento. O processamento do petróleo no sentido de remoção dos compostos de enxofre é necessário devido aos graves problemas de corrosão associados a presença destes compostos e aos grandes impactos ambientais provenientes da emissão desregrada de compostos de enxofre provenientes de fontes antropogênicas.³⁷⁻³⁸

O teor de enxofre tem-se a seguinte classificação para o óleo bruto: Petróleos “Doces” (*sweet*): teor de enxofre menor que 0,5 wt%; Petróleos “Ácidos” (*sour*): teor de enxofre maior que 0,5 wt%.

d) Ponto de fluidez

O ponto de fluidez é uma importante propriedade avaliada na etapa de produção e é determinada pelo método ASTM D5853³⁹. O método consiste em inserir uma quantidade de amostra de óleo cru em um tubo de vidro acoplado a um termômetro. A amostra é aquecida até 40°C por 20 min e, então, resfriada em um intervalo de 3°C em 3°C até que seja verificada a última temperatura antes da amostra parar de fluir. O ponto de fluidez também é definido como a menor temperatura na qual o óleo flui sob a ação gravitacional. Esta propriedade quando se encontra acima da temperatura ambiente pode ser indício da presença elevada de hidrocarbonetos parafínicos de cadeias longas (para óleos leves), ou seja, é utilizada para determinação da parafinicidade do petróleo, definindo assim as condições de transferência do petróleo nos dutos.⁴⁰ Sad e colaboradores⁴¹ enfatizam que o ensaio de ponto de fluidez é impreciso e apresenta baixa repetitividade e reprodutibilidade (o trabalho segue os critérios de repetitividade de 3°C para o ponto de fluidez máximo e 6°C para o ponto de fluidez mínimo, com 95% de confiança de acordo com a ASTM D5853).

e) Teor de nitrogênio total

Um outro parâmetro importante avaliado é o teor de nitrogênio total. A presença de substâncias nitrogenadas no petróleo afeta economicamente o seu processo de refino e altera a qualidade dos derivados.⁴² O teor de nitrogênio na maioria dos petróleos brasileiros é muito baixo, não excedendo 0,1% em massa. Em alguns óleos crus pesados, no entanto, o teor de nitrogênio pode chegar a até 2,0 wt%.⁴⁰ Os compostos nitrogenados no óleo aumentam sua capacidade de reter água em emulsão e tornam instáveis os produtos do refino, formando gomas e envenenando catalisadores. A presença destes compostos nos combustíveis também é danosa, visto que propiciam o aumento das emissões de poluentes, como os óxidos de nitrogênio.⁴³⁻⁴⁴

f) Viscosidade cinemática

A viscosidade cinemática é a principal propriedade física utilizada no transporte de petróleo. Esta descreve a resistência do fluido ao fluxo, sendo muito utilizada na caracterização do petróleo. A medição desta propriedade reológica possibilita prever o comportamento do escoamento do óleo durante a etapa de extração no reservatório, assim como o dimensionamento dos equipamentos e o planejamento das condições relacionadas às atividades de transferência e processamento.⁴⁵

2.2.RMN e petróleo

2.2.1. Ressonância Magnética Nuclear.

A espectroscopia de RMN como forma de espectroscopia que é, estabelece um exemplo de interação da matéria com a radiação eletromagnética e são os núcleos atômicos da amostra que possuem um momento magnético que intervêm nesse fenômeno.⁸³ Dentre as propriedades fundamentais que os núcleos possuem, o momento magnético (μ) e *spin* nuclear (I).são importantíssimos para a descrição da RMN, conferindo ao núcleo uma propriedade denominada magnetismo, ou seja, infere ao núcleo um desempenho semelhante a pequenos ímãs. 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.⁸²

O núcleo de hidrogênio possui *spin* igual a $\frac{1}{2}$ e existem dois níveis de energia com um excesso de população no estado de energia mais baixo, pois segundo a distribuição de Boltzmann $N_{\alpha} > N_{\beta}$. Os estados de energia α e β são separados por uma diferença de energia ΔE ($\Delta E = E_{\beta} - E_{\alpha}$), onde a energia de cada estado gerada pela aplicação do campo é mostrada na equação 2:⁸⁴

$$E = \frac{-m\gamma h B_o}{2\pi} \quad 2$$

Onde m é o número quântico de *spin* nuclear, h é a constante de Planck

e B_0 é a intensidade do campo magnético aplicado. A direção do campo magnético B_0 é convencionada como eixo z, conforme figura 2.

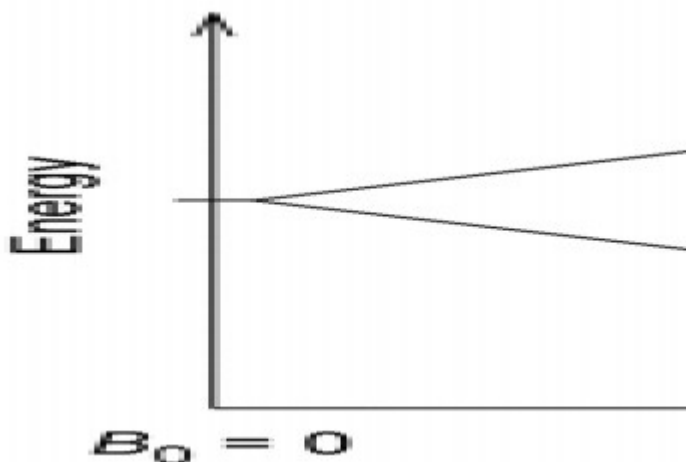


Figura 2: Níveis de energia dos núcleos de hidrogênio quando aplicado um campo magnético de magnitude B_0 .

A razão magnetogírica é uma constante nuclear fundamental. Após a aplicação do campo magnético o núcleo começa a mudar de direção sobre seu próprio eixo de rotação com frequência angular ω ; trata-se de um movimento de precessão. Um conjunto de núcleos equivalentes em movimento aleatório de precessão em torno do eixo-z (direção de B_0) tem magnetização resultante na direção z, sem resultante no plano xy.

Quando a radiofrequência (ν_1) aplicada equivale a frequência de precessão dos núcleos equivalentes (frequência Larmor ν_L) ocorre o efeito de ressonância e a relação básica da RMN pode ser descrita pela equação 3:

$$\nu_L = \nu_1 = \left(\frac{\gamma}{2\pi} \right) B_0 \quad 3$$

No espectrômetro pulsado, a amostra é submetida ao campo magnético e irradiada com pulsos de radiofrequência de alta potência. Esse pulso excita simultaneamente todos os núcleos da amostra, que após o relaxamento produz o decaimento livre de indução (FID), que é a soma da radiação de todos os núcleos durante o tempo da emissão. A informação contida no FID, que é em função do tempo, é convertida para o domínio da frequência pela transformada de *Fourier* e, então, o espectro no domínio da frequência é obtido.⁸⁴

2.2.2. RMN na caracterização de petróleos

Conhecer os parâmetros físico-químicos dos petróleos é importante para apontar possíveis mudanças que possam ocorrer na composição do óleo, afim de ajudar nas etapas de recuperação, processamento primário, transporte e avaliação do petróleo. Os ensaios convencionais demandam muito tempo, grandes volumes de amostra e elevados custos. Analisar de forma rápida se torna necessário para a agilidade de decisões operacionais.

A espectroscopia de Ressonância magnética nuclear de ^1H é um método de rápida análise, não destrutivo, que utiliza pouca quantidade de amostra¹⁴, abrangendo uma série de amostras de óleos, a partir da caracterização de frações mais leves a frações mais pesadas de petróleos (resinas, asfaltenos).¹³

É uma ferramenta analítica de imenso potencial e de grande utilidade para a análise dos produtos petrolíferos, especialmente da classe de combustíveis e está sendo cada vez mais empregada para o controle de qualidade do petróleo e seus produtos⁴⁸ por ser uma técnica que tem se adequadado para a identificação e quantificação de diferentes tipos de hidrocarbonetos, podendo fornecer informações sobre o conteúdo relativo de componentes alifáticos, olefínicos e aromáticos,¹¹ ou seja, proporciona, sozinha, uma grande quantidade de informações acerca da amostra em um único espectro.^{46,47}

Em 2012 houve um grande avanço com os estudos em espectroscopia de RMN e petróleos feitos por Poveda e colaboradores⁴⁸ ao quantificar novos grupos estruturais devido a presença de heteroátomos como: enxofre, nitrogênio e oxigênio em petróleos. Alguns parâmetros estruturais como fator de aromaticidade, número de anéis naftênicos e número de anéis total foram corrigidos ou introduzidos em um método completamente novo de cálculo, o qual promoveu uma melhoria no conjunto completo de parâmetros moleculares médios de frações pesadas de petróleos, asfaltenos, óleos crus, xisto e carvão. Os parâmetros de Poveda tornaram-se um estudo indispensável na pesquisa de RMN e petróleos por fornecerem informações detalhadas que auxiliam na elucidação das estruturas moleculares médias de sistemas aromáticos altamente condensados (Tabela 1). Desde então, alguns autores vêm utilizando

este estudo principalmente para análise composicional de asfaltenos.

Tabela 1. Parâmetros moleculares médios de RMN de ^1H ⁴⁸

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 ácido Carboxílico

No trabalho de Barros e colaboradores (2017)⁴⁹, uma amostra de óleo pesado foi submetida ao processo de degradação térmica objetivando compreender seu comportamento a nível molecular e avaliar suas propriedades físico-químicas em função do tempo de degradação. As análises de RMN ^1H e ^{13}C provaram que para todas as amostras os teores de hidrogênio e carbono alifáticos são maiores do que os teores de aromáticos. No entanto, houve redução de compostos alifáticos, sendo este comportamento principalmente evidenciada no final da degradação térmica.

Molina e colaboradores (2017)⁵⁰ demonstraram que a estrutura química média dos asfaltenos de uma amostra de óleos brutos é única em comparação com outros poços de petróleos através de ensaio de caracterização de RMN de ^1H e ^{13}C . Estudou-se a estrutura química dos asfaltenos para compreender a influência na cristalização de parafinas em óleos crus.

Barros e colaboradores (2018)¹⁴ caracterizaram frações de ácidos naftênicos de óleos brasileiros por RMN. A análise por RMN de ^1H revelou a predominância de alquilas evidenciado pela presença de H_β em frações, sugerindo que os compostos de ácidos naftênicos possuem cadeias longas e não-ramificadas.

De um modo em geral, os espectros de RMN são bem complexos por fornecerem uma grande quantidade de informações em um único ensaio. O uso de ferramentas sofisticadas para tratar e extrair as informações químicas mais relevantes se tornam necessárias para correlacionar os dados espectrais com as propriedades de interesse. Com isso, o progresso na interpretação dos espectros de RMN se deve ao uso de alguma metodologia de calibração multivariada

2.3.QUIMIOMETRIA

Quimiometria é a parte da química que utiliza métodos matemáticos e estatísticos para extrair o máximo de informação de dados químicos.⁵¹ O desenvolvimento das técnicas quimiométricas ganharam grande proporção juntamente com a química analítica.⁵²

Muito destaque tem sido dado aos sistemas multivariados, nos quais se pode medir muitas variáveis simultaneamente ao se analisar uma amostra qualquer. Nesses sistemas, a conversão da resposta instrumental no dado químico de interesse requer a utilização de técnicas de natureza multivariada que estão sendo consideradas como a melhor alternativa para a interpretação de dados e para o alcance do máximo de informação sobre o sistema.^{16,17} A calibração multivariada é uma das principais aplicações em quimiometria, e normalmente envolve o uso de um tipo de medida para estimar o valor de um parâmetro da propriedade de interesse.⁵³

Para a interpretação dos dados analíticos com muitas informações químicas algumas ferramentas quimiométricas são muito utilizadas, como a análise hierárquica de agrupamentos (HCA), a análise de componentes principais (PCA), a modelagem independente flexível por analogia de classe (SIMCA), a análise de discriminante linear (LDA), a regressão linear múltipla (RLM), a regressão em componentes principais (PCR) e os mínimos quadrados parciais (PLS).⁵³⁻⁵⁵ Há vários trabalhos com bons resultados na área de petróleo e seus derivados utilizando PCA e PLS como ferramentas quimiométricas para análise exploratória de dados e construção de modelos de calibração.⁵⁶⁻⁵⁹

O método de calibração multivariada mais utilizado em química analítica é a regressão por mínimos quadrados parciais (PLS) e sua teoria é facilmente encontrada na literatura.⁵⁴⁻⁵⁵ A ferramenta PLS é eficiente em dados que apresentam relação linear, normalmente obtida em matrizes mais simples. Embora o petróleo seja uma matriz bem complexa, o PLS tem fornecido bons resultados.^{3,5-8,19,60-61}

Em quimiometria, é sempre interessante melhorar o desempenho dos modelos e o uso de estratégias se tornam necessárias pois alguns fatores podem dificultar a obtenção de resultados satisfatórios em dados de PLS em matrizes complexas. Para contornar o problema, algumas adaptações do modelo PLS podem ser utilizadas, como o método OPLS (projeções ortogonais em estruturas latentes) e a seleção de variáveis.

2.3.1.PLS

O PLS é o método de calibração multivariada mais utilizado em química analítica para a construção de modelos quantitativos em matrizes complexas como o petróleo.³ É um método que pode ser usado para a construção de modelos que permitem estimar propriedades físico-químicas em petróleos 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.⁸²

O modelo PLS ajusta uma função linear dos scores da matriz X (dados instrumentais) aos scores do vetor y (propriedade de interesse) de modo a minimizar os desvios numa soma de 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 dado e, por outro lado, quanto maior o número de VLs, maior será o erro, resultando em um sobreajuste (*overfitting*). 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.⁸²

A construção de um modelo PLS consiste no compromisso entre a explicação da variância em X (dados instrumentais) e encontrar a correlação com y (propriedade de interesse). Os dois blocos X e y são decompostos simultaneamente em uma soma de variáveis latentes (VL) (como mostrado nas equações 4 e 5). A primeira VL descreve a direção de máxima variância que também se correlaciona com a propriedade de interesse e os fatores das VL são definidos de modo a manter um compromisso entre a explicação da variância em X e a previsão da variável de interesse. (Equações 4 e 5)

$$X = TP^T + E \quad (4)$$

$$y = Uq^T + f \quad (5)$$

Onde, T e U são os *scores*, P e q os *loadings* (pesos) e E e f os resíduos de X e y respectivamente.

O produto $T.P$ aproxima dos dados instrumentais e o produto $U.Q$ dos valores verdadeiros da propriedade de interesse. Um importante fator no PLS é que é possível obter uma matriz de escores que é comum para ambas as propriedades e medidas.

A matriz de scores T é estimada como uma combinação linear X com coeficientes ponderados por W (chamados pesos) de acordo com a equação 6.

$$T = XW \quad (6)$$

A partir de W , os coeficientes de regressão do modelo PLS podem ser estimados pela equação 7

$$b_{PLS} = W \hat{c} \quad (7)$$

O modelo linear PLS pode ser representado pela equação 8:

$$y = X.b_{PLS} \quad (8)$$

2.3.2.OPLS

O método OPLS é uma adaptação do método de pré-processamento de correção de sinal ortogonal (OSC)⁶² por meio da integração desta etapa à modelagem PLS. Tem como vantagem analisar a variação ortogonal em cada componente regular do PLS, ou seja, é uma modificação do método PLS acrescentando uma etapa que retira toda variação estruturada que não tem correlação (ortogonal) com a propriedade de interesse.⁶³⁻⁶⁴

O método OPLS separa a variação sistemática em X em duas partes, uma que é correlacionada (preditiva) a y e uma que não está correlacionada (ortogonal) a y . O OPLS pode não alterar o poder preditivo comparado com PLS quando os dados possuem alta linearidade⁶⁵ mas, pode facilitar a interpretação dos resultados permitindo a análise separada da variação não correlacionada.

No método OPLS informações da matriz X são decompostas em três partes distintas, como demonstrado na Equação 9:

$$X = T_p P_p^T + T_o P_o^T + E \quad (9)$$

Onde T_p é a matriz de score de X ; P_p é a matriz de *loading* de X ; T_o denota a matriz de score de X ortogonal a y ; P_o a matriz de *loading* de X ortogonal a y e E é a matriz de resíduo.

Aplicado em classificação o modelo OPLS é estimado em duas etapas. Primeiro, a variação Y -ortogonal é removido da matriz de X . Onde T_o é a matriz de score de Y -ortogonal para X e P_o é a matriz de *loading* de Y ortogonal para X . Onde X_p são os componentes preditivos do modelo OPLS, determinados a partir do conjunto de dados de treinamento⁶⁶ (Equação 10).

$$X_p = X - T_o P_o^T \quad (10)$$

Como foi proposto pela primeira vez em 2002⁶⁷, o OPLS e suas extensões (OPLS-DA, K-OPLS, etc.) foram utilizados em contextos distintos para múltiplas aplicações.⁶⁸ Por outro lado, OPLS para espectros de RMN de petróleo não foi relatado até agora.

2.3.3. Seleção de variáveis

Um modelo desejável é aquele que R^2 seja próximo de 1, os valores de RMSEC e RMSECV sejam mais próximos (e baixos) e RMSEP tenha o menor valor possível.⁶⁹ Entretanto, muitas vezes o modelo não apresenta resultados aceitáveis devido ao excesso de informação dos espectros de RMN.

Na construção de alguns modelos de calibração multivariada, como PLS e OPLS, usa-se a informação de toda faixa espectral (correlacionando com a propriedade de interesse). Considerando o grande número de variáveis fornecidas pela faixa espectral, algumas destas variáveis podem interferir na modelagem e tornar o tratamento dos dados mais lento. Nestes casos, uma forma de melhorar o desempenho de técnicas de calibração multivariada são os procedimentos apropriados para a seleção das regiões espectrais associadas às propriedade de interesse.^{23,70}

Com a seleção de variáveis é possível identificar um subconjunto de variáveis informativas que produzirão erros de previsão mais baixos, melhorando o desempenho e gerando modelos de fácil interpretação, especialmente onde um grande número de variáveis irrelevantes ou interferentes podem existir.⁷¹

2.3.3.1. iPLS e siPLS

A técnica iPLS é uma extensão PLS que se concentra nas principais regiões espectrais para prever a informação mais relevante nas diferentes subdivisões do espectro, removendo as interferências das demais regiões. Consequentemente é construído um novo modelo PLS a partir das variáveis selecionadas. O iPLS é desenvolvido em subintervalos espectrais de mesmo peso e o desempenho do modelo é baseado no RMSECV e R^2 .²³

O PLS por sinergismo de intervalos (siPLS) é uma extensão do iPLS.

Nos estudos de Norgaard e colaboradores foi desenvolvido um método para seleccionar regiões espectrais em modelos PLS nos quais o espectro é dividido em intervalos equidistantes. Uma combinação adequada de vários intervalos é então executada e se tem melhores previsões. Esse método foi chamado de intervalo de sinergia ou siPLS.^{3,23} No trabalho de Ferrão e colaboradores²⁴, os algoritmos iPLS e siPLS foram capazes de seleccionar a região espectral mais adequada para cada propriedade estudada do biodiesel (gravidade específica, teor de enxofre e ponto de fulgor) onde o algoritmo siPLS produziu modelos melhores do que o PLS. Estudos de seleção de variáveis iPLS e siPLS com espectros de RMN de ^1H de petróleos brutos ainda são escassos.^{24,62}

2.3.3.2.MPA

Na seleção de variáveis também foi usado o modelo de análise populacional (MPA) proposto pela primeira vez por Li e colaboradores (2010).²⁵ O MPA é uma estrutura para o desenvolvimento de um algoritmo para modelagem de sistemas analíticos complexos utilizando ferramentas estatísticas (coeficiente de regressão, erros de previsão, viés) para extrair informações importantes do modelo, fornecendo uma melhor compreensão e modelagem dos dados químicos, melhorando a previsão e a interpretação do modelo.²⁶⁻²⁸

Para dar uma visão geral do MPA o esboço é mostrado na Figura 2. O MPA trabalha as quatro etapas sucessivas:

- (1) amostragem aleatória: desenha aleatoriamente sub-conjuntos de dados.
- (2) para cada sub-conjunto de dados, um submodelo é construído;
- (3) calcular uma saída de interesse (por exemplo, erros de previsão) para cada submodelo;
- (4) a análise estatística é aplicada aos resultados dos submodelos.^{25,27}

É importante ressaltar que a chave do MPA é o quarto passo, pois ao analisar um parâmetro de interesse associado a um dos quatro espaços, um algoritmo de análise de dados pode ser desenvolvido. Como exemplo, o método de amostragem Monte Carlo que foi projetado para detectar *outliers* estudando a distribuição dos erros de previsão de cada amostra, que é um parâmetro associado ao espaço amostral.^{26,30}

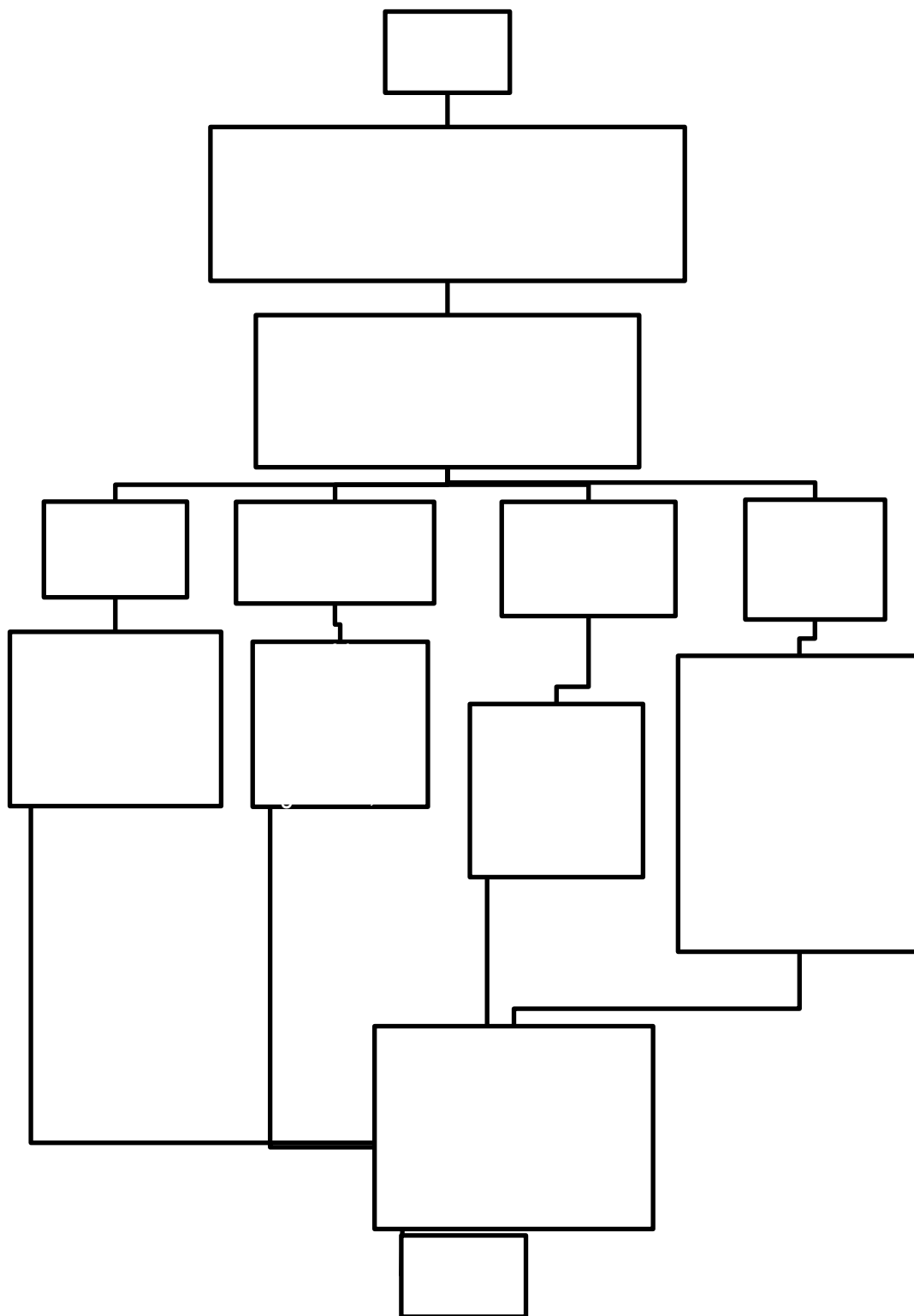


Figura 3. Esboço do algoritmo do MPA. Adaptado pela Autora.

Técnicas de amostragem aleatória para gerar uma população diversificada de submodelos, isto é, as técnicas de amostragem Monte Carlo (MCS) são usualmente aplicadas em MPA para obter diversidade de conjuntos de sub-amostras ou conjuntos de sub-variáveis.⁷² Outras técnicas também podem

ser usadas, como amostragem Jackknife (técnica de amostragem aleatória sem reposição)⁷³, amostragem bootstrap⁷⁴ (técnica de amostragem aleatória com reposição), amostragem por matriz binária⁷⁵ (realiza amostragem aleatória com a ajuda de uma matriz binária) e a permutação, que ao contrário de qualquer método de amostragem aleatória, cessa a conexão entre amostras e observações.⁷⁶

Na análise estatística tem-se como ideia central do MPA a análise das saídas de interesse para todos os submodelos (por exemplo, erros de previsão ou coeficientes de regressão). Diferentes propostas para a análise de diferentes saídas de interesse de todos os submodelos levarão a diferentes algoritmos. Como exemplo, a análise da distribuição de RMSEP mostrou-se eficaz na detecção de outliers³⁰ e úteis na seleção de variáveis.²⁶ Na análise estatística, com uma população de submodelos gerados por métodos de amostragem aleatória, a comparação de modelos é realizada usando uma distribuição empírica derivada dos resultados das saídas de interesse da seguinte forma:

- 1: As distribuições de erros de previsão para amostras normais e *outliers* seguem padrões diferentes
2. As distribuições de coeficientes de regressão para variáveis não informativas e variáveis informativas podem ser distinguidas;
3. A distribuição do RMSECV para combinações de variáveis pode ser usada para pesquisar o conjunto ótimo de variáveis;
4. A diferença entre as duas distribuições do RMSECV com e sem a variável específica pode ser usada para medir a importância da variável.

Para extrair as informações dessas distribuições, alguns critérios foram empregados, como média, desvio padrão, mediana, desvio mediano absoluto, intervalo de confiança de 95%, teste t-*Student*, teste U de Mann - Whitney, entre outros.²⁶⁻²⁷

2.3.4. Pré-tratamento de dados

Em uma análise de amostras que possuem matrizes complexas, como o petróleo, são necessários alguns passos de pré-tratamento. Um dos fundamentos de uma análise quantitativa é o preparo da amostra e posterior eliminação de interferentes.⁷⁷

Para o alinhamento dos deslocamentos químicos em dados de RMN foi utilizado o algoritmo icoshift.²⁰ Com o uso dessa ferramenta, o desempenho dos modelos melhorou consideravelmente obtendo valores de RMSECV mais baixos e uma melhor correlação entre os valores medidos e previstos.

O pré-tratamento utilizado nesse trabalho foi o SNV²² (variação normal padrão) que apresentou menores erros de predição e melhores valores de coeficiente de determinação quando comparados a outros pré-tratamentos testados (MSC - Correção do espalhamento multiplicativo, do inglês, *multiplicative scatter correction*; centragem na média; e derivada: a aplicação das derivadas remove uma parte da deformação da linha de base, e melhor a definição de bandas que se encontram sobrepostas em uma mesma região espectral, revelando picos de pequena absorbância). O SNV tem grande utilidade para a correção de linha de base e também atua corrigindo o efeito de espalhamento multiplicativo da luz.

2.3.5. Avaliação dos Modelos

Na avaliação da qualidade dos modelos de regressão multivariada alguns parâmetros como a raiz da média quadrática dos resíduos de calibração (RMSEC – Equação 11), a raiz da média quadrática dos resíduos de previsão (RMSEP – Equação 12) e o coeficiente de determinação (R^2 – Equação 13) são utilizados.⁷⁸

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_{i,cal} - \hat{y}_{i,cal})^2}{n_{cal} - VL - 1}} \quad (11)$$

$$RMSEP = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

Onde y_i e $\hat{y}_i$ correspondem aos valores de referência e previsão do modelo PLS, respectivamente, para a propriedade de interesse; cal e $pred$ correspondem ao número de amostras de conjuntos de calibração e previsão respectivamente e VL é o número de variáveis latentes usadas no modelo PLS. Para os coeficientes de determinação $\bar{y}$ é o valor médio do conjunto de previsão.

O valor de R^2 varia de 0 a 1, no qual, quanto mais próximo de 1, melhor o ajuste linear do modelo.⁶⁹

Os modelos construídos foram submetidos a testes estatísticos para verificação da presença de erros sistemáticos (teste de bias) (Equação 14) e para a avaliação de tendência nos resíduos do modelo (teste não paramétrico de permutação).¹⁷ Quando algum destes testes indicava erro significativo era realizada uma nova avaliação com a modificação da otimização dos modelos, como por exemplo, número de VL.

$$bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (14)$$

3.OBJETIVOS

3.1.Objetivos gerais

Desenvolver métodos analíticos para determinar as propriedades físico-químicas de petróleo bruto: enxofre total, nitrogênio total, fator de caracterização, parâmetros de solubilidade, ponto de fluidez e viscosidade cinemática à 20°C, 30°C, 40°C e 50°C utilizando espectroscopia de Ressonância Magnética Nuclear de hidrogênio associada aos métodos de calibração multivariada PLS, OPLS e PLS com seleção de variáveis iPLS, siPLS e MPA.

3.2.Objetivos específicos

Obter espectros de ressonância magnética nuclear de hidrogênio (RMN ^1H) de petróleo.

Construir alguns protocolos de ensaio para análise de petróleo por RMN objetivando uma análise quantitativa do petróleo.

Otimizar os modelos de calibração multivariados aplicados aos métodos quantitativos para petróleo.

Construir diferentes modelos quimiométricos para previsão de propriedades físico-químicas de petróleo.

Validar os modelos multivariados utilizados para determinar propriedades físico-químicas de petróleo.

4.METODOLOGIA

Neste trabalho foram utilizados 138 amostras de diferentes petróleos da bacia sedimentar brasileira. Os resultados dos ensaios de caracterização foram cedidos pelo Centro de Pesquisa Leopoldo Américo Miguez de Mello (CENPES/ Petrobras) que seguiu padrões técnicos de análises (Figura 3). Os espectros de RMN de ^1H foram realizados e disponibilizados pelo Laboratório de RMN do Núcleo de Competências em Química do Petróleo da Universidade Federal do Espírito Santo (NCQP-LabPetro/UFES).

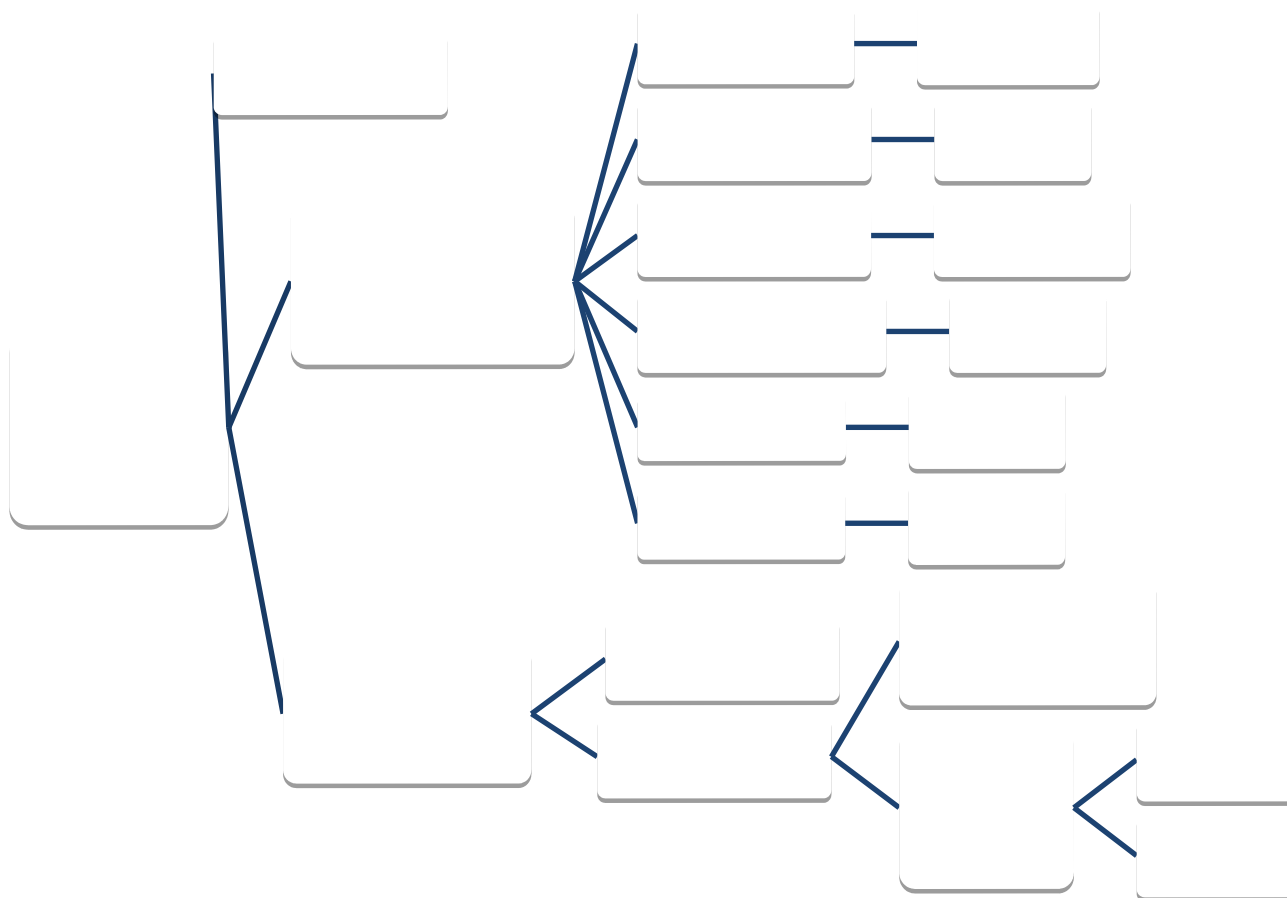


Figura 4. Esquema da obtenção e tratamento dos dados das amostras de petróleos

As propriedades físico-químicas foram determinadas de acordo com métodos padrão utilizados pela indústria de petróleo (Figura 3).

A viscosidade cinemática foi determinada pelo método padrão ASTM D7042⁷⁹ que consiste em injetar a amostra de petróleo bruto em um viscosímetro digital, geralmente na temperatura de 40°C. No entanto, para a análise de óleos pesados, os testes são geralmente realizados em

temperaturas mais altas (50°C e 60°C) e, em seguida, estimados em 40°C. Após a análise acima mencionada, a viscosidade cinemática foi estimada em 20 °C, 30°C e 40°C por regressão⁸⁰, de acordo com a Equação 15.

$$v_{trnas} = \log(\log(v + 0,7)) \quad (15)$$

Onde v_{trnas} é a viscosidade cinemática em relação a temperatura da viscosidade cinemática do óleo v .

4.1. Aquisição dos espectros de RMN de ^1H

Os ensaios de Ressonância Magnética Nuclear (RMN) de ^1H foram realizados em um espectrômetro da marca Varian MR-400 (Varian Inc., Palo Alto, CA (USA)), operando a 399,8 MHz sob as seguintes condições: janela espectral fixa de aproximadamente 6410,3 Hz; pulso de 90°; tempo de aquisição de 2,556s e número de transientes de 64 à 25 C.

As amostras foram preparadas dissolvendo-se uma massa correspondente a 5% em diclorometano- d_2 (CD_2Cl_2), totalizando um volume de 600µL de solução.

4.2. Modelos de calibração multivariada

Inicialmente, as amostras de petróleo foram separadas em conjunto de calibração e previsão (teste). Na seleção, aproximadamente 70% das amostras foram selecionadas para o conjunto de calibração e o restante para o conjunto de teste por meio do algoritmo de Kennard-Stone.³²

Para esse trabalho 138 amostras estavam disponíveis, mas nem todas foram utilizadas para a construção dos modelos quimiométricos, visto que algumas tiveram problemas na aquisição dos espectros e/ou na obtenção dos valores das propriedades de interesse, via metodologia de referência. Diante disso, um número diferente de amostras foi utilizado para a calibração e previsão das diferentes propriedades (Tabela 2).

Também foi realizado o procedimento de validação cruzada usando o método *venetian blind 5-fold* para otimizar o número de variáveis latentes (VL) para cada modelo.⁷⁸ O número de VL foi determinado pelo gráfico da taxa de erro de validação cruzada em função do número de variáveis latentes. O número ótimo de variáveis latentes representadas pelo gráfico indica a melhor exatidão.

Tabela 2. Quantidade de amostras de calibração e teste para modelagem quimiométrica

Propriedades de interesse	Amostras Calibração	Amostras Teste
Fator de caracterização	117	21
Nitrogênio total	94	22
Parâmetro de solubilidade	98	17
Enxofre total	95	16
Ponto de fluidez Mínimo	99	18
Ponto de fluidez Máximo	101	19
Viscosidade cinemática à 20 °C	110	20
Viscosidade cinemática à 30 °C	115	21
Viscosidade cinemática à 40 °C	100	17
Viscosidade cinemática à 50 °C	99	21

Para o desenvolvimento dos modelos de calibração multivariada os espectros de RMN de ¹H foram alinhados usando o algoritmo *icoshift*.²⁰ Além do alinhamento *icoshift*, o método de pré-processamento de variação normal padrão (SNV) foi utilizado para melhorar o ajuste na linha de base, pois atenua substancialmente as variações espectrais causadas principalmente por interferentes da matriz.²² Os modelos de calibração foram construídos empregando-se as técnicas de PLS, OPLS e seleção de variáveis por iPLS, siPLS e MPA com algoritmos produzidos ou adaptados pelo grupo de Quimiometria do Labpetro/UFES utilizando o software MATLAB® versão 7.1 (*The Mathworks, Natick, USA*).

Os resultados dos modelos foram avaliados pelo RMSEC, RMSEP e

coeficiente de determinação (R^2) tanto para a calibração (R^2_c) e teste (R^2_p).⁷⁸ Foram aplicados os testes estatísticos de viés para avaliação da presença de erros sistemáticos e o teste de permutação para avaliação de tendência nos resíduos.

5.RESULTADOS E DISCUSSÃO

5.1. Análise do espectro de RMN de ^1H de petróleo

Na Figura 4 estão apresentados os espectros de RMN de ^1H das amostras de petróleo, bem como o efeito do tratamento aplicado para correção do deslocamento químico. Os espectros de RMN de ^1H foram alinhados com o algoritmo *icoshift* antes da construção dos modelos de calibração. Como descrito na introdução desse trabalho, o alinhamento com *icoshift* é fundamental para se trabalhar com dados de RMN quando o intuito é modelagem. Com isso, o algoritmo se mostra eficiente para a finalidade de alinhamento (Figura 4b). Após o alinhamento, os deslocamentos químicos referentes aos sinais de interesse continuaram nas mesmas posições espectrais, ou seja, a relevância química é sempre mais importante que a matemática.

Também foi feito o pré-processamento SNV. Essa etapa contribui significativamente para o processo deixando os modelos mais robustos e precisos, pois as perturbações espectrais causadas por interferentes da matriz são substancialmente suavizadas.

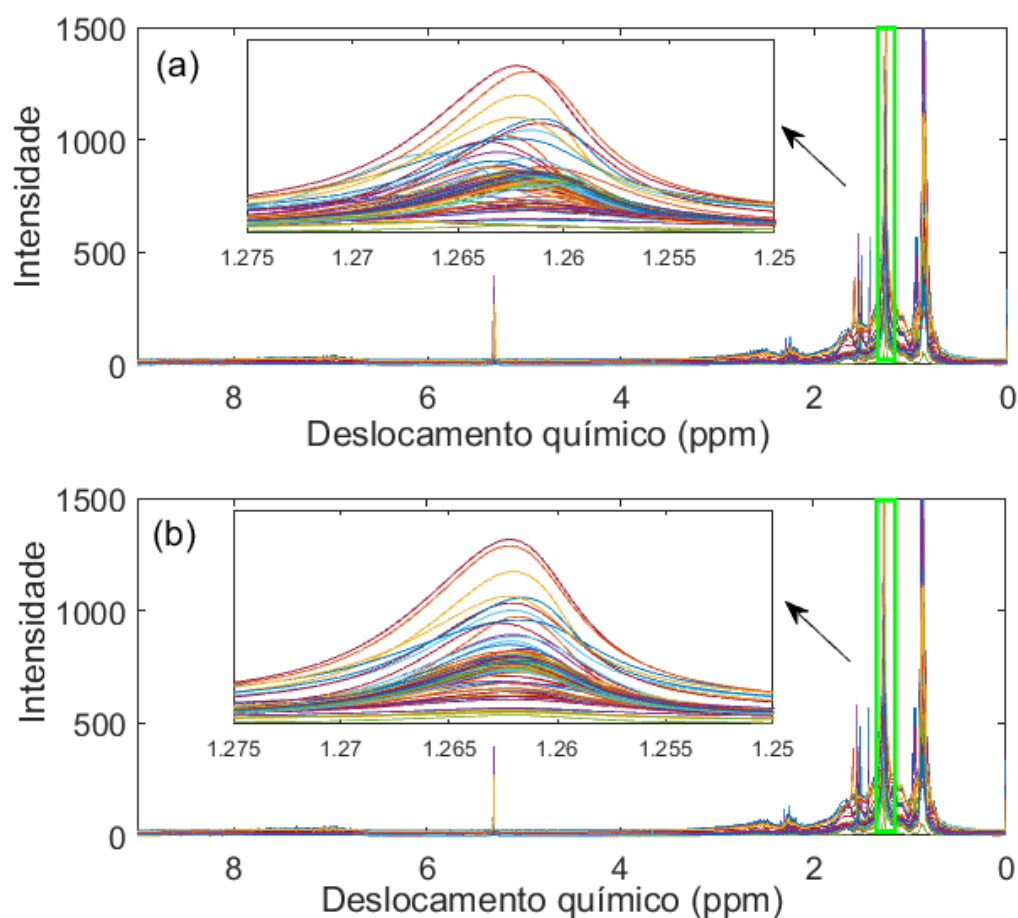


Figura 5. Espectro de RMN de ^1H : (a) sem alinhamento (b) depois do alinhamento pelo algoritmo *icoshift*.

O petróleo é constituído por uma enorme quantidade de compostos e mesmo utilizando uma técnica espectroscópica poderosa para elucidação estrutural, como a RMN, não é possível obter informações a nível molecular especificamente para cada constituinte da matriz.

Os espectros de RMN de ^1H obtidos podem ser discutidos de acordo com as principais regiões que trazem informação (Tabela 1), levando em consideração os principais tipos de hidrocarbonetos existentes no petróleo. A Figura 5 mostra o espectro de RMN de ^1H de três óleos diferentes, variando do leve (Figura 5a), médio (Figura 5b) ao pesado (Figura 5c). Analisando essa figura é possível perceber que os espectros do óleo leve e pesado são qualitativamente similares, apresentando um perfil semelhante. No espectro de óleo leve (Figura 5a) aparecem alguns picos a mais na região de 1-1,5 ppm, talvez decorrente da maior quantidade de hidrocarbonetos parafínicos (Tabela 1). Também é

possível verificar na Figura 5: a) a ausência de água na região de 4,6-4,9 ppm e b) o sinal que aparece em 5,3 ppm é do solvente CD_2Cl_2 (esse solvente foi escolhido por apresentar um sinal na região espectral com menor contribuição dos principais sinais).⁵

Logo, a técnica espectroscópica de RMN de ^1H é muito utilizada para quantificar e qualificar esses diferentes grupos em diferentes tipos de óleo. Com isso, os hidrocarbonetos parafínicos e aromáticos das amostras possivelmente podem ser identificados com os parâmetros moleculares médios mostrados na Tabela 1.

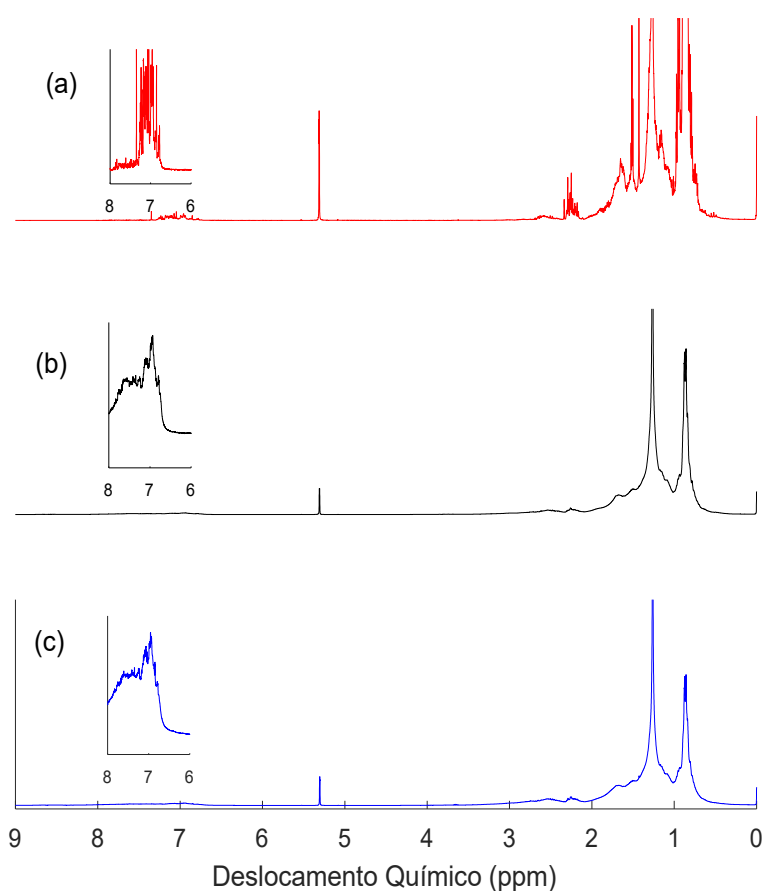


Figura 6. Espectro de RMN de ^1H de três petróleos: (a) leve, (b) médio e (c) pesado.

- Histogramas

Analisando os histogramas das amostras de petróleos para cada propriedade de interesse como mostrada na Figura 6, tem-se que para o fator de caracterização UOP as amostras apresentam, predominantemente, caráter naftênico seguido pelo caráter parafínico. O fator de caracterização UOP dá

informação quanto ao grupo predominante presente nas amostras de petróleos. Fator KUOP maior ou igual a 12 os petróleos são predominantemente parafínicos; KUOP menor ou igual a 10 os petróleos são predominantemente aromáticos e o KUOP entre 10 e 11,8 os petróleos possuem caráter naftênico (Figura 6a).³⁶

Na Figura 6b podemos observar poucos óleos com baixo teor de nitrogênio (menor que 0,1wt%).

Segundo Wiehe e Kennedy (2000)³⁴ o parâmetro de solubilidade de petróleos está situado em uma faixa entre 15,95 (MPa)^{1/2} e 18,20 (MPa)^{1/2}; no histograma mostrado na Figura 6c, as amostras estão majoritariamente nessa faixa. Tem-se 11 amostras abaixo de 16,35 (MPa)^{1/2}, ou seja, os asfaltenos permanecem insolúveis em *n*-alcanos e metilciclohexano e 104 amostras acima de 16,35 (MPa)^{1/2}, logo são solúveis a ciclohexano entre outros da escala de forma crescente de potencial de solubilidade.

Da mesma forma, para o teor de enxofre, a maioria das amostras estão dentro da distribuição de petróleos brasileiros que podem ser classificadas como petróleos “Doces” (*sweet*), pois em sua maioria o teor de enxofre é menor que 0,5 % de sua massa (Figura 6d).³⁷

Na Figura 6e o ponto fluidez mínimo varia de -60 °C a 9 °C e na Figura 6f o ponto de fluidez máximo varia de -54 °C a 21 °C. O ponto de fluidez é uma importante propriedade avaliada na etapa de produção. É definida como a menor temperatura na qual os óleos não fluem sob a ação gravitacional.

A viscosidade cinemática varia exponencialmente em relação à densidade, assim, os modelos de regressão foram construídos com o logaritmo da viscosidade cinemática para obter uma relação linear. Na construção do modelo da viscosidade cinemática para as temperaturas trabalhadas utilizou-se a viscosidade depois de submetida a duas operações logarítmicas. A faixa de viscosidade cinemática trabalhada foi de 2 a 631 mm².s⁻¹. (Figura 6g, 6h, 6i e 6j)

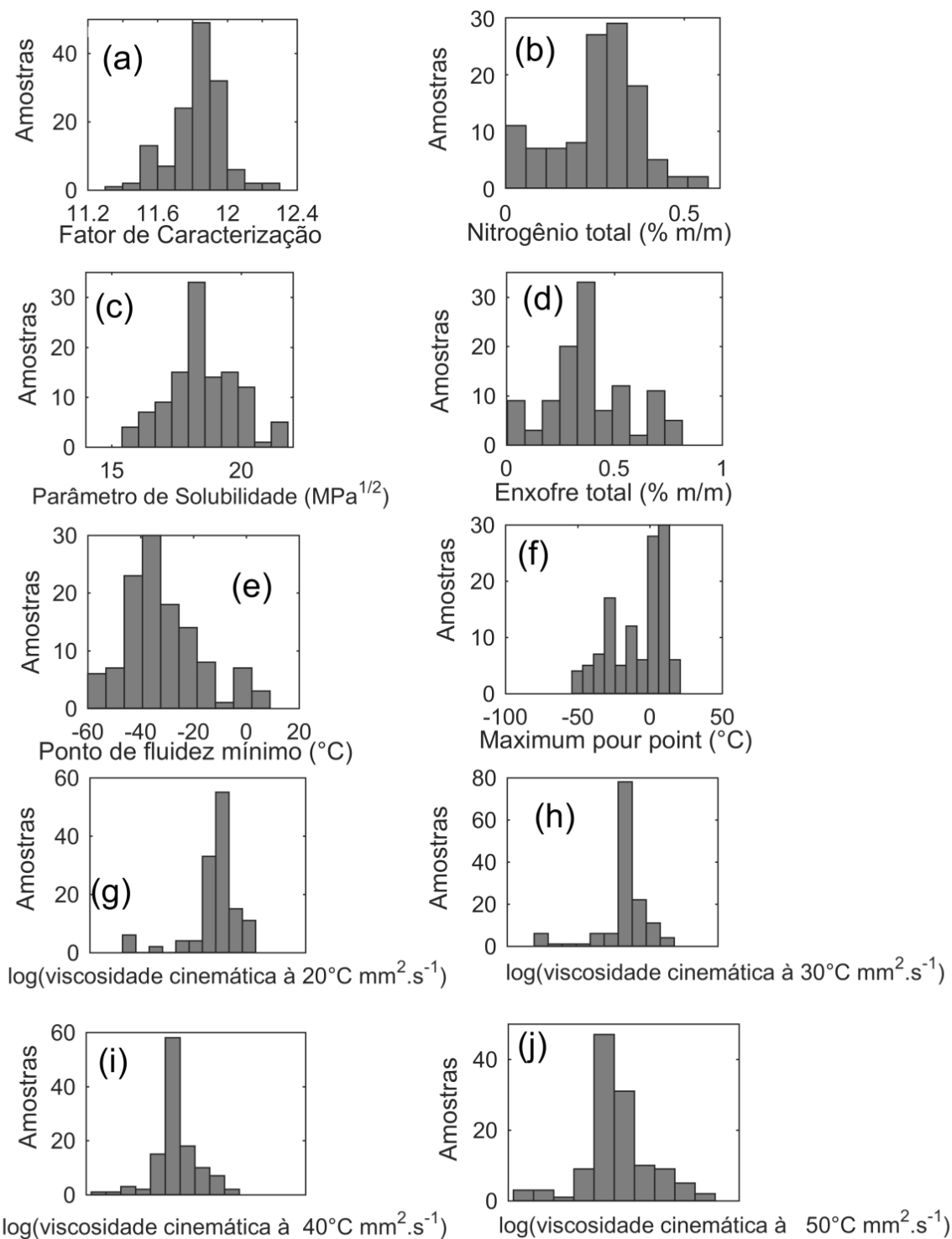


Figura 7. Histograma das amostras de petróleos: (a) fator de caracterização; (b) nitrogênio total; (c) parâmetro de solubilidade; (d) enxofre total; (e) ponto de fluidez mínimo; (f) ponto de fluidez máximo; viscosidade a (g) 20°C; (h) 30°C; (i) 40°C; (j) 50°C.

5.2. Calibração multivariada

A Tabela 4 mostra os resultados obtidos de RMSEC e RMSEP para os modelos PLS, iPLS, siPLS, OPLS e MPA. De acordo com a literatura supracitada, os resultados são considerados satisfatórios quando os valores de RMSEP são baixos, ou seja, quanto menor o valor de RMSEP melhor a exatidão média do modelo.

Em relação a essas análises, os resultados sugerem que o Fator de caracterização de UOP, teor de nitrogênio, viscosidade cinemática a 20 °C, 40 °C e 50 °C obteve o melhor resultado pelo modelo MPA. O teor de enxofre, ponto de fluidez máximo e viscosidade cinemática a 30 °C obteve o melhor resultado por PLS. O parâmetro de solubilidade obteve o melhor resultado pelo OPLS e o ponto de fluidez mínimo obteve o melhor resultado pelo iPLS.

A qualidade dos modelos de calibração foi verificada pela avaliação dos resíduos. Os testes realizados para identificar os erros sistemáticos (bias) e tendência (teste de permutação) revelaram que os modelos siPLS apresentaram erros de tendência nos resíduos de calibração para fator de caracterização (p-valor igual a 0,011), fluidez mínima (p-valor igual a 0,0159), viscosidade cinemática à 20 °C (p-valor igual a 0,0131) e os modelos iPLS para viscosidade cinemática à 30 °C (p-valor igual a 0,0320). O nível de significância adotado foi de 0,05.

De acordo com os valores de RMSEP e RMSEC para viscosidade cinemática à 40 °C, os modelos de calibração multivariada, com exceção do MPA, tem indício de superajuste por apresentarem um RMSEP três vezes maior que o RMSEC. O mesmo ocorre para o modelo OPLS para enxofre total. Nesse caso, uma alternativa ao problema é uma nova seleção das amostras de calibração e teste.

Tabela 3. Valores de RMSEC e RMSEP dos modelos PLS, iPLS, siPLS e OPLS de predição das propriedades de interesse a partir de espectros de RMN de ^1H .

Propriedades	PLS		iPLS		siPLS		OPLS		MPA
	RMSEC	RMSEP	RMSEC	RMSEP	RMSEC	RMSEP	RMSEC	RMSEP	RMSEP
Fator de caracterização KUOP	0,09	0,11	0,10	0,09	0,14	0,10	0,09	0,12	0,06
Nitrogênio total (wt%)	0,043	0,056	0,066	0,068	0,069	0,069	0,044	0,068	0,048
Parâmetro de solubilidade (MPa)^{1/2}	1,07	1,25	0,88	1,63	1,15	1,33	0,87	1,06	1,09
Enxofre total (wt%)	0,08	0,09	0,10	0,12	0,09	0,13	0,08	0,25	0,25
Ponto de fluidez mínimo (°C)	13,24	15,09	12,69	12,28	12,09	12,81	9,79	13,09	14,58
Ponto de fluidez máximo (°C)	10,14	13,24	13,12	14,37	12,01	14,82	10,21	14,09	19,62
log (Viscosidade cinemática à 20 °C mm²·s⁻¹)	0,093	0,137	0,081	0,114	0,099	0,123	0,083	0,159	0,045
log (Viscosidade cinemática à 30 °C mm²·s⁻¹)	0,032	0,034	0,122	0,172	0,106	0,056	0,072	0,187	0,037
log (Viscosidade cinemática à 40 °C mm²·s⁻¹)	0,058	0,275	0,065	0,237	0,051	0,269	0,043	0,275	0,036
log (Viscosidade cinemática à 50 °C mm²·s⁻¹)	0,061	0,050	0,052	0,077	0,049	0,083	0,059	0,050	0,042

Na Figura 7 tem-se uma relação dos valores de R^2 de previsão dos diferentes modelos desenvolvidos para cada propriedade de interesse. Vale destacar que alguns modelos resultaram em baixos valores de R^2 gerando resultados não muito satisfatórios.

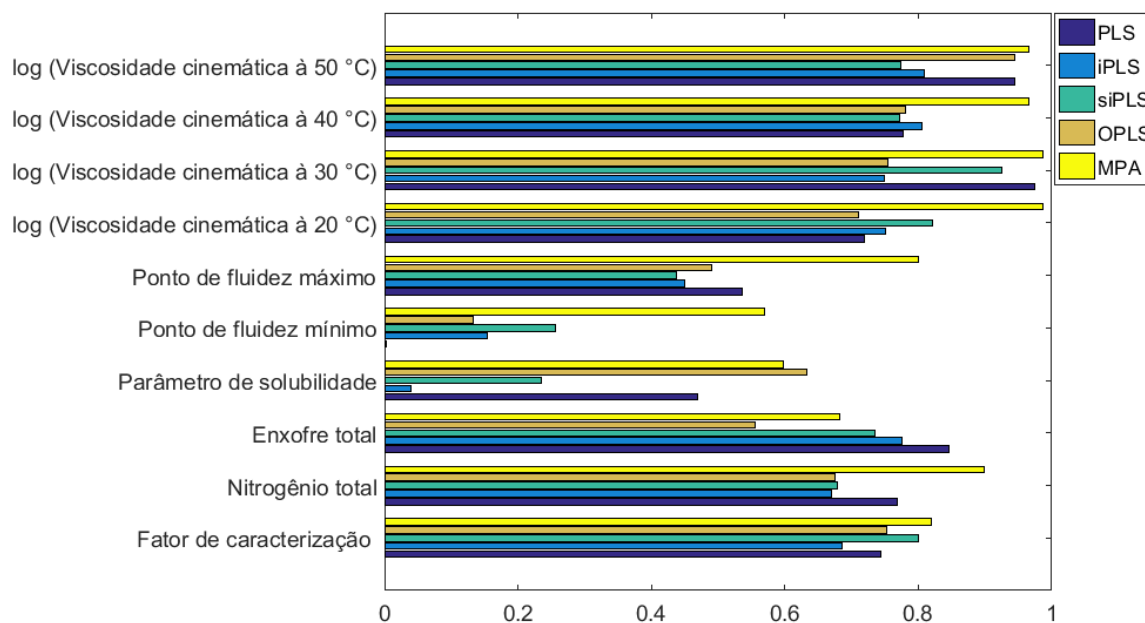


Figura 8. Valores de R^2 para as propriedades de interesse dos modelos PLS, iPLS, siPLS, OPLS e MPA: **1)** fator de caracterização, **2)** nitrogênio total, **3)** enxofre total, **4)** parâmetro de solubilidade, **5)** ponto de fluidez mínimo, **6)** ponto de fluidez máximo, viscosidade cinemática à **7)** 20°C, **8)** 30°C, **9)** 40°C e **10)** 50°C.

Fator de caracterização

O fator de caracterização é uma propriedade usada para classificar os óleos brutos em termos de conteúdo parafínico, naftênico e aromático. A ligação direta entre esta propriedade e as informações contidas nos espectros de RMN facilita a interpretação dos resultados, ou seja, apresentou nos modelos um erro padrão significativamente baixo e o coeficiente de determinação de 0,82 para o MPA (maior valor obtido), demonstrando uma alta linearidade entre os valores previstos e de referência para o MPA (Figura 8a). No trabalho de Masili e colaboradores (2012)¹⁹ foi utilizado PLS com espectros de RMN de ^1H de petróleos e obteve resultados com baixos erros médios e um R^2 de 0,96. O excelente desempenho alcançado se deve as estratégias adotadas para o preparo das amostras ao homogeneizá-las acima da temperatura do ponto de fluidez

antes da amostragem e, principalmente, com o pré-tratamento dos dados espectrais para caracterizar os petróleos brutos de maneira rápida, confiável e econômica. Antes da modelagem multivariada a área espectral foi dividida em regiões fixas, enquanto os dados entre 4,20 e 5,90 ppm foram omitidos para evitar a inclusão do sinal da água e uma região de ruídos. Tal procedimento foi preferido pois usa pontos dos dados originais para reduzir a dimensão dos dados da matriz a fim de assegurar a robustez do modelo.

As variáveis mais significativas para a construção do modelo MPA podem ser observadas na Tabela 4:

Tabela 4: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Fator de Caracterização.

Região do Espectro – ppm – 9 variáveis selecionadas	Parâmetros moleculares de Poveda
0,020 a 0,9260	H parafínico do tipo CH ₃ , em posição γ ou mais em relação ao sistema aromático
2,1740 a 2,2567	H parafínico e naftênico do tipo CH, CH ₂ e CH ₃ , ligados a sistemas aromáticos na posição α

Teor de Enxofre total

Na modelagem do teor de enxofre total foram detectados quatro *outliers* no conjunto teste. Os *outliers* foram detectados por apresentar um alto *leverage* simultaneamente a um resíduo elevado. Estes foram retiradas nos modelos PLS, OPLS, iPLS, siPLS, como consequência sua qualidade preditiva foi melhorada. O teor de enxofre também foi estimado por Masili e colaboradores (2012)¹⁹ e um R² de 0,97 foi obtido. Neste trabalho, no entanto, o melhor resultado obtido foi com PLS, com R² de 0,85 (Figura 8b)

Viscosidade cinemática à 20°C, 30°C, 40°C e 50°C

A propriedade viscosidade cinemática foi estimada por Rodrigues e colaboradores (2018)⁶ em cromatografia gasosa associada ao PLS obtendo um R² de 0,91 para a temperatura de 50 °C. No atual trabalho os modelos PLS, OPLS e MPA obtiveram-se um R² de 0,95, 0,95 e 0,96 (Figura 8d), respectivamente, mostrando que a metodologia adotada nestes modelos apresentou melhores

resultados. Filgueiras e colaboradores (2014)³ utilizaram espectros de FTIR-ATR para estimar a viscosidade cinemática a 40 °C atingindo um R² de 0,78 (PLS) e 0,85 (SVR). Neste trabalho, obteve-se os seguintes resultados de R² para essa propriedade 0,78 (PLS), 0,81(iPLS), 0,78 (OPLS) e 0,97 (MPA- Figura 8c), resultados superiores aos encontrados na literatura.

Analisando os modelos para o log da viscosidade cinemática, observou-se que os resultados mostraram uma grande discrepância, com isso, o erro relativo (ER) foi calculado através da Equação 16:

$$\mathfrak{R}(\%) = 100 \times \frac{|y_i - \hat{y}_i|}{y_i} \quad (16)$$

Para a viscosidade cinemática, o ER médio (%) foi de 25,6% (20 °C), 17,6% (30 °C), 21,9% (40 °C) e 19,5 (50 °C).

As variáveis mais significativas para a construção do modelo MPA podem ser observadas na Tabela 5, 6, 7 e 8. :

Tabela 5: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Viscosidade cinemática à 20°C

Região do Espectro – ppm – 87 variáveis selecionadas	Parâmetros moleculares de Poveda
0,8746 a 0,8800	H parafínico do tipo CH ₃ , em posição γ ou mais em relação ao sistema aromático
1,1756 a 1,4996	H parafínico ligados a sistemas aromáticos na posição β e alquílicos terminais
1,5001 a 1,5040	H naftênico do tipo CH ₂ ligados a sistemas aromáticos na posição β

Tabela 6: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Viscosidade cinemática à 30°C

Região do Espectro – ppm – 59 variáveis selecionadas	Parâmetros moleculares de Poveda
0,0015 a 0,8800	H parafínico do tipo CH ₃ , em posição γ ou mais em relação ao sistema aromático

1,1869 a 1,4996	H parafínico ligados a sistemas aromáticos na posição β e alquílicos terminais
1,5001 a 1,5035	H naftênico do tipo CH_2 ligados a sistemas aromáticos na posição β
2,1740 a 2,2567	H parafínico e naftênico do tipo CH , CH_2 e CH_3 , ligados a sistemas aromáticos na posição α

Tabela 7: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Viscosidade cinemática à 40°C

Região do Espectro – ppm – 222 variáveis selecionadas	Parâmetros moleculares de Poveda
0,0010 a 0,8805	H parafínico do tipo CH_3 , em posição γ ou mais em relação ao sistema aromático
1,0019 a 1,4781	H parafínico ligados a sistemas aromáticos na posição β e alquílicos terminais
1,5001 a 1,5388	H naftênico do tipo CH_2 ligados a sistemas aromáticos na posição β

Tabela 8: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Viscosidade cinemática à 50°C.

Região do Espectro – ppm – 28 variáveis selecionadas	Parâmetros moleculares de Poveda
0,8663	H parafínico do tipo CH_3 , em posição γ ou mais em relação ao sistema aromático
1,2397 a 1,4996	H parafínico ligados a sistemas aromáticos na posição β e alquílicos terminais
1,5001 a 1,5035	H naftênico do tipo CH_2 ligados a sistemas aromáticos na posição β
2,2479 a 2,2499	H parafínico e naftênico do tipo CH , CH_2 e CH_3 , ligados a sistemas aromáticos na posição α

Ponto de fluidez mínimo e máximo

As propriedades de ponto de fluidez máximo, ponto de fluidez mínimo e parâmetro de solubilidade apresentaram uma baixa performance ($R^2 < 0,7$) na maioria das ferramentas quimiométricas utilizadas. Nos estudos Sad e colaboradores (2014)⁴¹ para o ponto de fluidez, os óleos pesados geralmente apresentaram maiores valores de ponto de fluidez (mais positivos) sem diferença entre o ponto de fluidez máximo e mínimo. As limitações associadas aos ensaios de ponto de fluidez baseiam-se no fato de que esta propriedade não tem relação com nenhuma outra dos óleos crus. Por exemplo, espera-se que óleos

com maior teor de parafinas apresentem um ponto de fluidez maior, entretanto, este comportamento esperado pode variar em função das características químicas dos óleos.

No estudo de Rodrigues e colaboradores (2018)⁶ ao se estimar o ponto de fluidez por cromatografia gasosa em petróleos, algumas amostras apresentaram resultados discrepantes em relação ao ajuste esperado para dados em relação à reta identidade (não houve exclusão de amostras) e o resultado obtido para o coeficiente de determinação foi de 0,71 (PLS). Neste estudo, o melhor resultado alcançado foi um R^2 de 0,57 (MPA – Figura 8e) para o ponto de fluidez mínimo e 0,80 (MPA- Figura 8f) para o ponto de fluidez máximo.

As variáveis mais significativas para a construção do modelo MPA podem ser observadas nas Tabelas 9 e 10 para o ponto de fluidez mínimo e máximo:

Tabela 9: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Ponto de fluidez mínimo.

Região do Espectro – ppm – 7 variáveis selecionadas	Parâmetros moleculares de Poveda
0,8781 a 0,8795	H parafínico do tipo CH ₃ , em posição γ ou mais em relação ao sistema aromático
1,2476 a 1,2823	H parafínico ligados a sistemas aromáticos na posição β e alquílicos terminais

Tabela 10 Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Ponto de fluidez máximo.

Região do Espectro – ppm – 50 variáveis selecionadas	Parâmetros moleculares de Poveda
0,8301 a 0,8756	H parafínico do tipo CH ₃ , em posição γ ou mais em relação ao sistema aromático
1,2627 a 1,4291	H parafínico ligados a sistemas aromáticos na posição β e alquílicos terminais
1,5412 a 1,5549	H naftênico do tipo CH ₂ ligados a sistemas aromáticos na posição β
2,1750 a 2,2650	H parafínico e naftênico do tipo CH, CH ₂ e CH ₃ , ligados a sistemas aromáticos na posição α

Parâmetro de solubilidade

O objetivo principal da aplicação do parâmetro de solubilidade em óleos é prever a solubilidade ou afinidade entre os componentes das frações pesadas e frações leves.⁸¹ Os compostos utilizados no estudo desta propriedade são complexos e ainda não há relatos de parâmetros de solubilidade associados a ferramentas quimiométricas na literatura. O melhor coeficiente de determinação obtido foi de 0,64 para OPLS (Figura 8g).

Teor de nitrogênio total

A literatura científica sobre a propriedade do teor de nitrogênio total em óleos também é escassa (neste trabalho, o melhor resultado de R^2 foi o MPA (0,90 - Figura 8h).

As variáveis mais significativas para a construção do modelo MPA podem ser observadas na Tabela 11:

Tabela 11: Regiões do espectro em ppm para as variáveis mais significativas do MPA para a propriedade Teor de nitrogênio total.

Região do Espectro – ppm – 41 variáveis selecionadas	Parâmetros moleculares de Poveda
0,0001 a 0,9720	H parafínico do tipo CH ₃ , em posição γ ou mais em relação ao sistema aromático
1,0024 a 1,4321	H parafínico ligados a sistemas aromáticos na posição β e alquílicos terminais
1.5525 a 1.5539	H naftênico do tipo CH ₂ ligados a sistemas aromáticos na posição β

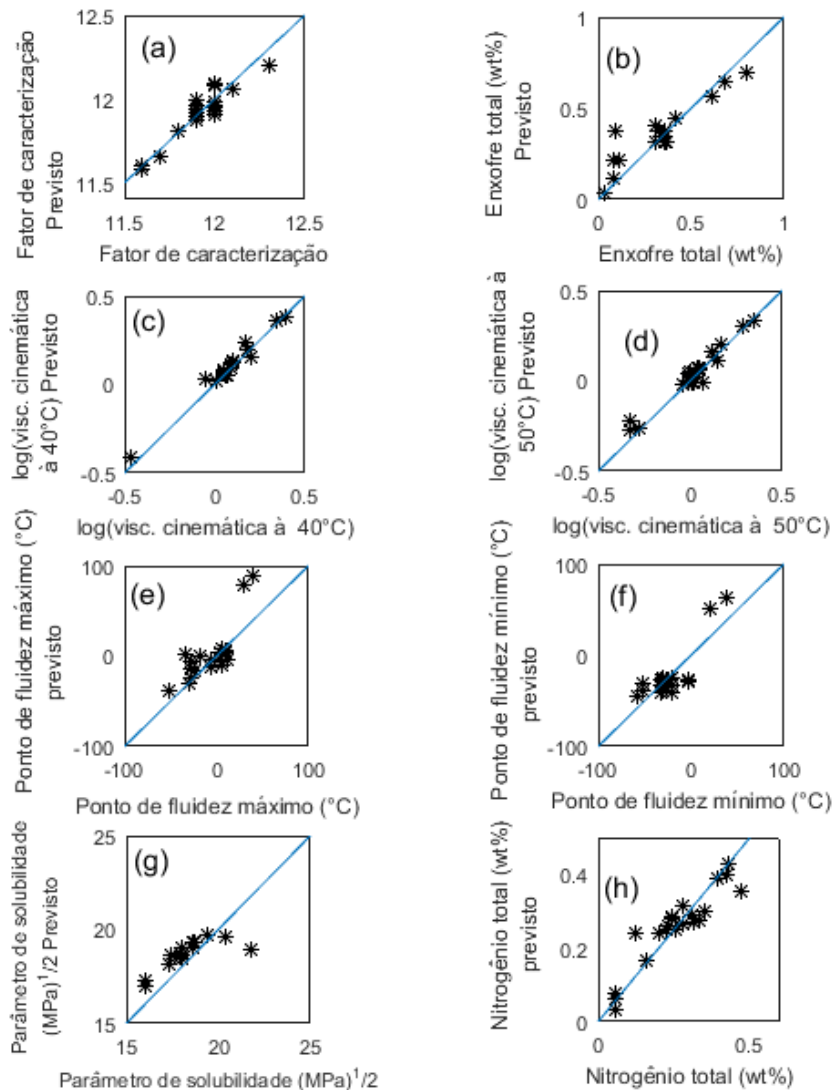


Figura 9. Gráfico dos valores de referência e previstos para as seguintes propriedades: (a) fator de caracterização; (b) enxofre total; (c) log viscosidade cinemática à 40 °C; (d) log viscosidade cinemática à 50 °C; (e) ponto de fluidez mínimo; (f) ponto de fluidez máximo; (g) parâmetro de solubilidade e (h) nitrogênio total.

MPA

Ao se utilizar a seleção de variáveis MPA observou-se uma boa relação linear pela maioria dos modelos com R^2 acima de 0,7. As propriedades teor de enxofre total, ponto de fluidez mínimo e parâmetro de solubilidade apresentaram um R^2 abaixo de 0,7. Foi feita a avaliação dos resíduos através do gráfico de leverage *versus* resíduos para observar possíveis *outliers*. Nos modelos de MPA não foram retirados *outliers* para nenhuma propriedade analisada. A modelagem MPA foi utilizada para analisar a qualidade preditiva do modelo com todas as amostras para todas as propriedades de interesse.

6.CONCLUSÃO

A técnica de RMN de ^1H associada a métodos de calibração multivariada demonstram eficácia na caracterização de petróleos.

O modelo de análise populacional (MPA) apresentou melhores resultados que o PLS, o OPLS, o siPLS e o IPLS para estimar a maioria das propriedades: fator de caracterização, nitrogênio total, ponto de fluidez mínimo, ponto de fluidez máximo, viscosidade cinemática à 20°C, 30°C, 40°C e 50°C. Pode-se considerar que o MPA é uma técnica quimiométrica recente e promissora.

Os coeficientes de determinação para as propriedades ponto de fluidez e parâmetro de solubilidade foram abaixo de 0,7, logo tais propriedades carecem de mais estudos.

7. REFERÊNCIAS BIBLIOGRÁFICAS

1. SPEIGHT, J. G.; WILEY, A. J. **Handbook of Petroleum Product Analysis**. 2002.
2. FILGUEIRAS, P. R. et al. Determination of Saturates, Aromatics, and Polars in Crude Oil by ^{13}C NMR and Support Vector Regression with Variable Selection by Genetic Algorithm. **Energy & Fuels**, v. 30, n. 3, p. 1972–1978, 2016
3. FILGUEIRAS, P. R. et al. Determination of API gravity , kinematic viscosity and water content in petroleum by ATR-FTIR spectroscopy and multivariate calibration. **Fuel**, v. 116, p. 123–130, 2014.
4. SILVA, S. L. et al. Chromatographic and spectroscopic analysis of heavy crude oil mixtures with emphasis in nuclear magnetic resonance spectroscopy: A review. **Analytica Chimica Acta**, v. 707, n. 1–2, p. 18–37, 2011.
5. DUARTE, L. M. et al. Determination of some physicochemical properties in Brazilian crude oil by ^1H NMR spectroscopy associated to chemometric approach. **Fuel**, v. 181, p. 660–669, 2016.
6. RODRIGUES, É. V. A. et al. Determination of crude oil physicochemical properties by high-temperature gas chromatography associated with multivariate calibration. **Fuel**, v. 220, n. January, p. 389–395, 2018.
7. NASCIMENTO, M. H. C. et al. Determination of flash point and Reid vapor pressure in petroleum from HTGC and DHA associated with chemometrics. **Fuel**, v. 234, n. July, p. 643–649, 2018.
8. ROCHA, J. T. C. et al. Sulfur Determination in Brazilian Petroleum Fractions by Mid-infrared and Near-infrared Spectroscopy and Partial Least Squares Associated with Variable Selection Methods. **Energy and Fuels**, v. 30, n. 1, p. 698–705, 2016.
9. RODRIGUES, R. R. T. et al. Evaluation of calibration transfer methods using the ATR FTIR technique to predict density of crude oil. **Chemometrics and Intelligent Laboratory**, n. 0, p. 7439, 2017.
10. AHMAD, I. et al. Characterization of Petroleum Crude Oils by Fourier Transform Infrared (FT-IR) and Gas Chromatography-Mass Spectrometrys. **Petroleum & Petrochemical Engineering**, p. 1–7, 2018.
11. NIELSEN, K. E. et al. Quantitative Analysis of Constituents in Heavy Fuel Oil by ^1H Nuclear Magnetic Resonance (NMR) Spectroscopy and Multivariate Data Analysis. **energy&fuels**, n. 9, p. 4070–4076, 2008.
12. POURABDOLLAH, K.; MOKHTARI, B. Application of ^1H NMR in the flow surveillance of oil wells. **Magnetic Resonance in Chemistry**, v. 50, n. 3, p. 208–215, 2012.

13. POVEDA, J.-C.; MOLINA, D. R.; PANTOJA-AGREDA, E. F. ^1H and ^{13}C NMR structural characterization of asphaltenes from vacuum residua modified by thermal cracking. **Journal Ciencia, Tecnologia y Futuro**, v. C, n. 4, p. 49–59, 2014.
14. BARROS, E. V. et al. Characterization of Naphthenic Acids in Thermally Degraded Petroleum by ESI(-)-FT-ICR MS and ^1H NMR after Solid-Phase Extraction and Liquid/Liquid Extraction. **Energy and Fuels**, v. 32, n. 3, p. 2878–2888, 2018.
15. MOLINA V., D.; NAVARRO URIBE, U.; MURGICH, J. Partial Least-Squares (PLS) Correlation between Refined Product Yields and Physicochemical Properties with the ^1H Nuclear Magnetic Resonance (NMR) Spectra of Colombian Crude Oils. **Energy & Fuels**, v. 21, n. 9, p. 1674–1680, 2007. Disponível em:
16. FERREIRA, M. M. C. et al. **Quimiometria I: calibração multivariada, um tutorial**. 1999
17. FILGUEIRAS, P. R. et al. Prediction of the distillation temperatures of crude oils using ^1H NMR and support vector regression with estimated confidence intervals. **Talanta**, v. 142, p. 197–205, 2015.
18. MOLINA, D.; NAVARRO, U.; MURGICH, J. Correlations between SARA fractions and physicochemical properties with ^1H NMR spectra of vacuum residues from Colombian crude oils. **Fuel**, v. 89, p. 185–192, 2010.
19. MASILI, A.; PULIGHEDDU, S.; SASSU, L. Prediction of physical – chemical properties of crude oils by ^1H NMR analysis of neat samples and chemometrics. **Magnetic Resonance in Chemistry**, n. July, p. 729–738, 2012.
20. SAVORANI, F.; TOMASI, G.; ENGELSEN, S. B. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. **Journal of Magnetic Resonance**, v. 202, n. 2, p. 190–202, 2010.
21. DUARTE, L. M. et al. Study of Distillation Temperature Curves from Brazilian Crude Oil by ^1H Nuclear Magnetic Resonance Spectroscopy in Association with Partial Least Squares Regression. **energy&fuels**, p. 1–6, 2017.
22. BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. **Applied Spectroscopy**, p. 772–777, 1989.
23. NORGAARD, L. et al. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. **Applied Spectroscopy**, v. 54, n. 3, p. 413–419, 2000.
24. FERRÃO, M. F. et al. Simultaneous determination of quality parameters of biodiesel / diesel blends using HATR-FTIR spectra and PLS , iPLS or siPLS

regressions. **Fuel**, v. 90, p. 701–706, 2011.

25. LI, H. et al. Model population analysis for variable selection. **Journal of Chemometrics**, n. May, p. 418–423, 2010.

26. LI, H.; LIANG, Y.; XU, Q. Model Population Analysis for Statistical Model Comparison. **Chemom. Pract. Appl.** 2012. p. 3–21.

27. DENG, B.; YUN, Y.; LIANG, Y. Model population analysis in chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 149, p. 166–176, 2015.

28. BIN, J. et al. An efficient variable selection method based on variable permutation and model population analysis for multivariate calibration of NIR spectra. **Chemometrics and Intelligent Laboratory Systems**, v. 158, p. 1–13, 2016.

29. DENG, B. et al. Model population analysis in model evaluation. **Chemometrics and Intelligent Laboratory Systems**, v. 172, p. 223–228, 2018. Disponível em: <<https://doi.org/10.1016/j.chemolab.2017.11.016>>.

30. CAO, D.S. et al. A new strategy of outlier detectin for QSAR/QSPR. 2010.

31. XU, Q.-S.; LIANG, Y.-Z.; DU, Y.-P. Monte Carlo cross-validation for selecting a model and estimating the prediction error in multivariate calibration. **Journal of Chemometrics**, 2004.

32. KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. **Technometrics**, 1969.

33. CASTRO, A. K. de. **Parâmetros de Solubilidade de Hildebrand dos petróleos e da mistura petróleo-heptano no limiar da precipitação dos asfaltenos empregando microscopia ótica e espectroscopia de infravermelho próximo (NIR)**. 2009. 0-107 f. Dissertação de mestrado. Universidade Federal do Maranhão, 2009.

34. WIEHE, I. A.; KENNEDY, R. J. The Oil Compatibility Model and Crude Oil Incompatibility. **energy&fuels**, p. 56–59, 2000.

35. HANSEN, C. M. 50 Years with solubility parameters - Past and future. **Progress in Organic Coatings**, v. 51, n. 1, p. 77–84, 2004.

36. UOP 375. Calculation of UOP Characterization factor and estimation of molecular weight of petroleum oils. **ASTM International**, 1986.

37. ASTM D4294. Standart Test Method for Sulfur in Petroleum and Petroleum Products by Energy Dispersive X-ray Fluorescence Spectrometry. **ASTM International**, 2010.

38. CANCIAN, R. V. **Estudo da distribuição de enxofre em frações**

destiladas de petróleo e análise multivariada das frações. 2010. Dissertação de mestrado. Universidade Federal do Espírito Santo, 2010.

39. ASTM D5853. Standard Test Method for Pour Point of Crude Oils. **ASTM International**, p. 1–8, 2011.

40. RIAZI, M. R. **Characterization and Properties of Petroleum Fractions.** 2005.

41. SAD, C. M. S. et al. Limitations of the Pour Point Measurement and the Influence of the Oil Composition on Its Detection Using Principal Component Analysis. **energy&fuels**, 2014.

42. RAINHA, K. P. **Estudo de alguns parâmetros físico-químicos de petróleo utilizando modelagem multivariada associada a espectroscopia no infravermelho.** 2014. Dissertação de mestrado. Universidade Federal do Espírito Santo, 2014.

43. ULLER, V. C. **Oportunidades e desafios da colocação de óleos brasileiros no mercado internacional: o refino e o mercado norte-americano de combustíveis.** 2007. Dissertação de mestrado. Universidade Federal do Espírito Santo, 2007.

44. MALACARNE, M. M. **Implantação e otimização de uma unidade manual de destilação a vácuo de petróleo no Labpetro/UFES.** 2011. Dissertação de mestrado. Universidade Federal do Espírito Santo, 2011.

45. TOZZI, F. C. **Estudo do efeito de mistura de petróleo na determinação da acidez, do teor de enxofre, do ponto de fluidez, da viscosidade e da gravidade API.** 2015. Dissertação de mestrado. Universidade Federal do Espírito Santo, 2015.

46. SILVERSTEIN, R. M.; WEBSTER, F. X.; KIEMLE, D. J. **Spectrometric Identification of Organic Compounds.** 7th. ed. [S.l: s.n.], 2007.

47. SOUZA, C. R. de. **Determinação do número de cetano no óleo diesel convencional através da espectroscopia de RMN de ¹H associada à calibração multivariada.** 2010. Dissertação de mestrado. Universidade Federal do Paraná, 2010.

48. 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**, v. 84–85, p. 1–7, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.petrol.2012.01.005>>.

49. BARROS, E. V. et al. Study of degradation of acid crude oil by high resolution analytical techniques. **Journal of Petroleum Science and Engineering**, v. 154, n. April, p. 194–203, 2017.

50. MOLINA, D.; ARIZA, E.; POVEDA, J. C. Structural differences among the

asphaltenes in Colombian light crudes from the Colorado Oil Field. **Energy and Fuels**, v. 31, n. 1, p. 133–139, 2017.

51. BRUNS, R. E.; FAIGLE, J. F. G. **Quimiometria**. . [S.l.: s.n.]. , 1984

52. VIEIRA, L. V. et al. Exploratory data analysis using API gravity and V and Ni contents to determine the origins of crude oil samples from petroleum fields in the Espirito Santo Basin (Brazil). **Microchemical Journal**, v. 124, p. 26–30, 2016.

53. BRERETON, R. G. **Applied Chemometrics for Scientists**. 2007.

54. HÖSKULDSSON, A. PLS regression methods. **Journal of Chemometrics**, v. 2, n. August 1987, p. 581–591, 1988.

55. WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: A basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109–130, 2001.

56. AMORIM, J. A. de. **Obtenção de propriedades físico-químicas de misturas de hidrocarbonetos em baixa e alta pressão visando a caracterização e modelagem**. 2007. Dissertação de mestrado..

57. BUENO, A. F. **Caracterização de petróleo por Espectroscopia no Infravermelho Próximo**. 2004.

58. WIZIACK, N. K. L. et al. A sensor array based on mass and capacitance transducers for the detection of adulterated gasolines. **Sensors and Actuators B: Chemical**, v. 140, n. 2, p. 508–513, 2009.

59. SANTOS, S. C. A. dos. **Aplicação de ferramentas quimiométricas na investigação de parâmetros para a definição do ponto de corte de interface em polidutos que transportam derivados de petróleo e álcool etílico anidro combustível**. 2013. Dissertação de mestrado. Universidade Federal da Bahia, 2013.

60. MOLINA, D.; PARRA, M.; MURGICH, J. MRL and PLS Correlations between the ¹H NMR Spectra of Vacuum Residues and the Deasphalting Conditions with the DMO Yields. **Petroleum Science and Technology**, n. September 2014, p. 37–41, 2011.

61. TERRA, L. A. et al. Petroleomics by electrospray ionization FT-ICR mass spectrometry coupled to partial least squares with variable selection methods: prediction of the total acid number of crude oils. **The Analyst**, v. 139, n. 19, p. 4908, 2014. Disponível em: <<http://xlink.rsc.org/?DOI=C4AN00538D>>.

62. WOLD, S. Orthogonal signal correction of near-infrared spectra. **Chemometrics and Intelligent Laboratory Systems**, p. 175–185, 1998.

63. SOUIHI, N. et al. OPLS in batch monitoring – Opens up new opportunities.

Analytica Chimica Acta, v. 857, p. 28–38, 2015.

64. SILVA, A. F. T. et al. Statistical process control of cocrystallization processes: A comparison between OPLS and PLS. **International Journal of Pharmaceutics**, v. 520, p. 29–38, 2017.

65. ERIKSSON, L. et al. Orthogonal PLS (OPLS) Modeling for Improved Analysis and Interpretation in Drug Design. **Molecular Informatics**, p. 414–419, 2012.

66. BYLESJÖ, M. et al. Orthogonal projections to latent structures as a strategy for microarray data normalization. **BMC Bioinformatics**, v. 10, p. 1–10, 2007.

67. TRYGG, J.; WOLD, S. Orthogonal projections to latent structures (O-PLS). **Journal of Chemometrics**, p. 119–128, 2002.

68. BOCCARD, J.; RUTLEDGE, D. N. A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. **Analytica Chimica Acta**, v. 769, p. 30–39, 2013.

69. PORTELA, N. A. et al. Quantification of biodiesel in petroleum diesel by ¹H NMR: Evaluation of univariate and multivariate approaches. **Fuel**, v. 166, p. 12–18, 2016.

70. MEHMOOD, T. et al. A review of variable selection methods in Partial Least Squares Regression. **Chemometrics and Intelligent Laboratory Systems**, v. 118, p. 62–69, 2012.

71. ROCHA, J. T. do C. **Aplicação de espectroscopia no infravermelho próximo (NIR) e médio (MIR) associada a métodos quimiométricos, para avaliação de parâmetros físicos-químicos em frações de petróleo**. 2016. Tese de doutorado. Universidade Federal de Juiz de Fora, 2016.

72. LI, H.; XU, Q.; LIANG, Y. libPLS: An integrated library for partial least squares regression and linear discriminant analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 176, n. March, p. 34–43, 2018.

73. MILLER, R. G. The jackknife - a review. **Oxford Journals**, 1974.

74. EFRON, B.; TIBSHIRANI, R. J. **An introduction to the Bootstrap**, 1994.

75. ZHANG, H. et al. Improving accuracy for cancer classification with a new algorithm for genes selection. **BMC Bioinformatics**, 2012.

76. XU, H. et al. A wavelength selection method based on randomization test for near-infrared spectral analysis. **Chemometrics and Intelligent Laboratory Systems**. [S.l: s.n.], 2009

77. SKOOG et al. **Fundamentos de Química Analítica**. [S.l: s.n.], 2005.

78. ASTM E1655. Standard Practices for Infrared Multivariate Quantitative Analysis. **ASTM International**, n. Reapproved 2012, p. 1–29, 2005.
79. ASTM D7042. Standard Test Method for Dynamic Viscosity and Density of Liquids by Stabinger Viscometer (and the Calculation of Kinematic Viscosity). **ASTM International**, p. 1–9, 2012.
80. DIAS, J. C. M.; AGUIAR, P. F. A statistical method for acceptance of crude oil viscosity - temperature curves. **Brazilian Journal of Petroleum and gas**, p. 5:19-24, 2011.
81. ACEVEDO, S. et al. Investigation of Physical Chemistry Properties of Asphaltenes Using Solubility Parameters of Asphaltenes and Their Fractions A1 and A2. **energy&fuels**, n. 28, p. 5921–5933, 2010.
82. OLIVEIRA, E. C. da S. **Utilização da RMN Associada a Métodos Quimiométricos na Caracterização de Petróleos e Derivados**. 2016. Tese de Doutorado. Universidade Federal do Espírito Santo, 2016.
83. GIL, Victor M. S., GERALDES, C. F. C. **Ressonância Magnética Nuclear. Fundamentos, métodos e aplicações**. 2ª Edição. Editora Fundação Calouste Gulbenkian, 2002
84. SILVERSTEIN, R.M., WEBSTER, F.X. & KIEMLE, D.J. **Identificação Espectrométrica de Compostos Orgânicos** 7th ed. LTC, 2007.