

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
DEPARTAMENTO DE QUÍMICA
CURSO DE QUÍMICA BACHARELADO

Andressa Lopez Soares

**DETERMINAÇÃO DO GRAU API E TEOR DE ENXOFRE E
NITROGÊNIO EM PETRÓLEO USANDO ATR-FTIR E MÉTODOS
QUIMIOMÉTRICOS**

Santa Maria, RS
2019

Andressa Lopez Soares

**DETERMINAÇÃO DO GRAU API E TEOR DE ENXOFRE E NITROGÊNIO EM
PETRÓLEO USANDO ATR-FTIR E MÉTODOS QUIMIOMÉTRICOS**

Trabalho de Conclusão de Curso apresentado ao Curso de Química Bacharelado, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Química**.

Orientador: Prof. Dr. Edson Irineu Müller

Santa Maria, RS
2019

RESUMO

DETERMINAÇÃO DO GRAU API E TEOR DE ENXOFRE E NITROGÊNIO EM PETRÓLEO USANDO ATR-FTIR E MÉTODOS QUIMIOMÉTRICOS

AUTOR: Andressa Lopez Soares

ORIENTADOR: Prof. Dr. Edson Irineu Müller

Grau API e a concentração de enxofre e de nitrogênio são propriedades importantes para determinar a qualidade do petróleo. As determinações do grau API, S e N foram feitas pelas normas ASTM D-7042, ASTM D-5453 e ASTM D-4629, respectivamente. Como alternativa aos métodos oficiais, a determinação desses parâmetros pode ser realizada utilizando a espectroscopia no infravermelho por transformada de Fourier com reflectância total atenuada (ATR-FTIR) associada à métodos de calibração multivariada. Foram avaliados os métodos de transformação de dados com 1ª derivada usando filtro de Savitzky-Golay e correção de espalhamento multiplicativo (MSC). A construção inicial dos modelos de calibração foi feita usando as amostras dos conjuntos de calibração e previsão (aleatoriamente selecionados) e algoritmo dos mínimos quadrados parciais (PLS). Para a seleção de variáveis foram avaliados os métodos iPLS, biPLS e siPLS (iToolbox para MATLAB) para cada propriedade do petróleo. O algoritmo iPLS produziu o melhor modelo de calibração para enxofre, utilizando 335 variáveis independentes e um valor de erro médio quadrático de previsão (RMSEP) de 517 mg kg⁻¹. O algoritmo siPLS apresentou os melhores modelos para o grau API e teor de nitrogênio total, com valores de RMSEP de 1,98 ° e 321 mg kg⁻¹ e número de variáveis iguais à 168 e 167, respectivamente. Os resultados obtidos demonstraram que a os modelos de calibração multivariada são uma alternativa viável para a determinação de API, S e N em petróleos. Além disso, a aplicação das ferramentas de seleção de variáveis permitiu a otimização do método, tornando-o mais robusto, simples e rápido.

Palavras-chave: ATR-FTIR. Calibração multivariada. Seleção de variáveis.

ABSTRACT

DETERMINATION OF API GRADE AND SULFUR AND NITROGEN CONTENT IN OIL USING ATR-FTIR AND CHEMOMETRICS METHODS

AUTHOR: Andressa Lopez Soares
ADVISOR: Prof. Dr. Edson Irineu Müller

Parameters such as API gravity and sulfur and nitrogen concentrations are important to determine crude oil quality. API, S and N grade determinations were evaluated by ASTM D-7042, ASTM D-5453 and ASTM D-4629 methods, respectively. As an alternative to official methods, the determination of these parameters can be performed using Fourier transform infrared spectroscopy with attenuated total reflectance (ATR-FTIR) associated with multivariate calibration methods. The first derivative with a Savitzky-Golay filter (SG) and multiplicative scattering correction (MSC) data transformation methods were evaluated. The initial construction of calibration models was set up using randomly selected samples for the calibration and prediction sets and partial least squares algorithm (PLS). The iPLS, biPLS and siPLS (iToolbox for MATLAB) were evaluated as variable selection methods for each oil property. The iPLS algorithm produced the best calibration model for sulfur using 335 independent variables and a root mean square error of prediction (RMSEP) of 517 mg kg⁻¹. The siPLS algorithm presented the best models for API gravity and total nitrogen content, with RMSEP values of 1.98 ° and 321 mg kg⁻¹ and number of variables equal to 168 and 167, respectively. The results showed that multivariate calibration models are a viable alternative for the determination of API, S and N in crude oils. In addition, the application of the variable selection tools allowed the optimization of the method, making it more robust, simpler and faster.

Keywords: ATR-FTIR. Multivariate Calibration. Variable selection.

AGRADECIMENTOS

À Universidade Federal de Santa Maria, pelo ensino, pesquisa e extensão, e todas as oportunidades de crescimento pessoal e acadêmico que me foram proporcionadas.

Ao professor Dr. Edson I. Müller, pela orientação deste trabalho. Obrigada por todo apoio e incentivo, desde o momento que solicitei a orientação. Sou grata por todos os ensinamentos, pela paciência e pelo exemplo de profissionalismo e dedicação.

Ao CENPES/PETROBRAS, pela doação das amostras.

À Sheila Oliveira, doutoranda da Universidade Federal do Paraná, responsável por me apresentar à Quimiometria e me contagiar com sua paixão pela pesquisa. Obrigada pela amizade e pelos conhecimentos compartilhados.

Ao Henrique Faccin, pela participação na banca, apoio e sugestões para a melhoria do trabalho. Obrigada também por todo suporte enquanto secretário do curso.

Aos meus pais, Sérgio e Maristela, pelo amor incondicional, suporte e incentivo durante todas as etapas da minha educação. Obrigada por acreditarem em mim e por tornar isto possível.

E à minha prima e melhor amiga, Laura Mayer, por toda cumplicidade, compreensão e apoio, durante todos os anos de graduação e em especial, durante a realização deste trabalho. Obrigada por me ouvir, me entender e tornar tudo mais fácil sempre!

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1 – Esquema representando os tipos de pré-tratamento de dados e os respectivos métodos mais utilizados | 22 |
| Figura 2 – Perfil dos espectros médios das amostras de petróleo obtidos por ATR-FTIR | 33 |
| Figura 3 – Principais etapas desenvolvidas para a obtenção do melhor modelo | 34 |
| Figura 4 – Conjunto de espectros médios obtidos por ATR-FTIR das amostras de petróleo na faixa de 650 - 4000 cm^{-1} | 37 |
| Figura 5 – Conjunto de espectros obtidos no infravermelho médio para as amostras de petróleo na faixa de 650 - 4000 cm^{-1} após pré-tratamento com MSC | 38 |
| Figura 6 – Regiões espectrais selecionadas no modelo siPLS60 | 40 |
| Figura 7 – Gráfico de correlação entre os valores medidos e previstos do grau API obtido a partir do modelo siPLS60 | 40 |
| Figura 8 – Gráfico de erro percentual de cada amostra do conjunto de previsão do grau API | 41 |
| Figura 9 – Conjunto de espectros obtidos no infravermelho médio para as amostras de petróleo na faixa de 651 - 3999 cm^{-1} após pré-tratamento com SG .. | 42 |
| Figura 10 – Região espectral selecionada no modelo iPLS10 | 44 |
| Figura 11 – Gráfico de correlação entre os valores medidos e previstos do teor de enxofre obtido a partir do modelo iPLS10 | 44 |
| Figura 12 – Gráfico de erro percentual de cada amostra do conjunto de previsão do teor de enxofre | 45 |
| Figura 13 – Regiões espectrais selecionadas no modelo siPLS60 | 47 |
| Figura 14 – Gráfico de correlação entre os valores medidos e previstos do teor de nitrogênio obtido a partir do modelo siPLS60..... | 48 |
| Figura 15 – Gráfico de erro percentual de cada amostra do conjunto de previsão do teor de nitrogênio | 48 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1 – Tipos de petróleo..... | 15 |
| Tabela 2 – Regiões espectrais do infravermelho | 19 |
| Tabela 3 – Resultados estatísticos dos modelos globais aplicando os pré-tratamentos SG e MSC e o método de regressão PLS para a propriedade grau API38 | |
| Tabela 4 – Resultados obtidos dos melhores modelos construídos para predição do grau API | 39 |
| Tabela 5 – Resultados estatísticos dos modelos globais aplicando os pré-tratamentos SG e MSC na regressão PLS para o teor de enxofre | 42 |
| Tabela 6 – Resultados obtidos dos melhores modelos construídos para predição do teor de enxofre | 43 |
| Tabela 7 – Resultados estatísticos dos modelos globais de aplicando os pré-tratamentos MSC e SG na regressão PLS para o teor de nitrogênio | 46 |
| Tabela 8 – Resultados obtidos dos melhores modelos construídos para predição do teor de nitrogênio | 47 |

LISTA DE QUADROS

| | |
|--|----|
| Quadro 1 – Compostos de enxofre presentes no petróleo..... | 17 |
|--|----|

LISTA DE ABREVIATURAS E SIGLAS

| | |
|----------|--|
| ANP | Agência Nacional do Petróleo |
| API | Instituto Americano do Petróleo (do inglês <i>American Petroleum Institute</i>) |
| ASCII | Código Americano para o Intercâmbio de Informação (do inglês <i>American Standard Code for Information Interchange</i>) |
| ASTM | Sociedade Americana de Ensaios e Materiais (do inglês <i>American Society for Testing and Materials</i>) |
| ATR | Reflexão Total Atenuada (do inglês <i>Attenuated Total Reflection</i>) |
| ATR-FTIR | Espectroscopia no infravermelho com Transformada de Fourier e Reflexão Total Atenuada (do inglês <i>Fourier Transform Infrared Spectroscopy with Attenuated Total Reflection</i>) |
| biPLS | Mínimos quadrados parciais por exclusão de intervalos (do inglês <i>Backward Interval Partial Least Squares</i>) |
| CENPES | Centro de Pesquisas e Desenvolvimento Leopoldo Américo Miguez de Mello |
| DTGS | Detector de sulfato de triglicina deuterada (do inglês <i>Deuterated Triglycine Sulfate</i>) |
| FIR | Infravermelho Distante (do inglês <i>Far Infrared</i>) |
| FTIR | Espectroscopia no Infravermelho com Transformada de Fourier (do inglês <i>Fourier Transform Infrared Spectroscopy</i>) |
| iPLS | Mínimos Quadrados Parciais por Intervalo (do inglês <i>Interval Partial Least Squares</i>) |
| MID | Infravermelho Médio (do inglês <i>Middle Infrared</i>) |
| MLR | Regressão Múltipla Linear (do inglês <i>Multiple Linear Regression</i>) |
| MSC | Correção de Espalhamento Multiplicativo (do inglês <i>Multiplicative Scatter Correction</i>) |
| NIR | Infravermelho Próximo (do inglês <i>Neer Infrared</i>) |
| PC | Componente Principal (do inglês <i>Principal Component</i>) |
| 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 | Coeficiente de correlação |
| RMSE | Raiz Quadrada Dos Erros Médios (do inglês <i>Root Mean Square Error</i>) |

| | |
|--------|---|
| RMSEC | Raiz Quadrada Do Erro Médio Da Calibração (do inglês Root Mean Square Error Of Calibration) |
| RMSECV | Raiz Quadrada Do Erro Médio Da Validação Cruzada (do inglês, Root Mean Square Error of Cross-Validation) |
| RMSEP | Raiz Quadrada Do Erro Médio De Previsão (do inglês, Root Mean Square Error of Prediction) |
| SG | Primeira derivada com filtro de Savitzky-Golay |
| SiPLS | Mínimos quadrados parciais por sinergismo de intervalos (do inglês, <i>Synergy Interval Partial Least Squares</i>) |
| SNV | Padronização normal de sinal (do inglês, <i>Standard Normal Variate</i>) |
| TAN | Índice de acidez total (do inglês, <i>Total Acid Number</i>) |
| VL | Variáveis Latentes |

SUMÁRIO

| | |
|---|-----------|
| 1 INTRODUÇÃO | 12 |
| 2 REFERENCIAL TEÓRICO | 14 |
| 2.1 PETRÓLEO | 14 |
| 2.1.1 Grau API..... | 14 |
| 2.1.2 Teor de enxofre | 16 |
| 2.1.3 Teor de nitrogênio..... | 18 |
| 2.2 ESPECTROSCOPIA NO INFRAVERMELHO | 18 |
| 2.3 ANÁLISE MULTIVARIADA | 21 |
| 2.3.1 Quimiometria..... | 21 |
| 2.3.2 Preparação dos dados | 21 |
| 2.3.2.1 Transformação dos dados | 23 |
| 2.3.2.1.1 Método da 1ª e 2ª derivada | 23 |
| 2.3.2.1.2 Método da correção do espalhamento multiplicativo (MSC) | 23 |
| 2.3.2.1.3 Padronização normal de sinal (SNV)..... | 24 |
| 2.3.2.2 Pré-processamento | 24 |
| 2.3.2.2.1 Centragem dos dados na média | 24 |
| 2.3.2.2.2 Autoescalamento..... | 24 |
| 2.3.3 Análise de componentes principais (PCA)..... | 25 |
| 2.3.4 Métodos de calibração multivariada | 26 |
| 2.3.5 Métodos de seleção de variáveis | 27 |
| 2.3.5.1 Regressão por mínimos quadrados parciais por intervalo (iPLS) | 28 |
| 2.3.5.2 Regressão por mínimos quadrados parciais por exclusão (biPLS) | 28 |
| 2.3.5.3 Regressão por mínimos quadrados parciais por sinergismo (siPLS) | 29 |
| 2.3.6 Avaliação dos modelos de regressão | 29 |
| 2.3.7 Testes de significância..... | 31 |
| 3 METODOLOGIA | 32 |
| 3.1 AMOSTRAS..... | 32 |
| 3.2 MEDIÇÃO DAS PROPRIEDADES PELO MÉTODO DE REFERÊNCIA | 32 |
| 3.3 ANÁLISE UTILIZANDO ESPECTROSCOPIA NO INFRAVERMELHO..... | 32 |
| 3.4 ANÁLISE MULTIVARIADA DOS DADOS..... | 33 |
| 4 RESULTADOS E DISCUSSÃO | 36 |
| 4.1 MODELO DE CALIBRAÇÃO PARA O GRAU API..... | 37 |
| 4.2 MODELO DE CALIBRAÇÃO PARA O TEOR DE ENXOFRE | 41 |
| 4.3 MODELO DE CALIBRAÇÃO PARA O TEOR DE NITROGÊNIO | 45 |
| 5 CONCLUSÃO | 49 |
| 6 REFERÊNCIAS | 51 |

1 INTRODUÇÃO

O petróleo é um recurso mineral de extrema importância nos dias de hoje, pois além de ser a fonte de energia mais utilizada no mundo, é também uma das matérias-primas mais importantes para a indústria. Seus derivados são usados em diversos setores essenciais na sociedade moderna, como meios de transporte e construção civil (SIMANZHENKOV; IDEM, 2003).

Apesar de ser constituído principalmente por hidrocarbonetos, observa-se também a presença de compostos contendo oxigênio, enxofre, nitrogênio e outros elementos em sua composição. Esses compostos podem causar uma série de problemas nas etapas de refinamento e processamento do petróleo, como envenenamentos dos catalisadores pelos compostos nitrogenados e corrosão dos equipamentos pelos compostos contendo enxofre (SPEIGHT, 2001).

Além disso, características como o grau API (Instituto Americano do Petróleo, do inglês *American Petroleum Institute*) afetam fortemente a viabilidade econômica dos campos produtores, pois, além de influenciar no valor do petróleo, auxiliam no projeto dos equipamentos utilizados na exploração e na produtividade do campo (FILGUEIRAS et al, 2014).

Devido a sua complexidade, a avaliação desses parâmetros é imprescindível para determinar sua qualidade. De acordo com Simanzhenkov e Idem (2003), petróleos de boa qualidade são definidos por altos valores de grau API e baixas concentrações de enxofre e nitrogênio. Com a crescente exploração e produção de petróleo no Brasil, a busca por métodos de determinação desses parâmetros de uma forma mais simples e robusta tem aumentado.

Atualmente a determinação do grau API é feita através da norma ASTM D-7042, utilizando um viscosímetro. As concentrações de enxofre e nitrogênio são determinadas pelas normas ASTM D-5453 e ASTM D-4629, respectivamente, e utilizam o método de combustão seguido da detecção do enxofre por fluorescência no ultravioleta e do nitrogênio total por quimiluminescência. Entretanto, esses métodos apresentam algumas desvantagens como elevado consumo de reagentes, tempo de análise e preparo de amostra.

Nesse sentido, a espectroscopia de infravermelho por transformada de *Fourier* com acessório de reflexão total atenuada (ATR-FTIR) apresenta algumas vantagens

em relação aos demais métodos analíticos, como a redução do uso de solventes, mínimo preparo da amostra e maior rapidez na análise.

A partir do avanço das técnicas computacionais e, conseqüentemente, dos métodos de análise multivariada de dados tornou-se mais fácil o tratamento dos dados espectrais obtidos por espectroscopia no infravermelho. Os métodos de análise multivariada são divididos em métodos de análise exploratória, de classificação e de regressão e são utilizados com o objetivo de obter maiores informações químicas através dos dados obtidos e facilitar a interpretação dos resultados (BRERETON, 2000).

Devido à vasta aplicabilidade das ferramentas quimiométricas, este trabalho tem como objetivo principal a construção de modelos de calibração robustos para predição dos valores de grau API e concentração de enxofre e de nitrogênio total em amostras de petróleo bruto, utilizando a espectroscopia no infravermelho médio com transformada de Fourier e reflexão total atenuada (ATR-FTIR) combinada às técnicas de calibração multivariada.

Para tanto, será utilizado o método de regressão por mínimos quadrados parciais (PLS, do inglês *Partial Least Squares*) para construir as curvas de calibração. Além disso, afim de otimizar os modelos, pretende-se avaliar as ferramentas de transformação de dados 1ª derivada com filtro de Savitzky-Golay (SG) e correção de espalhamento multiplicativo (MSC), assim como as ferramentas de seleção de variáveis, o método de regressão por mínimos quadrados parciais por intervalo (iPLS), por exclusão (biPLS) e por sinergismo (siPLS).

2 REFERENCIAL TEÓRICO

2.1 PETRÓLEO

O petróleo é caracterizado como uma mistura complexa de hidrocarbonetos, geralmente no estado líquido, que pode conter também compostos de enxofre, oxigênio, nitrogênio, metais e outros componentes (SPEIGHT, 2002). Sua composição varia de acordo com alguns critérios, como por exemplo, a localização, a idade e a profundidade de cada poço (SPEIGHT, 2001). Uma vez que as propriedades físico-químicas são cruciais na seleção do método de processamento, armazenamento e transporte do petróleo, é necessário classificar o petróleo de acordo as mesmas (SIMANZHENKOV; IDEM, 2003).

Antigamente, a qualidade do petróleo era mensurada simplesmente por suas características físico-químicas, como grau API e viscosidade. Essas características determinam a leveza do petróleo e isso implica na facilidade de processamento e na qualidade dos derivados obtidos. Atualmente, é sabido que outras características influenciam na qualidade e por isso, dados analíticos elementares, como a quantidade de enxofre e de nitrogênio, passaram a ser usados também para determiná-la (SPEIGHT, 2001).

A determinação rápida de parâmetros físico-químicos é crucial para o processamento do petróleo. Recentemente, a espectroscopia por infravermelho surgiu como uma ferramenta na análise quantitativa de petróleo e tem apresentado muitas vantagens, como a necessidade de pouca quantidade de amostra, procedimentos rápidos e pouco ou nenhum preparo de amostra. Com essa técnica, a conversão da resposta instrumental obtida em dados de interesse necessita o auxílio das técnicas de calibração multivariada (FILGUEIRAS et al., 2014).

2.1.1 Grau API

O grau API é uma medida relacionada à densidade do petróleo estabelecida pelo *American Petroleum Institute* (API) e serve para a identificação comercial dos diferentes tipos de petróleo. O valor, em termos de grau API, pode ser calculado de acordo com a Equação 1 (SPEIGHT, 2002).

$$API (^{\circ}) = \frac{141,51}{Densidade} - 131,50 \quad (1)$$

Sendo a densidade de uma substância dada pela relação entre a massa de um determinado volume de matéria e a massa de igual volume de água, medidos à temperatura de 20 °C.

De acordo com a Agência Nacional do Petróleo (ANP), o grau API é usado para classificar o petróleo em quatro categorias distintas: leve, mediano, pesado e extrapesado conforme mostrado na Tabela 1 (BRASIL, 2000).

Tabela 1 – Tipos de petróleo

| Tipo de petróleo | API |
|-------------------------|------------------------------------|
| Leve | $\geq 31^{\circ}$ |
| Mediano | $22^{\circ} \leq API < 31^{\circ}$ |
| Pesado | $10^{\circ} \leq API < 22^{\circ}$ |
| Extrapesado | $< 10^{\circ}$ |

Fonte: Adaptado de ANP, 2000.

O grau API afeta diretamente a viabilidade econômica dos campos de produção porque, além de petróleos com maiores valores de grau API serem mais caros, ele influencia também na escolha do equipamento a ser usado na etapa de exploração (FILGUEIRAS et al, 2014).

Tipicamente, os valores de grau API são obtidos por um procedimento padrão (ASTM D-5002) estabelecido pela Sociedade Americana de Ensaio e Materiais (ASTM, do inglês *American Society for Testing and Materials*) e utiliza um densímetro digital para realizar as medidas (SPEIGHT, 2002). Entretanto, publicações recentes mostram que esses valores podem ser previstos diretamente da composição molecular, através de dados espectrais obtidos por espectroscopia na região do infravermelho.

No método proposto por Filgueiras et al. (2014) foi possível determinar o grau API de amostras de petróleo utilizando espectroscopia de reflexão total atenuada no infravermelho com transformada de Fourier (ATR-FTIR) associada à calibração multivariada. O método apresentou resultados satisfatórios, com erros de predição

relativamente pequenos, na ordem de $0,38^{\circ}$ para o método de regressão por quadrados mínimos parciais (PLS), e algumas vantagens como rapidez da análise e simplicidade no preparo da amostra.

2.1.2 Teor de enxofre

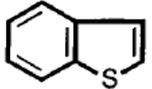
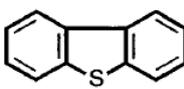
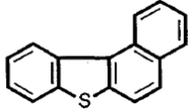
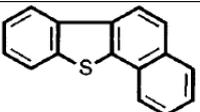
Os compostos de enxofre estão entre os principais constituintes heteroatômicos do petróleo. O enxofre está presente na forma de diversos compostos, conforme dispostos no Quadro 1, e as quantidades totais presentes no petróleo bruto variam de 0,1 a 8% (m/m) (SPEIGHT, 2002).

A presença de enxofre no petróleo bruto e suas frações pode causar um grande impacto nos processos de refino, resultando na corrosão de materiais metálicos e no entupimento dos gasodutos. Além disso, pode levar a formação de compostos de óxido de enxofre (SO_x) durante a combustão dos seus derivados. Por isso, para garantir a qualidade do petróleo bruto e suas frações é preciso monitorar frequentemente a quantidade de compostos de enxofre presentes nele (MÜLLER et al, 2012).

O método de referência para determinação de enxofre e petróleo é baseado na decomposição de amostras usando bombas ou tubos de combustão, seguindo o procedimento proposto pela norma ASTM D-129-00. Todavia, outros métodos têm surgido propondo análises mais rápidas e utilizando uma menor quantidade de amostra (MÜLLER et al, 2012).

Müller et al. (2012) propuseram um método alternativo que utiliza os dados espectrais obtidos por ATR-FTIR para determinar a quantidade de enxofre em amostras de resíduos da destilação do petróleo. O método combina a ATR-FTIR associado a métodos quimiométricos de calibração multivariada e seleção de variáveis, apresentando resultados satisfatórios e demonstrando-se uma alternativa viável e mais rápida para o método de referência. O melhor modelo utilizou o método de regressão por quadrados mínimos parciais por sinergismo (siPLS), combinando três intervalos e produziu valores de erro médio quadrático de validação cruzada (RMSECV) de 400 mg kg^{-1} e de erro médio quadrático de previsão (RMSEP) de 420 mg kg^{-1} , além de um coeficiente de correlação igual a 0,990.

Quadro 1 – Compostos de enxofre presentes no petróleo

| Composto | Nomenclatura |
|---|--------------------|
| RSH | Tióis |
| RSR' | Sulfetos |
|  | Sulfetos cíclicos |
| RSSR' | Dissulfetos |
|  | Tiofenos |
|  | Benzotiofenos |
|  | Dibenzotiofenos |
|  | Naftobenzotiofenos |
|  | |

Fonte: Adaptado de Speight (2001).

2.1.3 Teor de nitrogênio

No petróleo, o nitrogênio está presente na forma de compostos não básicos, como o pirrol, o indol e o carbazol, e na forma de compostos básicos como a piridina e seus derivados. Ambos estão presentes em maior quantidade nas frações de ponto de ebulição mais elevado (SPEIGHT, 2001).

Geralmente, a quantidade total de nitrogênio no petróleo bruto é baixa e está na faixa de 0,1 a 1,0 % (m/m) (SPEIGHT, 2002). Os compostos contendo nitrogênio podem ser responsáveis pelo envenenamento de catalizadores no processo de craqueamento do petróleo. Além disso, contribuem para a formação de goma em produtos como o óleo combustível doméstico (SPEIGHT, 2001). Desse modo, quanto menor o teor de nitrogênio, melhor a qualidade do petróleo e por isso é importante sua determinação.

O método de referência usado para medir a quantidade total de nitrogênio presente no petróleo é a norma ASTM D-4629, que combina técnicas de combustão oxidativa e detecção por quimiluminescência (SPEIGHT, 2002).

Na literatura, publicações recentes de métodos alternativos para a determinação do nitrogênio no petróleo podem ser encontradas. Ainda que seja apenas para a determinação do teor de nitrogênio básico, Müller et al. (2013) apresentaram um método usando ATR-FTIR combinado à calibração multivariada e à seleção de variáveis que permitiu determinar sua quantidade através dos dados espectrais das amostras.

Foram aplicadas as ferramentas de seleção de regressão por mínimos quadrados parciais por intervalo, por exclusão e por sinergismo (iPLS, biPLS e siPLS). O modelo obtido a partir do biPLS apresentou os melhores resultados, com um percentual de erro 10 % menor em relação ao método de referência e diversas vantagens, como menor custo e maior rapidez de análise.

2.2 ESPECTROSCOPIA NO INFRAVERMELHO

A radiação infravermelha corresponde a parte do espectro eletromagnético que está situada entre a região do visível e a região das micro-ondas (SILVERSTEIN; WEBSTER; KIEMLE, 2006). A região pertence ao intervalo de número de onda que vai de 10 a 12800 cm^{-1} e é subdividido em infravermelho próximo (NIR, do inglês *Near*

Infrared), médio (MID, do inglês *Middle Infrared*) e distante (FIR, do inglês *Far Infrared*). As subdivisões estão representadas na Tabela 2 (BARBOSA, 2007).

Tabela 2 – Regiões espectrais do infravermelho

| Região | Número de onda (cm ⁻¹) | Comprimento de onda (nm) | Frequência (Hz) |
|----------------|------------------------------------|--------------------------|---|
| Próximo (NIR) | 4.000 a 12.800 | 780 a 2.500 | 3,8 x 10 ¹⁴ a 1,2 x 10 ¹⁴ |
| Médio (MID) | 200 a 4.000 | 2.500 a 5.000 | 1,2 x 10 ¹⁴ a 6,0 x 10 ¹² |
| Distante (FIR) | 10 a 200 | 5.000 a 10.000 | 6,0 x 10 ¹² a 3,0 x 10 ¹¹ |

Fonte: Adaptado de Barbosa, 2007.

A espectroscopia no infravermelho é baseada no fato de que, quando as moléculas absorvem luz de determinada frequência, elas são excitadas a um nível mais alto de energia, causando alterações nos seus modos vibracionais e rotacionais (SKOOG, 2002). Desse modo, pode ser utilizada para quantificar componentes, identificar componentes desconhecidos, além de determinar a estrutura química da molécula (PAVIA; LAMPMAN; KRIZ, 1996).

Para que a molécula absorva a radiação infravermelha, é preciso que, como consequência do movimento vibracional ou rotacional, ela sofra uma alteração no seu momento dipolo. Só assim o campo elétrico alternado da radiação pode interagir com a molécula, provocando variações na amplitude de um de seus movimentos (SKOOG, 2002).

As vibrações e rotações em torno das ligações da molécula alteram continuamente as posições relativas dos átomos. As vibrações podem ser de dois tipos: estiramentos, que envolvem a variação contínua na distância entre os átomos ao longo do eixo de sua ligação; e deformações angulares, que são caracterizadas pela variação do ângulo entre duas ligações (SKOOG, 2002).

Dentre os instrumentos utilizados para obter espectros no infravermelho destacam-se os espectrômetros no infravermelho com transformada de *Fourier* (FTIR), que apresenta como componente óptico básico o interferômetro de *Michelson*. Esse instrumento apresenta uma série de vantagens, como maior potência da radiação que incide no detector, promovendo um aumento na relação sinal-ruído e a obtenção de espectros em poucos segundos (SKOOG, 2002).

Nessa técnica os espectros são obtidos posicionando a amostra entre o interferômetro e o detector, onde a radiação de alguns comprimentos de onda serão absorvidas. Em seguida, o procedimento matemático denominado transformada de *Fourier* é aplicado, afim de transformar os interferogramas em espectros em termos de frequência, tornando possível sua interpretação (BARBOSA, 2007; SILVERSTEIN; WEBSTER; KIEMLE, 2006).

Existem diversos modos de aquisição de espectros no infravermelho, destacando-se os de transmissão e reflexão. Em ambos os casos existem acessórios disponíveis comercialmente que facilitam a obtenção de espectros de amostras sólidas, líquida e gasosas. Os acessórios mais usados são os de reflexão externa ou especular, de reflexão total atenuada (ATR, do inglês *Attenuated Total Reflection*) e de reflexão difusa (BARBOSA, 2007).

A região do infravermelho médio (MID) possui a maior parte das aplicações da espectroscopia no infravermelho, sendo originalmente destinadas à identificação dos compostos orgânicos. Contudo, com o avanço dos instrumentos interferométricos e o surgimento da transformada de *Fourier*, surgiram também diversas aplicações na área de análises quantitativas (SKOOG, 2002).

Além disso, os picos de absorção espectrais da espectroscopia na região do MID são bandas fundamentais específicas, nítidas e que incluem as bandas de absorção provenientes de espécies aromáticas, enxofre e nitrogênio. Dessa forma, a espectroscopia MID é uma boa alternativa para analisar parâmetros do petróleo bruto, uma vez que a maioria das informações úteis podem ser extraídas por métodos quimiométricos (LI et al., 2011).

Em trabalhos recentes, a espectroscopia de infravermelho por transformada de *Fourier* com reflexão total atenuada (ATR-FTIR) tem sido utilizada para análises de petróleo bruto e suas frações, com a finalidade de minimizar o pré-tratamento e aumentar o rendimento da amostra. Como os espectros obtidos por infravermelho envolvem uma quantidade muito grandes de dados e tem um grau de complexidade elevado, é necessário utilizar técnicas quimiométricas para analisá-los (MÜLLER et al., 2013).

O uso da espectroscopia no infravermelho combinado à técnicas quimiométricas tem sido descrito na literatura para determinar parâmetros de petróleo bruto e seus derivados (BRERETON, 2003). Além disso, a análise multivariada foi

usada para determinar o índice de acidez total (TAN, do inglês *Total Acid Number*) em amostras de frações de petróleo por Parisotto et al. (2010).

2.3 ANÁLISE MULTIVARIADA

2.3.1 Quimiometria

Os avanços tecnológicos das últimas décadas, promoveram mudanças significativas em todas as áreas da ciência, dentre elas a química. O desenvolvimento de equipamentos como espectrômetros e cromatógrafos impulsionaram o crescimento na quantidade e na variedade de dados experimentais produzidos, surgindo assim a necessidade de tratar e interpretar esses dados a fim de obter informações químicas relevantes (FERREIRA, 2015). Essa necessidade foi a principal causa do desenvolvimento da quimiometria que, segundo Malinowski (1991 apud FERREIRA, 2015, p. 24), é o “uso de métodos matemáticos e estatísticos para o tratamento, a interpretação e a previsão de dados químicos”.

Apesar de ser um instrumento essencial na química analítica, a quimiometria está presente também em outras áreas da química, estabelecendo uma natureza multidisciplinar. Além disso, tem sido aplicada também nas indústrias química, farmacêutica e de alimentos, e na área médica, auxiliando na busca e identificação de marcadores de doenças (FERREIRA, 2015).

A quimiometria envolve um conjunto de métodos que vão desde a organização e a preparação dos dados, passando por uma análise exploratória até a construção de modelos de classificação ou de regressão (FERREIRA, 2015).

2.3.2 Preparação dos dados

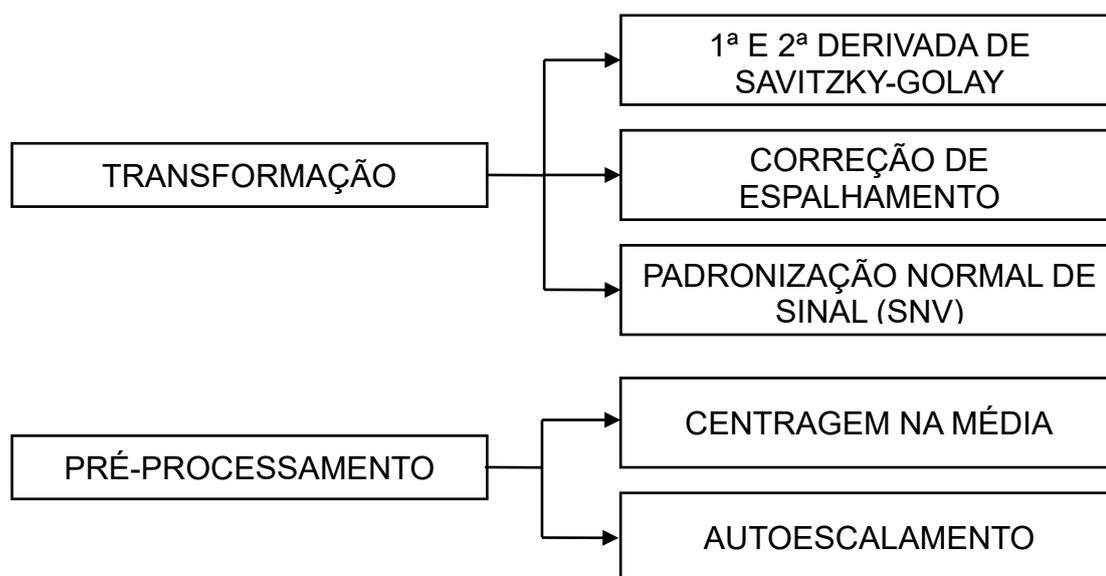
Na análise multivariada dos dados, a resposta instrumental é representada na forma de uma matriz \mathbf{X} , enquanto que a propriedade de interesse, medida por um método de referência, é representada por um vetor \mathbf{Y} . Na matriz \mathbf{X} , cada linha corresponde a uma amostra e cada coluna se refere a uma variável, ou seja, uma medida feita para todas as amostras (FERREIRA et al., 1999)

Antes da análise quimiométrica os dados devem ser pré-tratados para reduzir variações indesejáveis que não foram removidas durante a aquisição dos dados e que

podem influenciar nos resultados finais. Geralmente, são testados vários métodos para encontrar o pré-tratamento mais adequado à cada conjunto de dados (FERREIRA, 2015).

Representados na Figura 1, os métodos de pré-tratamento se dividem em dois tipos: a transformação dos dados, que é aplicada às amostras (linhas da matriz de dados) e o pré-processamento, aplicado às variáveis (colunas da matriz de dados). São várias as técnicas que podem ser utilizadas para reduzir as variações indesejáveis, no entanto, é preciso ter cuidado para que essas técnicas não acabem mascarando informações importantes do espectro (FERREIRA, 2015).

Figura 1 – Esquema representando os tipos de pré-tratamento de dados e os respectivos métodos mais utilizados



Fonte: Autoria própria.

O ruído aleatório, fonte de variações indesejadas, é constituído por uma mistura de erros experimentais de medidas, de amostragem e de outras fontes de variabilidade. Infelizmente, o ruído não pode ser eliminado completamente, mas pode ser reduzido quando adicionam medidas repetidas. Outra fonte de variações do sinal verdadeiro é o deslocamento constante da linha de base de um espectro, causado pelo instrumento ou pela própria amostra, que ocorre com frequência em espectros na região do infravermelho médio ou próximo (FERREIRA, 2015).

2.3.2.1 Transformação dos dados

Quando os esforços por adquirir espectros com a menor contribuição possível de variações indesejáveis são insuficientes, é possível recorrer à técnicas matemáticas para reduzir tanto variações aleatórias quanto sistemáticas. As técnicas de alisamento podem ser usadas para reduzir o ruído experimental, enquanto que as correções de linha de base e outras transformações podem ser usadas para eliminar ou reduzir as variações sistemáticas (FERREIRA, 2015).

Dentre os métodos de transformação mais usados destacam-se as técnicas de correção de linha de base como a 1ª e 2ª derivada pelo método de Savitzky-Golay, a correção multiplicativa de espalhamento (MSC, do inglês *Multiplicative Scatter Correction*), e a padronização normal de sinal (SNV, do inglês *Standard Normal Variate*).

2.3.2.1.1 Método da 1ª e 2ª derivada

A 1ª e a 2ª derivada do espectro podem ser usadas para remover os deslocamentos da linha de base e os efeitos de dispersão, sem perder os detalhes espectrais necessários para correlacionar as absorbâncias com as concentrações (PEREIRA et al., 2008).

O método de Savitzky-Golay é o mais utilizado para calcular as derivadas de um sinal. Segundo Ferreira (2015, p. 64), "o método de Savitzky-Golay é um filtro de média móvel em que se faz o ajuste, por quadrados mínimos, de um polinômio de grau n aos $(2n + 1)$ pontos da janela móvel".

2.3.2.1.2 Método da correção do espalhamento multiplicativo (MSC)

A transformação MSC corrige simultaneamente efeitos de espalhamento aditivos e multiplicativos. Esse espalhamento pode ser causado por diversos motivos mas, principalmente, pela falta de homogeneidade da amostra (SABIN; FERRÃO; FURTADO, 2004).

O método se baseia em usar a média dos espectros, deslocando e escalando cada espectro em torno dela, através de uma regressão linear. Os parâmetros de deslocamento corrigem os efeitos aditivos, enquanto os parâmetros de escalonamento corrigem os efeitos multiplicativos (FERREIRA, 2015).

2.3.2.1.3 Padronização normal de sinal (SNV)

A transformação SNV também corrige efeitos aditivos e multiplicativos e é apropriada para remover interferências de espalhamento e de tamanho de partícula. O método é muito parecido com a transformação MSC, diferindo apenas pelo de que, no SNV cada linha da matriz dos dados é autoescalada, ou seja, divide-se pelo desvio padrão (FERREIRA, 2015).

2.3.2.2 *Pré-processamento*

Dentre as técnicas de pré-processamento de dados mais utilizadas em análises multivariadas estão a centragem dos dados na média e o autoescalamento.

2.3.2.2.1 Centragem dos dados na média

Consiste em uma translação de eixos para o valor médio de cada um deles e, por isso, a estrutura dos dados é preservada. (MORGANO; QUEIROZ; FERREIRA, 1999). É o pré-processamento mais usual ao trabalhar com dados espectrais porque não altera a magnitude dos dados.

2.3.2.2.2 Autoescalamento

Esse método consiste em centrar os dados na média e dividi-los pelo desvio padrão, fazendo com que todas as variáveis passem a ter a mesma variância (SABIN; FERRÃO; FURTADO, 2004). O autoescalamento é recomendado quando as variáveis tem diferentes unidades ou quando os dados apresentam uma grande faixa de variação, porque iguala o impacto de cada uma delas.

Entretanto, em se tratando de dados espectrais, o efeito é que todos os comprimentos de onda terão o mesmo peso, tornando-se difícil diferenciar se ele representa um pico, um espalhamento ou um ruído de linha de base, e por isso não é recomendado (FERREIRA, 2015).

2.3.3 Análise de componentes principais (PCA)

A análise por componentes principais (PCA, do inglês *Principal Component Analysis*) é um método de análise exploratória de dados que foi introduzida em 1901 por Karl Pearson. Tem como objetivo principal reduzir a dimensionalidade do conjunto original de dados, preservando e ampliando as informações relevantes, tornando mais simples sua análise visual (FERREIRA, 2015).

Isso é possível através do estabelecimento de novas variáveis ortogonais entre si, denominadas componentes principais (PCs). As PCs são combinações lineares das variáveis originais e são organizadas em ordem decrescente de importância (CORREIA; FERREIRA, 2007). Nesse sentido, a primeira componente principal (PC1) é a combinação linear de máxima variância dos dados originais, a segunda (PC2) é ortogonal à PC1 e apresenta a maior variância neste subespaço (SCHNEIDER et al., 2018).

A aplicação da PCA torna possível a visualização gráfica de todo o conjunto de dados, aumentando sua compreensão e examinando a presença ou a ausência de agrupamentos entre as amostras (CORREIA & FERREIRA, 2007).

O método é baseado na decomposição da matriz \mathbf{X} de dados em uma matriz de escores (\mathbf{T}) e uma matriz pesos (\mathbf{P}), como descrito na Equação 2 (SCHNEIDER et al., 2018).

$$\mathbf{X} = \mathbf{P}^T \mathbf{T} + \mathbf{E} \quad (2)$$

Onde \mathbf{P}^T é a matriz transposta de pesos, \mathbf{T} é a matriz de escores e \mathbf{E} , a de resíduos.

Além disso, por ser baseada na decomposição do conjunto de dados em componentes principais, serve de base para outras técnicas de análise multivariada como a regressão por componentes principais (PCR, do inglês *Principal Component Regression*) e a regressão por quadrados mínimos parciais (PLS, do inglês *Partial Least Squares*).

2.3.4 Métodos de calibração multivariada

Os métodos de calibração multivariada são um conjunto de técnicas que permitem quantificar a correlação entre as propriedades de interesse e o sinal analítico, através da construção de modelos matemáticos empíricos e que, desse modo, possibilitam previsões estatisticamente confiáveis (FERREIRA, 2015).

Os métodos multivariados são os mais indicados porque permitem uma análise com várias espécies presentes ao mesmo tempo. Outra vantagem, é que é possível identificar eventuais problemas de interferentes ou linha de base nas amostras usadas no conjunto de calibração e nas novas amostras, do conjunto de previsão (FERREIRA, 1999).

São vários os métodos de regressão linear e cada um se adequa aos diferentes problemas que podem ser encontrados. Dentre os mais populares estão a regressão linear múltipla (MLR, do inglês *Multiple Linear Regression*), a regressão por componentes principais (PCR) e a regressão por quadrados mínimos parciais (PLS). (FERREIRA, 2015).

Entretanto, o modelo PLS é um dos mais utilizados devido a sua capacidade de minimizar problemas como a perda de resolução do sinal analítico (BORIN; POPPI, 2005). Este método não requer um conhecimento prévio de todos os componentes presentes na amostra, permitindo que a previsão de um determinado componente seja realizada mesmo na presença de interferentes, desde que estes estejam presentes também na etapa de construção do modelo.

Esse método de regressão é baseado na decomposição simultânea de uma matriz de dados \mathbf{X} (variáveis independentes) e de uma matriz de dados \mathbf{Y} (variáveis dependentes) em uma soma de produtos dos vetores de escores e de pesos. À esta decomposição soma-se mais uma matriz de erros que corresponde à parte não modelada (BRERETON, 2000).

A otimização simultânea causa pequenas distorções nas direções dos pesos, fazendo com que eles percam sua ortogonalidade e, conseqüentemente, gerando pequenas redundâncias que otimizam as relações lineares entre os escores de \mathbf{X} e de \mathbf{Y} . Como aqui os componentes principais não são mais ortogonais entre si, eles passam a se chamar variáveis latentes (VLs) (OTTO, 1999).

Ambas as matrizes, \mathbf{X} e \mathbf{Y} , são decompostas em suas respectivas matrizes de escores e pesos, como descrito nas Equações 3 e 4 (OTTO, 1999).

$$X = T P^T + E_X = \sum t_A p_A^T + E_X \quad (3)$$

$$Y = U Q^T + E_Y = \sum u_A q_A^T + E_Y \quad (4)$$

Onde \mathbf{X} é a matriz das respostas instrumentais, \mathbf{Y} é a matriz de respostas da propriedade de interesse obtida por metodologia padrão, \mathbf{T} e \mathbf{U} são os escores de \mathbf{X} e \mathbf{Y} , respectivamente, \mathbf{P} e \mathbf{Q} são os pesos de \mathbf{X} e \mathbf{Y} , respectivamente, \mathbf{E}_x e \mathbf{E}_y correspondem à matriz de resíduos e \mathbf{A} corresponde ao número de fatores (VLs).

A partir disso uma relação linear é obtida entre os escores de \mathbf{X} e os escores de \mathbf{Y} , como demonstrado na Equação 5 (OTTO, 1999).

$$\hat{u}_A = b_A \hat{t}_A \quad (5)$$

Ond \mathbf{b}_A é o vetor de coeficientes de regressão obtido na etapa de calibração.

2.3.5 Métodos de seleção de variáveis

Em se tratando de dados espectrais, o PLS, geralmente, realiza a calibração usando as informações de todo o espectro ao construir o modelo de regressão para determinar a propriedade de interesse (BORIN; POPPI, 2005). No entanto, muitas vezes são apenas algumas das regiões do espectro que influenciam diretamente na predição de determinada propriedade.

Nesse sentido, a utilização de ferramentas de seleção de variáveis pode melhorar significativamente a eficiência desta técnica de calibração, permitindo que sejam utilizados somente as variáveis relacionadas à propriedade de interesse, aumentando a robustez e a capacidade de predição dos modelos (OLIVEIRA et al, 2004).

Existem diversos modelos de seleção de variáveis, que incluem métodos de seleção manual, o emprego de redes neurais artificiais e de algoritmos genéticos. Dentre eles, destacam-se os algoritmos baseados na seleção de intervalos, como o método de mínimos quadrados parciais por intervalo (iPLS), método de mínimos quadrados parciais por exclusão (biPLS) e método de mínimos quadrados parciais por

sinergismo (siPLS) que têm apresentado resultados satisfatórios para dados espectrais (XIAOBO et al, 2010).

2.3.5.1 Regressão por mínimos quadrados parciais por intervalo (iPLS)

O iPLS é um método graficamente orientado, desenvolvido por Nørgaard et al. (2000), que pretende buscar e selecionar a região do espectro que proporcione o melhor modelo PLS. Essa técnica pode fornecer uma imagem geral de informações relevantes, concentrando-se nas regiões mais importantes e removendo a interferência de outras regiões (PEREIRA et al., 2008).

O modelo PLS por intervalo (iPLS) é baseado na divisão dos espectros em subintervalos equidistantes e na construção de modelos PLS de calibração para cada um deles (MÜLLER et al., 2013). São calculados os valores de raiz quadrada dos erros médios de validação cruzada (RMSECV) para cada modelo e em seguida comparados com o valor obtido para o modelo global, utilizando todo o espectro. São escolhidas as regiões que apresentam o menor valor de RMSECV (PARISOTTO et al., 2010).

Segundo Xiabo et al. (2010), a principal vantagem do método iPLS é a possibilidade de representar um modelo de regressão local em um gráfico, fornecendo uma visão geral dos dados espectrais e permitindo relacionar os melhores intervalos com as informações químicas presentes em cada um deles. Além disso, permite comparar os valores de RMSECV de cada intervalo com o valor de RMSECV do modelo global, tornando mais simples a interpretação dos dados.

A técnica iPLS foi usada por Borin e Poppi (2005) na quantificação de contaminantes em óleos lubrificantes para selecionar as regiões dos espectros obtidos por infravermelho médio, onde obtiveram bons resultados com baixos erros de predição. Além disso, foi aplicada para determinar parâmetros de qualidade de misturas de diesel/biodiesel usando infravermelho médio (FERRÃO et al., 2011) e determinação de enxofre usando ATR-FTIR (MÜLLER et al., 2012).

2.3.5.2 Regressão por mínimos quadrados parciais por exclusão (biPLS)

O método biPLS é uma ferramenta otimizada do iPLS e busca remover as regiões espectrais que não têm relevância para a propriedade a ser determinada. Ele se baseia em dividir o espectro em intervalos e o modelo PLS é calculado com cada intervalo deixado de fora. O intervalo deixado de fora é o que, quando está no modelo,

fornece o pior desempenho de RMSECV. O procedimento é repetido até que somente um intervalo permaneça. A combinação ideal é obtida com base no menor valor de RMSECV (WU et al., 2010).

Esse método foi aplicado para a determinação de nitrogênio básico em amostras de resíduo de petróleo usando ATR-FTIR (MÜLLER et al., 2013). A partir da seleção dos intervalos mais relevantes foram encontrados valores de RMSEP 10% menores em relação ao modelo de referência.

2.3.5.3 Regressão por mínimos quadrados parciais por sinergismo (siPLS)

Derivado do iPLS, o algoritmo siPLS calcula todos os modelos PLS das possíveis combinações de dois, três ou quatro intervalos. Depois, são calculados os valores de RMSECV de cada combinação de intervalos e a combinação com o menor valor é escolhida. Geralmente, os modelos siPLS fornecem melhores coeficientes de correlação e menores erros de predição em relação ao iPLS. Contudo, o tempo computacional depende do número de intervalos e de combinações, podendo ser bastante demorado (WU et al., 2010).

Müller et al. (2012) propuseram um método para determinar enxofre em amostras residuais de petróleo usando ATR-FT-IR onde o método de seleção de variáveis siPLS foi aplicado, apresentando melhores resultados em relação ao modelo obtido a partir de todo espectro.

2.3.6 Avaliação dos modelos de regressão

A escolha e a seleção dos modelos mais adequados podem ser feitas através da avaliação de alguns parâmetros resultantes da construção dos modelos de calibração multivariada.

A linearidade é um dos principais parâmetros a serem analisados ao construir um modelo de calibração. Ela corresponde à capacidade do método em fornecer resultados diretamente proporcionais à concentração da substância de interesse, dentro de um determinado intervalo de concentrações onde o método será aplicado. A linearidade geralmente pode ser avaliada pela observação do coeficiente de correlação, **R** (VALDERRAMA, 2009).

A dimensionalidade do modelo é outro fator relevante, uma vez que, para se obter um modelo relativamente simples, deve-se explicar o máximo de variância dos dados com o mínimo de variáveis latentes (VLs). Entende-se pelo número ótimo de VLs o número para o qual um erro mínimo de previsão seja obtido (GOMEZ-CARRACEDO, 2007). Ao construir um modelo cujo número de VLs seja inferior ao ideal resultará em um modelo subajustado, enquanto que um número superior ao ideal resultará em um modelo sobreajustado (GOMEZ-CARRACEDO, 2007).

Outro parâmetro bastante utilizado para comparar modelos de calibração é o cálculo dos erros médios quadráticos (RMSE), descrito na Equação 6. Esse cálculo varia de acordo com a aplicação e pode ser: raiz quadrada do erro médio quadrático de calibração (RMSEC, do inglês *root-mean-square error of calibration*), raiz quadrada do erro médio quadrático de validação cruzada (RMSECV, do inglês *root-mean-square error of cross validation*) e raiz quadrada do erro médio quadrático de validação (RMSEV, do inglês *root-mean-square error of validation*). O RMSECV é usado para determinar o número ótimo de VLs do modelo (BRERETON, 2000).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (6)$$

Onde y_i é o valor de referência para a i -ésima amostra, \hat{y}_i é o valor previsto para esta amostra e n é o número total de amostras utilizadas para a construção do modelo.

O RMSECV é baseado no método de validação cruzada (CV, do inglês *cross validation*), método que avalia a magnitude do erro de previsão de um dado modelo de calibração. A validação cruzada consiste em remover uma ou mais amostras do conjunto de calibração, construir um modelo de calibração sem elas e usar o modelo criado para prever o valor das amostras retiradas. O procedimento é repetido, removendo uma amostra do conjunto por vez, até que todas sejam retiradas e previstas. A partir do erro de previsão calculado para cada amostra é obtido o valor da raiz quadrada do erro médio quadrático de validação cruzada (RMSECV) (BRERETON, 2000).

Além disso, pode-se avaliar a capacidade de previsão do modelo de calibração através do cálculo da raiz quadrada do erro médio quadrático de previsão (RMSEP,

do inglês *root-mean-square error of prediction*), descrito na Equação 7 (BLANCO, 1999).

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i)^2}} \times 100 \quad (7)$$

Onde é y_i o valor de referência da i -ésima amostra e \hat{y}_i é o valor previsto para esta mesma amostra, sendo n o número total de amostras, utilizadas na construção do modelo.

2.3.7 Testes de significância

Ao propor uma nova metodologia é necessário verificar seu desempenho em relação ao método de referência. Nesse sentido, existem alguns testes estatísticos que servem para determinar se o método apresenta melhora significativa ou não.

Dentre eles destaca-se o teste F de significância que pode ser usada para comparar duas variâncias e determinar se dois métodos diferem em termos de precisão. De uma forma análoga, o teste F pode ser aplicado em modelos multivariados para calcular os erros de previsão de modelos distintos, conforme a Equação 8 (MEIER; ZÜND, 2000).

$$F = \left(\frac{RMSEP_2}{RMSEP_1} \right)^2, \quad RMSEP_2 > RMSEP_1 \quad (8)$$

Onde $RMSEP_1$ e $RMSEP_2$ são os valores de raiz quadrada do erro médio quadrático de previsão dos modelos A e B. O número de grau de liberdade é dado pelo número de amostras do conjunto menos um.

Os valores de F calculados são comparados com os valores de F crítico (valores tabelados), com um nível de confiança de 95%. Quando o valor de F calculado for maior que o F crítico os valores de RMSEP são considerados estatisticamente diferentes (GEMPERLINE, 2006).

3 METODOLOGIA

3.1 AMOSTRAS

As amostras de petróleo bruto utilizadas foram doadas pelo Centro de Pesquisas e Desenvolvimento Leopoldo Américo Miguez de Mello (CENPES, Petrobrás) e foram armazenadas em frascos de vidro âmbar à temperatura ambiente.

3.2 MEDIÇÃO DAS PROPRIEDADES PELO MÉTODO DE REFERÊNCIA

A determinação do grau API foi realizada de acordo com o método ASTM D-7042, utilizando um viscosímetro Stabinger (modelo SVM 3000, Anton Paar Inc.)

A determinação de enxofre e de nitrogênio total foi realizada utilizando um analisador por combustão de enxofre e nitrogênio total (modelo 9000 NS, Antek Instruments Inc.) e de detecção de enxofre por fluorescência no ultravioleta e de nitrogênio por quimiluminescência, segundo as normas ASTM D-5453 para enxofre e ASTM D-4629 para nitrogênio.

3.3 ANÁLISE UTILIZANDO ESPECTROSCOPIA NO INFRAVERMELHO

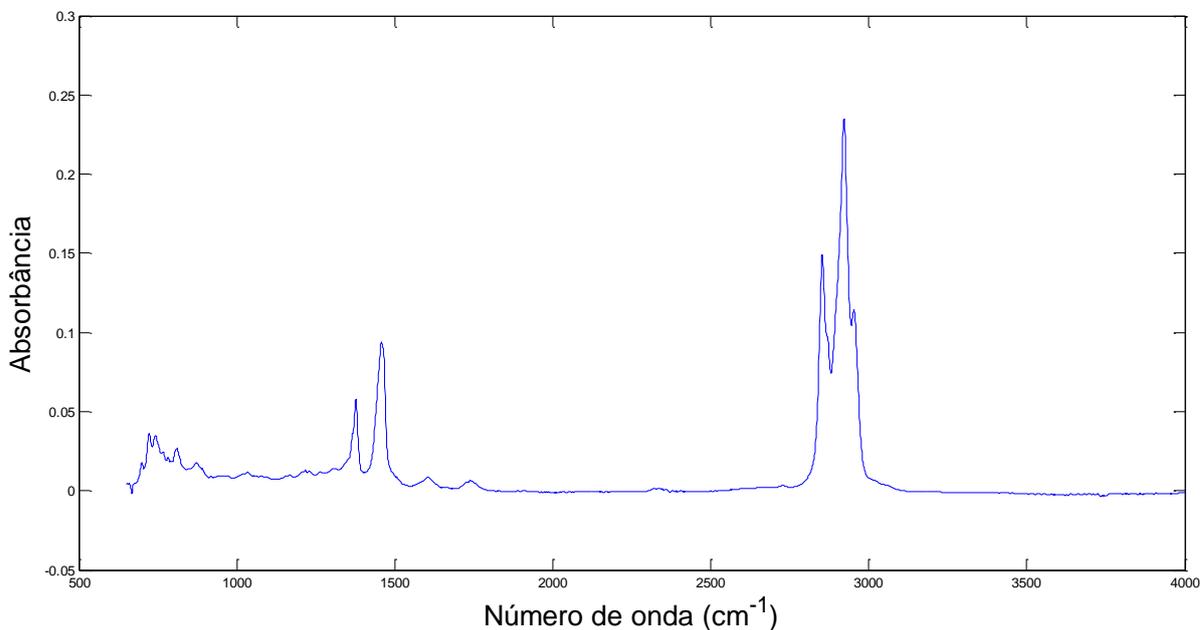
Os espectros das amostras de petróleo foram obtidos em um espectrômetro de infravermelho com transformada de *Fourier* (modelo Spectrum™ One FTIR, Perkin Elmer Inc.), provido de um acessório de reflectância total atenuada (ATR), detector de sulfato de triglicina deuterada (DTGS), na faixa de 650 a 4000 cm^{-1} , resolução de 4 cm^{-1} e 16 varreduras.

Primeiramente foi coletado o espectro de referência (*background*), sem a amostra na superfície do ATR, antes da aquisição do espectro de cada amostra. Em seguida, aproximadamente 50 μL de cada amostra de petróleo foram depositados na superfície do ATR.

Os espectros das amostras foram obtidos em triplicata e em seguida obteve-se o espectro médio de cada amostra, cujo perfil pode ser observado na Figura 2.

Depois da aquisição dos espectros de cada amostra, o compartimento do acessório foi cuidadosamente limpo com tolueno e seco com ar quente.

Figura 2 – Perfil dos espectros médios das amostras de petróleo obtidos por ATR-FTIR



Fonte: Autoria própria.

3.4 ANÁLISE MULTIVARIADA DOS DADOS

Foram utilizados três programas computacionais para o tratamentos dos dados: o The Unscrambler® versão 9.7 (Camo Analytics, <https://www.camo.com/unscrambler/>, NOR); o Origin Pro® versão 8.0 (Origin Lab Corporation, <https://www.originlab.com>, EUA) e o MATLAB® versão 7.0.1 (The Math Works, <http://www.mathworks.com>, EUA) provido do pacote iToolbox versão 2.0 (iToolbox for MATLAB, <http://www.model.kvl.dk>, EUA), para a seleção das variáveis (algoritmos iPLS, biPLS e siPLS). O tratamento dos dados foi feito através das etapas descritas na Figura 3.

Figura 3 – Principais etapas desenvolvidas para a obtenção do melhor modelo



Fonte: Autoria própria.

Antes da construção dos modelos foram selecionadas as amostras dos conjuntos de calibração e previsão para cada propriedade (grau API, teor de nitrogênio e de enxofre). As amostras do conjunto de calibração foram selecionadas de modo que incluísse as amostras com os maiores e menores valores de referência. Foram utilizadas 2/3 das amostras totais para o conjunto de calibração e 1/3 para o conjunto de previsão. A matriz de dados foi importada para o *software* The Unscrambler® onde foram identificadas e removidas as amostras anômalas de cada conjunto.

Todos os modelos foram obtidos pelo método de regressão PLS, utilizando a validação cruzada e com os dados centrados na média.

Ainda no The Unscrambler foram avaliados os métodos de pré-tratamento primeira derivada com filtro de Savitzky-Golay usando polinômio de segunda ordem e três pontos (SG) e correção de espalhamento multiplicativo (MSC), aplicando-os em cada conjunto de calibração e construindo um modelo PLS para cada um deles, afim de decidir qual deles seria utilizado em cada conjunto de acordo com os valores de coeficiente de correlação de calibração (R_{CAL}) e RMSECV.

Após a seleção do melhor método de pré-tratamento, os conjuntos de calibração e previsão foram transportados para o software Origin Pro® onde foram organizados em forma de matriz e exportados em formato ASCII (Código Americano para o Intercâmbio de Informação, do inglês *American Standard Code for Information Interchange*).

As matrizes de dados foram importadas para o *software* MATLAB® onde foram construídos os modelos globais, utilizando a regressão PLS e todas as variáveis do espectro. Por fim, foi realizada a etapa de seleção de variáveis usando os algoritmos iPLS, siPLS e biPLS disponíveis no pacote iToolbox.

Foram construídos os modelos de calibração dividindo o espectro em 10, 20, 40 e 60 intervalos e fazendo combinações até três intervalos no método siPLS. A partir das informações gráficas disponíveis no algoritmo foram selecionados os melhores intervalos, que apresentavam os menores valores de RMSECV e também o número ótimo de VLs a serem utilizadas nos modelos que seriam gerados.

Os modelos de iPLS, biPLS e siPLS gerados foram avaliados em termos de R_{CAL} e RMSECV. Os melhores modelos obtidos para cada um dos três métodos foram aplicados aos conjuntos de previsão para obter os valores de RMSEP. Dessa forma, os modelos PLS construídos a partir da seleção de variáveis foram comparados com os modelos globais em termos de número de variáveis, número de VLs, R_{CAL} , RMSECV e RMSEP.

4 RESULTADOS E DISCUSSÃO

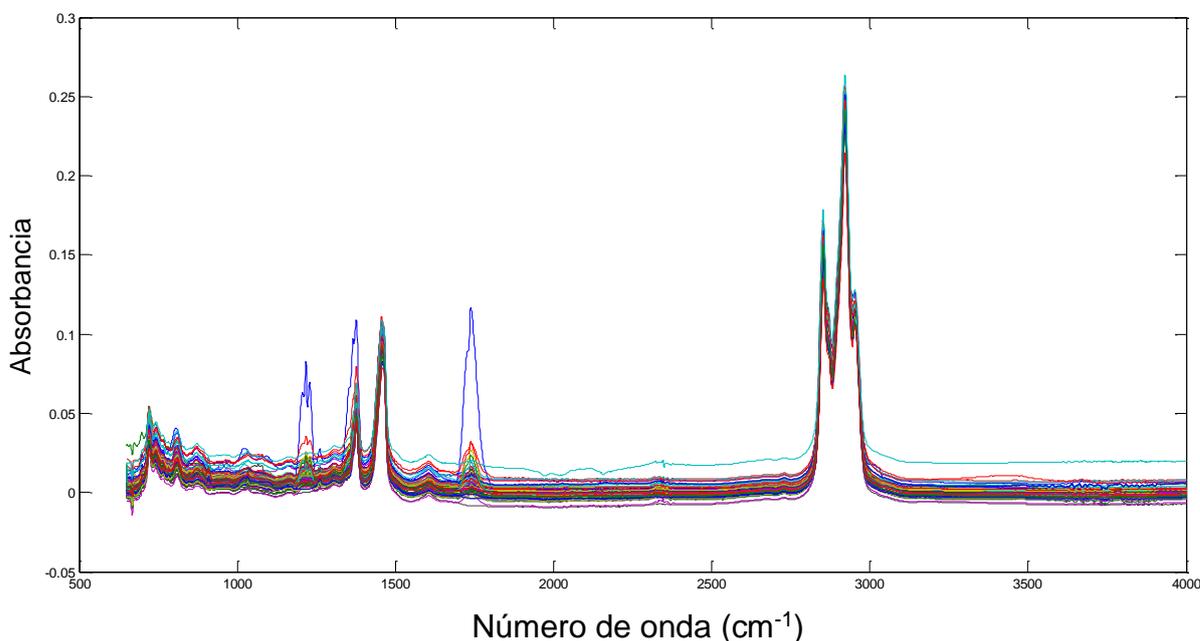
Foram obtidos os modelos de calibração para a predição de três propriedades do petróleo: grau API, teor de enxofre e de nitrogênio. Para isso, utilizou-se os espectros de infravermelho médio (MID) como conjunto de variáveis dependentes e os valores obtidos por métodos convencionais como variáveis independentes.

Na Figura 4 é possível visualizar o conjunto de espectros obtidos e suas bandas de absorção. Como o petróleo é uma mistura complexa, é de se esperar que a interpretação de seu espectro seja complicada. A região de 900 a 1300 cm^{-1} é uma região conhecida como região da “impressão digital” onde aparecem, geralmente, muitas bandas e modos de vibração acoplados, o que dificulta sua identificação (SILVERSTEIN; WEBSTER; KIEMLE, 2006).

O petróleo é constituído principalmente por hidrocarbonetos, em sua maioria alcanos e compostos aromáticos. A presença de alcanos pode ser confirmada no espectro pela existência de bandas nas regiões de 2800 a 3000 cm^{-1} , provenientes da deformação axial do C – H nos alcanos, e de bandas nas regiões de 1700 cm^{-1} e 1450 cm^{-1} , referentes às vibrações de deformação angular simétrica e assimétrica do C – H, respectivamente (SILVERSTEIN; WEBSTER; KIEMLE, 2006).

Já os hidrocarbonetos aromáticos podem ser confirmados pela ocorrência de bandas nas regiões de baixa frequência, entre 675 e 900 cm^{-1} , que resultam da deformação angular fora do plano das ligações C – H do anel, além da presença de bandas na região de 1000 a 1300 cm^{-1} , provenientes da deformação angular no plano (SILVERSTEIN; WEBSTER; KIEMLE, 2006).

Figura 4 – Conjunto de espectros médios obtidos por ATR-FTIR das amostras de petróleo na faixa de 650 - 4000 cm^{-1}



Fonte: Autoria própria.

Foram obtidos os modelos de calibração e previsão para as três propriedades (grau API, enxofre e nitrogênio) usando o método de regressão por mínimos quadrados parciais (PLS) e os resultados seguem abaixo.

4.1 MODELO DE CALIBRAÇÃO PARA O GRAU API

Após a retirada das amostras que apresentaram comportamentos anômalos nos modelos PLS previamente construídos, os conjuntos de calibração e previsão foram compostos por 74 e 38 amostras, respectivamente. Sendo que os valores de grau API das amostras variaram de 19,15 a 41,29 ° no conjunto de calibração e de 19,88 a 41,05 ° no conjunto de validação.

Depois da seleção das amostras foram avaliados dois tipos de pré-tratamento dos dados, SG e MSC, e os valores de R_{CAL} e R_{MECV} obtidos pela regressão PLS estão dispostos na Tabela 3. Observa-se que não há diferença significativa entre o valor de $RMSECV$ do modelo sem nenhum pré-tratamento e os valores de $RMSECV$ dos modelos aplicando os pré-tratamentos (teste F, com 95% de confiança).

Entretanto, optou-se pelo pré-tratamento usando MSC porque apresenta o menor número de variáveis latentes (VLs).

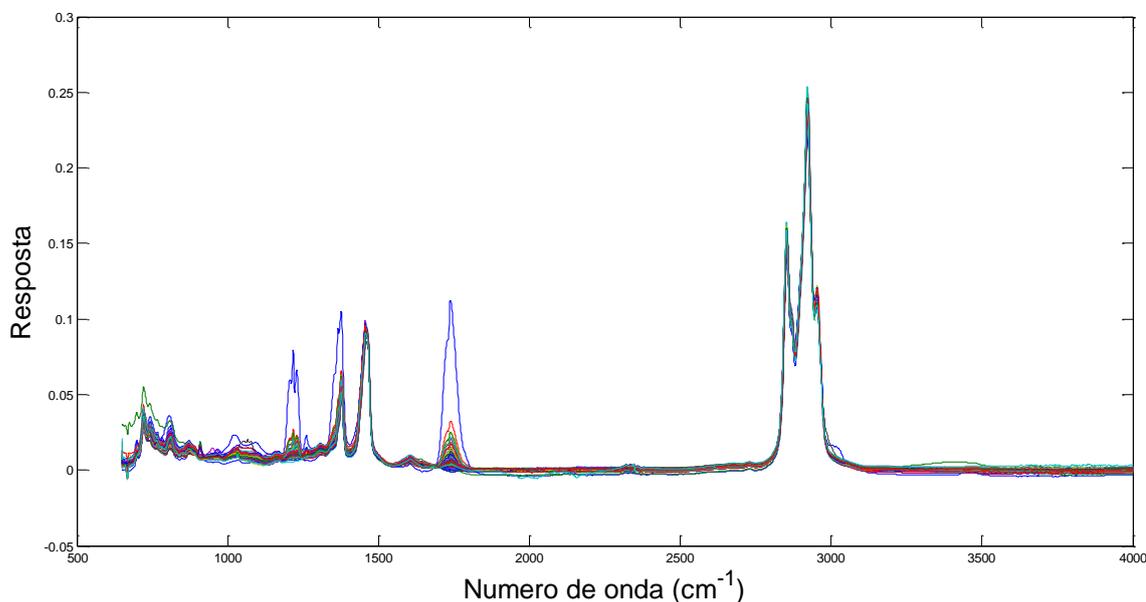
Tabela 3 – Resultados estatísticos dos modelos globais aplicando os pré-tratamentos SG e MSC e o método de regressão PLS para a propriedade grau API

| Pré-tratamento | VL | R _{CAL} | RMSECV (°) |
|----------------|----|------------------|------------|
| Nenhum | 5 | 0,9768 | 1,18 |
| MSC | 4 | 0,9788 | 1,10 |
| SG | 7 | 0,9888 | 1,25 |

Fonte: Autoria própria.

O método de correção de espalhamento multiplicativo (MSC) foi aplicado aos dados espectrais e o perfil espectral resultante pode ser visualizado na Figura 5.

Figura 5 – Conjunto de espectros obtidos no infravermelho médio para as amostras de petróleo na faixa de 650 - 4000 cm⁻¹ após pré-tratamento com MSC



Fonte: Autoria própria.

A partir disto foram construídos os modelos iPLS, biPLS e siPLS, dividindo o espectro em 10, 20, 40 e 60 intervalos e fazendo combinações de intervalos. Na Tabela 4 estão apresentados os resultados dos melhores modelos iPLS, biPLS e siPLS e os respectivos valores de R_{CAL}, RMSECV e RMSEP.

Nota-se que os valores de RMSECV e RMSEP dos modelos não apresentam diferença significativa em relação ao modelo global, de acordo com o teste F com 95% de confiança. Entretanto, levando em consideração que o modelo global utiliza 3351 variáveis e que o melhor modelo obtido (siPLS60), utiliza apenas 168 variáveis, tem-se que o segundo apresenta diversas vantagens principalmente em termos de velocidade computacional e maior robustez do modelo de calibração.

Tabela 4 – Resultados obtidos dos melhores modelos construídos para predição do grau API

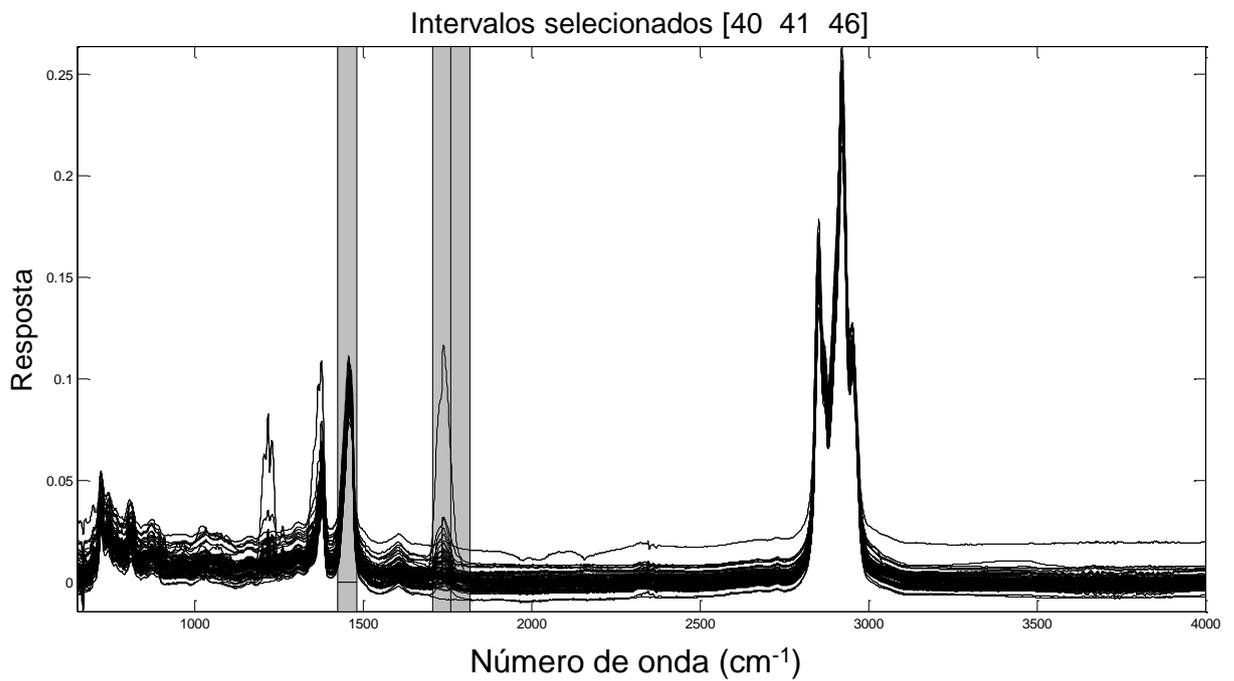
| Modelo | NV^a | Intervalos | VL | R_{CAL} | RMSECV (°) | RMSEP (°) |
|---------------|-----------------------|-------------------|-----------|------------------------|-------------------|------------------|
| global | 3351 | Todos | 4 | 0,9718 | 1,10 | 1,98 |
| iPLS10 | 335 | 8 | 4 | 0,9655 | 1,22 | 1,99 |
| biPLS60 | 223 | 10, 41, 46 e 47 | 4 | 0,9722 | 1,09 | 2,13 |
| siPLS60 | 168 | 40, 41 e 46 | 4 | 0,9763 | 1,01 | 1,65 |

Fonte: Autoria própria.

^aNV: número total de variáveis.

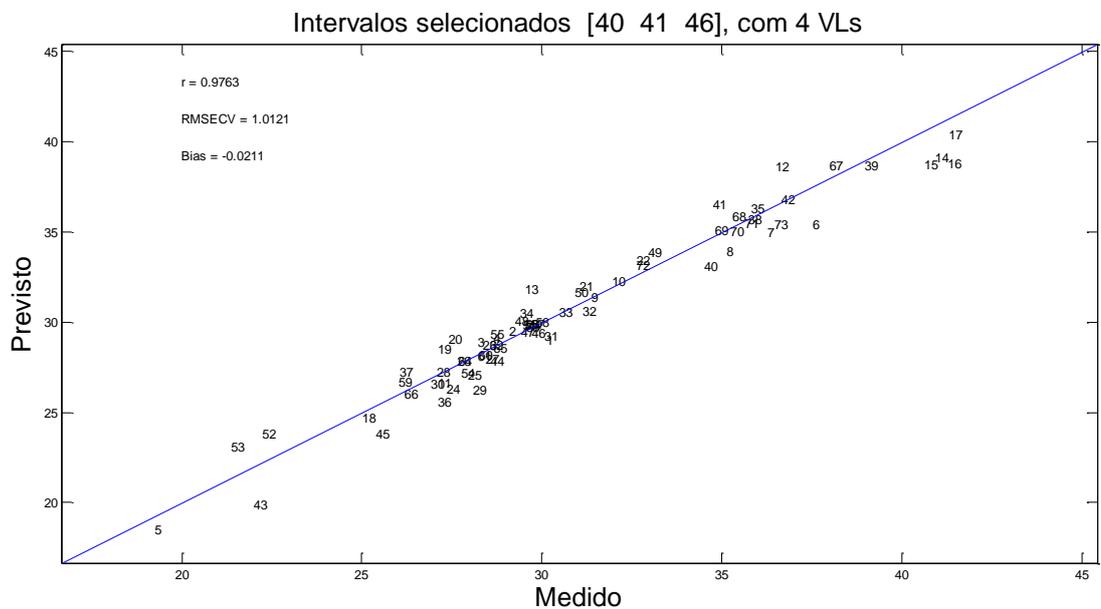
Na Figura 6 estão dispostas as regiões do espectro selecionadas no modelo siPLS60. As regiões abrangem os números de onda de 1425 a 1480 cm^{-1} (intervalo 41) e 1705 a 1816 cm^{-1} (intervalos 40 e 41). Segundo Silverstein et al. (1994), essas regiões estão associadas às vibrações de deformação angular assimétrica e simétrica do grupo metila, respectivamente. Na Figura 7 está representado o gráfico de correlação entre os valores medidos e previstos, com valores medidos variando de 19,15 a 41,29 °.

Figura 6 – Regiões espectrais selecionadas no modelo siPLS60



Fonte: Autoria própria.

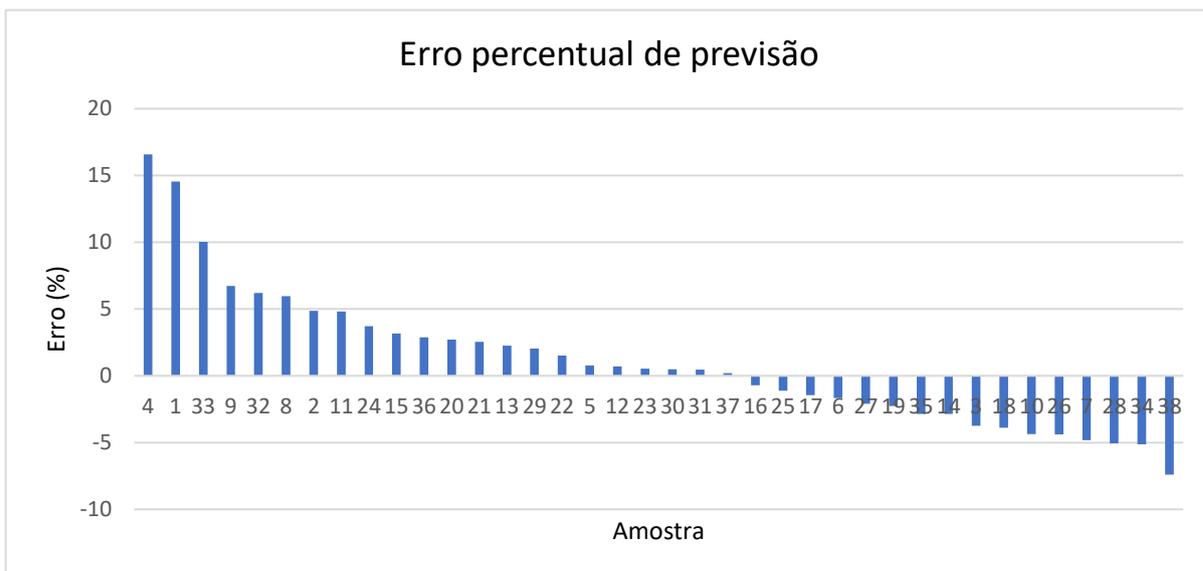
Figura 7 – Gráfico de correlação entre os valores medidos e previstos do grau API obtido a partir do modelo siPLS60



Fonte: Autoria própria.

Ao plotar os valores de percentual de erro de cada amostra do conjunto de previsão (Figura 8), pode-se observar que os valores variam para mais e para menos, numa faixa que vai de -8 à +17%.

Figura 8 – Gráfico de erro percentual de cada amostra do conjunto de previsão do grau API



Fonte: Autoria própria.

4.2 MODELO DE CALIBRAÇÃO PARA O TEOR DE ENXOFRE

Após a retirada das amostras com comportamento anômalo, os conjuntos de calibração e de previsão foram compostos pelos dados espectrais de 69 e 39 amostras, respectivamente. As concentrações de enxofre variaram de 660 a 8938 mg kg⁻¹ no conjunto de calibração e de 744 a 8030 mg kg⁻¹ para conjunto de previsão.

Foram comparados os modelos obtidos sem nenhum pré-tratamento, utilizando a correção de espalhamento multiplicativo (MSC) e a 1ª derivada com filtro de Savitzky-Golay (SG). Os resultados obtidos de R_{CAL} e RMSECV para cada um deles estão expostos na Tabela 5. De acordo com o teste F, os valores de RMSECV não apresentam diferença significativa. Todavia, optou-se por utilizar o pré-tratamento SG porque este apresentou os melhores valores de R_{CAL}.

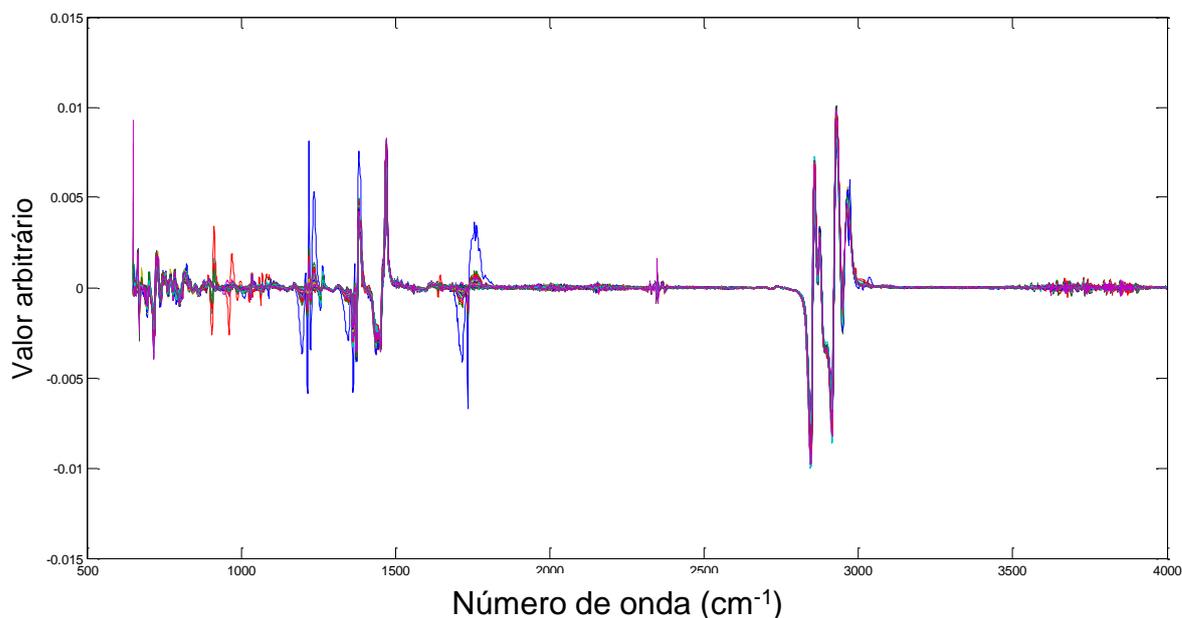
Na Figura 9 pode ser visualizado o perfil dos espectros após a aplicação do pré-tratamento.

Tabela 5 – Resultados estatísticos dos modelos globais aplicando os pré-tratamentos SG e MSC na regressão PLS para o teor de enxofre

| Pré-tratamento | VL | R _{CAL} | RMSECV (mg kg ⁻¹) |
|----------------|----|------------------|-------------------------------|
| Nenhum | 9 | 0,9724 | 546 |
| SG | 10 | 0,9945 | 416 |
| MSC | 8 | 0,9667 | 631 |

Fonte: Autoria própria.

Figura 9 – Conjunto de espectros obtidos no infravermelho médio para as amostras de petróleo na faixa de 651 - 3999 cm⁻¹ após pré-tratamento com SG



Fonte: Autoria própria.

Ao conjunto de calibração foram aplicadas as ferramentas de seleção de variáveis (iPLS, biPLS e siPLS), dividindo os espectros em 10, 20, 40 e 60 intervalos e fazendo combinações de intervalos. A partir disso foram obtidos os valores de R_{CAL} e RMSECV. Os melhores modelos obtidos de cada método e o modelo global (com todo o espectro) foram testados com o conjunto de previsão, obtendo assim os valores RMSEP.

Na Tabela 6 estão apresentados os resultados estatísticos dos melhores modelos de PLS obtidos a partir das diferentes ferramentas de seleção de variáveis. Os modelos empregando as ferramentas não apresentaram diferença significativa nos

valores de RMSECV e RMSEP em relação ao modelo global (teste F, com 95% de confiança).

Entretanto, o modelo iPLS, dividindo o espectro em 10 intervalos e utilizando apenas um deles apresentou resultados satisfatórios de R_{CAL} , RMSEC e RMSEP com um número de variáveis muito menor (335). Por isso, o modelo iPLS10, selecionando o intervalo número quatro e com um número de variáveis latentes igual a 10 foi escolhido para a predição do teor de enxofre.

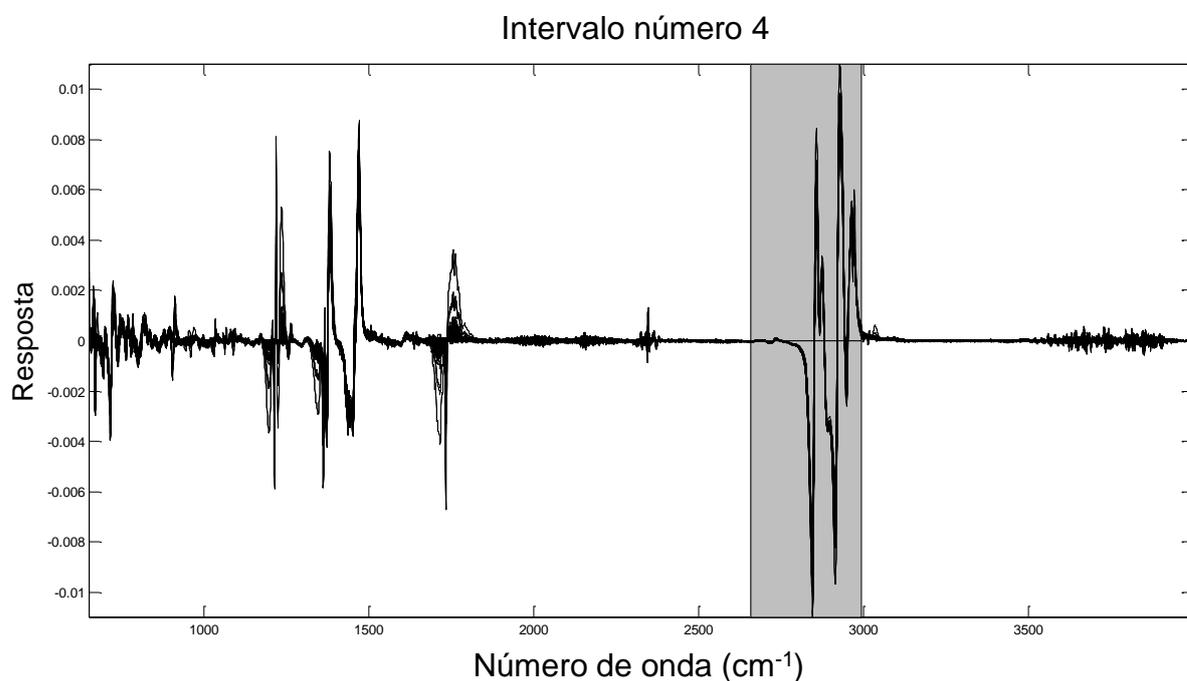
Tabela 6 – Resultados obtidos dos melhores modelos construídos para predição do teor de enxofre

| Modelo | NV^a | Intervalos | VL | R_{CAL} | RMSECV (mg kg⁻¹) | RMSEP (mg kg⁻¹) |
|---------------|-----------------------|---------------------|-----------|-----------------------------|--|---------------------------------------|
| Global | 3349 | Todos | 10 | 0,9602 | 416 | 703 |
| iPLS10 | 335 | 4 | 10 | 0,9432 | 484 | 517 |
| biPLS60 | 279 | 19, 20, 21, 46 e 59 | 10 | 0,9623 | 397 | 691 |
| siPLS20 | 503 | 7, 16 e 20 | 10 | 0,9641 | 392 | 670 |

Fonte: Autoria própria,
^aNV: número total de variáveis.

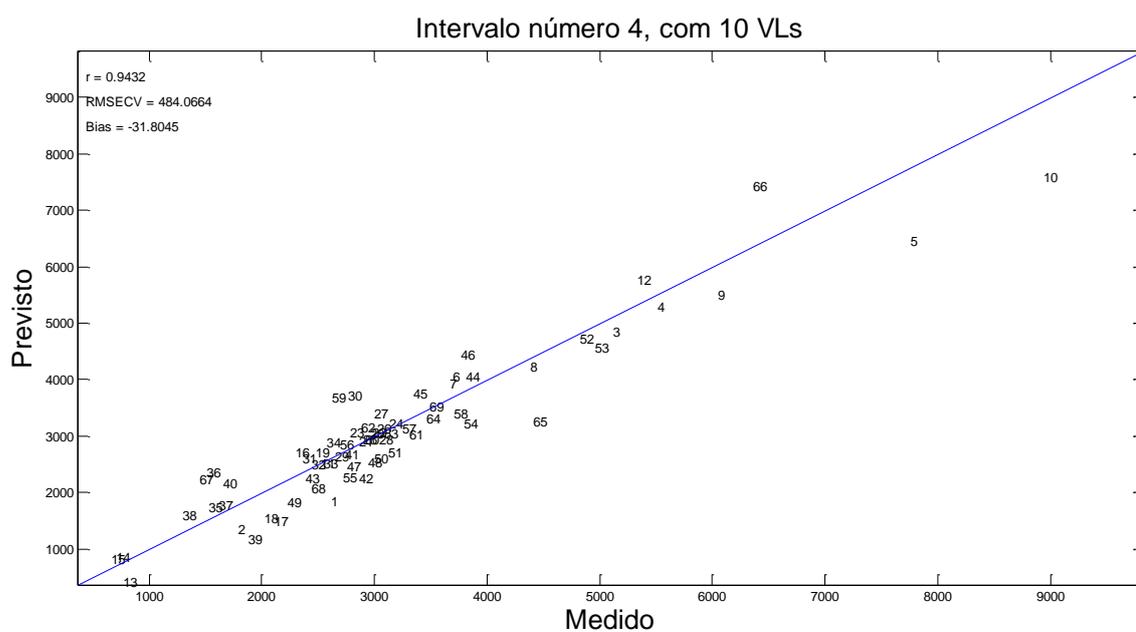
Na Figura 10 está destacado, em cinza, a região do espectro selecionada no modelo iPLS10. A escolha dessa região pelo algoritmo pode ser justificada pela presença de estiramentos axiais da ligação C – H nessa região, que vai de 2660 a 2994 cm^{-1} (SILVERSTEIN; WEBSTER; KIEMLE, 2006). Na Figura 11 pode ser visualizado o gráfico de correlação entre valores medidos e previstos, obtido pelo modelo iPLS10 e cuja concentração do enxofre, medida pelo método de referência, varia de 660 à 8938 mg kg^{-1} .

Figura 10 – Região espectral selecionada no modelo iPLS10



Fonte: Autoria própria.

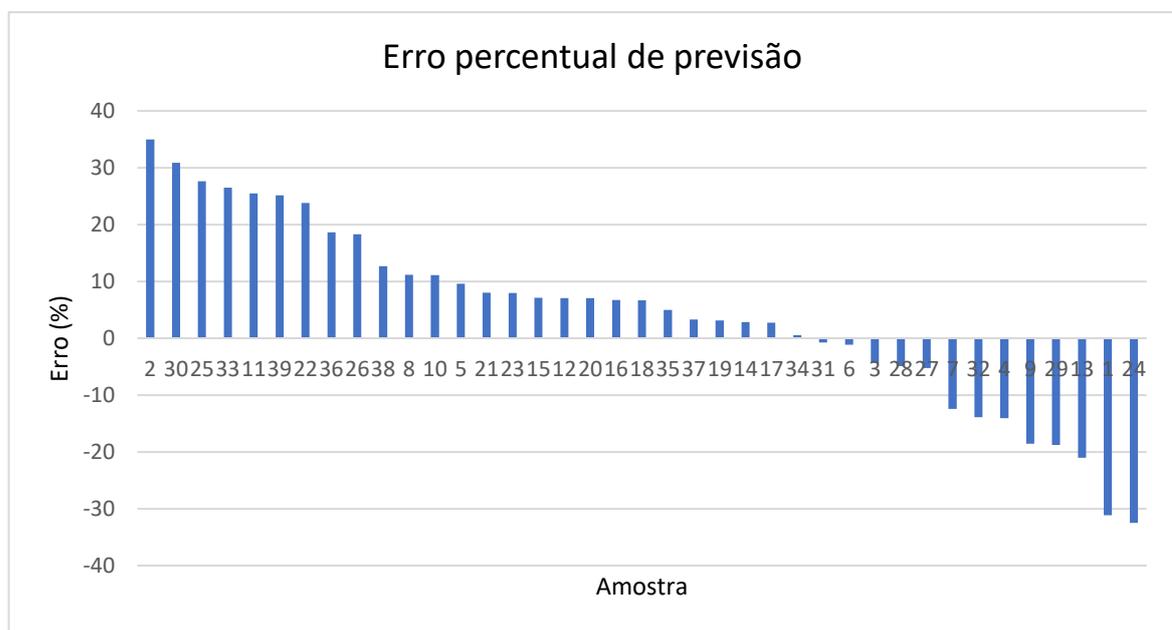
Figura 11 – Gráfico de correlação entre os valores medidos e previstos do teor de enxofre obtido a partir do modelo iPLS10



Fonte: Autoria própria.

Na Figura 12 está apresentado o gráfico de erro percentual por amostra do conjunto de previsão. Pode-se observar que há tanto erros para mais (eixo positivo), quanto para menos (eixo negativo). Os erros variam numa faixa que vai de -32 a +34%.

Figura 12 – Gráfico de erro percentual de cada amostra do conjunto de previsão do teor de enxofre



Fonte: Autoria própria.

4.3 MODELO DE CALIBRAÇÃO PARA O TEOR DE NITROGÊNIO

Ao realizar um estudo prévio do comportamento das amostras foram identificadas as amostras anômalas e estas foram retiradas do conjunto. Das 54 amostras selecionadas, 35 foram usadas para compor o conjunto de calibração e 19 foram usadas para o de previsão. As concentrações de nitrogênio variaram de 582 a 3603 mg kg⁻¹ no conjunto de calibração e de 666 a 3481 mg kg⁻¹ para conjunto de previsão.

Foram avaliados diferentes pré-tratamentos de dados aplicando-os ao conjunto de calibração e os resultados estatísticos estão dispostos na Tabela 7. Os valores de RMSECV dos modelos aplicando os métodos de pré-tratamento não apresentam

diferença significativa em relação aos valores de RMSECV do modelo sem nenhum pré-tratamento (teste F, com 95% de confiança). Entretanto optou-se por utilizar a 1ª derivada com filtro de Savitzky-Golay (SG) porque apresentou o melhor valor de correlação (R_{CAL}) entre os valores previstos pelo modelo e os valores de referência.

Tabela 7 – Resultados estatísticos dos modelos globais de aplicando os pré-tratamentos MSC e SG na regressão PLS para o teor de nitrogênio

| Pré-tratamento | VL | R_{CAL} | RMSECV ($mg\ kg^{-1}$) |
|----------------|----|-----------|--------------------------|
| Nenhum | 4 | 0,8907 | 507 |
| SG | 7 | 0,9857 | 419 |
| MSC | 3 | 0,8659 | 528 |

Fonte: Autoria própria.

Aplicando as ferramentas de seleção de variáveis (iPLS, biPLS e siPLS) ao conjunto de calibração, os espectros foram divididos em 10, 20, 40 e 60 intervalos, fazendo combinações de intervalos. Desse modo foram obtidos os valores de R_{CAL} e RMSECV. Em seguida, os melhores modelos obtidos de cada método foram testados usando as amostras do conjunto de previsão e obteve-se o valor de RMSEP.

Os resultados estão dispostos na Tabela 8 e avaliando os valores de RMSEP não se observa melhora dos modelos obtidos por seleção de variáveis em relação ao modelo global, utilizando todas as variáveis. Além disso, os modelos não apresentam diferença significativas em termos de valores de RMSECV em relação ao modelo global (teste F, com 95% de confiança). Todavia, o modelo siPLS60, combinando os intervalos 46, 54 e 58, apresentou uma maior correlação entre os valores previstos e os valores de referência (R_{CAL}), utilizando 167 variáveis (apenas 5% do número total).

Na Figura 13 estão destacados, em cinza, as regiões do espectro selecionadas no modelo siPLS60. As regiões abrangem os comprimentos de onda que vão de 761 à 815 cm^{-1} (intervalo 58), 981 à 1035 cm^{-1} (intervalo 54) e 1424 à 1479 cm^{-1} (intervalo 46). A região de 675 a 900 cm^{-1} , correspondente ao intervalo 58, ocorrem as deformações angulares da ligação C – H fora do plano de hidrocarbonetos aromáticos mononucleares e polinucleares. Já a região do intervalo 54 corresponde às vibrações de deformação angular das ligações C – C dos alcanos, que ocorrem entre 800 e 1200

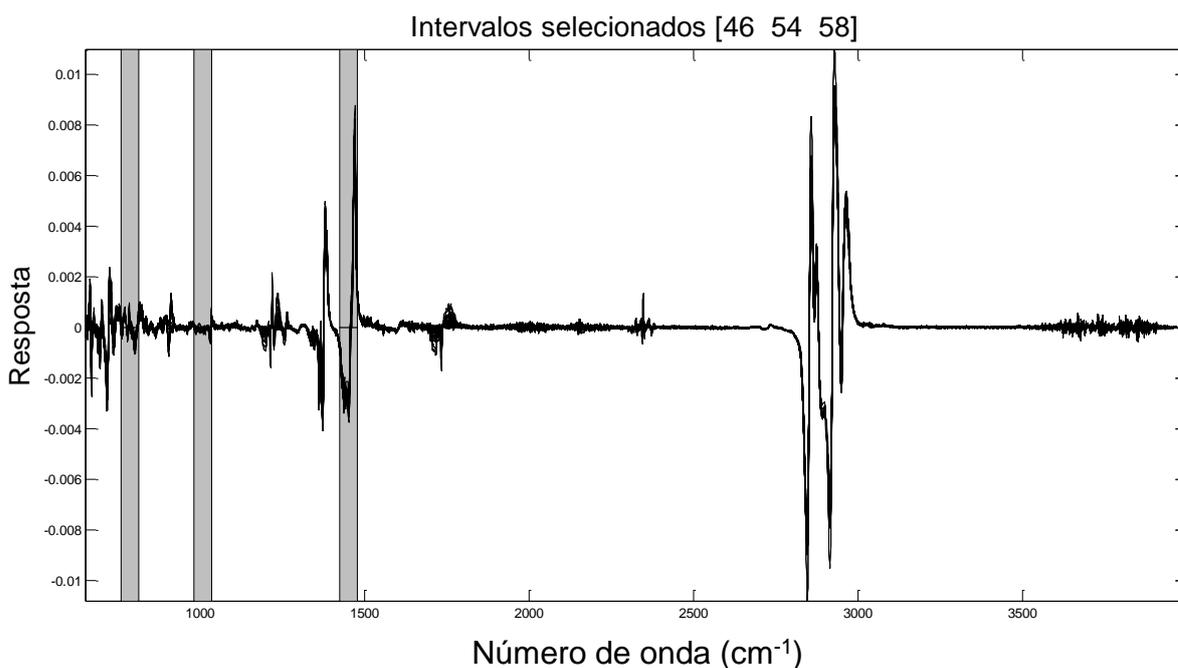
cm⁻¹. O intervalo 46 pode ser atribuído à banda de deformação angular simétrica do metileno (CH₂), que ocorre geralmente na posição de 1465 cm⁻¹ (SILVERSTEIN; WEBSTER; KIEMLE, 2006). A Figura 14 corresponde ao gráfico de correlação entre os valores medidos e previstos, obtido através do modelo siPLS60 e cuja concentração de nitrogênio, medida pelo método de referência, varia de 582 a 3603 mg kg⁻¹.

Tabela 8 – Resultados obtidos dos melhores modelos construídos para predição do teor de nitrogênio

| Modelo | NV ^a | Intervalos | VL | R | RMSECV (mg kg ⁻¹) | RMSEP (mg kg ⁻¹) |
|---------|-----------------|-----------------------------|----|--------|-------------------------------|------------------------------|
| Global | 3349 | Todos | 7 | 0,8877 | 419 | 321 |
| iPLS20 | 168 | 15 | 4 | 0,8839 | 421 | 464 |
| biPLS60 | 391 | 47, 19, 21, 60, 20, 56 e 41 | 7 | 0,8979 | 398 | 417 |
| siPLS60 | 167 | 46, 54 e 58 | 6 | 0,9284 | 337 | 388 |

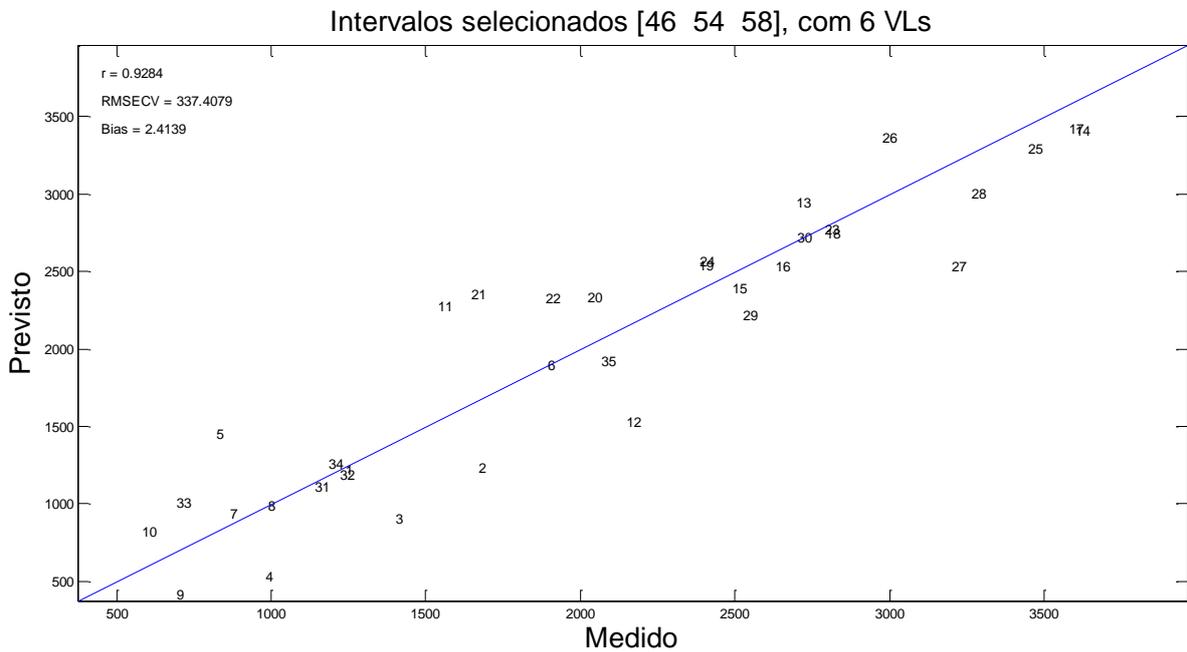
Fonte: Autoria própria.
^aNV: número total de variáveis.

Figura 13 – Regiões espectrais selecionadas no modelo siPLS60



Fonte: Autoria própria.

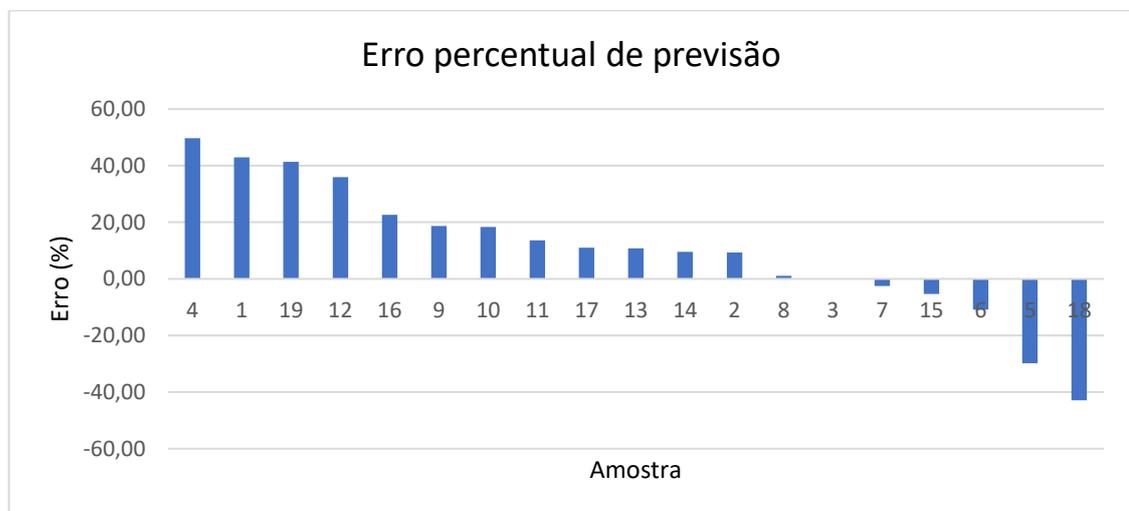
Figura 14 – Gráfico de correlação entre os valores medidos e previstos do teor de nitrogênio obtido a partir do modelo siPLS60



Fonte: Autoria própria.

Ao plotar os valores de percentual de erro de cada amostra do conjunto de previsão (Figura 15), pode-se observar que os valores variam para mais e para menos, numa faixa que vai de, aproximadamente, -43 à +50%.

Figura 15 – Gráfico de erro percentual de cada amostra do conjunto de previsão do teor de nitrogênio



Fonte: Autoria própria.

5 CONCLUSÃO

A utilização da espectroscopia no infravermelho com transformada de Fourier e reflexão total atenuada (ATR-FTIR) associada às técnicas de análise multivariada de dados demonstrou-se apropriada para a determinação do grau API e da concentração de enxofre e nitrogênio em amostras de petróleo bruto.

A avaliação das ferramentas de pré-tratamento permitiu decidir qual delas (SG ou MSC) era mais adequada para cada conjunto através dos resultados estatísticos obtidos pela regressão PLS, sendo adotada MSC para o conjunto de calibração do grau API e SG para os conjuntos de calibração das concentrações de enxofre e nitrogênio.

O modelo global para calibração do grau API, utilizando todas as 3351 variáveis independentes, forneceu um valor de RMSEP de 1,98 °, coeficiente de correlação de 0,9718 utilizando quatro variáveis latentes (VLs). Ao empregar o método de seleção de variáveis siPLS foi possível obter um modelo mais simples, com um valor de RMSEP igual a 1,65 °, coeficiente de correlação (R_{CAL}) igual a 0,9763, quatro variáveis latentes e utilizando somente 168 variáveis.

Para o modelo global de calibração da concentração de enxofre, utilizando 3349 variáveis, obteve-se um valor de RMSEP igual a 703 mg kg⁻¹ e coeficiente de correlação igual a 0,9602. Aplicando o método iPLS e dividindo o espectro em 10 intervalos, foram obtidos valores de RMSEP e R_{CAL} , iguais a 517 mg kg⁻¹ e 0,9432, respectivamente, utilizando apenas 335 variáveis independentes.

Já o modelo global para calibração da concentração de nitrogênio total, utilizando 3349 variáveis, apresentou valores de RMSEP e R_{CAL} , iguais a 321 mg kg⁻¹ e 0,8877, respectivamente, com número de VLs igual a sete. Enquanto isso, a aplicação do método siPLS, dividindo o espectro em 60 intervalos resultou em um valor de RMSEP de 388 mg kg⁻¹ e um valor de R_{CAL} igual a 0,9284, utilizando seis VLs e apenas 167 variáveis.

Dessa forma, pode-se concluir que os métodos iPLS, biPLS e siPLS não apresentaram melhorias significativas nos valores de R_{CAL} , RMSECV e RMSEP em relação aos modelos globais, entretanto resultaram na otimização dos modelos de calibração, tornando-os mais simples e conseqüentemente, possibilitando a predição dos valores em um menor tempo computacional.

Além disso, os resultados obtidos podem ser considerados razoáveis visto que os erros médios podem ser atribuídos às características das amostras e aos métodos de referência.

6 REFERÊNCIAS

BARBOSA, L. C. A. **Espectroscopia no infravermelho na caracterização de compostos orgânicos**. Viçosa: Editora UFV, 2007. 189 p.

BLANCO, M.; COELLO, J.; ITURRIAGA, H.; MASPOCH, S.; PAGÈS, J. **Calibration in nonlinear near infrared reflectance spectroscopy: a comparison of several methods**, *Anal. Chim. Acta*, v. 384, p. 207-214, 1999.

BORIN, A.; POPPI, R. **Application of Mid Infrared Spectroscopy and iPLS for the Quantification of Contaminants in Lubricating Oil**. *Vibrational Spectroscopy*, v. 37, p. 27–32, 2005.

BRASIL. Portaria n. 9, de 21 de janeiro de 2000. Regulamento Técnico de Reservas de Petróleo e Gás Natural, ANP nº 001/2000. **Diário Oficial da União**, Brasília, DF, 18 jan. 2000. Disponível em: <<http://legislacao.anp.gov.br/?path=legislacao-anp/portarias-anp/tecnicas/2000/janeiro&item=panp-9--2000>>. Acesso em 06 nov. 2019.

BRERETON, R. G. **Introduction to multivariate calibration in analytical chemistry**. *Analyst*, Bristol, v. 125, p. 2125-2154, 2000.

BRERETON, R. G. **Chemometrics: Data analysis for the laboratory and chemical plant**. Chichester: John Wiley & Sons, 2003. 497 p.

CORREIA, P. R. M.; FERREIRA, M. M. C. **Reconhecimento de padrões por métodos não supervisionados: explorando procedimentos quimiométricos para tratamento de dados analíticos**. *Quim. Nova*, v. 30, n. 2, p. 481-487, 2007.

FERRÃO, M. F.; VIERA, M. D. S.; PAZOS, R. E. P.; FACHINI, D.; GERBASE, A. E.; MARDER, L. **Simultaneous determination of quality parameters of biodiesel/diesel blends using HATR-FTIR spectra and PLS, iPLS or siPLS regressions**. *Fuel*, v. 90, n. 2, p. 701–706, 2011.

FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. Campinas: Editora Unicamp, 2015. 496 p.

FERREIRA, M. M. C., ANTUNES A. M., MELGO M. S., VOLPE L. O. **Quimiometria I: calibração multivariada, um tutorial**. *Quím. Nova*, São Paulo, v. 22, n. 5, p. 724-731, 1999.

FILGUEIRAS, P. R.; SAD, C. M. S.; LOUREIRO, A. R.; SANTOS, M. F. P.; CASTRO, E. V. R.; DIAS, J. C. M.; POPPI, R. J. **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.

GEMPERLINE, P. **Practical guide to chemometrics**. 2ª Ed. New York: CRC Press Taylor & Francis, p. 105-160, 2006.

GOMEZ-CARRACEDO, M. P.; ANDRADE, J. M.; RUTLEDGE, D. N.; FABER, N. M. **Selecting the optimum number of partial least squares components for the calibration of attenuated total reflectance-mid-infrared spectra of undersigned kerosene samples.** Anal. Chim. Acta v. 585, p. 253-265. 2007.

LI, J.; CHU, X.; TIAN, S.; LU, W. **Research on determination of nitrogen content in petroleum using mid-infrared spectroscopy.** China Petroleum Processing and Petrochemical Technology, v. 4, p. 1–7, 2011.

MEIER, P. C.; ZÜND, R. E. **Statistical Methods in Analytical Chemistry**, 2^a Ed. Canada: John Wiley & Sons, 2000. 407 p.

MORGANO, M.; QUEIROZ, S.; FERREIRA, M. **Application of exploratory analysis to the differentiation of vegetables.** Braz. J. Food Technol. v. 2, p. 73-79, 1999.

MÜLLER, A. L. H.; PICOLOTO, R. S.; MELLO, P. D. A.; FERRÃO, M. F.; DOS SANTOS M. D. F. P.; GUIMARÃES, R. C. L.; FLOR ES, E. M. M. **Total sulfur determination in residues of crude oil distillation using FT-IR/ATR and variable selection methods.** Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, v. 89, p. 82–87, 2012.

MÜLLER, A. L. H.; PICOLOTO, R. S.; GUIMARAES, R. C. L.; GUARNIERI, R. A.; FERREIRA, B. M. S.; DIAS, J. C. M.; MULLER, E. I. **Determination of Basic Nitrogen in Residues of Crude Oil Distillation Using ATR-FTIR and Chemometric Methods.** Analytical Letters, v. 46, n. 18, p. 2879–2889, 2013.

NORGAARD, L.; SAUDLAND A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L. ENGELSEN, S. B. **Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy.** Applied Spectroscopy, v. 54, n. 3, p. 413–419, 2000.

OLIVEIRA, F. C. C. et al. **A escolha da faixa espectral no uso combinado de métodos espectroscópicos e quimiométricos.** Quím. Nova, São Paulo, v.27, n. 2, p. 218-225, 2004.

OTTO, M. **Chemometrics: Statistics and Computer application in Analytical Chemistry.** Weinheim: John Wiley & Sons, 1999.

PARISOTTO, G.; FERRÃO, M. F.; MÜLLER, A. L. H.; MÜLLER, E. I.; SANTOS, M. F. P.; GUIMARÃES, R. C. L.; FLORES, E. M. M. **Total acid number determination in residues of crude oil distillation using ATR-FTIR and variable selection by chemometric methods.** Energy and Fuels, v. 24, p. 5474–5478, 2010.

PEREIRA, C. F.; PIMENTEL, M. F.; GALVÃO, R. K. H.; HONORATO, F. A.; STRAGEVITCH, L.; MARTINS, M. N. **A comparative study of calibration transfer methods for determination of gasoline quality parameters in three different near infrared spectrometers.** Analytica Chimica Acta, v. 611, n.1, p. 41–47, 2008.

PAVIA, D. L.; LAMPMAN, G. M.; KRIZ, G. S. **Introduction to spectroscopy.** Fort Worth: Harcourt Brace College, 1996.

SABIN, J. G.; FERRÃO, M. F.; FURTADO, J. C. **Análise multivariada aplicada na identificação de fármacos antidepressivos.** Parte II: Análise por componentes principais (PCA) e o método de classificação SIMCA. *Revista Brasileira de Ciências Farmacêuticas*, v. 40(3), p. 387–396, 2004.

SCHNEIDER, M.; VACARO, B. B.; MARCELO, M. C. A.; FERRÃO, M. F. **Métodos de Análise Exploratória Aplicadas ao Ensino de Espectroscopia na Região do Infravermelho.** *Rev. Virtual Quim.*, v. 10, n.2, 2018.

SILVERSTEIN, R. M.; WEBSTER, F. X.; KIEMLE, D. J., **Identificação Espectrométrica de Compostos Orgânicos.** 7ª ed. Rio de Janeiro: LTC Editora, 2006. 490 p.

SIMANZHENKOV, V.; IDEM, R. **Crude Oil Chemistry.** New York, EUA: Marcel Dekker, 2003. 410 p.

SKOOG, D. A.; HOLLER, F. J.; NIEMAN, T. A. **Princípios de Análise Instrumental.** 5ª ed. São Paulo: Bookman, 2002. 836 p.

SPEIGHT, J. G. **Handbook of Petroleum Analysis.** New York, USA: John Wiley & Sons, 2001. 474 p.

SPEIGHT, J. G. **Handbook of Petroleum Product Analysis.** New Jersey, USA: John Wiley & Sons, 2002. 461 p.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. **Estado da arte de figura de mérito em calibração multivariada.** *Quím. Nova*, v. 32, n. 5, p. 1-10, 2009.

WU, D.; HE, Y.; NIE, P.; CAO, F.; BAO, Y. **Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice.** *Analytica Chimica Acta*, v. 659, p. 229-237, 2010.

XIAOBO, Z. et al. **Variables selection methods in near-infrared spectroscopy.** *Analytica Chimica Acta*, v. 667, n. 1–2, p. 14–32, 2010.