

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

**Transferência de Calibração de Modelos Multivariados para  
Previsão de Propriedades Físico-Químicas em Petróleo  
Bruto**

**Rayza Rosa Tavares Rodrigues**

**Dissertação de Mestrado em Química**

**VITÓRIA  
2017**

**Rayza Rosa Tavares Rodrigues**

**Transferência de Calibração de Modelos Multivariados para  
Previsão de Propriedades Físico-Químicas em Petróleo Bruto**

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

**Área de Concentração:** Química.

**Linha de Pesquisa:** Química do Petróleo e Biocombustíveis.

Orientador: Prof. Dr. Eustáquio Vinícius  
Ribeiro de Castro

**VITÓRIA  
2017**

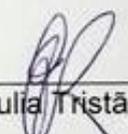
# Transferência de Calibração de Modelos Multivariados para Previsão de Propriedades Físico-Químicas em Petróleo Bruto

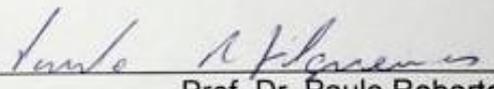
Rayza Rosa Tavares Rodrigues

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

Aprovado em 27/03/2017 por:

  
\_\_\_\_\_  
Prof. Dr. Eustáquio Vinicius Ribeiro de Castro  
Universidade Federal do Espírito Santo  
Orientador

  
\_\_\_\_\_  
Profa. Dra. Julia Tristão do Carmo Rocha  
UNESC  
Examinadora Externa

  
\_\_\_\_\_  
Prof. Dr. Paulo Roberto Filgueiras  
Universidade Federal do Espírito Santo  
Examinador Interno

Universidade Federal do Espírito Santo

Vitória, março de 2017

Dados Internacionais de Catalogação-na-publicação (CIP)  
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

---

R696t Rodrigues, Rayza Rosa Tavares, 1991-  
Transferência de calibração de modelos multivariados para  
previsão de propriedades físico-químicas em petróleo bruto /  
Rayza Rosa Tavares Rodrigues. – 2017.  
86 f. : il.

Orientador: Eustáquio Vinícius Ribeiro de Castro.  
Dissertação (Mestrado em Química) – Universidade Federal  
do Espírito Santo, Centro de Ciências Exatas.

1. Calibração. 2. Fontes de infravermelho. 3. Petróleo. I.  
Castro, Eustáquio Vinícius Ribeiro de. II. Universidade Federal do  
Espírito Santo. Centro de Ciências Exatas. III. Título.

CDU: 54

---

À minha mãe Edilma, por todo amor e companherismo  
E ao meu pai Ivar, por ser meu maior incentivador.

## **AGRADECIMENTOS**

À Deus, por me guiar pelos melhores caminhos.

À minha mãe Edilma, minha companheira de todas as horas.

Ao meu pai Ivar, pela presença mesmo à distância.

Ao meu irmão Diego e à minha cunhada Priscila, pelo otimismo e pelo carinho.

À minha família, por serem meu porto seguro.

Ao meu orientador, professor Eustáquio, pela oportunidade e pela confiança.

Ao professor Paulo, pelas valiosas contribuições ao longo deste trabalho.

Aos amigos do LabPetro/Ufes e em especial a Karla, Betina, Jessica e Bárbara pela companhia, colaboração, amizade e por dividirem comigo as frustrações do cotidiano.

Aos colegas do PPGQUI/Ufes, professores e alunos, pelo incentivo e colaboração para meu desenvolvimento.

Aos amigos do Atitude, alunos e professores, por me lembrarem sempre da importância de sonhar juntos e fazer a diferença.

À banca de qualificação, os professores Paulo e Wanderson pelas correções, contribuições e sugestões.

À banca de defesa, inclusive os membros suplentes, professores Eustáquio, Paulo, Julia, Wanderson, Juliano e Álvaro, por aceitarem o convite.

Ao Cenpes/Petrobras, pelas amostras e análises que possibilitaram a existência deste trabalho.

À UFSM, pelas análises realizadas.

Ao Labpetro/UFES, pelos recursos que permitiram a elaboração deste trabalho.

À Capes, pelo suporte financeiro.

A todos que de alguma forma contribuíram para a conclusão deste trabalho e me estimulam a sonhar sempre mais.

*“Se você pode sonhar, você pode realizar”*

Walt Disney

## LISTA DE FIGURAS

<b>Figura 1.</b> Espectro MIR típico de uma amostra de petróleo (a) sem qualquer pré-processamento e pré-processado por (b) SNV, (c) Derivada com suavização de Savitsk-Golay e (d) airPLS. ....	30
<b>Figura 2.</b> Ilustração do conceito de erro de previsão em função da complexidade do modelo de calibração. A seta indica o ponto ótimo. (Adaptado de Martens, 1991)...	33
<b>Figura 3.</b> Esquema demonstrativo da separação do conjunto de dados de petróleo bruto. ....	42
<b>Figura 4.</b> Espectros de MIR para amostras de petróleo bruto obtidos nos instrumentos (a) primário e (b) secundário.....	44
<b>Figura 5.</b> Histograma de densidade API para amostras de petróleo utilizadas em modelo de calibração. ....	45
<b>Figura 6.</b> Modelo PCA para espectros brutos dos instrumentos primário e secundário pré-processados por derivada com suavização. ....	45
<b>Figura 7.</b> Espectros de MIR dos instrumentos primário e secundário (a) brutos e (b) pré-tratados por SNV para modelagem por OPLS. ....	47
<b>Figura 8.</b> Valores previstos com relação aos valores medidos de densidade API para amostras externas pelos espectros de MIR do instrumento secundário antes e após SBC.....	49
<b>Figura 9.</b> Espectro de uma amostra externa dos instrumentos primário e secundário (a) após DS, e (b) após PDS com janela de 5 pontos. ....	50
<b>Figura 10.</b> Modelo PCA após transferências DS e PDS. ....	51
<b>Figura 11.</b> Resultado do teste de acurácia da previsão de amostras externas por meio de espectros de instrumento secundário pelo modelo PDS-OPLS (a) comparado ao modelo original e (b) comparado ao modelo FR.....	52
<b>Figura 12.</b> Resultados dos modelos de aplicação de espectros de um instrumento secundário para prever densidade API de amostras de petróleo bruto. As linhas horizontais representam os valores referentes ao modelo PLS original e os valores percentuais são o aumento de RMSEP. ....	53
<b>Figura 13.</b> Histograma de distribuição das amostras conforme (a) densidade API; (b) TIAC; (c) NAT; (d) NAN.....	62
<b>Figura 14.</b> Gráfico que correlaciona a densidade API obtida de acordo com a norma ISO 12185 e os valores correspondentes previstos pelo modelo.....	63

<b>Figura 15.</b> Gráfico de correlação entre TIAC obtida por procedimentos normatizados e os valores correspondentes previstos pelo modelo PLS para o instrumento primário. ....	65
<b>Figura 16.</b> Gráfico de correlação entre valores de referência e valores correspondentes previstos pelo modelo PLS para o instrumento primário para as propriedades NAT (a) e NAN (b). ....	66
<b>Figura 17.</b> Espectros brutos de 10 amostras de transferência de petróleo bruto coletados em dois instrumentos de MIR: (a) primário; (b) secundário. ....	68
<b>Figura 18.</b> Conjunto de espectros de petróleo bruto de amostras de transferência após tratamento com airPLS nos dois instrumentos de MIR: (a) primário; (b) secundário. ....	69
<b>Figura 19.</b> Gráfico de escores dos dois primeiros componentes principais de um modelo PCA construído com espectros MIR dos instrumentos primário e secundário tratados com airPLS. ....	70
<b>Figura 20.</b> Espectros das dez amostras externas obtidos (a) no instrumento primário e após pré-processamento por airPLS; e (b) no instrumento secundário e após interpolação, transferência PDS e tratamento por airPLS. ....	71
<b>Figura 21.</b> Gráfico de escores dos dois primeiros componentes principais de um modelo PCA construído com espectros primários tratados com airPLS e secundários transferidos por PDS e também tratados com airPLS. ....	72
<b>Figura 22.</b> Espectro secundário transferido superposto ao espectro primário correspondente para as amostras externas 3 (a) e 8 (b). ....	73
<b>Figura 23.</b> Resultados da previsão de densidade API para amostras dez externas do modelo PLS original com espectros primários e com espectros secundários submetidos à transferência de calibração PDS em comparação com valores de referência. ....	74
<b>Figura 24.</b> Resultados da previsão de TIAC para dez amostras externas do modelo PLS original com espectros primários e com espectros secundários submetidos à transferência de calibração PDS em comparação com valores de referência. ....	76
<b>Figura 25.</b> Resultados da previsão de NAT para nove amostras externas do modelo PLS original com espectros primários e com espectros secundários submetidos à transferência de calibração PDS em comparação com valores de referência. ....	77

**Figura 26.** Resultados da previsão de NAN para três amostras externas do modelo PLS original com espectros primários e com espectros secundários submetidos à transferência de calibração PDS em comparação com valores de referência. ....78

## LISTA DE TABELAS

<b>Tabela 1.</b> Composição elementar de petróleo .....	20
<b>Tabela 2.</b> Classificação de petróleo segundo densidade API .....	21
<b>Tabela 3.</b> Parâmetros do modelo original PLS de predição de densidade API a partir de espectros MIR obtidos no instrumento primário. ....	46
<b>Tabela 4.</b> Parâmetros do modelo OPLS de predição de densidade API a partir de espectros MIR obtidos no instrumento primário. ....	48
<b>Tabela 5.</b> Resultados do modelo FR de predição de densidade API a partir de espectros MIR obtidos no instrumento secundário.....	48
<b>Tabela 6.</b> Resultados dos modelos de transferência DS, PDS e PDS-OPLS de predição de densidade API a partir de espectros MIR obtidos no instrumento secundário.....	51
<b>Tabela 7.</b> Porcentagem média de similaridade entre amostras externas do instrumento secundário submetidas a diversos procedimentos em relação aos dados do instrumento primário.....	54
<b>Tabela 8.</b> Quantidade de amostras nos conjuntos de modelagem por PLS para cada propriedade .....	60
<b>Tabela 9.</b> Diferenças nos números de onda entre espectros obtidos nos diferentes instrumentos de MIR .....	61
<b>Tabela 10.</b> Dados dos modelos de calibração por PLS com espectros de instrumento MIR primário para as propriedades API, TIAC, NAT e NAN .....	64
<b>Tabela 11.</b> Resultados dos modelos generalizados, aqueles que tiveram amostras de validação externa incorporadas, em comparação aos modelos iniciais .....	67
<b>Tabela 12.</b> Porcentagem média de similaridade entre amostras de transferência interpoladas do instrumento secundário submetidas a diversos procedimentos em relação aos dados do instrumento primário .....	69
<b>Tabela 13.</b> Valores de referência das propriedades API, TIAC, NAT e NAN para as dez amostras que compõem o conjunto externo .....	73
<b>Tabela 14.</b> Resultado do teste estatístico de comparação de acurácia entre modelos PLS originais construídos para instrumento primário e modelos de transferência PDS de espectros secundários aplicados aos modelos originais .....	75

## LISTA DE ABREVIATURAS E SIGLAS

- airPLS - *adaptive iteratively reweighted penalized least squares* (método iterativamente adaptativo por mínimos quadrados ponderados e penalizados)
- ANP – Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
- API, densidade – *American Petroleum Institute* (medida de densidade relativa de petróleo proposta pelo Instituto Americano de petróleo)
- ATR - *attenuated total reflectance* (reflectância total atenuada )
- ASTM – *American Society for Test and Materials* (Sociedade americana para teste e materiais)
- CENPES - Centro de Pesquisas Leopoldo Americo Miguez de Mello
- CPM – *cross polar microscopy* (microscopia de luz polarizada )
- CV – *cross validation* (validação cruzada)
- DS – *direct standardization* (padronização direta)
- DSC – *differential scanning calorimeter* (calorimetria exploratória diferencial)
- FTIR – *Fourier transform infrared* (infravermelho com transformada de Fourier)
- FR – *full recalibration* (recalibração total)
- ISO – *International Organization for Standardization* (Organização Internacional para padronização)
- LabPetro - Laboratório de Pesquisa e Desenvolvimento de Metodologias para Análise de Óleos Pesados
- MIR – *mid-infrared* (infravermelho médio)
- MLR – *multiple linear regression* (regressão linear múltipla)
- MSC – *multiplicative scatter correction* (correção do espalhamento multiplicativo)
- NAN – número de acidez naftênica
- NAT – número de acidez total
- NCQP - Núcleo de Competências em Química do Petróleo
- NIR – *near infrared* (infravermelho próximo)
- OPLS – *orthogonal projections to latent structures* (projeções ortogonais por estruturas latentes)
- OSC (*orthogonal signal correction*, correção ortogonal de sinal)
- PCA – *principal component analysis* (análise de componentes principais)
- PCR – *Principal Component Regression* (regressão por componentes principais)
- PDS – *piecewise direct standardization* (padronização direta por partes)

PETROBRAS - Petróleo Brasileiro S/A

PLS – *partial least squares* (mínimos quadrados parciais)

$R^2$  – coeficiente de determinação

RMSEC – *root mean square errors of calibration* (raiz quadrada do erro quadrático médio de calibração)

RMSECV – *root mean square error of cross validation* (raiz quadrada do erro quadrático médio de validação cruzada)

RMSEP – *root mean square error of prediction* (raiz quadrada do erro quadrático médio de predição)

SBC – *slope and bias correction* (correção de inclinação e intercepto)

SVN – *standard normal variate* (variação normal padrão)

SVR – *support vector regression* (regressão por vetor de suporte)

TIAC – temperatura de início do aparecimento de cristais

VL – variáveis latentes

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	16
<b>1.1 Objetivos</b> .....	18
<b>1.2 Estrutura da dissertação</b> .....	18
<b>2. PETRÓLEO</b> .....	20
<b>2.1 Espectroscopia aplicada a petróleo</b> .....	21
<b>3. QUIMIOMETRIA</b> .....	25
<b>3.1 Mínimos quadrados parciais (PLS)</b> .....	26
<b>3.2 Projeções ortogonais em estruturas latentes (OPLS)</b> .....	27
<b>3.3 Pré-processamento de dados</b> .....	28
<b>3.4 Transferência de calibração</b> .....	30
<b>3.5 Avaliação de modelos quimiométricos</b> .....	32
<b>3.6 Comparação de exatidão de modelos de transferência</b> .....	36
<b>3.7 Análise da sobreposição entre espectros das mesmas amostras</b> .....	37
<b>4. AVALIAÇÃO DE MÉTODOS DE TRANSFERÊNCIA DE CALIBRAÇÃO</b> .....	38
<b>4.1 Introdução</b> .....	38
<b>4.2 Objetivos</b> .....	39
4.2.1 <i>Objetivo geral</i> .....	39
4.2.2 <i>Objetivos específicos</i> .....	39
<b>4.3 Metodologia</b> .....	39
4.3.1 <i>Caracterização físico-química</i> .....	40
4.3.2 <i>Instrumentação</i> .....	40
4.3.3 <i>Quimiometria</i> .....	41
a) <i>Modelos de predição de densidade API</i> .....	42
b) <i>Avaliação dos modelos</i> .....	43
<b>4.4 Resultados e discussão</b> .....	44
<b>4.5 Conclusões</b> .....	54
<b>5. TRANSFERÊNCIA DE CALIBRAÇÃO PARA MODELOS PLS</b> .....	55
<b>5.1 Introdução</b> .....	55
<b>5.2 Objetivos</b> .....	56
5.2.1 <i>Objetivo geral</i> .....	56
5.2.2 <i>Objetivos específicos</i> .....	56
<b>5.3 Metodologia</b> .....	56

5.3.1	<i>Caracterização físico-química</i>	57
5.3.2	<i>Instrumentação</i>	58
5.3.3	<i>Quimiometria</i>	58
a)	<i>Modelos de predição de densidade API, TIAC, NAT E NAN</i>	59
b)	<i>Transferências de calibração</i>	60
<b>5.4</b>	<b>Resultados e discussão</b>	62
<b>5.5</b>	<b>Conclusões</b>	79
<b>6.</b>	<b>CONCLUSÕES GERAIS</b>	80
<b>7.</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	81

## RESUMO

A calibração multivariada associada à técnica de infravermelho é uma alternativa aos métodos tradicionais para determinação de parâmetros físico-químicos em petróleo. Entretanto, o modelo multivariado construído é aplicável somente para o instrumento no qual os espectros foram obtidos. A transferência de calibração de modelos multivariados é de fundamental importância para a indústria petrolífera, levando-se em conta que aumenta a aplicabilidade de modelos e permite estimar com rapidez e com poucos gastos diversas propriedades físico-químicas. Este trabalho foi dividido em duas seções: a primeira é dedicada à avaliação de modelos de transferência entre dois instrumentos de infravermelho a região média (MIR) para modelos PLS (mínimos quadrados parciais) e OPLS (projeções ortogonais em estruturas latentes); e a segunda visa aplicar a transferência PDS (padronização direta por partes) a modelos PLS para predição de densidade API, TIAC (temperatura de início de aparecimento de cristais), NAT (número de acidez total) e NAN (número de acidez naftênica). Dentre os métodos quimiométricos SBC (correção de declive e viés), FR (recalibração total), DS (padronização direta) e PDS avaliados para previsão de densidade API, o modelo de transferência PDS aplicado ao modelo OPLS resultou na melhor capacidade preditiva (RMSEP de 1,48). A aplicação direta de PDS a um modelo PLS original conta com a vantagem de aproveitar modelos já consolidados, e os resultados da primeira seção indicam que isto é uma possibilidade. Na segunda seção, as propriedades API, TIAC, NAT e NAN foram modeladas e validadas por PLS para o instrumento primário. Espectros secundários, diferentes daqueles da primeira seção, passaram por uma interpolação polinomial a fim de igualar as variáveis às do instrumento primário, antes da aplicação da transferência PDS. Os espectros transferidos e processados por airPLS, técnica iterativa adaptativa, se tornaram indistinguíveis por PCA (análise por componentes principais) dos espectros primários. A técnica de PDS associada a modelos PLS mostrou-se promissora, especialmente sendo capaz de estimar com boa exatidão o valor da densidade API de amostras de petróleo classificadas como medianas.

**Palavras-chave:** Transferência de calibração. MIR. Petróleo bruto.

## ABSTRACT

Multivariate calibration combined with infrared technique is an alternative to traditional methods of determination of physicochemical parameters in crude oil. However, a multivariate model can only be applied for the instrument in which the spectra were measured. In case of equipment upkeep or change of instrument, transferring the calibration model is necessary for the new instrumental condition or new instrument. Transfer of calibration of multivariate models have fundamental importance for petroil industry, considering that it increases the aplicability of models and enables quick and unexpensive estimatives of various physico-chemical parameters. This work is divided into two sections: the first one is dedicated to evaluation of transfer models between two mid-infrared instruments (MIR) for PLS (partial least squares) models and OPLS (orthogonal projections to latent structures); and the second aims to apply the PDS (piecewise direct standardization) transfer technique to PLS models for prediction of API gravity, WAT (wax appearance temperature, in portuguese TIAC), TAN (total acid number, in portuguese NAT) and NAN (naphtenic acid number). Among the evatuated chemometric methods SBC (slope and bias correction), FR (full recalibration), DS (direct standardization) and PDS, for prediction of API density, the PDS transfer model applied to the OPLS model resulted in the best predictive ability (RMSEP of 1.48). Direct application of PDS to a PLS original model has the advantage of reusing consolidated models, and the results from the first section indicate that this is a possibility. At the second section, the properties API, WAT, TAN and NAN were modeled and validated by PLS. The secondary espectra, other than those of the first section, were subjected to a polynomial interpolation to equalize the variables to those of the primary instrument, prior to the application of the PDS transfer. The transferred and airPL (adaptative iteratively reweighted penalized least squares) processed spectra became indistiguishable from the primary spectra by PCA (principal component analysis). The PDS technique associated to PLS models was shown to be promissing, specially for being capable of estimating with good accuracy the value of API gravity of petroleum samples classified as medium oil.

**Keywords:** Calibration transfer. MIR. Crude oil.

## 1. INTRODUÇÃO

O desenvolvimento da quimiometria, e em especial os métodos de calibração multivariada, promoveu grandes avanços à química, especialmente na área de química analítica. Por meio de técnicas quimiométricas, dados complexos do ponto de vista químico, estatístico e matemático podem ser tratados de maneira a criar modelos para estimar uma propriedade de interesse a partir de uma medida instrumental. A modelagem é feita relacionando dados instrumentais, frequentemente obtidos de maneira simples, rápida, pouco onerosa e consumindo pequeno volume de amostra, a valores correspondentes de uma propriedade de interesse (BRERETON, 2003).

Assim como outros cientistas de várias partes do mundo, o grupo de quimiometria da Ufes trabalha há alguns anos com modelagem quimiométrica, com especial destaque à utilização de espectroscopia no infravermelho com transformada de Fourier (FTIR) nas regiões próxima (NIR) e média (MIR) para previsão de propriedades físico-químicas de amostras de petróleo e derivados. Devido à complexidade desse tipo de amostra, a caracterização necessária à indústria envolve um grande número de ensaios analíticos, que são dispendiosos e demorados (SPEIGHT, 2002; RIAZI, 2005). De fato, é possível ultrapassar a marca de 500 ensaios, levando em alguns casos mais de um ano para finalização (DIAS, 2005). Para tomada de decisões comerciais e industriais, uma alternativa que permita obter estimativas de propriedades físico-químicas acerca da amostra é, portanto, de grande valor.

O método dos mínimos quadrados parciais (PLS, *partial least squares*) é a principal maneira de construir modelos de calibração multivariada a partir de uma matriz contendo um vetor de dados para cada amostra, isto é, dados de primeira ordem. A importância deste método está relacionada à conveniência de dispensar o conhecimento exato de todos os constituintes da amostra, desde que os dados aplicados ao modelo sejam semelhantes àqueles utilizados na calibração durante a modelagem (BRERETON, 2003; SOARES et al., 2008).

Modelos PLS são interessantes para amostras de petróleo, uma vez que devido a grande variabilidade não se conhece sua composição química exata. Vários estudos anteriores indicam a eficiência da aplicação da espectroscopia NIR ou MIR para a previsão de algumas propriedades físico-químicas em petróleo, como Abbas

et al (2012) que aplicaram PLS a espectros de MIR para prever densidade API e razão entre compostos alifáticos e aromáticos em amostras de petróleo bruto e Filgueiras et al (2014) que criaram modelos de PLS e também pelo método não-linear de regressão por vetor de suporte (SVR – *support vector regression*) para prever densidade API, viscosidade cinemática e teor de água em petróleo bruto por meio de espectros de MIR destacando a vantagem de necessitar de apenas um espectro para determinar as três propriedades.

O modelo multivariado PLS tem caráter local, isto é, se aplica somente ao instrumento no qual os espectros utilizados na modelagem foram obtidos e para amostras similares às usadas na calibração. No caso de manutenção, troca de instrumento ou mesmo utilização do modelo por outros laboratórios faz-se necessária a transferência do modelo construído para a nova condição instrumental ou novo aparelho (FEUDALE, 2002). Este estudo visa propor maneiras de viabilizar a generalização de modelos por meio da utilização de modelagem mais robusta ou transferência de calibração, aumentando a aplicabilidade de modelos e evitando assim que uma grande quantidade de medidas precise ser feita para cada instrumento.

Uma modificação da técnica PLS é o OPLS (projeções ortogonais em estruturas latentes), que requer a ortogonalização de todas as amostras que serão preditas, uma vez que para reduzir a complexidade do modelo é removida a variação sistemática nos espectros que não são correlacionadas à propriedade sendo prevista. Desta forma, o OPLS proporciona um modelo robusto que em alguns casos pode ser utilizado diretamente com dados de um instrumento secundário, diferente daquele utilizado para calibrar o modelo (GABRIELSSON, 2006).

Para transferência de modelos, as técnicas SBC (*slope and bias correction*, correção de declive e viés), DS (*direct standardization*, padronização direta) e PDS (*piecewise direct standardization*, padronização direta por partes) foram estudadas e aplicadas a casos reais a fim de atender necessidade da indústria possibilitando aplicação mais ampla de modelos construídos com dados de determinado instrumento analítico.

Neste trabalho foram utilizados dados espectrais de infravermelho médio (MIR) provenientes de três laboratórios distintos no desenvolvimento de modelos PLS e OPLS de previsão das propriedades físico-químicas densidade API (*American Petroleum Institute*, medida de densidade relativa do Instituto Americano de

Petróleo), TIAC (temperatura de início do aparecimento de cristais), NAT (número de acidez total) e NAN (número de acidez naftênica) e transferência de calibração.

### **1.1 Objetivos**

- Construir e avaliar modelos de transferência de calibração para predição de propriedades físico-químicas de petróleo bruto entre dois instrumentos de infravermelho médio.
- Prever propriedades físico-químicas de petróleo bruto em um instrumento MIR secundário por meio de modelos PLS validados para um instrumentos primário.

### **1.2 Estrutura da dissertação**

Esta dissertação apresenta algumas técnicas de transferência de calibração para possibilitar a generalização de modelos quimiométricos construídos por PLS utilizando espectros de infravermelho médio relacionados a parâmetros de qualidade de amostras de petróleo brutos representativas da produção brasileira. Os óleos utilizados neste trabalho apresentam ampla variabilidade, e a distribuição das propriedades físico-químicas abordadas estão apresentadas em ao longo deste texto.

O capítulo 2 é dedicado à abordagem teórica acerca de petróleo, discorrendo sobre constituição e classificação, bem como descrevendo alguns aspectos ligados à aplicação da espectroscopia nesta matriz química.

No capítulo 3 é feito um embasamento teórico sobre quimiometria visando especificamente uma compreensão sobre as técnicas de mínimos quadrados parciais (PLS) e projeções ortogonais em estruturas latentes (OPLS) para regressão de dados. Além da abordagem de algumas técnicas de pré-processamento e de métodos estatístico de avaliação de modelos, são explanadas as técnicas de transferência de calibração aplicadas em capítulos subsequentes.

O capítulo 4 traz uma avaliação de métodos de transferência de calibração, comparando a exatidão entre a previsão das mesmas amostras pelas técnicas SBC, DS, PDS relacionando ainda aos métodos de regressão PLS e OPLS.

Uma transferência de calibração envolvendo equiparação de espectros de diferentes instrumentos de MIR pela técnica PDS foi feita e apresentada no capítulo

5 para aplicação em modelos PLS construídos para quatro propriedades físico-químicas de petróleo bruto, densidade API, TIAC, NAT e NAN.

Por fim, o capítulo 6 resume as conclusões gerais acerca das abordagens de transferência de calibração realizadas.

## 2. PETRÓLEO

Diversos produtos usados cotidianamente são derivados do petróleo, como gasolina, óleo diesel, querosene, asfalto além de matérias-primas de várias indústrias químicas, como nafta, propano e butano. O petróleo tem uma constituição muito complexa e variável, contendo milhares de compostos diferentes de massa molar variando de  $16 \text{ g.mol}^{-1}$  a mais de  $2000 \text{ g.mol}^{-1}$  (SPEIGHT, 2002; RIAZI, 2005).

O rendimento e as propriedades dos derivados de petróleo dependem das características físicas e químicas do óleo original, como a concentração de hidrocarbonetos e constituintes minoritários (SPEIGHT, 2001). A diversidade de petróleos, isto é, a ampla variação na composição química, depende da origem do petróleo, da idade do reservatório e da profundidade em que o óleo é extraído do poço (PARISOTTO, 2007 apud ANDRADE, 2009).

A Tabela 1 mostra a variação dos elementos químicos mais frequentes em petróleo. Embora mais de 90% seja constituída por carbono e hidrogênio, a presença de compostos nitrogenados, oxigenados e sulfurados interfere significativamente nas propriedades físico-química do petróleo, afetando por exemplo, a acidez, corrosão, reatividade e fluidez do óleo. A presença de elementos traço, como vanádio e níquel podem afetar a eficiência catalítica dos processos de refino de petróleo.

**Tabela 1.** Composição elementar de petróleo (SPEIGHT, 2001).

<b>Elemento</b>	<b>Porcentagem em massa (%)</b>
<b>Carbono</b>	83,0 - 87,0
<b>Hidrogênio</b>	10,0 - 14,0
<b>Nitrogênio</b>	0,1 - 2,0
<b>Oxigênio</b>	0,05 - 1,5
<b>Enxofre</b>	0,05 - 6,0

A medida da densidade do petróleo é um dos principais fatores usados para comercialização, uma vez que funciona como indicador de qualidade, sendo que óleos menos densos fornecem produtos de maior valor comercial. A ciência do petróleo utiliza uma escala hidrométrica de densidade relativa desenvolvida pelo *American Petroleum Institute* (API), dada pela Equação 1.

$$^{\circ}API = \frac{141,5}{densidade\ específica\ (60^{\circ}F)} - 131,5 \quad \text{Eq. 1}$$

Os petróleos podem então ser classificados em leves, médios e pesados, conforme Tabela 2. Petróleos mais leves em geral fornecem maior quantidade de produtos de maior valor comercial, como gasolina e óleo diesel, enquanto petróleos mais pesados contém maior quantidade de metais, que podem causar problemas durante etapas do refino envolvendo catálise, e geram produtos de menor valor agregado.

**Tabela 2.** Classificação de petróleo segundo densidade API. Fonte: ANP

<b>Tipo de petróleo</b>	<b>°API</b>
<b>Leve</b>	≥ 31
<b>Médio</b>	≥ 22 e < 31
<b>Pesado</b>	< 22

Após a extração do petróleo no campo produtor, o óleo é transportado por oleodutos até as refinarias, onde é destilado para obtenção de derivados. O valor comercial de um petróleo varia conforme a facilidade de transporte e a qualidade para refino, de acordo com sua capacidade de gerar produtos que atendam à demanda do mercado. A caracterização de diversas propriedades físico-químicas é fundamental para avaliar o petróleo, auxiliando na elaboração de estratégias de transporte e refino, além de informarem sobre potenciais derivados (SPEIGHT, 2002).

Uma importante propriedade determinada em petróleo é a temperatura de início de aparecimento de cristais, TIAC, que mede o momento em que os primeiros cristais de parafina deixam de fazer parte da solução e modificam o comportamento reológico do petróleo (VENKATESAN et al., 2005). A determinação da TIAC pode ser feita por diferentes técnicas, entre elas: microscopia de luz polarizada (CPM – *cross polar microscopy*), reômetro/viscosímetro e calorimetria exploratória diferencial (DSC). Na técnica CPM, a TIAC é determinada como a temperatura na qual os primeiros cristais aparecem sob luz polarizada; no método do reômetro/viscosímetro, a determinação é feita pelo ponto de aumento repentino da viscosidade com

redução da temperatura, neste caso o termo foi denominado “temperatura de cristalização” (MARCHESINI et al., 2012); a técnica que utiliza DSC funciona pela obtenção de um perfil de calor liberado durante cristalização da parafina, e em geral fornece boa precisão. (KOK, M. V., LETOFFE, J. M., CLAUDY, P., 1999 e JAPPER-JAAFAR, A., BHASKOROB, P. T. e MIORC, Z. S., 2016) Na indústria de petróleo é fundamental conhecer a TIAC para prever o processo de deposição orgânica, permitindo estimar a tendência de um óleo à formação de depósitos orgânicos que possam comprometer tanto o transporte do óleo entupindo oledutos quanto os processos das refinarias.

Dentre as diversas propriedades que caracterizam uma amostra de petróleo, destaca-se o número de acidez total (NAT), cuja unidade é dada em miligramas de KOH por grama de óleo. Amostras mais ácidas em geral formam emulsões mais estáveis, além de provocar corrosão em oledutos, portanto é desejável que o valor de NAT seja baixo; no entanto a existência de compostos ácidos no óleo é importante pois está relacionada à formação de um filme protetor contra corrosão (SLAVCHEVA et al., 1999).

O principal problema associado à elevada acidez do petróleo tem relação com a presença de ácidos naftênicos, estruturas com anéis fundidos de ciclopentano e ou ciclo-hexano ou ácidos acíclicos com saturação pertencente à função orgânica dos ácidos carboxílicos (GRUBER et al., 2012). Sendo assim, uma técnica não normatizada porém de grande importância é a determinação do número de acidez naftênica, que inclui etapa de extração de ácidos naftênicos previamente à medida por MIR. A extração de ácidos naftênicos é uma tarefa difícil devido à complexidade de uma amostra de petróleo bruto (GRUBER et al., 2012).

## **2.1 Espectroscopia aplicada a petróleo**

A utilização de métodos espectroscópicos vem sendo explorada por diversos grupos de pesquisa como alternativa para estimativa de propriedades físico-químicas de petróleos associada a técnicas quimiométricas. (MACHO et al., 2002; FALLA et al., 2006; PEINDER et al., 2010; LAXALDE et al., 2011; ABBAS et al., 2012; JINGYAN et al., 2012; FILGUEIRAS et al., 2014b) Técnicas espectroscópicas são capazes de fornecer informações sobre o óleo em nível molecular de maneira prática e rápida, e informações quantitativas em geral podem ser obtidas

indiretamente, especialmente em amostras muito complexas como é o caso do petróleo. (FILGUEIRAS et al., 2011; PEINDER et al., 2009).

A espectroscopia de infravermelho é uma técnica que possibilita ensaios de forma rápida, barata, simples e precisa, que em geral não requer preparo de amostra além de não ser destrutiva. (SKOOG e LEARY, 1992) A região do infravermelho inclui números de onda entre 12800 e 10  $\text{cm}^{-1}$ , sendo dividida em próxima (12800 a 4000  $\text{cm}^{-1}$ ), média (4000 a 200  $\text{cm}^{-1}$ ) e distante (200 a 10  $\text{cm}^{-1}$ ). Este trabalho trata somente da espectroscopia de infravermelho médio (MIR).

A determinação da presença de grupos funcionais é uma aplicação frequente da MIR, especialmente ligações químicas com maiores variações no momento de dipolo, como C=O, N-H e O-H, olefinas alifáticas, ligação C-H em compostos aromáticos e vibrações C-C. O espectro obtido é considerado uma impressão digital da amostra, uma vez que cada amostra apresenta um perfil único pois pequenas diferenças na composição e estrutura das moléculas geram grandes modificações na distribuição das bandas de absorção. (SKOOG e LEARY, 1992)

Espectroscopia de reflexão interna é uma técnica que foi desenvolvida para obter espectros de amostras difíceis de trabalhar, como sólidos pouco solúveis, filmes e adesivos. A fonte de radiação é direcionada a passar por um material mais denso para outro menos denso com um ângulo tal que ocorre reflexão total do feixe. (SKOOG e LEARY, 1992) Foi provado que durante a reflexão total o feixe de radiação, denominado onda evanescente, penetra no meio menos denso em uma pequena distância. (FAHRENFORT, 1961) O meio mais denso é um material cristalino, por exemplo feito de seleneto de zinco (ZnSe), enquanto o menos denso é a própria amostra. Desta forma, a radiação é atenuada pela amostra, por isto o fenômeno é chamado reflectância total atenuada (ATR), e o processo permite a obtenção de informações únicas da amostra.

Em um ATR simples (*single* ATR) o feixe de radiação é refletido uma única vez na amostra e segue para o detector. Já no multi ATR (*multi-bounce* ATR) a radiação é refletida várias vezes, com absorção e atenuação, antes de alcançar o detector. (SKOOG e LEARY, 1992) O caminho ótico percorrido pela amostra é maior e mais pontos da amostra são analisados, o que resulta em bandas melhor resolvidas, por isso amostras de petróleo e derivados são preferencialmente analisadas por este método.

Associando a técnica de MIR a métodos quimiométricos, várias propriedades podem ser previstas simultaneamente a partir de um único espectro de infravermelho (FILGUEIRAS et al., 2014b). Embora esta metodologia não seja substitutiva aos métodos tradicionais, é possível obter informações preliminares que agilizam o processo de tomada de decisão na indústria petroquímica.

### 3. QUIMIOMETRIA

O avanço e a popularização de técnicas analíticas, como a espectroscopia de infravermelho com transformada de Fourier (FTIR), permitiu que grandes conjuntos de dados pudessem ser gerados rapidamente. Desta forma, foi estimulando o desenvolvimento de métodos para obter o maior número de informações possível a partir destes dados. Neste contexto foi criado o campo da Quimiometria, o qual usa conceitos de Química, Matemática e Estatística por meios computacionais para extrair informações químicas relevantes a partir de dados medidos e representar e apresentar estas informações. (WOLD e SJÖSTRÖM, 1998)

A quimiometria tem sido aplicada em diversas matrizes químicas, com especial destaque em amostras com alto grau de complexidade, como petróleo e derivados. Uma análise tradicional completa de uma amostra de petróleo, fundamental para a indústria, é dispendiosa, demorada, requer grande volume de amostra além de necessitar de técnicos qualificados para realização de cada análise. (SPEIGHT, 2001) Várias propriedades físico-químicas, tais como densidade API, teor de enxofre, viscosidade e número de acidez, precisam ser determinadas para que o petróleo tenha seu valor estabelecido. Além disso, o conhecimento acerca das propriedades do óleo permite a escolha de tratamentos adequados à obtenção de produtos de alto valor comercial, como gasolina e óleo diesel, que atendam às exigências de agências de controle ambiental. Neste contexto, a quimiometria se torna uma alternativa interessante para obtenção de resultados com menor tempo e custo. No entanto, há uma redução na exatidão em comparação aos métodos hoje utilizados na caracterização de petróleo, uma vez que às incertezas inerentes ao método tradicional de análise, combinam-se as incertezas devidas à inferência estatística.

Métodos lineares como a regressão por mínimos quadrados parciais (PLS) vêm sendo muito explorados na avaliação de petróleos e derivados. (SKROBOT et al., 2005; PEREIRA et al., 2006; SKROBOT et al., 2007) Um desenvolvimento deste método é o OPLS (*Orthogonal Projections to Latent Structures* – projeções ortogonais em estruturas latentes) que modifica o processo de mínimos quadrados parciais por iterações não-lineares (NIPALS – *Non-linear Iterative Partial Least Squares*) para melhorar variações sistemáticas do conjunto de dados instrumental

excluindo informações que não são correlacionadas ao valor de referência, criando modelo de menor complexidade e maior robustez. (BOUVERESSE et al., 1996)

### 3.1 Mínimos quadrados parciais (PLS)

Mínimos quadrados parciais (PLS – *Partial Least Squares*) é a técnica de regressão mais usada em calibração multivariada aplicada a diversas áreas. As previsões da propriedade de interesse é feita associando técnica instrumental a valores de referência para a propriedade. Essas previsões são essencialmente estatísticas e a presença de interferentes não tem influência nos resultados desde que as amostras de calibração contenham características similares em relação às amostras a serem previstas pelo modelo (vantagem de primeira ordem) (ABBAS et al., 2012).

Em 1986, GELADI e KOWALSKI destacaram pontos fracos dos métodos regressão linear múltipla (MLR – *Multiple Linear Regression*) e regressão por componentes principais (PCR – *Principal Component Regression*), como a impossibilidade de aplicação caso o número de variáveis seja maior do que o número de amostras, além da alta colinearidade dos dados provocar dificuldades, o que limita a aplicação em dados espectroscópicos. Estes autores desenvolveram um algoritmo completo e elegante delineando o método PLS para remediar essas fraquezas.

Medidas instrumentais (matriz  $\mathbf{X}$ , neste trabalho espectros MIR) são relacionadas com a propriedade de interesse correspondente (vetor  $\mathbf{y}$ ). Uma vez que as predições são essencialmente estatísticas, a presença de interferentes não tem influência nos resultados desde que as amostras de calibração contenham características similares em relação às amostras a serem previstas pelo modelo (vantagem de primeira ordem) (BRERETON, 2000).

A matriz de dados instrumentais é decomposta em matrizes de *scores* ( $\mathbf{T}$ ) e *loadings* ( $\mathbf{P}$ ) simultaneamente à decomposição do vetor  $\mathbf{y}$  em *scores* e *loadings* ( $\mathbf{q}$ ), sendo as matrizes  $\mathbf{E}$  e  $\mathbf{f}$  os resíduos associados, conforme Equações 2 e 3 (MARTENS e NAES, 1991).

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad \text{Eq. 2}$$

$$\mathbf{y} = \mathbf{Tq}^T + \mathbf{f} \quad \text{Eq. 3}$$

Esta regressão é feita de modo a minimizar a covariância entre  $X$  e  $y$ . O sistema é então descrito por um número de variáveis latentes (VL), que deve estar relacionado ao número de fontes de variação nos espectros do conjunto de calibração. (ASTM, 2012)

Espectros de infravermelho associados a técnicas quimiométricas têm um potencial promissor para estimar propriedades físicas e químicas em petróleo bruto e derivados, com diversas aplicações bem sucedidas já reportadas em vários artigos científicos. (BREITKREITZ et al., 2003; SATYA et al., 2007; PEINDER et al., 2010; PARISOTTO et al., 2010; LAXALDE et al., 2011; MÜLLER et al., 2012; ABBAS et al., 2012; MELÉNDEZ et al., 2012; JINGYAN et al., 2012; FILGUEIRAS et al., 2014b) BREITKREITZ et al. (2003) utilizaram, dentre outras técnicas, o PLS associado à espectroscopia de infravermelho na região próxima (NIR) para prever teor total de enxofre em diesel combustível. ABBAS et al. (2012) aplicaram PLS a espectros de MIR para prever densidade API e razão entre compostos alifáticos e aromáticos em amostras de petróleo bruto, concluindo que a técnica pode ser usada como método preliminar para estimar rapidamente as propriedades estudadas. FILGUEIRAS et al. (2014b) criaram modelos de PLS e também pelo método não-linear de regressão por vetor de suporte (SVR – *Support Vector Regression*) que possibilitaram prever densidade API, viscosidade cinemática e teor de água em petróleo bruto por meio de espectros de MIR destacando a vantagem de necessitar de apenas um espectro para determinar as três propriedades. ROCHA et al. (2016) determinaram o teor de enxofre em frações de petróleos brasileiros por MIR e NIR usando diferentes métodos de seleção de variáveis e regressão por PLS.

### **3.2 Projeções ortogonais em estruturas latentes (OPLS)**

O método das projeções ortogonais em estruturas latentes (*orthogonal projections to latent structures* – OPLS) é uma modificação do método PLS que inclui uma etapa que filtra toda variação estruturada que não tem correlação (ortogonal) com a propriedade de interesse. (SOUJHI et al., 2015)

Segundo TRYGG e WOLD (2002) a remoção de variação sistemática não-correlacionada por meio do método OPLS torna a interpretação dos modelos mais simples, podendo reduzir a complexidade enquanto mantém a capacidade preditiva. A variação não-correlacionada à propriedade de interesse pode ser proveniente, por exemplo, de espalhamento de luz e diferenças no caminho ótico ou ainda por nem

todas as variáveis serem relacionadas àquela propriedade. O OPLS é um método desenvolvido como uma modificação do método de pré-processamento OSC (*orthogonal signal correction*, correção ortogonal de sinal) por meio da integração desta etapa à modelagem PLS, facilitando a interpretação de resultados e permitindo a análise separada da variação não-correlacionada.

VERRON et al. (2004) estudaram propriedades teóricas da técnica OPLS e a aplicaram em espectros NIR de amostras de solo, mostrando graficamente as propriedades que defendem, especialmente sobre a escolha do número ótimo de componentes a ser usado na modelagem. SOUIHI et al. (2015) trabalharam no monitoramento de reação de hidrogenação em série, conseguindo melhor entendimento do processo e detecção precoce de falhas. Além de ser utilizado para criação de modelos, métodos que incluem correção ortogonal têm sido explorados para a área de transferência de calibração (TRYGG e WOLD, 2002).

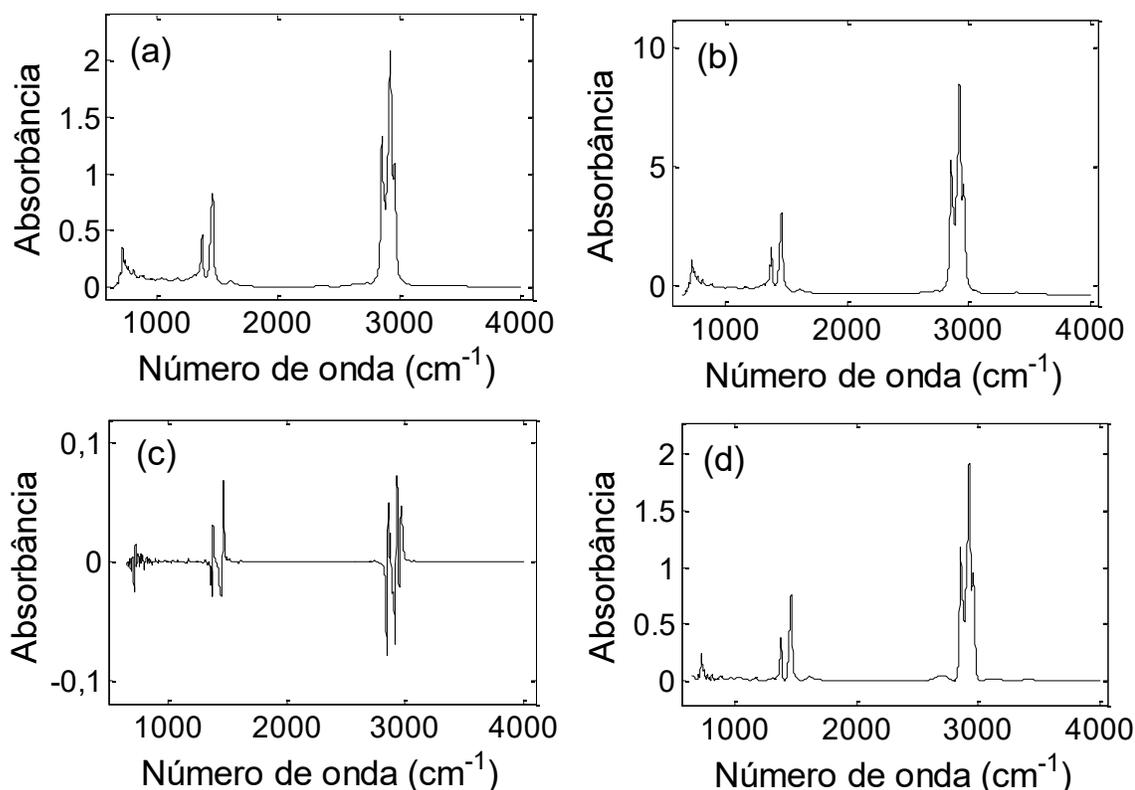
### **3.3 Pré-processamento de dados**

A fim de facilitar a interpretação e a obtenção de informações a partir de espectros de infravermelho, é fundamental promover transformações antes do tratameto quimiométrico para corrigir a distribuição, deslocamentos de linha de base e eventuais espalhamentos de luz. Este tipo de transformação, chamada pré-processamento, pode ser feita por meio de variados métodos, como correção de espalhamento multiplicativo (MSC – *multiplicative scatter correction*) (BLANCO et al., 1997; FEARN et al., 2009), variação normal padrão (SNV – *standard normal variate*) (BARNES, HANOA e LISTER, 1989; BLANCO et al., 1997; FEARN et al., 2009) e aplicação de derivadas (WINDIG e STEPHENSON, 1992).

Os métodos MSC e SNV promovem correção do espalhamento multiplicativo da luz produzindo resultados semelhantes, e por isso, costumam ser consideradas permutáveis. A diferença básica entre os métodos consiste no fato de que a SNV utiliza a média e o desvio padrão de cada espectro individual, enquanto o MSC relaciona cada espectro a um espectro de referência. (BLANCO et al, 1997)

O pré-processamento pela aplicação de primeira ou segunda derivada visa remover possíveis deformações na linha de base, bem como melhorar a definição de bandas de absorção sobrepostas. Desta forma, bandas de pequena absorbância podem ser reveladas e a resolução espectral é aumentada. (BRERETON, 2003) O método consiste na aplicação da derivada a uma janela de pontos a partir de um polinômio. Três parâmetros precisam ser definidos: o tamanho da janela de variáveis, o grau do polinômio e a ordem da derivada. O mais usual é a aplicação de derivada de primeiro grau, quando há desvios lineares na linha de base, ou de derivada de segundo grau, para eliminação de desvio não lineares. O tratamento por derivada é frequentemente seguido de uma suavização de Savitzky-Golay para evitar o aumento de ruído espectral (SAVITZKY e GOLAY, 1964).

Uma maneira de corrigir desvios da linha de base de espectros de infravermelho é utilizar o método iterativamente adaptativo por mínimos quadrados ponderados e penalizados, o airPLS. Este método consiste em um ajuste da linha de base feito de maneira rápida e flexível, por meio de um algoritmo. A soma dos erros quadrados entre uma linha de base previamente montada e os sinais originais permitem calcular pesos a cada iteração e cada nova rodada o procedimento é reponderado de maneira a aproximar-se a linha de base complexa (GAN et al., 2006). Enquanto a derivada de Savitzky-Golay resulta em um gráfico completamente diferente do espectro original, o método airPLS mantém a aparência do espectro, assim como a escala da absorbância original, facilitando a interpretação das modificações realizadas. A Figura 1 mostra o efeito dos vários tipos de pré-processamento em um espectro MIR típico de petróleo.



**Figura 1.** Espectro MIR típico de uma amostra de petróleo (a) sem qualquer pré-processamento e pré-processado por (b) SNV, (c) Derivada com suavização de Savitsk-Golay e (d) airPLS.

### 3.4 Transferência de calibração

A aplicação de técnicas de transferência de calibração é uma alternativa para possibilitar a generalização de modelos multivariados de calibração, levando em conta que podem ocorrer variações na resposta instrumental ao longo do tempo ou ainda pode ser necessário utilizar um novo instrumento para obtenção de dados. Dentro destas técnicas, são destaques na literatura a correção de inclinação e intercepto (*slope and bias correction - SBC*), a padronização direta (*direct standardization - DS*) e a padronização direta por partes (*piecewise direct standardization - PDS*).

O SBC é um método descrito por BOUVERESSE et al. (1994) em que dados obtidos em um instrumento secundário das amostras a serem previstas são inseridos diretamente no modelo original construído para um instrumento primário e posteriormente os valores de respostas são corrigidos por um modelo linear univariado de forma a se aproximar do valor real. Este método tem potencial especialmente para transferências entre instrumentos de mesma marca e modelo,

pois as respostas obtidas no instrumento secundário em geral se comportam de maneira linear quando aplicadas ao modelo de calibração.

O método DS faz a transferência integral dos espectros lidos no instrumento secundário para o primário utilizando uma matriz de transferência. (HONORATO et al., 2007; LIU, HAN e YANG, 2011; WANG, VELTKAMP e KOWALSKI, 1991) Neste método, proposto em 1991 por WANG, VELTKAMP e KOWALSKI, um algoritmo é utilizado para transferir os espectros das amostras de transferência do instrumento secundário de forma a fazê-los parecidos aos espectros do instrumento primário e para isso utiliza todo o espectro. A Equação 4 mostra que é preciso obter uma transformação linear  $F$ , o que é feito projetando as matrizes espectrais por PLS e então os escores ( $U$ ) calculados são empregados na Equação 5 (HONORATO et al., 2007). Qualquer espectro secundário externo pode então ser transformado pela Equação 4 e aplicado ao modelo original.

$$X_{prim} = X_{sec}F \quad \text{Eq. 4}$$

$$F = (U_{sec})^{-1} \cdot U_{prim} \quad \text{Eq. 5}$$

O método PDS, também definido por WANG, VELTKAMP e KOWALSKI (1991), funciona de maneira similar ao DS, porém a transferência dos espectros é feita por janelas. A matriz  $F$  é formada pela concatenação de vetores estimados pelo cálculo de um fator de regressão por PLS para cada janela (HONORATO et al., 2007). São criados modelos lineares locais que relacionam a resposta do instrumento a ser padronizado ao longo de um intervalo de frequências às respostas do instrumento primário a uma dada frequência. (FEUDALE et al., 2002) Devido ao caráter local deste método, é possível a correção simultânea de diferenças de intensidade, número de onda, deslocamentos e alargamento de picos. (FEUDALE et al., 2002)

Em um cenário ideal, os métodos SBC, DS e PDS utilizam uma quantidade pequena de amostras para transferência de calibração, pois os espectros destas amostras precisam ser obtidos em ambos os instrumentos primário e secundário. Isto pode gerar dificuldades em construir um modelo representativo caso essas amostras sejam muito diferentes entre si (FEUDALE et al., 2002), por isso a etapa de seleção de amostras de transferência é crucial para garantir boa capacidade preditiva.

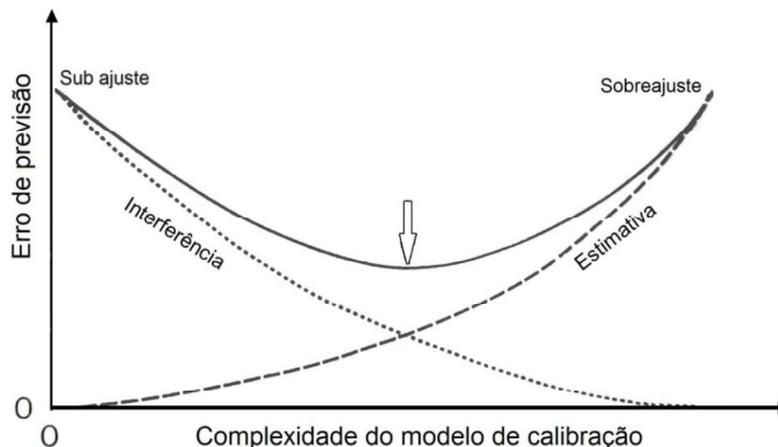
Honorato et al. (2007) fizeram um estudo de caso com amostras de milho usando, inclusive, modelos de transferência DS e PDS por NIR (HONORATO et al.,

2007), enquanto Liu, Han e Yang (2011) fizeram transferências de espectros de NIR usando divesas técnicas, como SBC, DS e PDS, destacando diferenças de capacidade preditiva dependendo dos tipos de instrumentos utilizados. Grelet et al. (2015) aplicaram PDS em espectros MIR de amostras de leite a fim de reduzir a variabilidade espectral entre instrumentos permitindo a aplicação de modelos consolidados por meio da técnica de transferência. A literatura acerca da aplicação de transferência de calibração utilizando espectros MIR é muito escassa, por isso ressalta-se a importância de elucidar a aplicabilidade das técnicas em dados desta natureza.

### **3.5 Avaliação de modelos quimiométricos**

O método PLS é linear nas variáveis de previsão mas não linear nas observações, sendo assim ao invés de minimizar a soma dos quadrados dos resíduos é utilizado um estimador não linear, denominado coeficiente **b**. (VAN DER VOET, 1999) A redução dos dados é feita por meio da regressão em variáveis latentes (VL) que representam a maior variância possível.

A definição do número ótimo de VL é necessária para evitar sub ajuste ou sobreajuste do modelo. Se forem usados menos fatores, a relação entre os valores previstos e medidos é insatisfatória, pois faltam componentes ou fontes de variações sistemáticas, isto é, há um sub ajuste do modelo. Por outro lado, se o modelo tiver excesso de fatores, o ajuste da curva com relação aos dados de calibração é muito bom, no entanto as previsões se tornam ruins, pois houve adição de ruído instrumental ao modelo de calibração, um sobreajuste. (MARTENS, 1991) A Figura 2 representa as duas principais contribuições para o erro de previsão, interferência e estimativa, indicando a solução ótima de complexidade do modelo para que não haja sub ajuste ou sobreajuste.



**Figura 2.** Ilustração do conceito de erro de previsão em função da complexidade do modelo de calibração. A seta indica o ponto ótimo. (Adaptado de Martens, 1991).

A habilidade de previsão de um modelo para uma amostra desconhecida ou conjunto de amostras externas está relacionada ao erro de previsão que deve, evidentemente, ser minimizado para obtenção de resultados com maior exatidão. Em geral, é utilizada a raiz quadrada do erro quadrático médio de predição (RMSEP – *root mean square error of prediction*) por que este é dado na mesma unidade da variável independente  $y$ , calculado pela Equação 6. (FILGUEIRAS et al., 2014a)

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n_{prev}}} \quad \text{Eq. 6}$$

onde  $y_i$  é o valor de referência da  $i$ -ésima variável dependente,  $\hat{y}_i$  seu valor estimado e  $n_{prev}$  é o número de amostras de previsão.

Cálculo similar pode ser realizado com os dados de calibração, gerando o RMSEC (raiz quadrada do erro quadrático médio de calibração – *root mean square error of calibration*) atentando que perde-se graus de liberdade correspondentes às variáveis latentes e também ao centrar os dados na média (Equação 7).

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - VL - 1}} \quad \text{Eq. 7}$$

A estimativa do número ótimo de VL pode ser feita por validação interna, utilizando validação cruzada (CV – *cross validation*). A ideia por trás da CV *leave one out*, é repetir a calibração  $i$  vezes, tratando o  $i$ -ésimo termo como objeto a ser previsto, até que todos os objetos sejam tratados como objeto a ser previsto. Alternativamente um bloco de amostras pode ser utilizado ao invés de um único

objeto, procedimento denominado *k-fold*, mais apropriado para conjuntos de dados composto por um número grande de amostras que é menos variável do que a *leave-one-out* CV. (MEVIK e CEDERKVIST, 2004) Pode então ser determinada a raiz quadrada do erro quadrático médio de validação cruzada, RMSECV (*root mean square error of cross validation*, Equação 8), em que  $n_{cal}$  é o número de amostras no conjunto de calibração. Em relação ao número de variáveis latentes, conforme o número de VL aumenta o RMSECV diminui, até que se atinja um mínimo, a partir daí, o aumento de VL implicará em um aumento do RMSECV. (MARTENS, 1991) O valor ótimo de VL é então determinado quando as diferenças de RMSECV não são mais significativas, sendo assim um RMSECV pode ser menor com mais VL, no entanto considera-se que isto seria devido a ruído presente no espectro. Um modelo cujas amostras de calibração e de previsão sejam representativas, terá um valor de RMSEP similar ao RMSECV.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n_{cal}}} \quad \text{Eq. 8}$$

A presença de erros aleatórios e sistemáticos é comum a todo experimento e pode comprometer os resultados. Erros aleatórios em geral seguem uma distribuição normal centrada em zero. (TAYLOR, 1997) Estimativas são feitas para que as previsões tenham a exatidão desejada, por exemplo, neste estudo os resultados reportados têm um intervalo de confiança ( $IC = \hat{y}_i \mp lim$ ) de 95%. O limite de confiança ( $lim$ ) do IC é determinado conforme a norma ASTM E 1655-05 pela Equação 9.

$$lim = t_{\alpha,df} RMSEC \sqrt{(1 + h_i)} \quad \text{Eq. 9}$$

onde  $t$  é o valor da distribuição  $t$  de Student com  $df$  graus de liberdade e nível de significância  $\alpha$ , e  $h_i$  é o *leverage* para cada amostra, que representa a distância de cada amostra ao centro dos dados (distância de Mahalanobis). O *leverage* é calculado pela Equação 10.

$$h_i = t_{VL,i}^T (T_{VL}^T T_{VL})^{-1} t_{VL,i} \quad \text{Eq. 10}$$

onde  $t_{VL,i}^T$  é o vetor de escores transposto para a amostra  $i$  com  $VL$  variáveis latentes e  $T$  é a matriz de escores do modelo PLS para o conjunto de amostras. É adotado um valor limite de  $h_i = 3VL/n$ , sendo  $n$  o número de amostras no conjunto investigado, acima do qual o erro de predição da amostra é investigado e pode levar à eliminação dos dados por ser esta amostra anômala.

Se, entretanto, existirem erros sistemáticos significativos, o modelo apresenta uma viés (*bias*) e os erros não são mais centrados na distribuição gaussiana de erros. (TAYLOR, 1997) Um método bastante usado para detectar viés envolve o teste-*t* de Student bicaudal. Inicialmente o viés é calculado conforme a Equação 11 (FILGUEIRAS et al., 2014a).

$$bias = \frac{\sum_{i=1}^l y_i - \hat{y}_i}{n} \quad \text{Eq. 11}$$

Em seguida, é calculado o desvio padrão dos erros (SDE – *standard deviation of the errors*) por meio da Equação 12.

$$SDE = \frac{\sum_{i=1}^l (y_i - \hat{y}_i - bias)^2}{n - 1} \quad \text{Eq. 12}$$

A avaliação da presença de erros sistemáticos é feita testando as seguintes hipóteses:

- $H_0$ : viés = 0;
- $H_1$ : viés  $\neq$  0;

O valor de *t* para os dados no modelo é calculado conforme Equação 13:

$$t = \frac{|bias| \cdot \sqrt{n}}{SDE} \quad \text{Eq. 13}$$

O valor de *t* obtido deve ser menor do que um *t*-tabulado ( $t_{tab}$ ) calculado com *n*-1 graus de liberdade e com um valor de significância de 5% para que a hipótese nula seja aceita e ausência de erros sistemáticos significativos no modelo seja confirmada. (TAYLOR, 1997) Se *t* for maior do que o  $t_{tab}$ , a hipótese nula é válida e há presença de viés, portanto o modelo é rejeitado. A presença de erros sistemáticos pode ser devida a um sobreajuste ou sub ajuste do modelo, e neste caso é recomendado alterar o número de VL e reconstruir do modelo.

A forma de detectar tendências nos erros é por meio do teste não-paramétrico de permutação. (TAYLOR, 1997) Testes paramétricos partem do princípio que os dados do modelo seguem uma distribuição conhecida, enquanto os testes não-paramétricos não fazem suposição acerca da distribuição dos dados. (FILGUEIRAS et al., 2014a) O teste de permutação consiste em reordenar repetidamente os *n* valores da variável **y**, gerando novas variáveis de resposta devido à troca de posições. As novas variáveis **y** permutadas não devem ter associação com os resíduos de calibração. O teste de permutação aplicado neste trabalho faz uma

avaliação dos resíduos assumindo que estes são distribuídos segundo a função polinomial:

$$e_i = b_n y_i^n + b_{n-1} y_i^{n-1} + \dots + b_1 y_1 + b_0 \quad \text{Eq. 14}$$

A avaliação da presença de tendência nos resíduos de calibração é feita por meio da definição das hipóteses a serem testadas:

- $H_0$ : os resíduos  $e_i$  são independentes de  $y_1$ .
- $H_1$ : os resíduos  $e_i$  são dependentes de  $y_1$  conforme a Equação 15:

$$e_i = g(y_i) + \varepsilon_i \quad \text{Eq. 15}$$

onde  $\varepsilon_i$  é um erro aleatório independente e  $g(y_i)$  é uma função polinomial que pode modelar a relação entre resíduos e valores de referência. O resultado do teste de permutação é dado em um *p-valor*, que deve ser superior a 0,05 para que a hipótese nula seja aceita, o que indica ausência de erros de tendência nos resíduos de validação. FILGUEIRAS et al. (2014a) fizeram uma abordagem mais detalhada do algoritmo utilizado para verificação de tendências nos resíduos.

### 3.6 Comparação de exatidão de modelos de transferência

A fim de avaliar a eficiência dos modelos de transferência, é fundamental adotar um procedimento para comparar suas capacidades preditivas. Neste trabalho foi adotado um *t*-teste randômico para verificar a igualdade entre a exatidão de dois modelos multivariados. Desta forma, evita-se a conclusão injustificada da superioridade ou inferioridade de um método em relação a outro. (VAN DER VOET, 1994) O teste realizado neste trabalho analisa a distribuição dos erros de predição (RMSEP), por meio das seguintes etapas (FILGUEIRAS et al., 2014a):

- (i) Calcular a diferença média quadrática real dos resíduos dos modelos a serem comparados,  $\bar{d}^*$ ;
- (ii) Permutar aleatoriamente o vetor com os resíduos de um dos modelos mantendo a ordem dos resíduos do outro;
- (iii) Calcular as diferenças médias das permutações,  $\bar{d}_i$ ;
- (iv) Comparar  $\bar{d}^*$  com  $\bar{d}_i$ ;
- (v) Repetir as etapas (ii) a (iv)  $k$ -vezes.

A avaliação do teste é feita considerando as seguintes hipóteses acerca dos resíduos de previsão dos dois modelos (A e B) a serem comparados:

- $H_0: RMSEP_A^2 = RMSEP_B^2$ ;
- $H_1: RMSEP_A^2 \neq RMSEP_B^2$ ;

O  $p$ -valor do teste é determinado pela proporção do número de vezes em que  $\bar{d}^* > \bar{d}_i$ , sendo que sendo menor do que 0,05 (para 95% de confiança) não há evidência significativa para aceitar  $H_0$  e pode-se considerar que a exatidão dos modelos é a mesma. Por outro lado, os modelos são considerados estatisticamente iguais se o  $p$ -valor é superior ao nível de significância adotado com a aceitação da hipótese  $H_1$ .

A realização deste teste possibilita uma comparação justa entre modelos e evita conclusão equivocada sobre o desempenho de um modelo em reação ao outro.

### 3.7 Análise da sobreposição entre espectros das mesmas amostras

A fim de analisar a sobreposição entre espectros das mesmas amostras de espectros brutos obtidos em instrumento primário e espectros transferidos ou tratados obtidos em instrumento secundário, a Equação 16, uma modificação de uma proposta de AMIGO et al. (2008), foi aplicada.

$$fit(\%) = 100x \left( 1 - \sqrt{\frac{\sum (X_{prim,i} - X_{sec,i})^2}{\sum X_{prim,i}^2}} \right) \quad \text{Eq. 16}$$

onde  $X$  refere-se a espectros do instrumento primário (*prim*) e secundário (*sec*) e  $i$  varia até contemplar todas as amostras. O resultado é uma média das  $i$  amostras e indica que há maior semelhança entre espectros quanto mais próximo for de 100%.

## 4 AVALIAÇÃO DE MÉTODOS DE TRANSFERÊNCIA DE CALIBRAÇÃO

### 4.1 Introdução

A calibração multivariada associada à técnica de infravermelho é uma alternativa aos métodos tradicionais para determinação de parâmetros físico-químicos em petróleo. Entretanto, o modelo multivariado construído é aplicável somente para o instrumento no qual os espectros foram obtidos. Em caso de manutenção ou troca de instrumento, fatos que modificam significativamente os espectros obtidos no equipamento, deve-se realizar a transferência do modelo de calibração construído originalmente para a nova condição instrumental ou novo aparelho.

A transferência de calibração é uma opção para melhorar a aplicabilidade dos modelos, permitindo uso mais generalizado. Desta forma, um único modelo validado bastaria para previsão de propriedades físico-químicas e assim espectros em instrumentos diferentes daquele em que as medidas foram originalmente feitas, evitando a repetição de um grande número de medidas e o trabalho computacional envolvido em um novo processo de modelagem.

Conjuntos de espectros de infravermelho na região do médio (MIR) obtidos em dois instrumentos diferentes para várias amostras de petróleo foram estudados. Modelos de calibração multivariada foram investigados por PLS (mínimos quadrados parciais), a técnica mais usual, e também por OPLS (projeções ortogonais em estruturas latentes), que elimina informações ortogonais ao sistema. A técnica PLS foi também aplicada às mesmas amostras de um instrumento MIR secundário, o chamado modelo de recalibração total (FR) para permitir comparação mais justa das transferências de calibração.

Para possibilitar a previsão da densidade API por espectros secundários, foram analisados modelos de transferência por meio de técnicas de nesta área, a saber: SBC (correção de declive e viés), que corrige as respostas obtidas pela aplicação direta de espectros secundários a um modelo; DS (padronização direta), que modifica os espectros secundários de maneira a assemelhá-los aos espectros primários antes da aplicação no modelo original; e PDS (padronização direta por partes), similar ao DS porém padroniza por meio de janelas do espectro.

## 4.2 Objetivos

### 4.2.1 *Objetivo geral*

Avaliar modelos de transferência de calibração para predição de propriedades físico-químicas de petróleo bruto entre dois instrumentos de infravermelho na região média.

### 4.2.2 *Objetivos específicos*

- Analisar amostras de petróleo em dois diferentes instrumentos de infravermelho;
- Construir modelo de calibração multivariada para determinação de propriedades físico-químicas de petróleo utilizando os métodos de mínimos quadrados parciais (PLS) e projeções ortogonais por estruturas latentes (OPLS);
- Criar modelo de transferência de calibração usando correção de declive e viés (SBC), padronização direta (DS) e padronização direta por partes (PDS);
- Comparar a exatidão dos métodos que possibilitam a predição de propriedades por meio de espectros de MIR de um instrumento secundário;

## 4.3 Metodologia

Neste estudo, 96 amostras de petróleo bruto brasileiro foram disponibilizadas pelo Centro de Pesquisas Leopoldo Americo Miguez de Mello (CENPES/ Petrobras) bem como seus respectivos valores de densidade API. Esses dados foram associados aos espectros de infravermelho obtidos em dois instrumentos na região do médio (MIR).

Em um primeiro momento, modelos de calibração PLS e OPLS para predição de densidade API foram construídos e validados com dados de um instrumento primário. Além disso, um modelo de recalibração total (FR – full recalibration) por PLS, que consiste na remodelagem dos dados para o instrumento secundário.

As técnicas SBC, DS e PDS foram aplicadas para transferência e então a capacidade preditiva dos modelos foi comparada por meio de testes com base na distribuição de erros de predição, para verificação de viés e tendências nos resíduos dos modelos.

#### 4.3.1 *Caracterização físico-química*

Na indústria do petróleo as propriedades físico-químicas são determinadas conforme técnicas normatizadas usadas rotineiramente para caracterização das amostras de petróleo.

Para este estudo, a densidade API foi determinada pela injeção de cada amostra no densímetro digital automático Anton Paar modelo DMA 5000, e a densidade API foi calculada conforme a Equação 1 para 60°F seguindo a técnica padrão de análise para densidade API (ISO 12185-96).

#### 4.3.2 *Instrumentação*

Os espectros de infravermelho na região média foram obtidos em um instrumento denominado primário, localizado na Universidade Federal do Espírito Santo, para as 96 amostras. Essas mesmas análises foram realizadas em um instrumento MIR denominado secundário, localizado na Universidade Federal de Santa Maria. Ambos instrumentos são espectrômetros da marca Perkin Elmer com cristal de ZnSe e acessório de múltipla reflectância total atenuada - ATR-MIR. Devido à limitação imposta pelo cristal de ZnSe, a varredura foi feita entre 4000 e 650  $\text{cm}^{-1}$ . Os espectros foram lidos na região entre 4000 e 650  $\text{cm}^{-1}$  com 32 escaneamentos e 4  $\text{cm}^{-1}$  de resolução. A umidade relativa e a temperatura ambiente das salas foram mantidas em torno de 35% e 23°C, respectivamente. Amostras com densidade API inferior a 30 e amostras parafínicas foram aquecidas antes da análise.

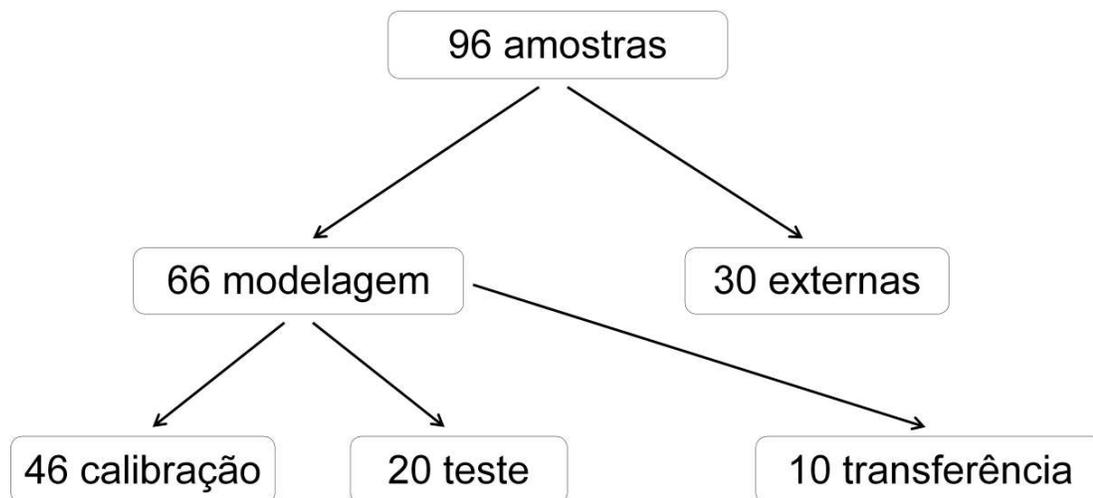
Variações da composição química e conseqüentemente das propriedades físico-químicas de amostras de petróleos são um fator delicado que precisa ser levado em conta. Para este trabalho, este risco foi minimizado por meio do fracionamento de amostras em cinco recipientes que foram mantidos sob condições adequadas de armazenamento. Naturalmente há perda de parte dos compostos voláteis no transporte entre a plataforma de produção e o laboratório, mas entre laboratórios foram evitadas perdas adicionais por meio da manutenção de condições favoráveis. Ademais, todas as amostras passaram por uma verificação por meio de teste de viscosidade em ambos os laboratórios.

### 4.3.3 Quimiometria

A análise quimiométrica consistiu na modelagem de densidade API por meio das técnicas PLS, OPLS, SBC, DS e PDS e comparação entre os valores previstos por cada técnica e também com os valores de referência. Os dados foram processados no software MATLAB versão 7.12.0 (MathWorks Inc., Natick, MA). Todos os espectros foram pré-processados pelos métodos de derivada com suavização de Savitzky–Golay (primeira derivada, polinômio de segundo grau e janela de 7 pontos) (SAVITZKY e GOLAY, 1964) ou variação normal padrão (SNV - *Standard Normal Variate*) (BLANCO et al., 1997).

Na construção dos modelos PLS e OPLS os dados foram centrados na média e o procedimento de validação cruzada usando o método *venetian blind 5-fold* foi aplicado para otimizar o número de variáveis latentes (VL) para cada modelo. (ASTM, 2012) Dessa forma, o número de VL a ser utilizado na construção do modelo foi determinado pelo gráfico da taxa de erro de validação cruzada em função do número de variáveis latentes, ou seja, foi selecionado aquele com o número de variáveis latentes correspondente ao valor onde as diferenças de RMSECV não são mais significativas com relação à taxa de erro.

As 96 amostras com dados espectrais disponíveis de ambos os instrumentos de MIR e densidade API foram separadas para modelagem, conforme esquema mostrado na Figura 3. O conjunto original foi separado em conjuntos externo e de modelagem, com 30 e 66 amostras respectivamente por meio do algoritmo de Kennard-Stone. (KENNARD e STONE, 1969). Este conjunto externo tem como objetivo possibilitar a comparação entre os diferentes modelos. Em seguida, as amostras do conjunto de modelagem foi separado em conjuntos de calibração (46 amostras) e teste (20 amostras para validação externa) também utilizando o método proposto por Kennard-Stone. Para possibilitar a transferência de calibração, 10 amostras representativas da modelagem foram selecionadas.



**Figura 3.** Esquema demonstrativo da separação do conjunto de dados de petróleo bruto.

Os resultados dos modelos foram avaliados quanto a RMSEC, RMSEP de amostras de teste e externas e os coeficientes de determinação ( $R^2$ ) de calibração, teste e externo. Para avaliar os modelos foram aplicados dois testes estatísticos. O teste de vies (*bias*) foi utilizado para verificar ausência de erros sistemáticos e o teste de permutação para avaliar tendência nos resíduos do modelo. (FILGUEIRAS et al., 2014a) Os algoritmos utilizados se baseiam na ASTM E1655 e foram aplicados com um nível de significância de 0,05.

#### a) Modelos de predição de densidade API

A propriedade escolhida para modelagem foi a densidade API, devido a sua importância para a indústria do petróleo (SPEIGHT, 2001), já que é fator determinante na valoração de petróleo extraído de determinado produtor, sendo fundamental para a tomada de decisões comerciais. Além disso, a densidade API já foi modelada previamente, por exemplo no trabalho de Filgueiras e colaboradores (2014).

Um modelo PLS foi construído para previsão de densidade API a partir dos espectros MIR do instrumento primário e denominado modelo original. Este modelo foi utilizado como referência para avaliar a acurácia das predições realizadas pelos espectros do instrumento secundário.

Uma alternativa para possibilitar a previsão de uma propriedade de interesse por espectros secundários a um modelo original é construir um modelo robusto em que possa ser aplicado diretamente, sem etapa de transferência, por meio da técnica

OPLS. (SOUHI et al., 2015) Esta técnica inclui uma etapa que filtra toda variação estruturada que não tem correlação (ortogonal) com a propriedade de interesse. (SOUHI et al., 2015)

O modelo de recalibração total (FR) foi construído com espectros do instrumento secundário utilizando todas as amostras conforme o modelo original, usando regressão PLS e o mesmo número de VL. Apesar da necessidade de uma grande quantidade de medidas no instrumento secundário, esta remodelagem é importante para avaliar a capacidade preditiva do instrumento secundário em comparação ao modelo primário.

#### b) *Transferência de calibração*

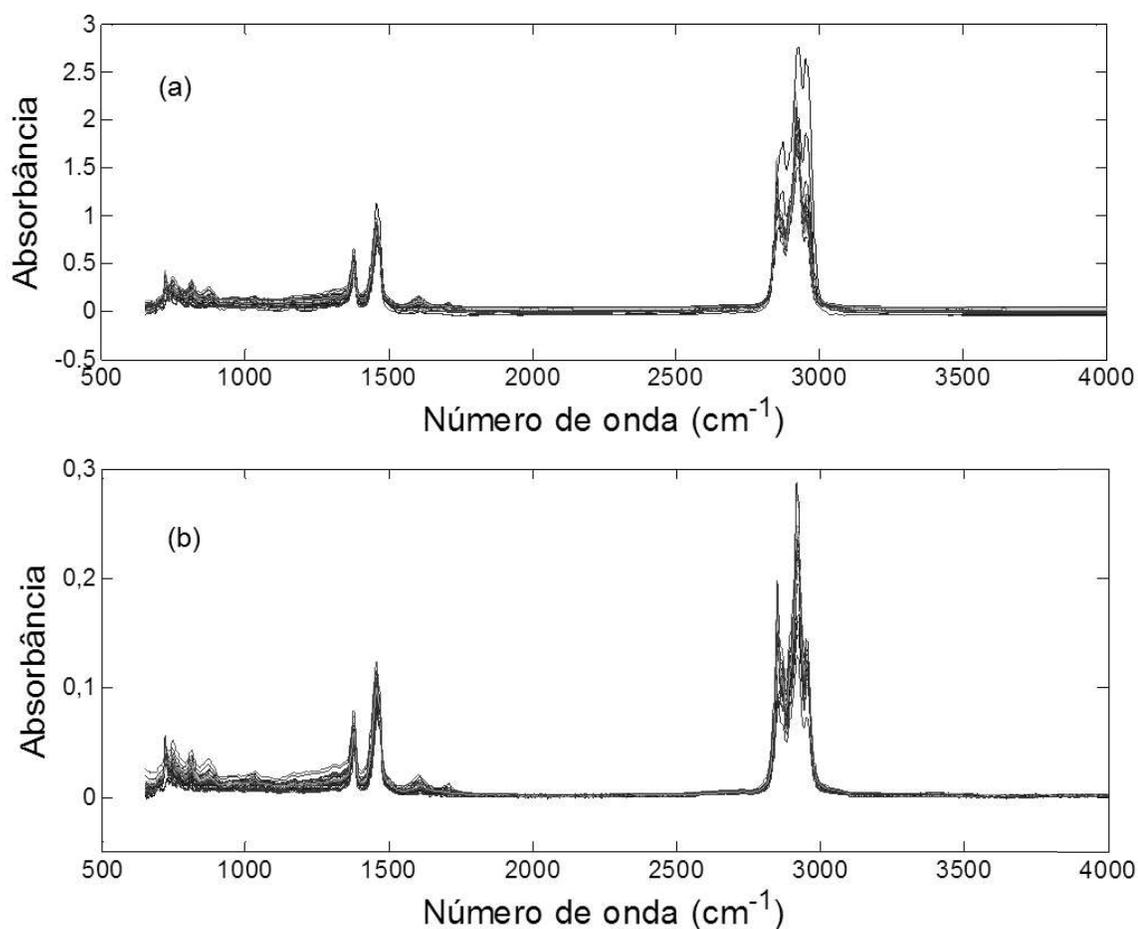
Para aplicação dos métodos SBC, DS e PDS, uma etapa crucial é a seleção das amostras de transferência, que foi realizado por meio do algoritmo de Kennard-Stone (KENNARD e STONE, 1969). As primeiras amostras selecionadas pelo algoritmo de Kennard-Stone foram usadas somente no processo de modelagem, enquanto as dez últimas amostras, ou seja, aquelas com espectros mais similares ao espectro médio, formaram o conjunto de transferência aplicado nas três técnicas de transferência investigadas.

Neste estudo as técnicas SBC, DS e PDS foram associadas ao modelo de calibração PLS e a técnica PDS foi ainda associada ao modelo de calibração OPLS. Para comparar a acurácia dos modelos foi utilizado um *t*-teste estatístico randômico para avaliar se a distribuição de erros quadrados é equivalente (VAN DER VOET, 1994).

Foram feitos modelos qualitativos PCA (*principal component analysis*, análise por componentes principais) para verificar a presença de grupos entre amostras obtidas em um mesmo instrumento. Este procedimento foi feito tanto antes, quanto após procedimentos de transferência para inferir sobre a capacidade de distinção de espectros obtidos em diferentes instrumentos. Adicionalmente foi aplicada uma equação para verificar a superposição média entre espectros brutos e após pré-processamentos e transferências, conforme descrito pela Equação 16.

#### 4.4 Resultados e discussão

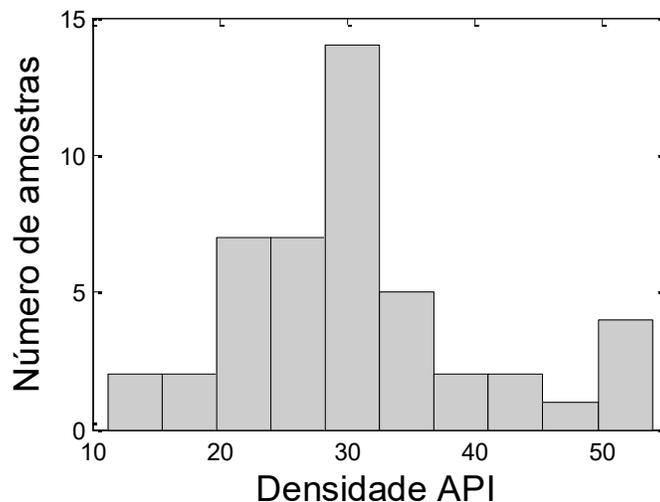
A Figura 4 mostra espectros brutos de ATR-MIR para amostras de petróleo tomados nos instrumentos primário e secundário. O perfil espectral é notavelmente similar, no entanto o espectro do instrumento primário apresenta maior variação da linha de base em relação ao instrumento secundário. Ademais, a intensidade das absorvâncias difere grandemente, por um fator de cerca de dez vezes, entre espectros de instrumentos diferentes. Tanto variação de linha de base quanto intensidade das bandas de absorção são características intrínsecas de cada espectrômetro e são passíveis de correção durante as etapas de pré-processamento e modelagem.



**Figura 4.** Espectros de MIR para amostras de petróleo bruto obtidos nos instrumentos (a) primário e (b) secundário.

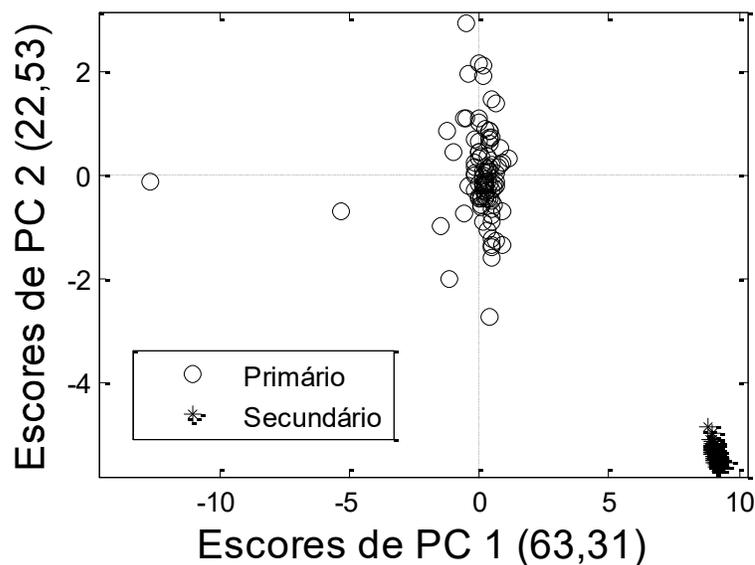
As amostras de calibração compreendem desde petróleos muito pesados (°API 11,2) até petróleos muito leves (°API 54), com distribuição mostrada na Figura 5. Modelos multivariados construídos com estes dados poderão ser aplicados a

amostras de petróleo com valor de densidade API entre os valores mínimo e máximo, ou seja, após validação o modelo será capaz de estimar a densidade API de óleos muito diferentes entre si.



**Figura 5.** Histograma de densidade API para amostras de petróleo utilizadas em modelo de calibração.

A Figura 6 contém o gráfico de escores para as duas primeiras componentes principais de um modelo qualitativo PCA onde nota-se clara distinção entre espectros dos dois instrumentos. Este modelo explica mais de 75% da variância dos dados, o que indica a impossibilidade de aplicar espectros secundários diretamente ao modelo original PLS construído e validado para espectros obtidos no instrumento MIR primário.



**Figura 6.** Modelo PCA para espectros brutos dos instrumentos primário e secundário pré-processados por derivada com suavização.

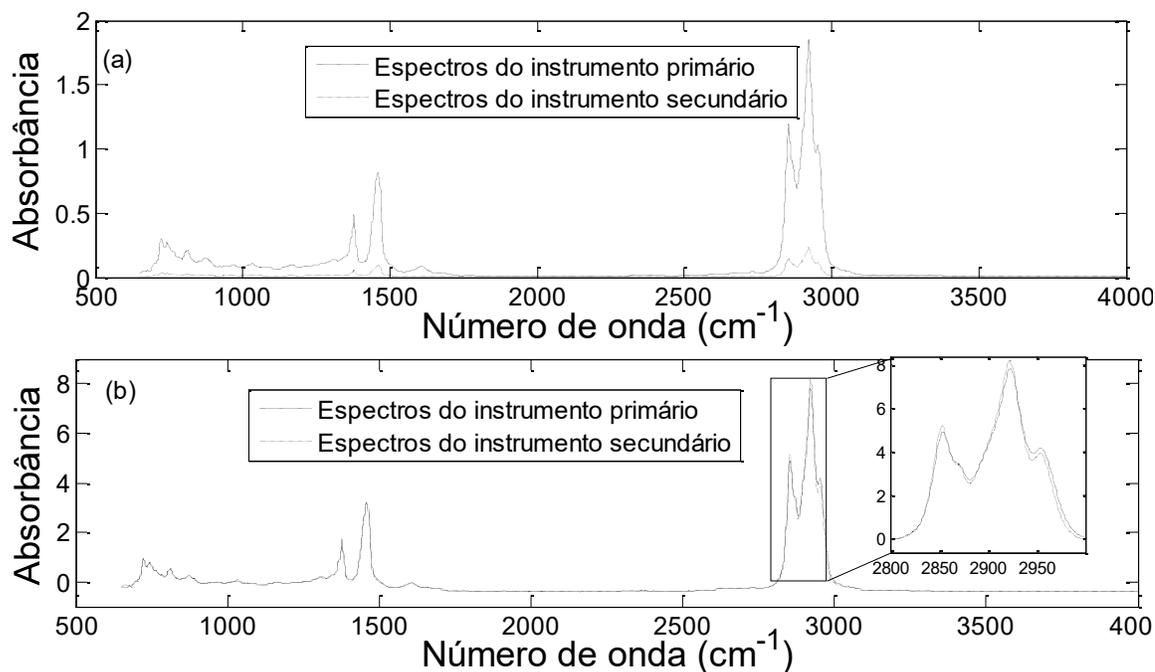
O modelo de calibração original foi construído por PLS por meio dos espectros do instrumento primário pré-tratados por derivada com suavização Savitzky-Golay. (SAVITZKY e GOLAY, 1964) Os parâmetros do modelo estão descritos na Tabela 3 e mostram que espectros de MIR do instrumento primário podem ser usados para estimar com boa exatidão a densidade API de amostras de petróleo bruto. Em especial destaque está o valor de RMSEP tanto para amostras dos conjuntos de teste e externo, respectivamente 1,6 e 1,02, em unidade de densidade API. Esses resultados acentuam a boa capacidade de predição do modelo, uma vez que a indústria de petróleo aceita valores de densidade API com incerteza de uma unidade, e é natural que modelos estatísticos tenham maior erro, uma vez que se acumulam aos erros associados à prática analítica os erros inerentes ao processo de modelagem. Os valores de coeficiente de determinação ( $R^2$ ) foram maiores do que 0,90 para todos os conjuntos de dados, o que indica comportamento linear.

**Tabela 3.** Parâmetros do modelo original PLS de predição de densidade API a partir de espectros MIR obtidos no instrumento primário.

<b>Parâmetro</b>	<b>Resultado</b>
<b>Variáveis latentes</b>	4
<b>Variância explicada em X</b>	96,37 %
<b>Variância explicada em y</b>	93,94 %
<b>RMSEC</b>	2,60
<b>RMSEP (teste)</b>	1,60
<b>RMSEP (externo)</b>	1,02
<b><math>R^2c</math></b>	0,9394
<b><math>R^2p</math> (teste)</b>	0,9051
<b><math>R^2p</math> (externo)</b>	0,9339

A Figura 7 mostra os dois conjuntos de espectros antes e após o pré-tratamento SNV, que gera sobreposição quase exata dos espectros secundários sobre os primários, ao contrário dos espectros brutos (Figura 4). Isto ocorre devido ao fato de após ser centrado na média, cada espectro individual é escalado com relação a seu desvio padrão. (BLANCO et al., 1997) A aplicação de espectros pré-tratados por SNV em modelo PLS provocou erro de tendência nos resíduos, por isso este procedimento foi descartado. Por outro lado, para modelagem por OPLS está pré-

tratamento gerou os melhores resultados sem que haja erros sistemáticos ou tendenciosos.



**Figura 7.** Espectros de MIR dos instrumentos primário e secundário (a) brutos e (b) pré-tratados por SNV para modelagem por OPLS.

Os resultados obtidos para o modelo OPLS, para previsão das amostras externas tanto por meio de espectros primários quanto pelos espectros secundários, mostram que é obtida melhor previsão (menor RMSEP) ao utilizar dados do instrumento primário (Tabela 4). Teste estatístico de comparação de acurácia permitem concluir que a igual acurácia com um p-valor superior a 0,05, e neste caso foi provado que as previsões das amostras externas possuem acurácias iguais entre si (p-valor=0,18), mas diferente do modelo original (p-valor= $9,0 \times 10^{-5}$ ), ou seja este modelo irá estimar a densidade API de uma mesma amostra com exatidão inferior à do modelo PLS mostrado anteriormente, tanto para espectros primários quanto secundários.

**Tabela 4.** Parâmetros do modelo OPLS de predição de densidade API a partir de espectros MIR obtidos nos instrumentos primário e secundário.

<b>Parâmetro</b>	<b>OPLS</b>
<b>Variáveis latentes</b>	4
<b>RMSEC</b>	1,61
<b>RMSEP (teste)</b>	1,15
<b>RMSEP (ext. prim.)</b>	1,53
<b>RMSEP (ext. sec.)</b>	1,98
<b>R<sup>2</sup>c</b>	0,9740
<b>R<sup>2</sup>p (teste)</b>	0,9501
<b>R<sup>2</sup>p (ext. prim.)</b>	0,8294
<b>R<sup>2</sup>p (ext. sec.)</b>	0,9179

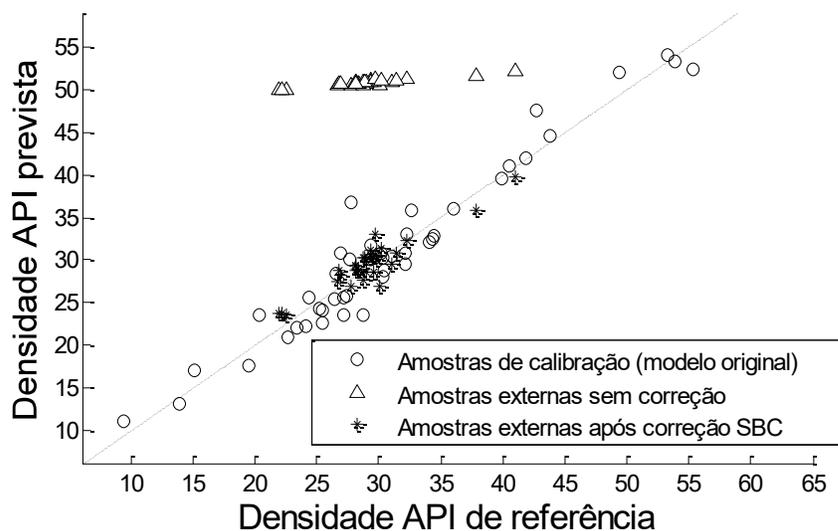
O procedimento PLS feito para o modelo primário foi também realizado com espectros MIR obtidos no instrumento secundário, gerando o modelo FR, que consiste na reconstrução do modelo, e demanda que todas as amostras sejam analisadas em ambos equipamentos. Este procedimento foi adotado a fim de comparar plenamente os modelos de transferência, de acordo com a capacidade de previsão associada ao instrumento secundário. A Tabela 5 mostra que houve piora na exatidão em relação ao modelo original. Entretanto, o teste estatístico de acurácia provou que a exatidão é igual para amostras de teste (p-valor=0,75), porém diferente para externas (p-valor<0,001) e igual ao modelo OPLS (p-valor=0,12).

**Tabela 5.** Resultados do modelo FR de predição de densidade API a partir de espectros MIR obtidos no instrumento secundário.

<b>Parâmetro</b>	<b>Modelo FR</b>
<b>Variáveis latentes</b>	4
<b>Variância explicada em X</b>	95,24
<b>Variância explicada em y</b>	96,05
<b>RMSEC</b>	2,10
<b>RMSEP (teste)</b>	1,68
<b>RMSEP (externo)</b>	1,53
<b>R<sup>2</sup>c</b>	0,9605
<b>R<sup>2</sup>p (teste)</b>	0,9053
<b>R<sup>2</sup>p (externo)</b>	0,8586

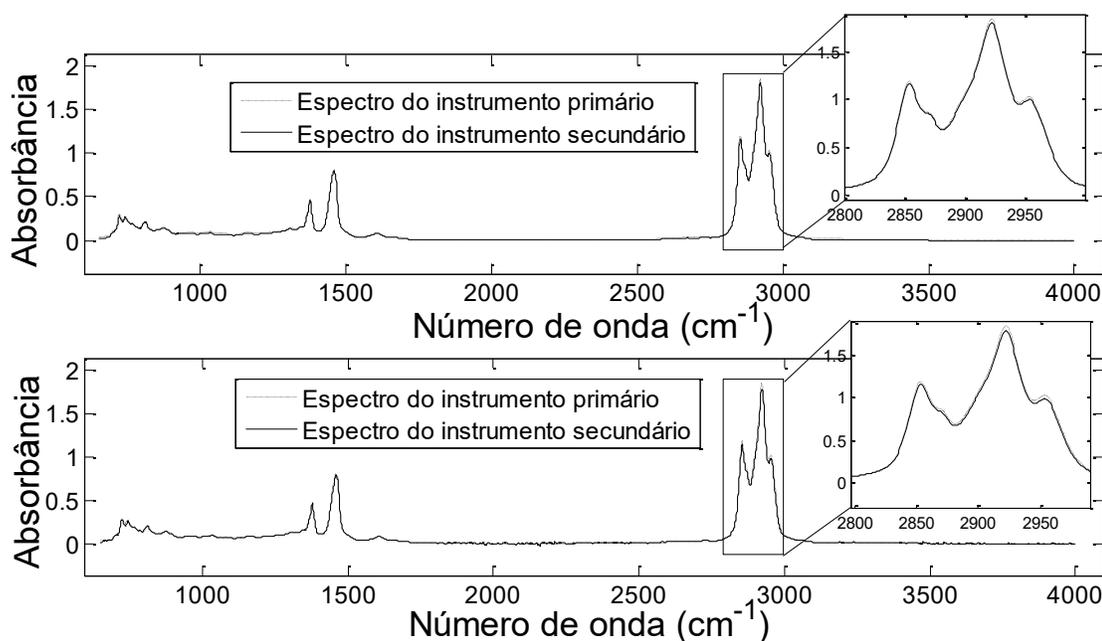
Bouveresse e colaboradores (1996) afirmam que este o método de Kennard-Stone (KENNARD e STONE, 1969) é um procedimento satisfatório na área de transferência de calibração para selecionar amostras representativas para os métodos DS e PDS, que corrigem diferenças entre espectros, no entanto não é interessante para o método SBC, que corrige os valores preditos. Neste trabalho, as dez últimas amostras selecionadas pelo método de Kennard-Stone, ou seja, aquelas com espectros mais similares ao espectro médio, compuseram o conjunto de transferência aplicado nas três técnicas de transferência investigadas. As amostras de transferência são representativas de toda a gama de valores da propriedade de interesse do conjunto de dados original disponível, garantindo que o modelo possa ser aplicado à maioria das amostras que necessitem ser estimadas.

Devido à semelhança entre os espectros de ambos os instrumentos, aplicou-se diretamente os espectros do instrumento secundário ao modelo original e observou-se perfil linear de previsão, no entanto com valores discrepantes. Esses dados foram então corrigidos e os parâmetros de correção originaram o modelo SBC, que nas amostras externas apresentou um RMSEP 1,46 e um  $R^2_p$  0,8614. A Figura 8 mostra os valores previstos das amostras externas em relação aos valores de referência antes e após SBC. Visualmente os resultados se assemelham aos dados de calibração do modelo PLS original, porém o teste estatístico acusou diferença de acurácia em relação ao modelo original ( $p$ -valor=0,01), indicando que o modelo SBC tem menor capacidade preditiva, e igual ao modelo FR ( $p$ -valor=0,38), portanto tem a mesma capacidade preditiva.



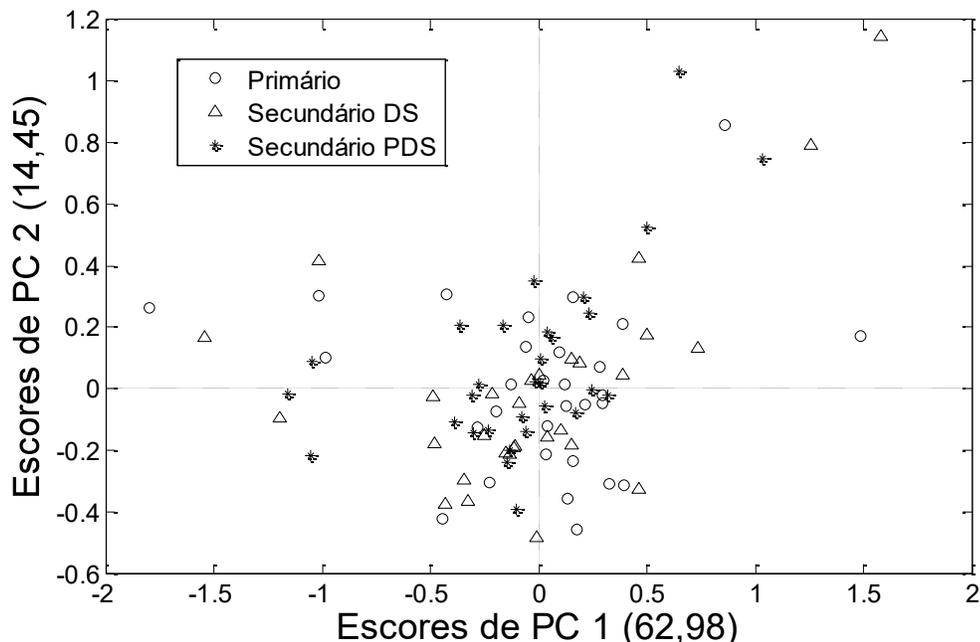
**Figura 8.** Valores previstos com relação aos valores medidos de densidade API para amostras externas pelos espectros de MIR do instrumento secundário antes e após SBC.

Métodos amplamente presentes na literatura, DS e PDS, são usados para transferir os espectros secundários de forma que se assemelhem aos espectros primários. Desta forma, os espectros transferidos por essas técnicas podem ser aplicados ao modelo PLS originalmente construídos para o instrumento primário. As amostras de transferência são utilizadas para calibração e o modelo é então aplicado às amostras externas, sendo que no DS o algoritmo transfere o espectro como um todo e no PDS são utilizadas janelas de variáveis para a transferência, conforme Equação 4. Após a realização de ambos os processos de transferências, DS e PDS, ocorreu boa sobreposição dos espectros secundários aos primários, conforme pode se visto na Figura 9.



**Figura 9.** Espectro de uma amostra externa dos instrumentos primário e secundário (a) após DS, e (b) após PDS com janela de 5 pontos.

O modelo qualitativo não supervisionado PCA demonstra que a distinção de espectros de diferentes instrumentos antes notória (Figura 6) torna-se inexistente após transferência DS ou PDS (Figura 10), indicando que os espectros secundários foram satisfatoriamente transformados e se comportam como se tivessem sido obtidos no instrumento primário.



**Figura 10.** Modelo PCA após transferências DS e PDS.

Os espectros das amostras externas medidas no instrumento secundário e transferidas por DS e PDS foram aplicadas no modelo de calibração original. Os resultados mostrados na Tabela 6 indicam que a exatidão do modelo PDS é melhor do que o modelo DS e ambos apresentam coeficiente de determinação acima de 0,8. Ambos os modelos DS e PDS têm acurácia diferente do modelo original (p-valor=0,02 e p-valor= $7,0 \times 10^{-5}$ , respectivamente) e igual ao FR (p-valor=0,28 e p-valor=0,40, respectivamente).

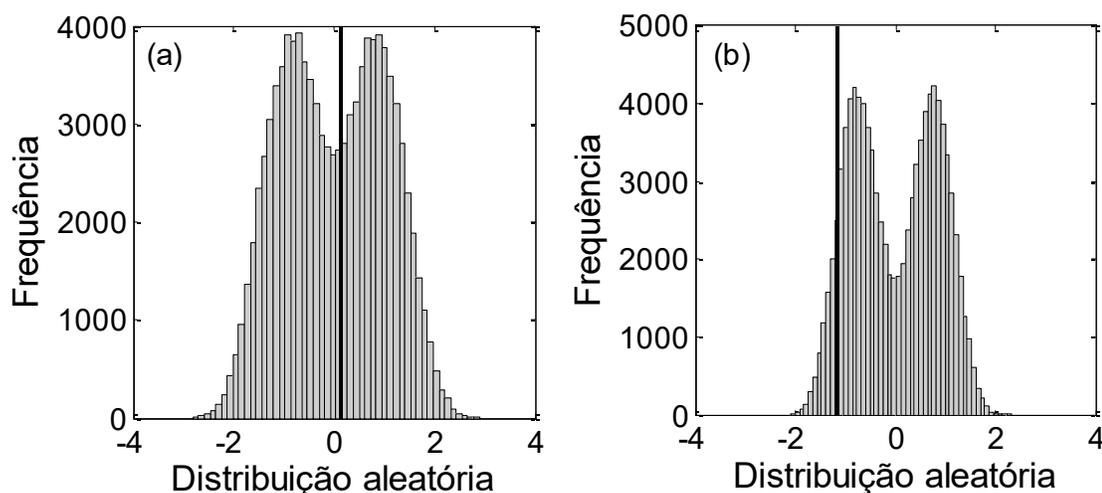
**Tabela 6.** Resultados dos modelos de transferência DS, PDS e PDS-OPLS de predição de densidade API a partir de espectros MIR obtidos no instrumento secundário.

	DS	PDS	PDS-OPLS
RMSEP	1,6914	1,6085	1,4857
R <sup>2</sup> <sub>p</sub>	0,8396	0,8291	0,8459

A última estratégia adotada neste estudo foi aplicar espectros secundários transferidos por PDS ao modelo OPLS, em uma tentativa de obter resultados com acurácia igual ao modelo original. Ao aplicar estes espectros ao modelo OPLS, que é mais robusto, é provável que se obtenha melhor exatidão em relação à aplicação direta dos espectros secundários. De fato, o modelo PDS-OPLS, também mostrado na Tabela 6, representou menor valor de RMSEP para as amostras externas do

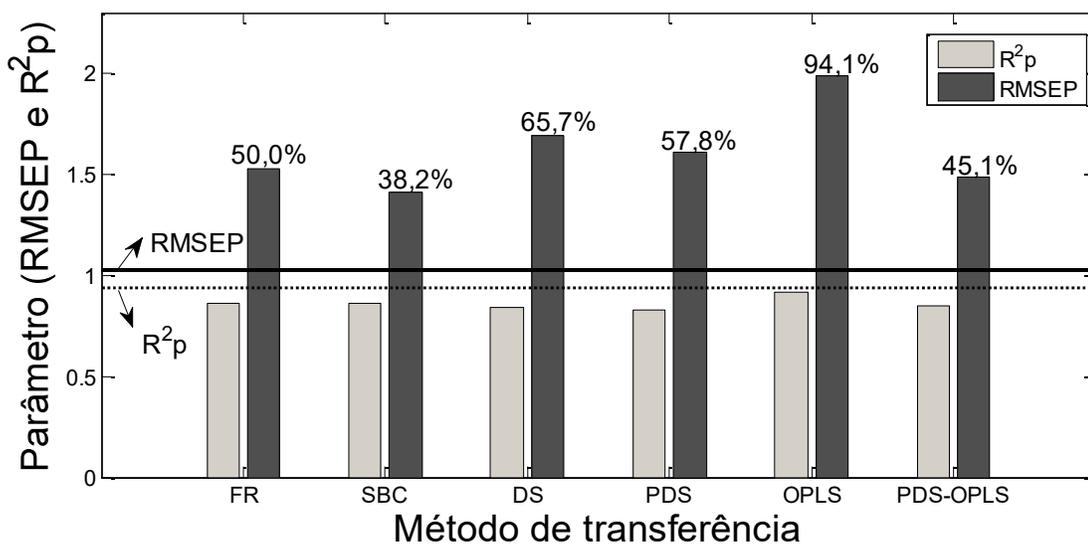
instrumento secundário, indicando ser melhor proceder uma transferência PDS antes de aplicar no modelo OPLS.

Ao contrário dos modelos FR, SBC, DS, PDS e OPLS, o modelo PDS-OPLS apresentou acurácia igual ao modelo original ( $p$ -valor=0,16), o que prova mesma capacidade de prever a densidade API utilizando espectros do instrumento secundário, conforme mostrado na Figura 11a. A Figura 11b mostra que a acurácia do modelo PDS-OPLS também é estatisticamente igual ( $p$ -valor=0,94) à do modelo FR.



**Figura 11.** Resultado do teste de acurácia da previsão de amostras externas por meio de espectros de instrumento secundário pelo modelo PDS-OPLS (a) comparado ao modelo original e (b) comparado ao modelo FR.

A Figura 12 mostra aplicação de espectros secundários no modelo de recalibração e em cada modelo de transferência, sendo que as linhas de referência correspondem às respostas do modelo PLS original, prevendo a densidade API por meio de espectros do instrumento primário. Observa-se que o método SBC foi o que gerou menor RMSEP e maior  $R^2_p$ , no entanto os valores previstos apresentaram diferença estatística na acurácia com relação ao modelo original. O segundo destaque é o modelo PDS-OPLS, este sim com acurácia estatisticamente igual ao original, sendo melhor para prever a propriedade de interesse por meio de espectros do instrumento secundário. No caso de modelos PLS já construídos e validados, é interessante a possibilidade de aplicar espectros secundários transferidos por PDS ou ainda realizar correção das respostas obtidas com aplicação direta de espectros secundários por meio de SBC, com ligeiro aumento no valor de RMSEP.



**Figura 12.** Resultados dos modelos de aplicação de espectros de um instrumento secundário para prever densidade API de amostras de petróleo bruto. As linhas horizontais representam os valores referentes ao modelo PLS original e os valores percentuais são o aumento de RMSEP.

São reportados na Tabela 7 os valores médios obtidos conforme Equação 16 na comparação entre espectros secundários brutos ou após os procedimentos realizados neste trabalho, sempre em relação aos espectros originais das mesmas amostras medidas no instrumento primário. Os espectros brutos variam tão grandemente que o resultado obtido foi -229,72%, isto porque conforme visto na Figura 4 até mesmo a intensidade das absorvâncias apresenta valores muito discrepantes. Os espectros pré-tratados por SNV apresentaram similaridade superior a 75%, indicando que o procedimento torna o conjunto de dados mais parecido. A transferência de calibração por DS mostrou uma similaridade mais elevada (80,62%) do que a aplicação do SNV, e a transferência por PDS aumentou ainda mais a semelhança entre espectros secundários e primários, com porcentagem superior a 82%. No entanto, em concordância com os testes estatísticos que indicaram ser melhor o modelo PDS-OPLS, os espectros transferidos por PDS e pré-tratados por SNV apresentaram melhor similaridade, 91%, em relação aos espectros brutos.

**Tabela 7.** Porcentagem média de similaridade entre amostras externas do instrumento secundário submetidas a diversos procedimentos em relação aos dados do instrumento primário.

<b>Procedimento</b>	<b>Similaridade (%)</b>
Nenhum	-229,73
SNV	77,30
DS	80,62
PDS	82,64
PDS + SNV	91,00

#### **4.4 Conclusões**

Os métodos quimiométricos SBC, FR, DS e PDS possibilitam a aplicação de espectros MIR obtidos em um instrumento secundário em modelos de calibração construídos para um instrumento primário. Em especial, para prever densidade API de amostras de petróleo bruto, o modelo de transferência PDS aplicado a um modelo original OPLS resulta em excelente capacidade preditiva (RMSEP de 1,48) em relação aos modelos construídos neste trabalho, com aumento percentual de 45% no valor de RMSEP porém exatidão estatisticamente igual ao modelo original ( $p$ -valor=0,16), construído por PLS para dados do instrumento primário.

A aplicação de dados de um instrumento secundário diretamente a um modelo OPLS para instrumento primário, embora gere aumento de RMSEP (1,98, aumento de 94%) e diferença de exatidão em relação ao modelo de referência, também é uma alternativa interessante por dispensar a necessidade de amostras de transferência.

No caso de necessidade de manter um modelo PLS original, é possível ainda aplicar amostras secundárias transferidas por PDS diretamente, neste caso com um aumento de 58% do RMSEP. Alternativamente, é possível manter o modelo PLS original, aplicar diretamente espectros secundários e corrigir os valores previstos por meio do método SBC.

## 5 TRANSFERÊNCIA DE CALIBRAÇÃO PARA MODELOS PLS

### 5.1 Introdução

A fim de atender à demanda de permitir a aplicação de espectros secundários a modelos PLS construídos para um instrumento primário, foi proposta a construção de modelos PLS para as propriedades API, TIAC, NAT e NAN. Após avaliação e validação destes modelos, seria possível testar a eficácia do método de transferência PDS.

A temperatura de início de aparecimento de cristais, TIAC, é a temperatura em graus célsius em que os primeiros cristais de parafina deixam de fazer parte da solução e provocam mudanças no comportamento reológico do petróleo. Este tipo de informação é muito útil na previsão do processo de deposição orgânica, uma vez que permite concluir se um petróleo apresenta tendência à formação de depósitos orgânicos.

O teor de acidez de uma amostra de petróleo é determinado pelo método titulométrico chamado número de acidez total (NAT), que é expresso em miligramas de hidróxido de potássio suficientes para neutralizar a acidez de um grama de petróleo (PARISOTTO, 2007). Este método pode ser um indicativo do grau de corrosividade da amostra, mas não traz informação sobre a qualidade e natureza dos ácidos (GRUBER et al., 2012).

Os ácidos naftênicos apresentam corrosividade elevada, sendo um problema para refino e craqueamento de petróleo. Óleos com acidez elevada, isto é, com valores superiores a 0,5 mg KOH/g, podem corroer linhas de transferência, colunas e condensadores da destilação (GRUBER et al., 2012). A estrutura química de um ácido naftênico é a de um ácido carboxílico contendo anéis fundidos que apresentam saturação. Este termo vem adicionalmente aplicado para designar outros ácidos orgânicos, tais quais aqueles que possuem anéis aromáticos, cadeias lineares e ramificadas (COLATI, 2012; GRUBER et al., 2012).

Por outro lado, o número de acidez naftênica (NAN) é obtido por um método mais específico, porém sem normatização, que consiste em uma análise feita por MIR dos ácidos naftênicos obtidos após extração em fase sólida. Esta etapa de extração é dificultada pela alta complexidade do petróleo (GRUBER et al., 2012).

## 5.2 Objetivos

### 5.2.1 *Objetivo geral*

Avaliar a eficiência da predição de propriedades físico-químicas de petróleo bruto em um instrumento MIR secundário por meio de modelos PLS validados para um instrumentos primário.

### 5.2.2 *Objetivos específicos*

- Analisar amostras de petróleo em dois diferentes instrumentos de infravermelho;
- Construir modelos de calibração multivariada por PLS para determinação de propriedades físico-químicas de petróleo;
- Criar modelos de transferência de calibração usando a técnica PDS;
- Verificar a qualidade das predições realizadas após a transferência de calibração;

## 5.3 Metodologia

Foi realizada a determinação de diversas propriedades físico-químicas petróleo bruto por meio de dados de aproximadamente 160 amostras de óleo representativo da produção brasileira. Em especial, foram estudadas as propriedades API, TIAC, NAT e NAN, importantes na indústria do petróleo quanto à valoração do óleo e comportamento durante as etapas envolvidas em seu processamento. Essas propriedades foram inicialmente relacionadas aos espectros de MIR obtidos em um instrumento primário por meio de modelagem PLS.

Antes da modelagem, dez amostras foram excluídas do conjunto para comporem um conjunto externo para validar a transferência de calibração que seria feita. Essas dez amostras, bem como outras dez selecionadas para o procedimento de transferência espectral, foram medidas em um instrumento MIR secundário. Não houve necessidade de obter o espectro das outras 140 amostras no instrumento secundário, uma vez que estas foram utilizadas somente para calibração do modelo.

Os espectros obtidos nos instrumentos primário e secundário apresentam diferença na quantidade de variáveis bem como no valor destas. Por isso, foi necessário incluir uma etapa de interpolação polinomial antes da transferência. Esta etapa consistiu na identificação dos 5 pontos mais próximos de cada variável do

instrumento primário, construção de um regressor e aplicação no ponto mais próximo do instrumento secundário.

Nesta etapa do trabalho, apenas a técnica PDS foi aplicada para transferência de espectros entre os instrumentos primário e secundário, e então a capacidade preditiva dos modelos foi comparada por meio de teste com base na distribuição de erros de predição.

### 5.3.1 *Caracterização físico-química*

As propriedades API, TIAC, NAT e NAN das 160 amostras de petróleo bruto foram determinadas conforme as técnicas normatizadas ou internas conforme feito rotineiramente pela indústria petrolífera para caracterização das amostras de petróleo.

A densidade foi determinada pela injeção de cada amostra no densímetro digital automático Anton Paar modelo DMA 5000, sendo medida a 50°C, devido a alta viscosidade de algumas amostras. A densidade foi então estimada a 20° de acordo com valores Tabelados, para o cálculo da densidade API conforme a Equação 1 (ISO 12185-96).

A temperatura inicial de aparecimento de cristais, TIAC, dada em °C, foi determinada por calorimetria exploratória diferencial (DSC), que é uma análise térmica em que tanto amostra quanto referência são submetidas a um calorímetro diferencial, o qual mede a quantidade exata de calor de transição entre esses materiais (KOK, M. V., LETOFFE, J. M., CLAUDY, P., 1999 e JAPPER-JAAFAR, A., BHASKOROB, P. T. e MIORC, Z. S., 2016). A taxa de aquecimento adotada foi de 1°C por minuto.

O número de acidez total foi obtido experimentalmente seguindo a norma ASTM D 664, em que é feita uma titulação potenciométrica com titulador automático Metrohm com uma unidade de dosagem, agitador magnético e o eletrodo Solvotrode easyClean. Uma solução alcoólica de hidróxido de potássio 0,1 mol.L<sup>-1</sup> foi usada como titulante, e o eletrodo utilizado foi o de membrana com solução de cloreto de lítio saturada em etanol. As amostras foram solubilizadas em uma mistura de solventes contendo tolueno, propan-2-ol (isopropanol) e uma pequena porção de água. Cada análise consumiu cerca de 40 gramas de amostra. Os resultados são dados em miligramas de KOH por grama de amostra.

O método utilizado para determinação do número de acidez naftênica não pode ser pormenorizado por motivo de sigilo industrial. É sabido que a quantificação de ácidos naftênicos em petróleo ou frações envolve leitura de absorção em infravermelho médio após etapa de extração em fase sólida (SPE) dos ácidos naftênicos contidos na amostra. Uma curva de calibração é construída em torno do máximo de absorvidade no comprimento de onda  $1710\text{ cm}^{-1}$  e então é possível calcular o valor da propriedade, sendo os resultados expressos na mesma unidade do número de acidez total, mg KOH/g.

### 5.3.2 Instrumentação

Os espectros de infravermelho na região do médio (MIR) foram obtidos em um instrumento primário, localizado na Universidade Federal do Espírito Santo, para as 160 amostras. Este instrumento é da marca Perkin Elmer contendo cristal de ZnSe e acessório de múltipla reflectância total atenuada - ATR-MIR. Os espectros foram lidos na região entre  $4000\text{ cm}^{-1}$  e  $650\text{ cm}^{-1}$  com 32 escaneamentos e  $4\text{ cm}^{-1}$  de resolução. Durante a obtenção dos espectros a temperatura ambiente foi mantida em  $23^{\circ}\text{C}$  e a umidade relativa em torno de 35%. As 20 amostras que compõem os conjuntos externos e de transferência foram também medidas em um instrumento MIR secundário, localizado no Rio de Janeiro. Os espectros no MIR secundário foram obtidos nas mesmas condições utilizadas no instrumento primário, com absorvância medida entre  $4000\text{ cm}^{-1}$  e  $400\text{ cm}^{-1}$ .

### 5.3.3 Quimiometria

Inicialmente foi feita modelagem quimiométrica por meio da técnica PLS para o MIR primário para as propriedades API, TIAC, NAT e NAN. Após validação destes modelos, dez espectros de transferência obtidos no MIR secundário foram relacionadas aos espectros correspondentes do MIR primário por meio da técnica de modelagem PDS para transferência de calibração. Os dados foram processados no *software* MATLAB versão 7.12.0 (MathWorks Inc., Natick, MA). Todos os espectros foram pré-processados pelo método airPLS, que corrige desvios da linha de base de espectros de infravermelho por meio da soma dos erros quadrados entre uma linha de base previamente montada e os sinais originais, que permitem calcular pesos a cada iteração; a cada

nova rodada o procedimento é reponderado de maneira a aproximar-se a linha de base complexa (GAN et al., 2006). Uma característica do método airPLS é a manutenção do perfil dos espectros, bem como da escala da absorbância original, facilitando a interpretação das modificações realizadas e comparação entre espectros.

Os dados foram ainda centrados na média e como procedimento de validação cruzada foi utilizado o método *venetian blind 5-fold* a fim de otimizar o número de variáveis latentes (VL) para cada modelo (ASTM, 2012). O número ótimo de VL em cada caso foi determinado pelo gráfico da taxa de erro de validação cruzada em função do número de VL, ou seja, foi selecionado a quantidade de VL correspondente à menor variação do valor da taxa de erro considerando a adição de uma nova variável.

Foram construídos um total de quatro modelos PLS, um para cada propriedade de interesse, e os resultados dos modelos foram avaliados quanto a RMSEC, RMSEP de amostras de teste e externas e os coeficientes de determinação ( $R^2$ ) de calibração, teste e externo. A avaliação dos modelos foi feita por meio dos testes estatísticos de viés e permutação. O teste de vies (*bias*) verifica a presença de erros sistemáticos e o teste de permutação avalia possíveis tendências nos resíduos do modelo (FILGUEIRAS, 2014a). Os algoritmos utilizados se baseiam na ASTM E1655 com um nível de significância de 0,05.

a) *Modelos de predição de densidade API, TIAC, NAT e NAN*

Utilizando dados do instrumento primário, modelos PLS foram construídos, otimizados e validados para previsão de densidade API, TIAC, NAT e NAN. Em contraste ao procedimento adotado no Capítulo 4 desta dissertação, os procedimentos realizados neste Capítulo 5 não foram feitos modelos de recalibração FR, tampouco modelos de transferência SBC e DS, mas somente foi aplicado o método PDS. Pelo princípio da parcimônia, não foi realizada modelagem por OPLS neste procedimento, uma vez que os modelos PLS, mais simples e rápidos, tiveram resultados satisfatórios.

A quantidade de amostras disponíveis para estes modelos foi bastante superior (Tabela 8) àquela do capítulo anterior, cuja modelagem contou com 66 amostras. No capítulo atual, foram utilizadas apenas dez amostras de transferência que podem ou não estar contidas no conjunto de modelagem e dez amostras externas,

necessariamente ausentes do conjunto de modelagem, para teste e comparação entre modelos originais e transferidos. Devido a limitações dos procedimentos de caracterização físico-química por consequência da grande variação entre as amostras de petróleo, para cada propriedade os conjuntos para modelagem, transferência e externo continham número diferente de amostras. As amostras de transferências e as externas incluem as mesmas utilizadas para modelagem de densidade API desde que os dados da propriedade de interesse estivesse disponível.

**Tabela 8.** Quantidade de amostras nos conjuntos de modelagem por PLS para cada propriedade.

<b>Propriedade</b>	<b>Modelagem</b>
API	148
TIAC	142
TAN	145
NAN	56

Para obtenção dos modelos originais PLS, os dados disponíveis para modelagem foram divididos em dois conjuntos: um para calibração e outro para validação externa (denominado conjunto de teste). A separação do conjunto de modelagem foi feita por meio da ordenação da propriedade de interesse de maneira crescente e seleção de amostras de maneira segmentada na proporção 3:1 entre calibração e validação externa. Posteriormente as amostras de validação externa foram incorporadas à calibração a fim de obter um modelo com maior quantidade de informação e levando em consideração a existência de um conjunto de amostras externas medidas em dois instrumentos MIR.

#### b) *Transferência de calibração*

Uma análise preliminar dos espectros mostrou que a quantidade de variáveis espectrais (número de onda) dos espectros secundários difere daquela dos espectros primários, conforme mostrado na Tabela 9. Para resolver este problema, foi aplicada uma interpolação polinomial, visando igualar os pontos dos espectros secundários àqueles do instrumento primário. Esta interpolação polinomial é feita identificando os 5 pontos mais próximos do ponto a ser estimado e então é

construído um regressor de grau 4 que é aplicado ao número de onda a ser estimado.

**Tabela 9.** Diferenças nos números de onda entre espectros obtidos nos diferentes instrumentos de MIR.

<b>Instrumento</b>	<b>Variável inicial (cm<sup>-1</sup>)</b>	<b>Variável final (cm<sup>-1</sup>)</b>	<b>Total de variáveis</b>
Primário	650	4000	3351
Secundário	399,2	4001,6	1869

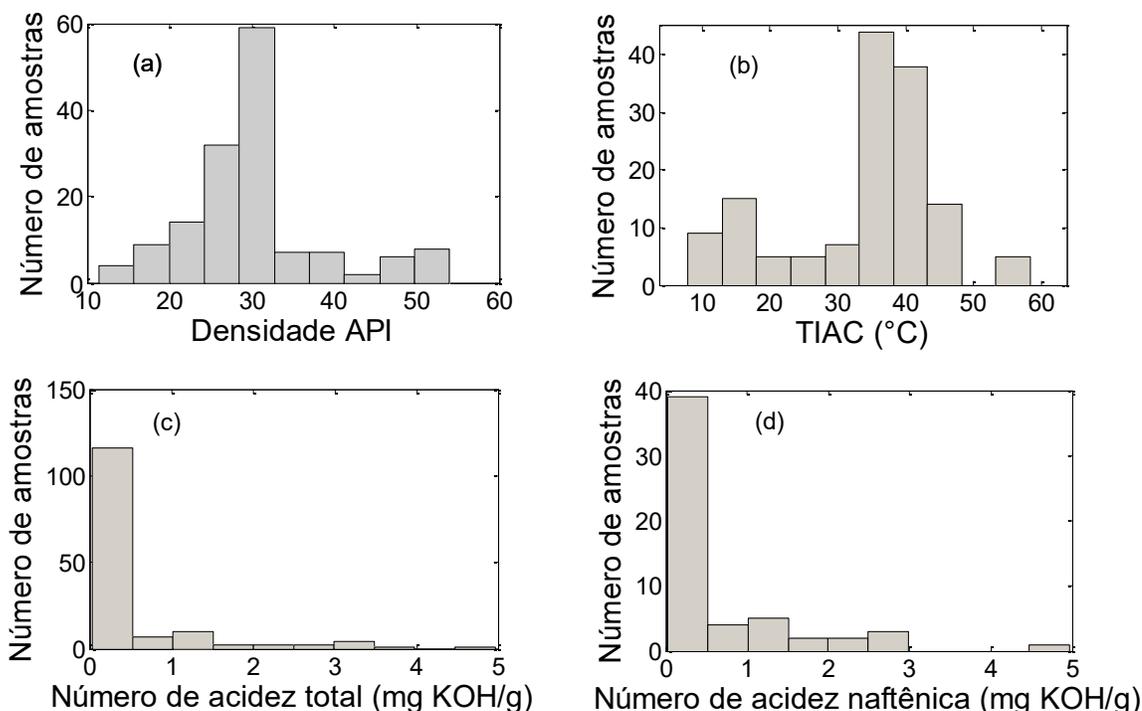
Com os dois conjuntos de dados tendo as mesmas variáveis, é possível aplicar o algoritmo PDS para transferir os espectros, tornando-os mais semelhantes aos espectros primários. Após a etapa de transferência, os espectros foram pré-processados por airPLS, método de correção da linha de base que utiliza iterações adaptativas, assim como foi feito aos espectros primários no procedimento de modelagem, e enfim puderam ser aplicados ao modelo PLS original. Os resultados obtidos após transferência foram comparados aos modelos de referência e também aos resultados das mesmas amostras no modelo PLS original.

Um parâmetro levado em conta tanto durante a calibração do modelo quanto durante a reflexão sobre a previsão obtida para cada amostra externa foi o *leverage*, que representa a distância de cada amostra ao centro dos dados. Amostras contidas no conjunto de calibração que apresentem alto *leverage* simultaneamente a um resíduo elevado podem ser consideradas amostras anômalas, devendo ser eliminada para evitar qualquer influência negativa nos modelos de calibração (VALDERRAMA, BRAGA e POPPI, 2007). Neste trabalho foi estabelecido um valor limite de *leverage* correspondente a três vezes o número de VL sobre o número de amostras de calibração. Amostras de previsão com valores acima deste limite não podem ser avaliadas pelo modelo.

A acurácia dos modelos foi avaliada por meio de um *t*-teste estatístico randômico que verifica a equivalência da distribuição de erros quadrados (VAN DER VOET, 1994) na previsão de amostras externas.

## 5.4 Resultados e discussão

A Figura 13 mostra a distribuição das amostras segundo as propriedades a serem modeladas: API, TIAC, NAT e NAN. O histograma de distribuição de densidade API (Figura 13 a) inclui as 148 amostras disponíveis para modelagem, sendo que a amostra mais pesada apresenta °API 11,4 e a mais leve tem °API 54,0, sendo que a maior parte das amostras apresenta valores intermediários.



**Figura 13.** Histograma de distribuição das amostras conforme (a) densidade API; (b) TIAC; (c) NAT; (d) NAN.

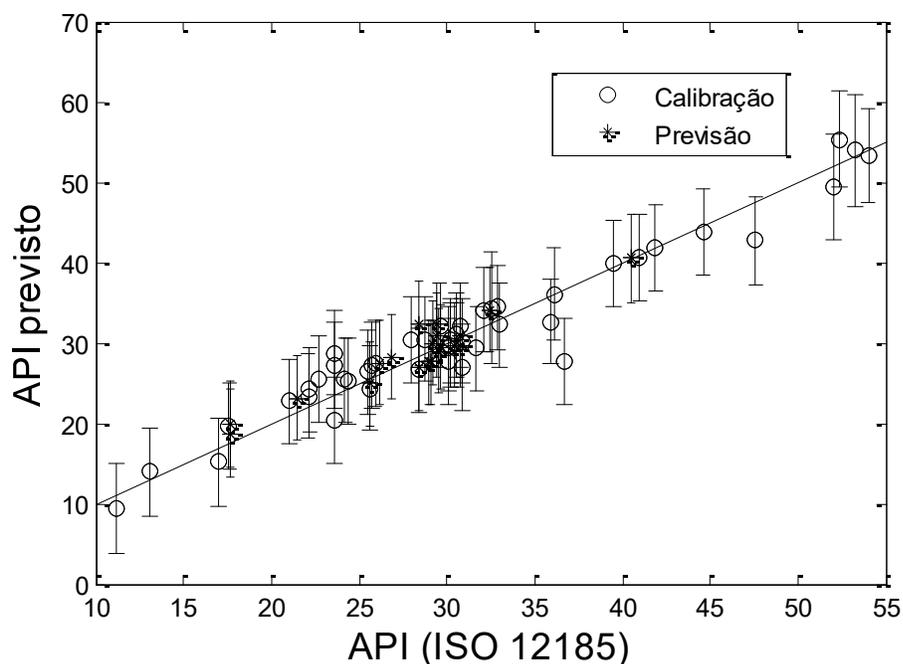
A TIAC (Figura 13 b) das amostras varia entre 7,96 e 58,38°C, com 83 das 142 amostras com TIAC entre 33,2 e 43,3°C. A aparência dos histogramas de NAT e NAN (Figura 13 c e d), com a maior parte das amostras com baixa acidez com valores inferiores a 0,5 mg KOH/g, o que é compatível com o fato da acidez total refletir em sua maioria a presença de ácidos naftênicos. No entanto, a quantidade de amostras varia grandemente, já que é comum que o NAN seja determinado somente para amostras de petróleo bruto com elevado NAT, tendo apenas valores de 56 amostras para NAN e 145 para NAT.

A modelagem realizada com os dados do instrumento primário aplicou PLS utilizando espectros de MIR com linha de base corrigida por airPLS para previsão de

densidade API. O conjunto de dados foi separado por meio do ordenamento das amostras de maneira crescente e seleção segmentada, uma a cada três para o conjunto de validação externa e as demais para calibração e validação interna.

O modelo contou com 102 amostras de calibração usadas para validação interna por meio de validação cruzada *5-fold*. Cinco variáveis latentes descrevem satisfatoriamente os dados sem sobreajuste. Os testes de viés e permutação não acusaram erros sistemáticos ou tendenciosos.

A Figura 14 mostra o gráfico que correlaciona valores de referência e previstos pelo modelo, para os conjuntos de calibração e previsão. É possível visualizar que as previsões das amostras de calibração e previsão considerando as incertezas associadas contém o valor de referência, com poucas exceções de amostras que se afastam da reta em que valores previstos coincidem com valores de referência.



**Figura 14.** Gráfico que correlaciona a densidade API obtida de acordo com a norma ISO 12185 e os valores correspondentes previstos pelo modelo.

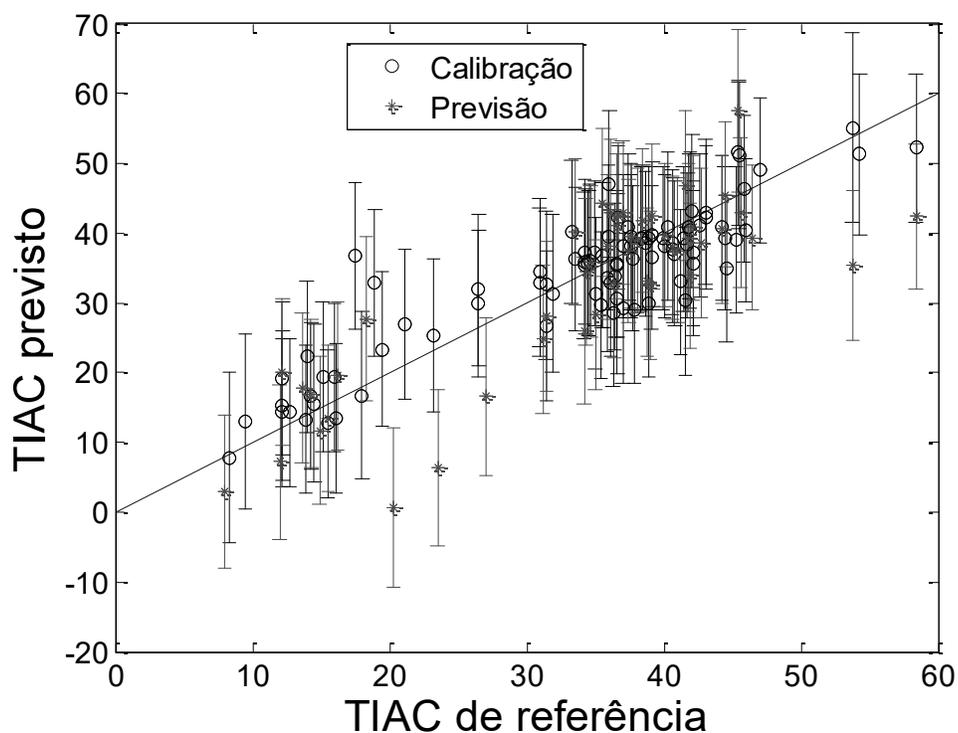
A Tabela 10 mostra que os parâmetros de exatidão (RMSEC e RMSEP) do modelo API foram inferiores a 2, e os coeficientes de determinação superiores a 0,94 indicam que ambos os conjuntos de dados de calibração e validação externa se comportam de maneira linear. Esses resultados concordam com as conclusões obtidas por inspeção visual da Figura 14.

**Tabela 10.** Dados dos modelos de calibração por PLS com espectros de instrumento MIR primário para as propriedades API, TIAC, NAT e NAN.

<b>Propriedades</b>	<b>Abrangência do modelo</b>	<b>VL<sup>(a)</sup></b>	<b>RMSEC<sup>(b)</sup></b>	<b>RMSEP<sup>(c)</sup></b>	<b>R<sup>2(d)</sup> calibração</b>	<b>R<sup>2(d)</sup> previsão</b>
<b>Densidade API</b>	11,4-54,0	5	1,37	1,97	0,977	0,948
<b>TIAC</b>	7,96-58,38	9	5,12	7,25	0,807	0,694
<b>NAT</b>	0,03-4,96	7	0,23	0,27	0,843	0,911
<b>NAN</b>	0,02-4,96	5	0,32	0,64	0,914	0,680

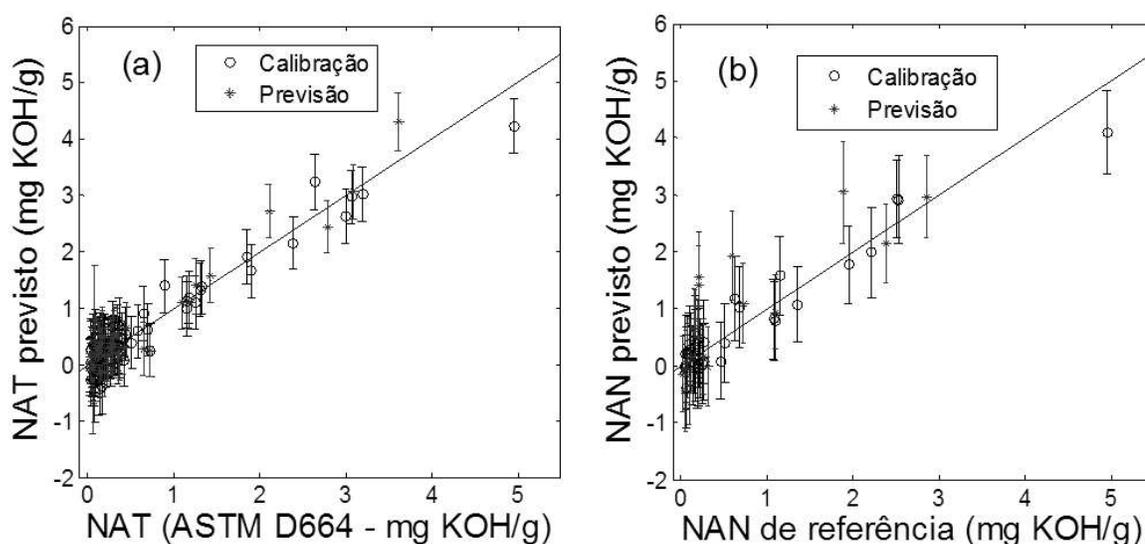
<sup>(a)</sup>VL: variáveis latentes; <sup>(b)</sup>RMSEC: *root mean square errors of calibration* (raiz quadrada do erro quadrático médio de calibração); <sup>(c)</sup>RMSEP: *root mean square errors of prediction* (raiz quadrada do erro quadrático médio de previsão); <sup>(d)</sup>R<sup>2</sup>: coeficiente de determinação.

A modelagem da TIAC foi feita de maneira análoga, estando presente no conjunto de calibração 94 amostras com valores entre 7,96 e 58,38. A etapa de validação interna mostrou que o número ótimo de VL é 9, um valor relativamente alto, porém como há grande complexidade na amostra não houve sobreajuste. Testes estatísticos de viés e permutação não apontaram erros sistemáticos nos valores previstos tanto para calibração quanto para previsão, ou tendência nos resíduos. Houve uma diferença relativamente grande nos valores de RMSE de calibração e previsão, porém conforme pode ser visualizado na Figura 15, os intervalos de confiança são muito altos, pois este é dependente do valor de RMSEC (Eq. 9) que está acima de 5°C. Além disso, o ajuste à reta em que valores previstos coincidem com os valores de referência foram baixos tanto para calibração quanto para previsão, ficando abaixo de 0,7 no segundo caso, indicando que o modelo poderia ser melhorado para fornecer estimativas mais confiáveis, no entanto técnicas como seleção de variáveis ou modelagens não lineares fogem do escopo deste trabalho.



**Figura 15.** Gráfico de correlação entre TIAC obtida por procedimentos normalizados e os valores correspondentes previstos pelo modelo PLS para o instrumento primário.

A Figura 16 mostra os gráficos de correlação entre NAT e NAN obtidos pelo modelo PLS por dados de MIR do instrumento primário em relação aos valores correspondentes dos procedimentos de referência. Vale ressaltar que as duas propriedades químicas em questão estão intimamente relacionadas, uma vez que a maior parte dos ácidos presentes nas amostras de petróleo bruto que são neutralizados por KOH pelo método ASTM D664 para obter o NAT são ácidos naftênicos. A abrangência dos modelos é muito próxima, já que as amostras de calibração incluem amostras de NAT e NAN variando desde 0,03 ou 0,02 até 4,96mg KOH/g. Os valores de RMSEC e RMSEP do modelo de NAT variam em apenas 0,04, e os coeficientes de determinação indicam boa linearidade para amostras de calibração e validação externa, pelo modelo construído com 7 VL. Já o modelo de NAN, com 5 VL, apresentou valores de RMSE mais elevados, especialmente para o conjunto de teste, com RMSEP sendo o dobro do RMSEC; e o coeficiente de determinação foi ruim, indicando que o modelo não é satisfatório. Uma justificativa plausível para a diferença tão grande entre os modelos de NAT e NAN sendo os dados tão semelhantes é a quantidade de amostras disponíveis ser muito menor no NAN, sendo 145 para NAT (96 de calibração e 49 de validação) e apenas 56 para NAN (37 para calibração e 19 para validação). De qualquer maneira ambos os modelos PLS foram aprovados quanto à ausência de viés e tendência nos resíduos da calibração e teste.



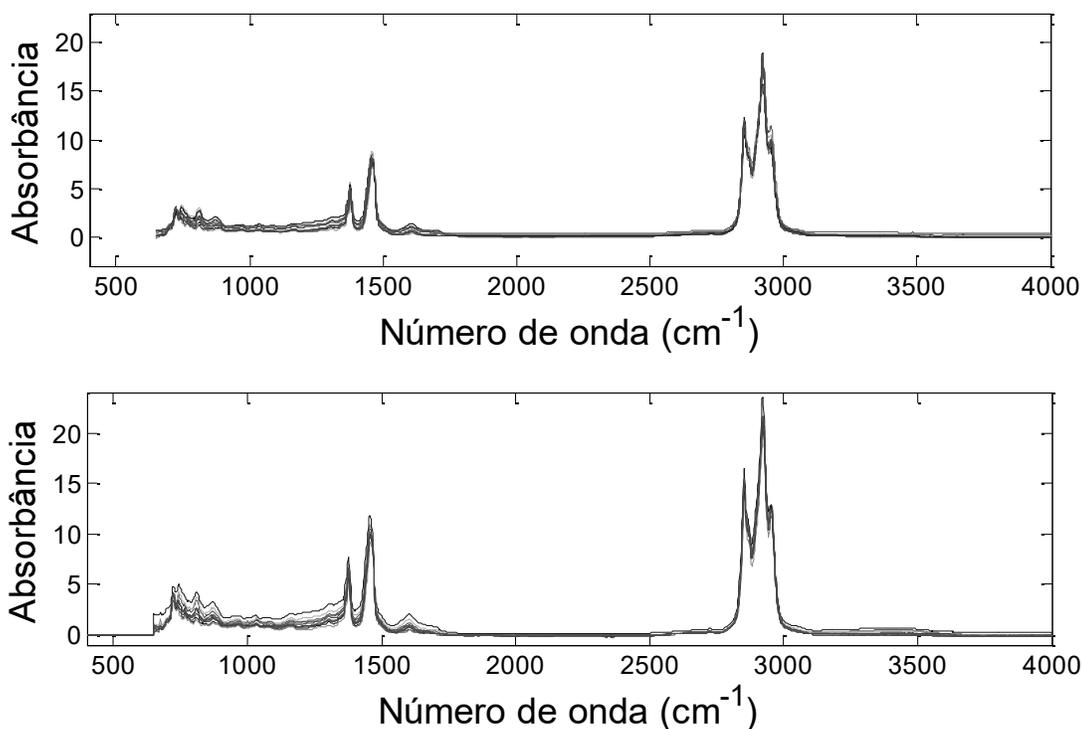
**Figura 16.** Gráfico de correlação entre valores de referência e valores correspondentes previstos pelo modelo PLS para o instrumento primário para as propriedades NAT (a) e NAN (b).

Após a construção de todos os modelos apresentados até o momento, a chamada modelagem inicial, as amostras separadas para validação externa foram incorporadas para, assim, o modelo PLS conter maior variabilidade. A Tabela 11 mostra os resultados obtidos. Enquanto os modelos para TIAC e NAN apresentaram quase o mesmo valor de RMSEC, os modelos de API e NAT sem as amostras de validação tinham um RMSEC menor, indicando que poderiam levar a previsões errôneas e enfatizando a importância de incluir o máximo de informações químicas na calibração. A diferença no valor de RMSEC dos modelos para API, no entanto, não foi muito elevada, apenas 11,7%.

**Tabela 11.** Resultados dos modelos generalizados, aqueles que tiveram amostras de validação externa incorporadas, em comparação aos modelos iniciais.

	Modelagem inicial		Generalização	
	RMSEC	Amostras de calibração	RMSEC	Amostras de calibração
<b>API</b>	1,37	98	1,53	147
<b>TIAC</b>	5,12	94	5,13	142
<b>NAT</b>	0,23	96	0,37	146
<b>NAN</b>	0,32	37	0,35	56

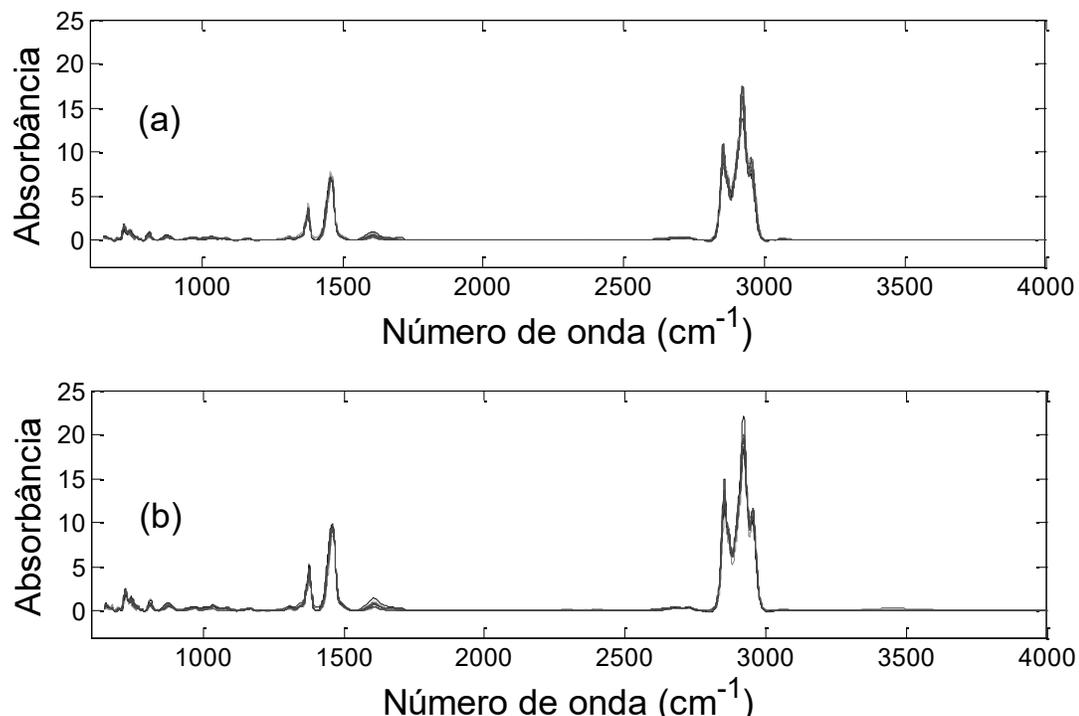
De posse dos modelos PLS construídos e validados para o instrumento primário, iniciou-se o procedimento para possibilitar a aplicação de espectros de um instrumento secundário. A Figura 17 mostra um paralelo entre espectros brutos obtidos nos instrumentos da Ufes e de um laboratório localizado no Rio de Janeiro. O aspecto visual mostra perfis espectrais semelhantes. Os espectros do instrumento secundário, entretanto, apresentam intensidades ligeiramente superiores.



**Figura 17.** Espectros brutos de 10 amostras de transferência de petróleo bruto coletados em dois instrumentos de MIR: (a) primário; (b) secundário.

Um fator complicador é a diferença nos pontos espectrais entre os instrumentos, uma vez que os espectros exportados do instrumento secundário continham apenas 1869 números de onda enquanto os modelos originais foram construídos com 3351 variáveis. Para igualar as variáveis e possibilitar a aplicação do algoritmo PDS foi necessária uma etapa de interpolação dos espectros secundários, adequando um polinômio de quarto grau aos 5 pontos mais próximos do número de onda desejável. O objetivo deste procedimento é manter a aparência do perfil espectral deixando espectros secundários com valores de absorvância exatamente nos mesmos números de onda do que os espectros primários.

Uma vez que os espectros de calibração dos modelos PLS foram pré-processados por airPLS para corrigir a linha de base, os espectros secundários após a interpolação foram também tratados pelo método previamente à aplicação no modelo PLS original. A Figura 18 mostra que o efeito do pré-processamento é similar nos conjuntos de dados provenientes dos instrumentos primários e secundários, tornando praticamente inexistente a variação de linha de base entre as amostras e acentuando as variações espectrais que caracterizam as diferenças amostrais nas regiões de maior absorvância detectada.



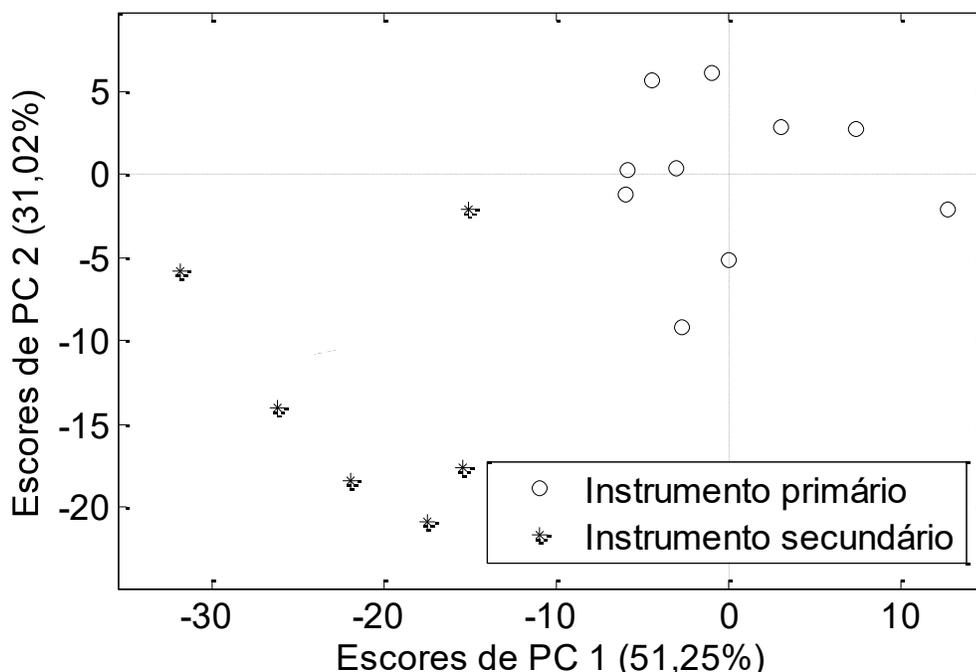
**Figura 18.** Conjunto de espectros de petróleo bruto de amostras de transferência após tratamento com airPLS nos dois instrumentos de MIR: (a) primário; (b) secundário.

A fim de verificar se a grande semelhança visual entre os espectros é passível de distinção por meio de métodos quimiométricos, foi aplicada a Equação 16. A Tabela 12 mostra que a porcentagem de similaridade de espectros secundários interpolados aos espectros das mesmas amostras sem qualquer tratamento é de 49,8%. Após alinhamento da linha de base por airPLS a similaridade continuou inferior a 50%, o que pode indicar a ineficiência de aplicação direta de espectros secundários ao modelo PLS original construído para o instrumento primário.

**Tabela 12.** Porcentagem média de similaridade entre espectros de transferência interpoladas secundários submetidas a diversos procedimentos em relação aos espectros primários.

Procedimento		Similaridade (%)
Espectros primários	Espectros secundários	
Nenhum	Nenhum	49,8%
airPLS	airPLS	44,4%
Nenhum	PDS	88,5%
airPLS	PDS+airPLS	85,1%

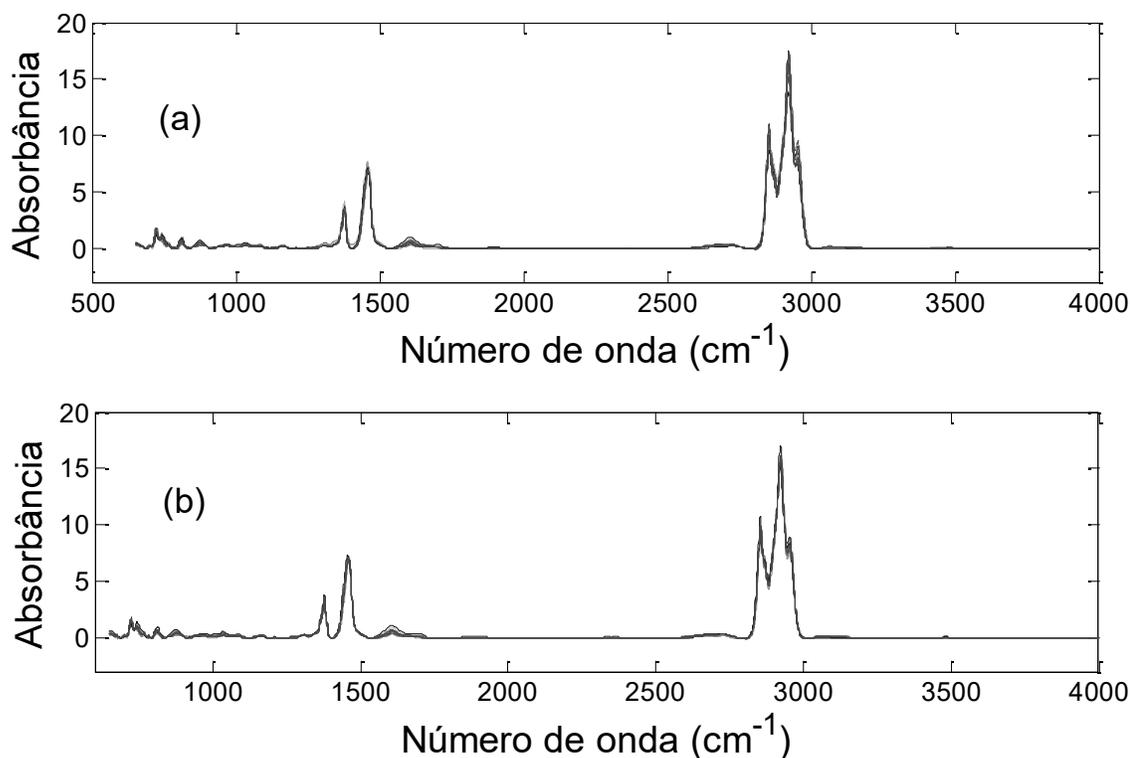
Foi também elaborado um modelo PCA, cujo gráfico de escores para as duas primeiras componentes principais, que explicam mais de 82% da variância dos dados, é mostrado na Figura 19. É possível notar o agrupamento nítido entre amostras de cada instrumento e portanto, apenas pré-processar as amostras por airPLS não é suficiente para possibilitar a aplicação de espectros secundários ao modelo PLS original.



**Figura 19.** Gráfico de escores dos dois primeiros componentes principais de um modelo PCA construído com espectros MIR dos instrumentos primário e secundário tratados com airPLS.

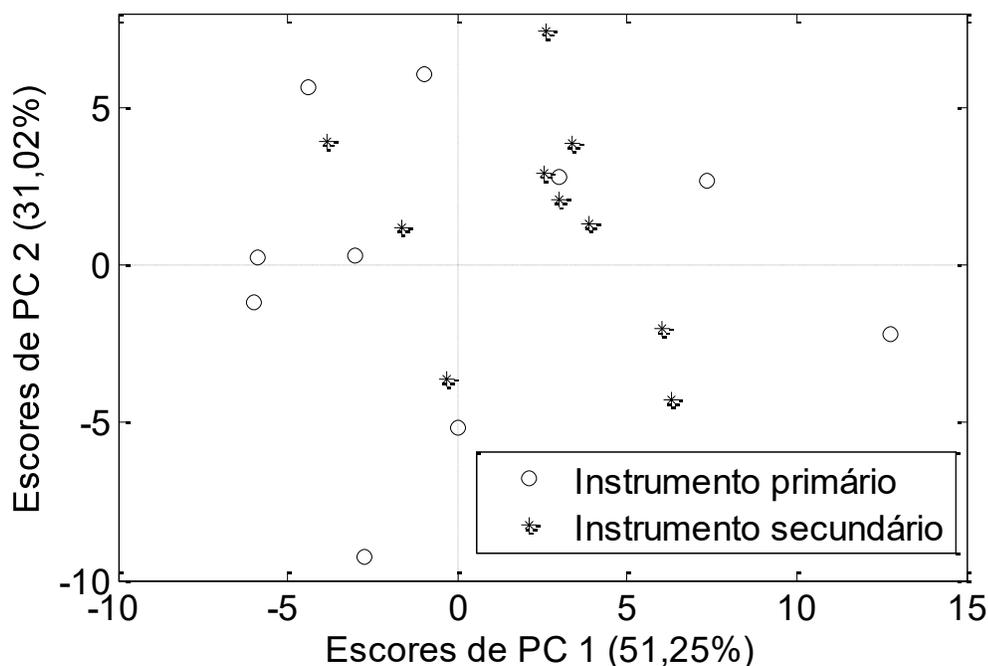
Considerando o desempenho moderadamente satisfatório do método de transferência PDS registrado no Capítulo 4 desta dissertação, os espectros secundários interpolados foram submetidos a este procedimento, que padroniza diretamente a espectros primários utilizando janelas de 5 pontos. Em seguida, os espectros foram tratados por airPLS, condição do modelo de calibração PLS onde se deseja aplicar os espectros.

A Figura 20 mostra que os espectros provenientes dos diferentes instrumentos neste ponto apresentam similaridade visual muito grande, então foi aplicada a Equação 16 para quantificar esta semelhança. A similaridade dos espectros transferidos por PDS em relação aos espectros do instrumento primário foi de 88,5%, conforme Tabela 12, um aumento significativo com relação aos espectros brutos. Após transferência e tratamento por airPLS houve uma ligeira redução da similaridade, no entanto esta permanece superior a 85% em comparação aos espectros primários tratados por airPLS, o que indica possibilidade de obter bons resultados na aplicação nos modelos originais para obter estimativa das propriedades modeladas.



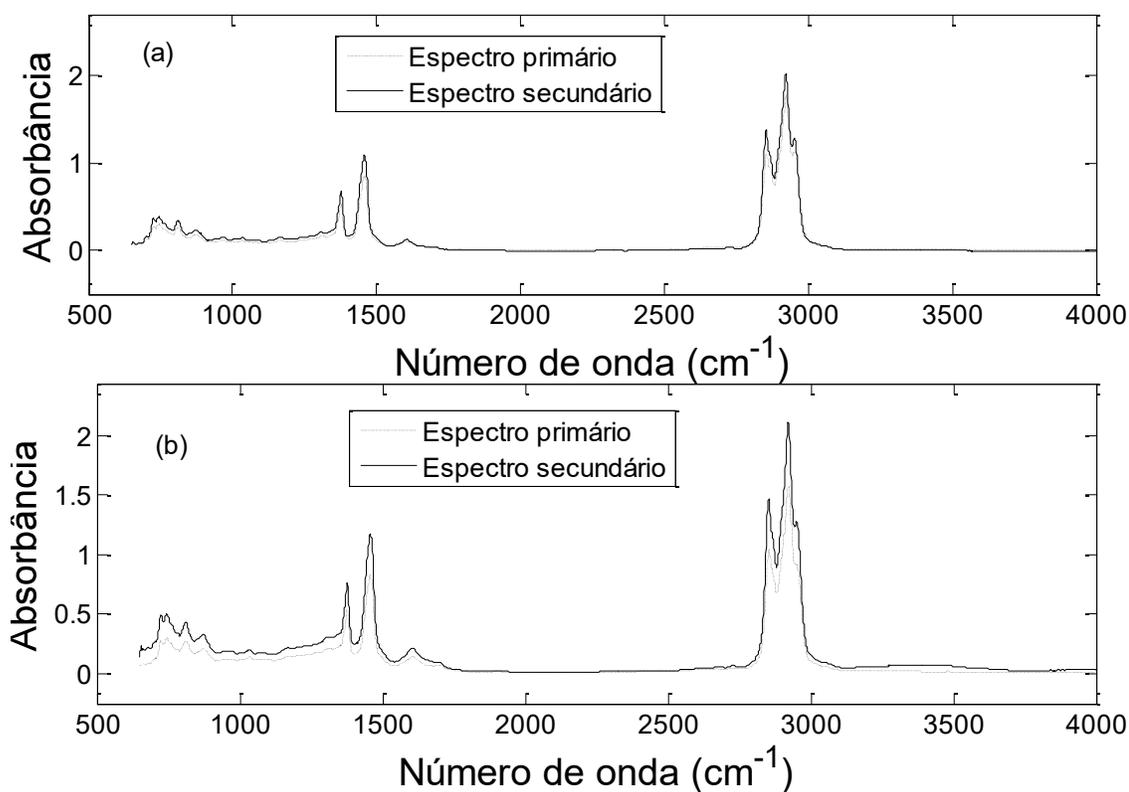
**Figura 20.** Espectros das dez amostras externas obtidos (a) no instrumento primário e após pré-processamento por airPLS; e (b) no instrumento secundário e após interpolação, transferência PDS e tratamento por airPLS.

Foi novamente construído um modelo PCA após a transferência e correção da linha de base por airPLS para avaliar a distinção entre espectros provenientes de diferentes instrumentos. Neste caso, o gráfico de escores (Figura 21) dos dois primeiros componentes principais mostra que amostras de um mesmo instrumento não se agruparam, mas todas encontram-se misturadas sendo que a maior parte da informação das amostras (82%) que está contida nos espectros foi explicada.



**Figura 21.** Gráfico de escores dos dois primeiros componentes principais de um modelo PCA construído com espectros primários tratados com airPLS e secundários transferidos por PDS e também tratados com airPLS.

A Figura 22 mostra a superposição entre espectros de mesma amostra após os espectros secundários terem sido transferidos por PDS e ambos primários e secundários tratados por airPLS. Fica evidenciado que para algumas amostras a superposição foi quase total (Figura 22a), enquanto para outras a diferença foi mais notável (Figura 22b). A Tabela 13 mostra os valores de referência das propriedades físico-químicas modeladas para as dez amostras externas. Ambas as amostras mostradas na Figura 22 têm valores intermediários das propriedades de interesse conforme indicado na Tabela 13, dentro do intervalo de maior frequência se observados os histogramas da Figura 11.

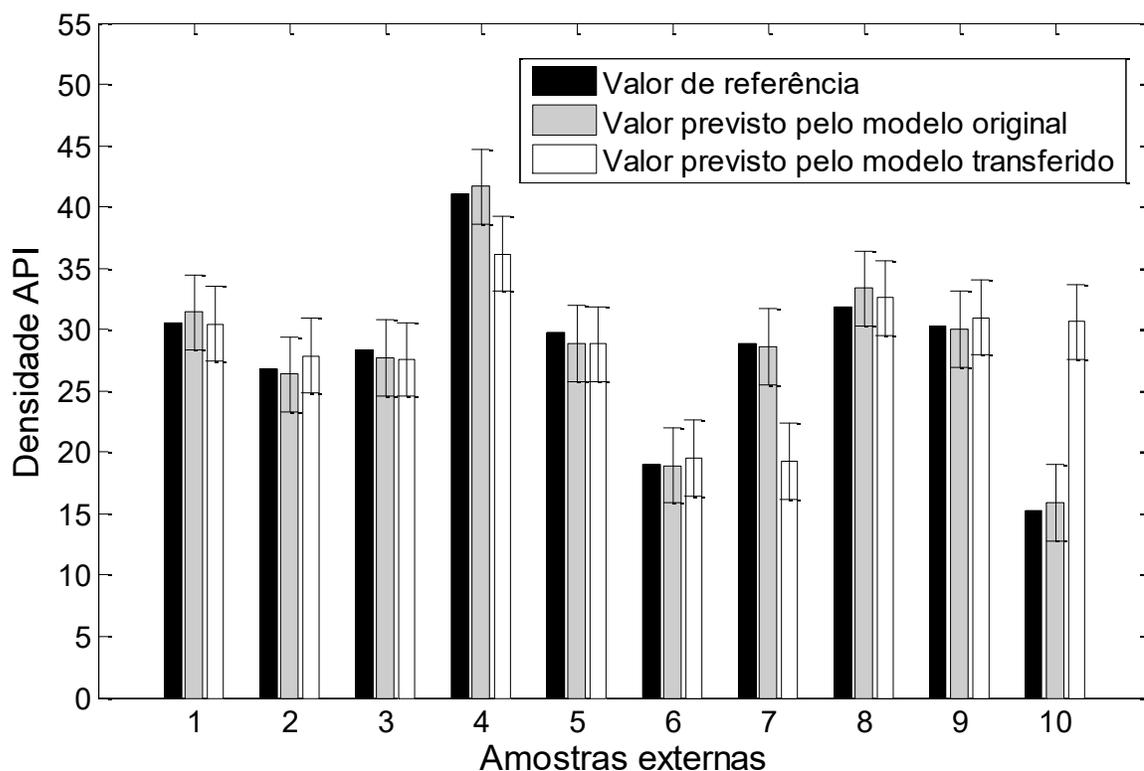


**Figura 22.** Espectro secundário transferido superposto ao espectro primário correspondente para as amostras externas 3 (a) e 8 (b).

**Tabela 13.** Valores de referência das propriedades API, TIAC, NAT e NAN para as dez amostras que compõem o conjunto externo.

Amostra	API	TIAC (°C)	NAT (mg KOH/g oleo)	NAN (mg KOH/g oleo)
1	30,6	38,32	0,11	0,02
2	26,8	41,80	0,29	0,21
3	28,4	44,24	0,17	-
4	41,0	31,46	< 0,03	-
5	29,8	40,59	0,04	-
6	19,0	15,06	1,84	1,14
7	28,8	39,4	0,27	-
8	31,9	39,54	0,20	-
10	15,2	50,7	2,02	-

A Figura 23 compara valores de referência, valores previstos pelo modelo PLS original por meio de espectros primários e valores previstos com aplicação dos espectros secundários transferidos por PDS para densidade API. Os resultados para espectros primários e secundários foram equivalentes, à exceção de duas amostras externas. A amostra 10, a mais pesada do conjunto, foi prevista muito mais próxima ao valor de referência pelo espectro primário, tendo apresentado grande discrepância quando da aplicação de espectro secundário transferido, quase o dobro do real. A amostra 4, a mais leve, apresentou previsão equivalente dentro do intervalo de confiança para os dois modelos, no entanto a estimativa feita pelo modelo transferido apresentou um erro absoluto de quase 5 unidades em relação ao valor de referência. O fato destas amostras terem valores de densidade API extremos é um possível fator da dificuldade de modelagem. No entanto, os valores de *leverage* de todas as amostras foram bastante inferiores ao limite determinado na modelagem. O modelo transferido não estimou com boa exatidão a amostra 7, embora esta seja classificada como mediana.



**Figura 23.** Resultados da previsão de densidade API para amostras dez externas do modelo PLS original com espectros primários e com espectros secundários submetidos à transferência de calibração PDS em comparação com valores de referência.

O teste estatístico de comparação de acurácia (Tabela 14) mostrou que a previsão de densidade API feita por meio dos espectros secundários transferidos por PDS é equivalente àquela feita para os espectros primários para as 10 amostras externas (p-valor=0,15).

**Tabela 14.** Resultado do teste estatístico de comparação de acurácia entre modelos PLS originais construídos para instrumento primário e modelos de transferência PDS de espectros secundários aplicados aos modelos originais.

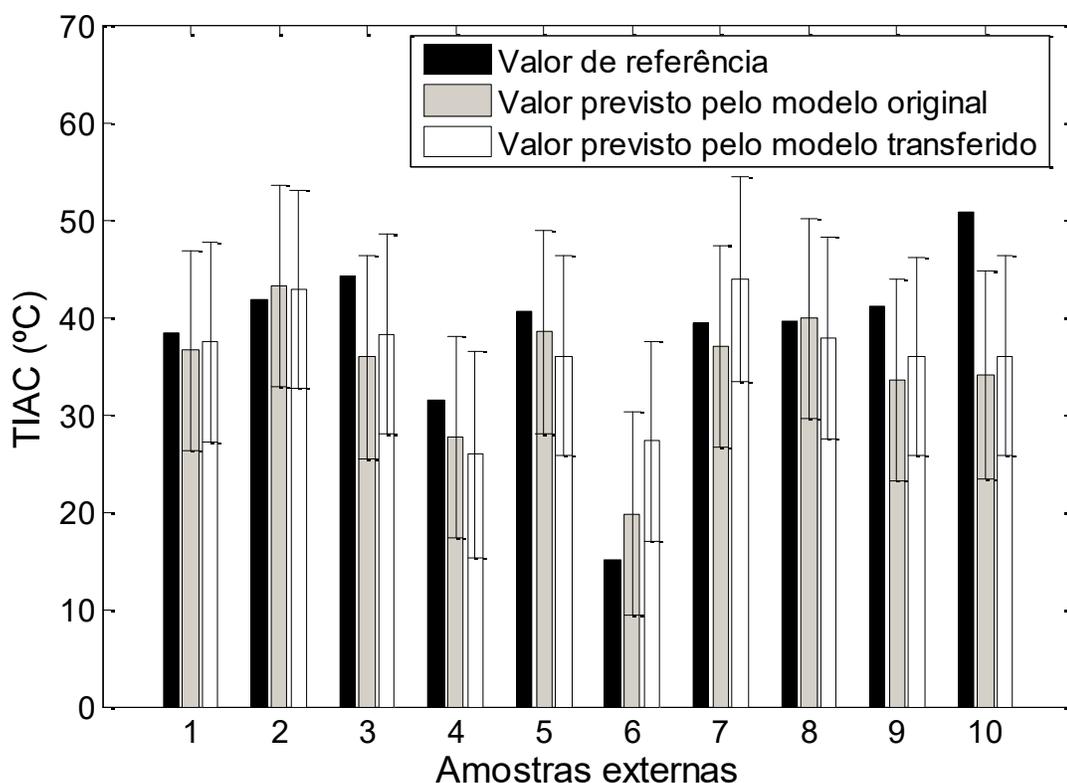
<b>Propriedade</b>	<b>p-valor*</b>
<b>API</b>	0,15
<b>TIAC</b>	0,82
<b>NAT</b>	0,06
<b>NAN</b>	0,75

\* Um p-valor maior do que 0,05 indica que a acurácia dos modelos comparados é estatisticamente igual com 95% de confiança.

A comparação da acurácia do modelo transferido da propriedade TIAC em relação ao modelo PLS construído com espectros primários, conforme resultado mostrado na Tabela 14, indica equivalência do modelo transferido ao modelo PLS original para as 9 amostras externas com dados disponíveis (p-valor = 0,82) com 95% de confiança.

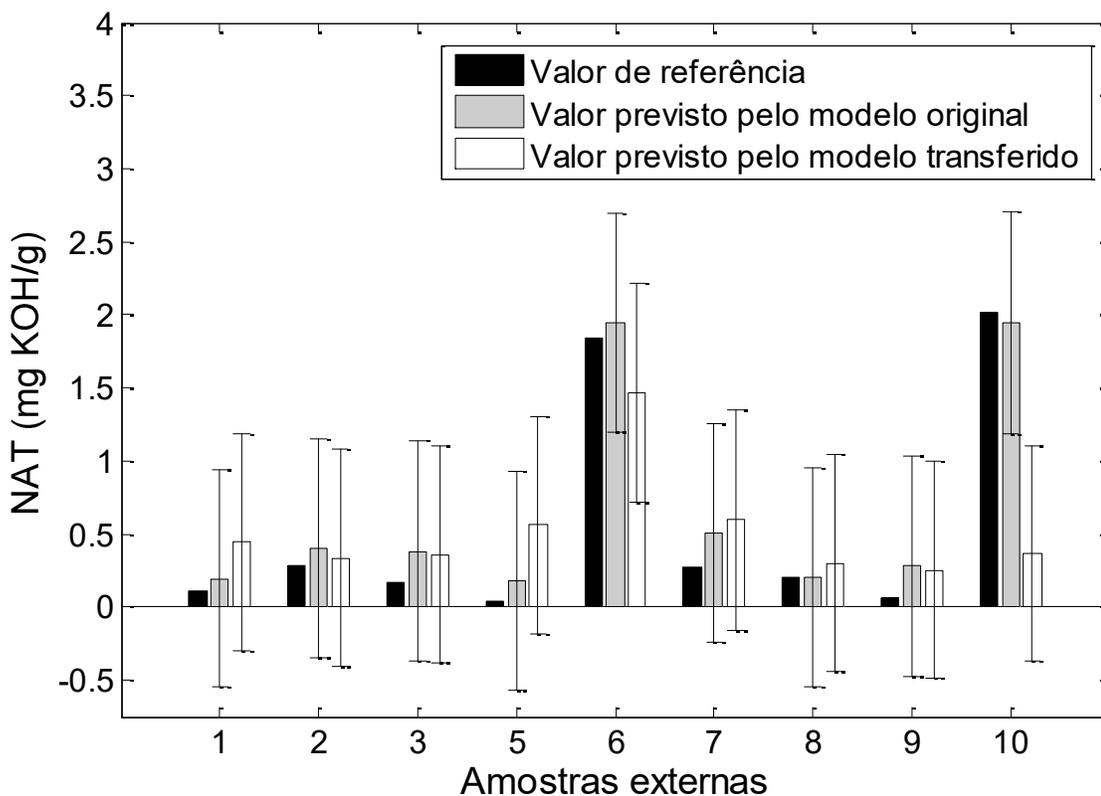
Os valores previstos, em °C, para a propriedade TIAC tanto no modelo original quanto no modelo previsto com os respectivos intervalos de confiança calculados estão mostrados na Figura 24 juntamente com os valores de referência para as dez amostras do conjunto externo.

A amostra 3 apresentou previsão ruim para ambos os modelos, com um erro absoluto de cerca de 6°C em relação ao valor de referência, no entanto o valor de referência está contemplado nos intervalos de confiança. As amostras 4, 5 e 6 foram melhor previstas pelo modelo original, sendo que a amostra 6 apresenta o valor de referência fora do intervalo de confiança previsto para o modelo transferido. As amostras 9 e 10 tiveram resultados ruins para modelo original e transferido, sendo que para a amostra 9 os dois modelos englobam o valor de referência dentro dos intervalos de confiança previstos enquanto para a amostra 10 ambos estão fora, e portanto representam erros dos modelos. A amostra 7, cuja densidade API não foi estimada com exatidão pelo modelo transferido, apresentou erro de 3,5°C na previsão de TIAC, com previsão incluindo o valor de referência nos intervalos de confiança dos modelos primário e transferido.



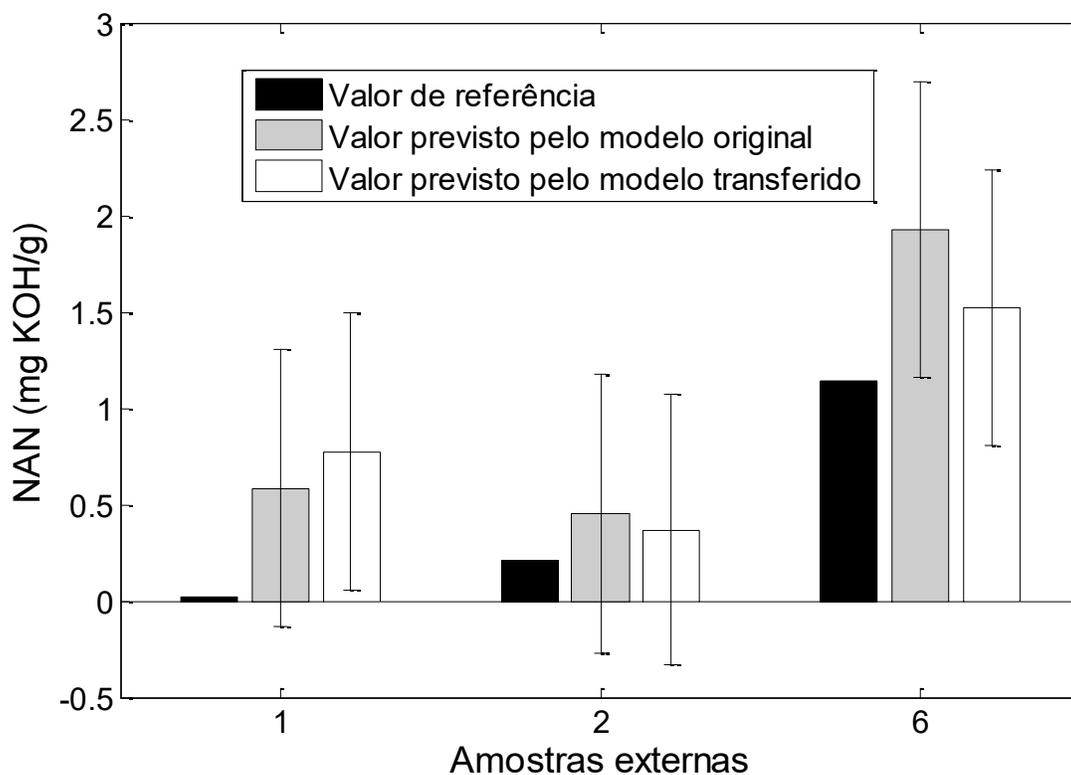
**Figura 24.** Resultados da previsão de TIAC para dez amostras externas do modelo PLS original com espectros primários e com espectros secundários submetidos à transferência de calibração PDS em comparação com valores de referência.

Os modelos para NAT têm seus resultados para amostras externas apresentados na Figura 25. Nota-se que as amostras com elevada acidez (6 e 10) apresentaram estimativas de NAT mais distantes do valor de referência pelo modelo transferido, especialmente com o valor de referência fora do intervalo de confiança na amostra 10. Além disso, as amostras 1, 3, 5 e 7 tiveram erro mais elevado com espectros secundários transferidos, com diferença de acidez da ordem de 0,3 mg KOH/g, embora o intervalo de confiança englobe o valor de referência. A amostra 9 apresentou resíduo alto em ambos os modelos, porém com previsão correta dentro do intervalo de confiança. As amostras 2 e 8 foram previstas com erros absolutos mais baixos do que as demais pelo modelo transferido. Não foi feita análise de NAT para a amostra 4, por isto não foi possível comparar os resultados de previsão. O teste estatístico de comparação de acurácia (Tabela 14), indica que os modelos são equivalentes quanto a distribuição quadrática dos erros das 9 amostras externas com dados disponíveis (p-valor=0,06).



**Figura 25.** Resultados da previsão de NAT para nove amostras externas do modelo PLS original com espectros primários e com espectros secundários submetidos à transferência de calibração PDS em comparação com valores de referência.

Para a propriedade NAN (Figura 26), apenas três das dez amostras externas tinham dados disponíveis. A amostra de NAN de valor mais baixo apresenta resultados ruins para modelos original e transferidos, tendo o modelo original contemplado o valor de referência dentro do intervalo de confiança previsto, ao contrário do modelo transferido. A amostra 2 apresentou previsão equivalente para os dos modelos, ambos com valor de referência incluído no intervalo de confiança previsto. A amostra 6 foi melhor prevista pelo modelo transferido, sendo que o modelo original apresenta erro absoluto mais elevado e o intervalo de confiança não inclui o valor de referência. O teste estatístico que compara a distribuição quadrática dos erros indica com 95% de confiança que não há diferença entre os modelos para as 3 amostras (Tabela 14,  $p$ -valor=0,75), no entanto a pequena quantidade de amostras compromete conclusões generalizadas.



**Figura 26.** Resultados da previsão de NAN para três amostras externas do modelo PLS original com espectros primários e com espectros secundários submetidos à transferência de calibração PDS em comparação com valores de referência.

## 5.5 Conclusão

Dentre os modelos PLS construídos e validados para previsão das propriedades densidade API, TIAC, NAT e NAN, aqueles com melhores resultados foram para API e NAT, com RMSE baixos e alta linearidade. Já os modelos para TIAC e NAN apresentaram grande diferença de exatidão entre conjuntos de calibração e previsão (41,6% e 100%, respectivamente).

A generalização dos modelos PLS com incorporação das amostras de validação externa ao conjunto de calibração provocou aumento no RMSEC nos modelos de API e NAT. Isto evidencia a importância de incluir o máximo de informações químicas na calibração, evitando previsões errôneas.

A transferência PDS de espectros secundários após interpolação polinomial e o processamento airPLS gerou espectros indistinguíveis daqueles obtidos no instrumento primário por PCA. A similaridade dos espectros pré-tratados aumentou de 44,4% para 85,1% após a aplicação desta transferência.

O modelo transferido apresentou erro elevado para amostras de API extremos, mas foi excelente para amostras de petróleo classificadas como medianas pela ANP, com exceção de uma amostra. Os modelos para TIAC apresentaram erros elevados, porém apenas 2 das 10 amostras externas apresentaram previsão errônea dentro do intervalo de confiança para o modelo transferido - uma delas também foi prevista erroneamente pelo modelo original. O modelo de NAT transferido apresentou erro elevado para muitas amostras, com exceção de duas, no entanto apenas uma não inclui o valor de referência nos intervalos de confiança previstos. Para NAN, haviam apenas 3 amostras externas com dados disponíveis, comprometendo conclusão sobre a transferência. Um teste estatístico mostrou que os modelos transferidos são equivalentes aos originais para as 4 propriedades trabalhadas, com 95% de confiança.

Para instrumentos MIR que não apresentam grandes diferenças espectrais é possível fazer a transferência de modelos multivariados somente associando a técnica PDS ao PLS original. No entanto, resultados apenas razoáveis para modelos originais comprometem seriamente a aplicação de espectros secundários a um modelo construído para um instrumento primário. Ainda assim, os resultados encontrados para a técnica foram promissores, sendo capazes de estimar com boa exatidão especialmente o valor da densidade API para amostras de petróleo classificadas como medianas.

## 6 CONCLUSÕES GERAIS

A avaliação de modelos de transferência entre dois instrumentos MIR mostram as diferentes possibilidades de aplicação de espectros obtidos em instrumento secundário em modelo PLS ou OPLS. O modelo de transferência PDS aplicado ao modelo OPLS resulta em melhor capacidade preditiva de densidade API para petróleo bruto (RMSEP de 1,48). A aplicação direta de PDS a um modelo PLS original conta com a vantagem de aproveitar modelos já consolidados, tendo apresentado aumento no valor de RMSEP de 57,8%.

Modelos PLS para previsões de densidade API, TIAC, NAT e NAN foram transferidos para utilização em instrumento secundário por meio de PDS.

Ao serem aplicados aos modelos PLS, os espectros transferidos geraram estimativas razoáveis para as quatro propriedades físico-químicas, tendo um teste estatístico de comparação de acurácia mostrado com 95% de confiança que os modelos são equivalentes aos originais.

A transferência PDS associada a modelos PLS mostrou-se promissora, sendo capaz especialmente de estimar com boa exatidão o valor da densidade API de amostras de petróleo classificadas como medianas.

A transferência de calibração é uma alternativa plausível para aumentar a aplicabilidade de modelos multivariados a instrumentos diferentes daquele para o qual esses modelos foram originalmente construídos. Estas aplicações são importantes para a indústria do petróleo permitindo estimar rapidamente e de forma pouco onerosa várias propriedades físico-químicas de petróleo bruto em diferentes laboratórios, evitando a necessidade de reconstruir modelos já validados e permitindo rápida tomada de decisões comerciais.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

ABBAS, O.; REBUFA, C.; DUPUY, N.; PERMANYER, A.; KISTER, J. PLS regression on spectroscopic data for the prediction of crude oil quality: API gravity and aliphatic/aromatic ratio. *Fuel*, **98**, 5-14, 2012.

AMIGO, J. M.; SKOV, T.; COELLO, J.; MASPOCH, S.; BRO, R. Solving GC-MS problems with PARAFAC2. *TrAC, Trends Anal. Chem.*, **27**(8), 714-725, 2008.

ANP. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Disponível em: <<http://www.anp.gov.br/>>. Acessado em 18 out 2016.

ASTM International. **ASTM Standard E1655-05**: Standard practices for infrared multivariate quantitative analysis. *ASTM International*: West Conshohocken, PA, USA, 2012.

BARNES R. J.; DHANOA M. S.; LISTER S. J.; Standard Normal Variate transformation and De-trending of Near-infrared Diffuse Reflectance Spectra. *Appl. Spectrosc.*, **43** (5), 772-777, 1989.

BLANCO, M.; COELLO, J.; ITURRIAGA, H.; MASPOCH, S.; de la PEZUELA, C. Effect of data preprocessing methods in near-infrared diffuse reflectance spectroscopy for the determination of the active compound in a pharmaceutical preparation. *Appl. Spectrosc.*, **51** (2), 240-246, 1997.

BOUVERESSE, E.; HARTMANN, C.; MASSART, D. L.; LAST, I. R.; PREBBLE, K. A. Standardization of Near-Infrared Spectrometric Instruments. *Anal. Chem.*, **68**, 982-990, 1996.

BREITKREITZ, M. C.; RAIMUNDO, I. M.; ROHWEDDER, J. J. R.; PASQUINI, C.; DANTAS FILHO, H. A.; JOSÉ, G. E.; ARAÚJO, M. C. U. Determination of total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration. *Analyst*, **128**, 1204-1207, 2003.

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

BRERETON, R. G. **Chemometrics**: Data analysis for the laboratory and chemical plant. Chichester: John Wiley & Sons Ltda, 2003. 489 p.

COLATI, K. A. P. **Extração e Caracterização de Ácidos Naftênicos por Espectrometria de Massas de Altíssima Resolução e Exatidão (ESI-FT-ICR MS)**. Dissertação de Mestrado. Departamento de Química da Universidade Federal do Espírito Santo, Vitória, 2012. 19 p.

DIAS, J. C. M. **Estudo exploratório das propriedades de caracterização de petróleos brasileiros**. Dissertação de Mestrado. Departamento de Química da Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005. 10 p.

FALLA, F. S.; LARINI, C.; LE ROUX, G. A. C.; QUINA, F. H.; MORO, L. F. L.; NASCIMENTO, C. A. O. Characterization of crude petroleum by NIR. *J. Pet. Sci. Eng.*, **51**, 127-137, 2006.

FEUDALE, R. N.; WOODY, N. A.; TAN, H.; MYLES, A. J.; BROWN, S. D.; FERRÉ, J. Transfer of multivariate calibration models: a review. *Chemom. Intell. Lab. Syst.*, **64**, 181-192, 2002.

FILGUEIRAS, P. R.; ALVES, J. C. L.; SAD, C. M. S.; CASTRO, E. V. R.; POPPI, R. J.; Evaluation of trends in residuals of multivariate calibration models by permutation test. *Chemom. Intell. Lab. Syst.*, **133**, 33–41, 2014a.

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*, **116**, 123-130, 2014b.

FEARN, T.; RICCIOLI, C.; GARRIDO-VARO, A.; GUERRERO-GINEL, J. E. On the geometry of SNV and MSC. *Chemom. Intell. Lab. Syst.*, **96**, 22–26, 2009.

GABRIELSSON, J.; JONSSON, H.; AIRIAU, C.; SCHIDT, B.; ESCOTT, R.; TRYGG, J. OPLS methodology for analysis of pre-processing effects on spectroscopic data. *Chemom. Intell. Lab. Syst.*, **84**, 153–158, 2006.

GAN; FENG; RUAN; GUIHUA; JINYUAN, M. Baseline correction by improved iterative polynomial fitting with automatic threshold. *Chemom. Intell. Lab. Syst.*, **82** (1-2), 59, 2006.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. *Anal. Chim. Acta*, **185**, 1-17, 1986.

GRELET, C.; FERNÁNDEZ PIERNA, J. A.; DARDENE, P.; BAETEN, V.; DEHARENG, F. Standardization of milk mid-infrared spectra from a European dairy network. *J. Dairy Sci.*, **98**, 2150-2160, 2015.

GRUBER L. D. A.; DAMASCENO F. C.; CARAMÃO E. B.; JACQUES R. A.; GELLER A. M.; CAMPOS M. C. V. Ácidos Naftênicos no Petróleo. *Quim. Nova*, **35**, 1423-1433, 2012.

HONORATO, F. A.; BARROS, B. N.; MARTINS, M. N.; GALVÃO, R. K. H.; PIMENTEL, M. F. Transferência de calibração em métodos multivariados. *Quim. Nova*, **5**, vol. 30, 1301-1312, 2007.

ISO 12185. Crude petroleum and petroleum products -- Determination of density -- Oscillating U-tube method. Geneva: **International Organization for Standardization**, 1996.

JAPPER-JAAFAR, A.; BHASKOROB, P. T.; MIORC, Z. S. A new perspective on the measurements of wax appearance temperature: Comparison between DSC,

thermomicroscopy and rheometry and the cooling rate effects. *J. Pet. Sci. Eng.*, **147**, 672–681, 2016.

JINGYAN, L.; XIAOLI, C.; SONGBAI, T. Research on determination of total acid number of petroleum using mid-infrared attenuated total reflection spectroscopy. *Energy Fuels*, **26**, 5633–5637, 2012.

KOK, M. V.; LETOFFE, J. M.; CLAUDY, P. DSC and Rheometer investigations of crude oils. *J. Therm. Anal. Calorim.* **56**, 959–965, 1999.

LAXALDE, J.; RUCKEBUSCH, C.; DEVOS, O.; CAILLOL, N.; WAKL, F.; DUPONCHEL, L. Characterization of heavy oils using near-infrared spectroscopy: optimization of pre-processing methods and variable selection. *Anal. Chim. Acta*, **705**, 227-234, 2011.

LIU, X.; HAN, L. J.; YANG, Z. L. Transfer of near infrared spectrometric models for silage crude protein detection between different instruments. *J. Dairy Sci.*, **94**, 5599–5610, 2011.

MACHO, S.; LARRECHI, M. S. Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry. *Trends in Analytical Chemistry*, **21**, 779–806, 2002.

MARCHESINI, F. H.; ALICKE, A. A.; MENDES, P. R. S.; ZIGLIO, C. M. Rheological Characterization of Waxy Crude Oils: Sample Preparation. *Energy Fuels*, **26**, 2566-2577, 2012.

MARTENS, H.; NAES, T. **Multivariate Calibration**. Guildford: John Wiley & Sons Ltda, 1991. 419 p.

MELÉNDEZ, L. V.; LACHE, A.; ORREGO-RUIZ, J. A.; PACHÓN, Z.; MEJÍA-OSPINO, E. Prediction of the SARA analysis of Colombian crude oils using ATR-FTIR spectroscopy and chemometric methods. *J. Pet. Sci. Eng.*, **90-91**, 56–60, 2012.

MEVIK, B. H.; CEDERKVIST, H. R. Mean squared error of prediction (MSEP) estimates for Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR). *J. Chemom.*, **18**, 422-429, 2004.

MÜLLER, A. L. H.; PICOLATO, R. S.; MELLO, P. A.; FERRÃO, M. F.; SANTOS, M. F. P.; GUIMARÃES, R. C. L.; MÜLLER, E. I.; FLORES, E. M. M. Total sulfur determination in residues of crude oil distillation using FT-IR/ATR and variable selection methods. *Spectrochim. Acta, Part A*, **89**, 82-87, 2012.

PARISOTTO, G. **Determinação do número de acidez total em resíduos de destilação atmosférica e de vácuo do petróleo empregando a espectroscopia no infravermelho (ATR-FTIR) e calibração multivariada**. Dissertação de Mestrado. Departamento de Química da Universidade Federal de Santa Maria, Santa Maria, 2007. 4-5 p.

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

PEINDER, P.; VISSER, T.; WAGEMANS, R.; BLOMBERG, J.; CHAABANI, H.; SOULIMANI, F.; WECKHUNYSEN, B. M. Sulfur speciation of crude oils by partial least squares regression modeling of their infrared spectra. *Energy Fuels*, **24**, 557-562, 2010.

PEREIRA, R. C. C.; SKROBOT, V. L.; CASTRO, E. V. R.; FORTES, I. C. P.; PASA, V. M. D. Determination of gasoline adulteration by principal components analysis-linear discriminant analysis applied to FTIR spectra. *Energy Fuels*, **20**, 1097-1102, 2006.

RIAZI, M. R. **Characterization and properties of petroleum fractions**. 1 ed. Baltimore: American Society for Testing and Materials (ASTM), 2005.

ROCHA, J. T. C.; OLIVEIRA, L. M. S. L.; DIAS, J. C. M.; PINTO, U. B.; MARQUES, M. L. S. P.; OLIVEIRA, B. P.; FILGUEIRAS, P. R.; CASTRO, E. V. R.; OLIVEIRA, M. A. L. Sulfur Determination in Brazilian Petroleum Fractions by Mid-infrared and Near-infrared Spectroscopy and Partial Least Squares Associated with Variable Selection Methods. *Energy Fuels*, **30**, 698-705, 2016.

SATYA, S.; ROEHNER, R. M.; DEO, M. D.; HANSON, F. V. Estimation of properties of crude oil residual fractions using chemometrics. *Energy Fuels*, **21**, 998-1005, 2007.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, **36**, 1627-1639, 1964.

SKOOG, D. A.; LEARY, J. J. **Principles of instrumental analysis**. Orlando: Saunders College Publishing, 4 ed., 1992. 700 p.

SKROBOT, V. L.; CASTRO, E. V. R.; PEREIRA, R. C. C.; PASA, V. M. D.; FORTES, I. C. P. Identification of Adulteration of gasoline applying multivariate data analysis techniques HCA and KNN in chromatographic data. *Energy Fuels*, **19**, 2350-2356, 2005.

SKROBOT, V. L.; CASTRO, E. V. R.; PEREIRA, R. C. C.; PASA, V. M. D.; FORTES, I. C. P. Use of principal component analysis (PCA) and linear discriminant analysis (LDA) in gas chromatographic (GC) data in the investigation of gasoline adulteration. *Energy Fuels*, **21**, 3394-3400, 2007.

SLAVCHEVA, E.; SHONE, B.; TURNBULL, A. Review of naphthenic acid corrosion in oil refining. *Br. Corros. J.*, **34**, No. 2, 125-131, 1999.

SOARES, I.P.; REZENDE, T.F.; SILVA, R.C.; CASTRO, E.V.R.; FORTES, I.C.P. Multivariate Calibration by Variable Selection for Blends of Raw Soybean

Oil/Biodiesel from Different Sources Using Fourier Transform Infrared Spectroscopy (FTIR) Spectra Data. *Energy & Fuels*, **22**, 2079-2083, 2008.

SOUJHI, N.; LINDEGREN, A.; ERIKSSON, L.; TRYGG, J. OPLS in batch monitoring - Opens up new opportunities. *Anal. Chim. Acta*, **857**, 28-38, 2015.

SPEIGHT, J. G. **Handbook of Petroleum Analysis**. John Wiley & Sons, Inc., New York, 2001, p. 43.

TAYLOR, J. R. **An introduction to error analysis**: the study of uncertainties in physical measurements. Sausalito: University Science Books, 2. ed., 1997, p. 327.

TRYGG, J.; WOLD, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.*, **16**, 119-128, 2002.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J.; Estado da arte de figuras de mérito em calibração multivariada. *Química Nova*, **32**, 1278-1287, 2009.

VAN DER VOET, H. Pseudo-degrees of freedom for complex predictive models: the example of partial least squares. *J. Chemom.*, **13**, 195-2008, 1999.

VENKATESAN, R.; NAGARAJAN, N. R.; PASO, K.; SASTRY, A. M.; FOGLER, H.S. The strength of paraffin gels formed under static and flow conditions. *Chem. Eng. Sci.*, **60**, 3587-3598, 2005.

VERRON, T.; SABATIER, R.; JOFFRE, R. Some theoretical properties of the O-PLS method. *J. Chemom.*, **18**, 62-68, 2004.

WANG, Y.; VELTKAMP, D. J.; KOWALSKI, B. R. Multivariate instrument standardization. *Anal. Chem.*, **63**, 2750-2756, 1991.

WINDIG, W.; STEPHENSON, D. A. Self-modeling mixture analysis of second-derivative near-infrared spectral data using the simplisma approach. *Anal. Chem.*, **6**, 2735-2742, 1992.

WOLD, S.; SJÖSTRÖM, M. Chemometrics, present and future success. *Chemom. Intell. Lab. Syst.*, **44**, 3-14, 1998.