

ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO

GUILHERME CARMELO UNGAR

Separação e Caracterização de Frações de Petróleo (SARA) para  
Modelagem de Unidade de Coqueamento Retardado

São Paulo

2012

ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO

GUILHERME CARMELO UNGAR

Separação e Caracterização de Frações de Petróleo (SARA) para  
Modelagem de Unidade de Coqueamento Retardado

Trabalho de conclusão de curso  
apresentado como parte dos  
requisitos para obtenção do título de  
Bacharel em Engenharia Química

Orientadoras:  
Rita Maria de Brito Alves, Ph.D.  
Maria Anita Mendes, Ph.D.

São Paulo

2012

## Índice

Tabelas e Figuras.....	5
Resumo.....	8
<b>Abstract</b> .....	9
<b>1. Introdução</b> .....	10
<b>2. Revisão Bibliográfica</b> .....	12
<b>2.1. Química do petróleo</b> .....	12
<b>2.2. Coqueamento Retardado</b> .....	16
<b>2.3. Cromatografia HPLC</b> .....	19
<b>2.4. Espectrometria MALDI-TOF</b> .....	21
<b>2.5. Espectrometria de Fluorescência</b> .....	23
<b>2.6. Regressão em Componentes Principais (PCR)</b> .....	25
2.6.1. Cálculo dos Componentes Principais.....	25
2.6.2. Regressão multilinear.....	27
<b>2.7. Regressão por Mínimos Quadrados Parciais (PLSR)</b> .....	29
<b>3. Procedimento Experimental</b> .....	32
<b>3.1. Separação dos Óleos em Frações (SARA)</b> .....	33
3.1.1. Precipitação dos Asfaltenos.....	34
<b>3.2. Análise por Espectrometria MALDI-TOF</b> .....	35
<b>3.3. Caracterização por Espectrometria de Fluorescência</b> .....	36
<b>4. Análise de Dados</b> .....	36
<b>4.1. Preparação de Dados e Seleção de Variáveis</b> .....	36
4.1.1. Dados dos Espectros LDI-TOF.....	37
4.1.2. Dados de Fluorescência em Estado Estacionário.....	37
<b>4.2. Regressão por Componentes Principais (PCR)</b> .....	38
<b>4.3. Regressão por Mínimos Quadrados Parciais (PLSR)</b> .....	39
<b>5. Resultados e Discussão</b> .....	39
<b>5.1. Espectros LDI-TOF</b> .....	39
<b>5.2. Regressões e Correlações</b> .....	41
5.2.1. PCR para os dados de Fluorescência.....	41
5.2.2. PLSR através do Matlab.....	44
5.2.3. PLSR para os dados de Fluorescência.....	46

5.2.4. PLSR para os dados do LDI-TOF .....	48
6. Conclusões .....	51
7. Referências Bibliográficas .....	52
8. Anexos .....	55
8.1. Anexo 1: Amostras utilizadas.....	55
8.2. Anexo 2: Espectros de MALDI-TOF das amostras de Petróleo, RAT e RV de 25/02/11, 02/03/11, 29/03/11 e amostras de RV de .....	57
8.3. Anexo 3: Resultados do Minitab para regressão do PC1 e PC2 (Fluorescência) e predição do °API.....	67
8.5. Anexo 5: Resultados do PLSR/Fluorescência com o uso de 10 componentes.....	70
8.6. Anexo 6: Resultados do PLSR/LDI-TOF com o uso de 13 componentes	75

## Tabelas e Figuras

Tabela 1: Frações típicas de petróleo .....	13
Tabela 2: Classificação dos petróleos segundo a API .....	16
Tabela 3: Mapeamento dos espectros a serem comparados.....	39
Tabela 4: Auto-valores calculados para os dados de fluorescência. ....	41
Tabela 5: Matriz dos scores calculados dos componentes do PCR para as 31 amostras.....	42
Tabela 6: Detecção de outliers no PCR para predição do grau API.....	44
Tabela 7: Matrizes de variáveis preditoras e de resposta carregadas no Matlab .....	45
Tabela 8: Coeficientes de determinação ( $R^2$ ) para PLSR/Fluorescência e 5 componentes .....	47
Tabela 9: Coeficientes de determinação ( $R^2$ ) para PLSR/Fluorescência e 10 componentes .....	48
Tabela 10: Coeficientes de determinação ( $R^2$ ) para PLSR/LDI-TOF com 5, 8 e 13 componentes.....	49
Tabela 11: Amostras utilizadas nas modelagens .....	55
Tabela 12: Amostras utilizadas apenas para comparação dos espectros LDI-TOF. ...	56
Figura 1: Fluxograma de processo de coqueamento retardado com as principais operações.....	19
Figura 2: Esquema de funcionamento de um equipamento de HPLC.....	21
Figura 3: Fluxograma do processo de Destilação e Coqueamento Retardado da RPBC .....	32
Figura 4: Esquema de separação Saturados, Aromáticos, Resinas e Asfaltenos (SARA) .....	34
Figura 5: Espectros de Fluorescência em Estado Estacionário para amostras 1, 2, 3, 15 e 18. ....	38
Figura 6: Histograma de distribuição de resíduos para o PCR em relação ao grau API. ....	43
Figura 7: MSEP em função do número de componentes (PLSR/Fluorescência) .....	46
Figura 8: Variância acumulada em Y em função do número de componentes (PLSR/Fluorescência).....	47
Figura 9: MSEP em função do número de componentes (PLSR/LDI-TOF).....	49
Figura 10: 25/02/11 – Petróleo UC (Amostra 46).....	57
Figura 11: 25/02/11 – RAT UC (Amostra 26) .....	57
Figura 12: 25/02/11 – RV (Amostra 22) .....	57
Figura 13: 25/02/11 – Petróleo UN (Amostra 19) .....	58
Figura 14: 25/02/11 – RAT UN (Amostra 23) .....	58
Figura 15: 25/02/11 – RV (Amostra 22) .....	58
Figura 16: 25/02/11 – Petróleo UV (Amostra 25) .....	59
Figura 17: 25/02/11 – RAT UV (Amostra 21) .....	59
Figura 18: 25/02/11 – RV (Amostra 22) .....	59
Figura 19: 02/03/11 Petróleo UC (Amostra 47) .....	60
Figura 20: 02/03/11 RAT UC (Amostra 44) .....	60

Figura 21: 02/03/11 RV (Amostra 70) .....	60
Figura 22: 02/03/11 Petróleo UN (Amostra 48) .....	61
Figura 23: 02/03/11 RAT UN (Amostra 69) .....	61
Figura 24: 02/03/11 RV (Amostra 70) .....	61
Figura 25: 02/03/11 Petróleo UV (Amostra 38) .....	62
Figura 26: 02/03/11 RAT UV (Amostra 73) .....	62
Figura 27: 02/03/11 RV (Amostra 70) .....	62
Figura 28: 29/03/11 Petróleo UC (Amostra 60) .....	63
Figura 29: 29/03/11 RAT UC (Amostra 30) .....	63
Figura 30: 29/03/11 RV (Amostra 29) .....	63
Figura 31: 29/03/11 Petróleo UN (Amostra 56) .....	64
Figura 32: 29/03/11 RAT UN (Amostra 32) .....	64
Figura 33: 29/03/11 RV (Amostra 29) .....	64
Figura 34: 29/03/11 Petróleo UV (Amostra 53) .....	65
Figura 35: 29/03/11 RAT UV (Amostra 34) .....	65
Figura 36: 29/03/11 RV (Amostra 29) .....	65
Figura 37: 28/10/10 RV (Amostra 10) .....	66
Figura 38: 22/10/10 RV (Amostra 12) .....	66
Figura 39: 25/02/11 Petróleo (Amostra 19) .....	66
Figura 40: Variância acumulada na matriz Y (PLSR/Fluorescência) .....	70
Figura 41: Variância acumulada na matriz X (PLSR/Fluorescência) .....	70
Figura 42: MSEP em função do número de componentes (PLSR/Fluorescência) .....	71
Figura 43: API calculado em função do API experimental (PLSR/Fluorescência) com 10 componentes .....	71
Figura 44: Teor de asfaltenos calculado em função do experimental (PLSR/Fluorescência) com 10 componentes .....	72
Figura 45: RC calculado em função do RC experimental (PLSR/Fluorescência) com 10 componentes .....	72
Figura 46: Distribuição de Resíduos para o modelo do API (PLSR/Fluorescência) com 10 componentes .....	73
Figura 47: Distribuição de Resíduos para o modelo do teor de asfaltenos (PLSR/Fluorescência) com 10 componentes .....	73
Figura 48: Distribuição de Resíduos para o modelo do RC (PLSR/Fluorescência) com 10 componentes .....	74
Figura 49: Variância acumulada na matriz Y (PLSR/LDI-TOF) .....	75
Figura 50: Variância acumulada na matriz X (PLSR/LDI-TOF) .....	75
Figura 51: MSEP em função do número de componentes (PLSR/LDI-TOF) .....	76
Figura 52: API calculado em função do API experimental (PLSR/LDI-TOF) com 13 componentes .....	76
Figura 53: Teor de asfaltenos calculado em função do experimental (PLSR/LDI-TOF) com 13 componentes .....	77
Figura 54: RC calculado em função do RC experimental (PLSR/LDI-TOF) com 13 componentes .....	77
Figura 55: Distribuição de Resíduos para o modelo do API (PLSR/LDI-TOF) com 13 componentes .....	78

Figura 56: Distribuição de Resíduos para o modelo do teor de asfaltenos (PLSR/LDI-TOF) com 13 componentes .....	78
Figura 57: Distribuição de Resíduos para o modelo do RC (PLSR/LDI-TOF) com 13 componentes .....	79

## Resumo

O projeto em que este Trabalho de Conclusão de Curso se inclui tem por objetivo o desenvolvimento de modelo matemático para avaliação da Unidade de Processo de Coqueamento Retardado da RPBC (Refinaria Presidente Bernardes). A dificuldade em obter uma modelagem e simulação com boa representatividade da unidade industrial está na qualidade dos dados e grau de detalhe dos modelos utilizados, os quais são altamente empíricos. A proposta do projeto é estudar as incertezas, validades de correlações e novas estratégias de modelagem e de otimização de uma unidade de coqueamento retardado de maneira a atender satisfatoriamente os rendimentos e qualidade de produtos requeridos.

Assim, foram realizadas análises espectrométricas das principais correntes da unidade de modo a caracterizá-las com maior nível de detalhe molecular, em especial, a fração de componentes pesados denominados asfaltenos. Através destas análises foi possível correlacionar as propriedades químicas da carga com as características específicas do produto desejado.

O trabalho compreendeu a pesquisa bibliográfica, para melhor conhecimento de assuntos relevantes (por exemplo: características do petróleo, espectrometria, estatística multivariada, etc.), seguida da preparação das amostras segundo a metodologia estudada, da obtenção dos espectros pela técnica de Espectrometria LDI-TOF (para os asfaltenos) e Espectrometria de Fluorescência (para amostras de petróleo), e da obtenção de correlações entre os espectros obtidos e as propriedades das amostras medidas diretamente na refinaria (densidade API, teor de asfaltenos e Resíduo de Carbono).

Os modelos obtidos pela técnica do PCR para os dados de fluorescência não forneceram correlações satisfatórias. Já os modelos obtidos por PLSR, tanto para Fluorescência quanto para LDI-TOF, apresentaram correlações bastante interessantes.

## Abstract

The project which includes this final paper aims to develop a mathematical model to evaluate the Delayed Coking process unit at RPBC (Presidente Bernardes Refinery). The challenge to obtain a very representative modeling and simulation depends on data quality and degree of models' detail, factors essentially empirical. So, in order to satisfactorily comply with the income and the products quality required, this proposal is based on uncertainty study, correlations validity and new strategies for modeling and optimization of a delayed coking unit.

Thus, some unit's streams spectrometric analyses were performed for the purpose of characterizing them with a higher level of molecular detail, mainly the heavy fraction of compounds known as asphaltenes. Through this analysis it should be possible to correlate the bulk chemical properties of streams with the specific characteristics of the desired product.

This work embraced a library research, to improve the knowledge about relevant related issues (for example: oil characteristics, spectroscopy, multivariate statistics analysis, etc.). Then, this steps were followed: samples preparation according to the studied methodology (SARA), spectra obtainment using LDI-TOF Mass Spectrometry (for asphaltenes) and Fluorescence Spectrometry (for oil samples only) and, finally, achievement of correlations between the obtained spectra and the samples properties measured directly at the refinery (API density, asphaltene percentage and of Carbon Residue).

The models obtained through the PCR technique to Fluorescence data didn't provide satisfactory correlations. Nevertheless, the models obtained through PLSR for both Fluorescence and LDI-TOF showed very interesting correlations.

## 1. Introdução

A utilização de petróleos pesados tem se tornado a cada dia mais importante, uma vez que o suprimento mundial de petróleo convencional vem diminuindo. Na literatura é possível encontrar comparações entre as composições dos óleos convencionais e as composições de óleos pesados. Os óleos convencionais apresentam uma maior quantidade de saturados (acima de 30%) e baixíssima quantidade de asfaltenos (menor que 10%) (1).

A conversão destes óleos em combustíveis de alta qualidade e maior valor agregado depende de uma combinação de processos, dentre eles o hidrotreatamento, o hidrocraqueamento e o coqueamento retardado. Os catalisadores utilizados nos processos catalíticos sofrem desativação mais rapidamente no processamento destes óleos mais pesados, devido à presença de grandes quantidades de compostos de nitrogênio residuais de etapas anteriores. Dessa forma, é preciso adaptar os processos de refino, para que as especificações necessárias para a alimentação dos processos seguintes sejam atingidas.

É preciso então definir as diferenças entre petróleos convencionais e não convencionais quanto às suas propriedades químicas. O conhecimento dessas composições permite, além de otimizar as etapas de refino, uma melhor avaliação da qualidade dos produtos formados. Um dos problemas no controle da refinaria é a considerável variação verificada nas propriedades do petróleo alimentado e, conseqüentemente, a variação da qualidade dos produtos obtidos. Isto ocorre pois são utilizados óleos crus de diversas fontes, misturados e sujeitos à estratificação nos tanques de armazenamento (2).

A caracterização de óleos também é necessária por outros motivos. O comportamento do petróleo durante seu armazenamento, transporte e processo de refino depende fortemente da natureza do óleo cru (3), a presença de componentes incompatíveis pode ocasionar inúmeros problemas como entupimento de dutos, precipitação, deposição, assim como quebra de equipamentos de processo (4).

Para a caracterização do petróleo, primeiramente é necessário separá-lo em suas frações. Uma metodologia denominada SARA é bastante utilizada

com este intuito (1,5,6). As frações obtidas por esta separação são compostas predominantemente por compostos saturados, aromáticos, resinas e asfaltenos, respectivamente. Os asfaltenos são compostos aromáticos de grande cadeia molecular, que se encontram na emulsão do petróleo. São insolúveis em n-heptano e, dessa forma, a primeira etapa da separação consiste na precipitação dos asfaltenos em n-heptano. As frações que permanecem solubilizadas, denominadas maltenos, são posteriormente separadas em saturados, aromáticos e resinas através de Cromatografia Líquida de Alta Performance (HPLC – High-performance Liquid Chromatography) (1,5).

Após a separação, as frações SARA são submetidas a diversas técnicas de caracterização. Para a determinação de pesos moleculares (MW) e distribuições de pesos moleculares (MWD) pode ser utilizada a GPC – Gel Permeation Chromatography, baseada na exclusão das moléculas por tamanho quando estas passam por leitos com poros de diferentes tamanhos. Também são frequentemente realizadas diferentes técnicas de Espectrometria (como a LDI-TOF e a Fluorescência) e a RMN – Ressonância Magnética Nuclear que possibilita a determinação de estruturas moleculares (1,5). Uma boa combinação destas técnicas e da interpretação de seus resultados possibilita a obtenção de parâmetros moleculares importantes.

Dessa forma, deseja-se com este trabalho caracterizar os óleos retirados da refinaria e relacionados com o processo de coqueamento retardado, com o objetivo de criar modelos e correlações matemáticas entre os resultados da caracterização e propriedades físicas facilmente medidas nas plantas (densidade API, teor de asfaltenos, Resíduo de Carbono). Para isso, amostras de óleos de pontos diferentes da refinaria, anteriores ao coqueamento, foram retiradas em datas e horários diferentes.

## 2. Revisão Bibliográfica

### 2.1. Química do petróleo

Petróleo é o nome dado ao mineral formado por uma mistura de hidrocarbonetos e impurezas, utilizada há muito tempo pela humanidade. Para se ter uma idéia, arqueólogos encontraram indícios de sua extração e utilização entre 6000 a.C. e 5000 a.C.. Gregos e romanos o utilizavam para fins bélicos, os egípcios para embalsamar mortos e pavimentar estradas. O óleo era retirado de exsudações naturais. Somente com a descoberta de um grande poço de petróleo da Pensilvânia em 1859, pelo coronel Edwin Drake, o desenvolvimento das técnicas de extração e refino intensificou-se e o petróleo passou a ser realmente interessante do ponto de vista comercial.

Esta mistura de substâncias apresenta caráter oleoso, inflamável, é menos densa que a água e apresenta odor característico (7). Uma propriedade muito importante do petróleo é a sua capacidade de queima, entrando na categoria dos minerais combustíveis de origem biológica (caustobiólitos), como o carvão mineral e a turfa.

A composição elementar do petróleo encontra-se nas seguintes faixas: carbono (84-87%); hidrogênio (12-14%); oxigênio, nitrogênio e enxofre (1-2%). Também é possível encontrar até 0,3% de metais (7). Em termos moleculares, o petróleo é composto principalmente de hidrocarbonetos saturados (parafinas e naftenos), aromáticos, resinas e asfaltenos. Estes compostos dificilmente conseguem ser purificados e, assim, a destilação consegue separá-los apenas dentro de certas faixas de temperaturas de ebulição. A Tabela 1, adaptada de (7, 8) mostra as frações típicas em que os compostos do petróleo conseguem ser separados em pressão atmosférica:

Tabela 1: Frações típicas de petróleo

<b>Fração</b>	<b>Temp. de Ebulição (°C)</b>
<b>Gás Residual; GLP</b>	Até 40
<b>Gasolina</b>	40 – 175
<b>Querosene</b>	175 – 235
<b>Gasóleo Leve</b>	235 – 305
<b>Gasóleo Pesado</b>	305 – 400
<b>Lubrificantes</b>	400 – 510
<b>Resíduo</b>	Acima de 510

As parafinas são hidrocarbonetos de cadeias lineares (normal) ou ramificadas (isômeros), que não possuem ligações duplas entre os átomos de carbono (cadeias saturadas). As parafinas mais leves e de cadeia linear apresentam-se em gases e ceras parafínicas. Os isômeros, de cadeias ramificadas, são encontrados mais comumente em frações mais pesadas retiradas dos óleos brutos. Em relação ao estado físico, são gasosos os compostos parafínicos com 1 a 4 carbonos, líquidos quando possuem de 5 a 17 carbonos e sólidos os que possuem cadeias ainda maiores. Os parafínicos são pouco propensos a participar das principais reações químicas de interesse industrial, sendo aproveitados principalmente em reações de oxidação e isomerização ou sulfurização catalítica.

Os naftenos também são hidrocarbonetos saturados, muito abundantes nos brutos (25 a 75%). Diferenciam-se das parafinas por serem compostos cíclicos. Os naftenos monocíclicos são os mais investigados, pois são o que se apresentam em maior quantidade nas frações leves do petróleo. Na gasolina, por exemplo, estão presentes tanto ciclopentano quanto ciclohexano substituídos. Embora se saiba que existam naftenos com até oito anéis, técnicas analíticas conseguiram determinar com certeza a estrutura dos compostos com no máximo cinco anéis. A reatividade dessas substâncias é

semelhante a das parafinas. Dessa forma, os naftenos sofrem reações de quebra da ligação C-C, desidrogenação ou aromatização.

Aromáticos são compostos também cíclicos, formados por um ou mais anéis benzênicos. Por exemplo, o naftaleno é uma substância formada por dois anéis de benzeno condensados. Compostos com três ou mais anéis são encontrados principalmente nas frações mais pesadas do petróleo bruto. Por apresentarem ligações insaturadas entre carbonos, sua reatividade é maior do que as parafinas e naftenos. A quantidade de aromáticos nos brutos varia de 15% a 50%, de forma que se apresenta em maior quantidade nas frações do petróleo com altas temperaturas de ebulição. São compostos que possuem anéis benzênicos substituídos, e por volta de 30 carbonos nos grupos substituintes.

Os asfaltenos são substâncias escuras solúveis em solventes aromáticos, mas insolúveis em parafínicos. São compostos extremamente complexos e difíceis de analisar, devido principalmente à sua tendência de agregação e às suas propriedades paramagnéticas (9). As resinas são estruturalmente parecidas com os asfaltenos, porém apresentam pesos moleculares menores. Tanto as resinas quanto os asfaltenos são compostos polares, característica interessante principalmente pensando-se na separação cromatográfica. As frações de compostos saturados, aromáticos e as resinas são denominadas maltenos (10).

Nitrogênio, enxofre e oxigênio estão presentes no petróleo principalmente como heteroátomos em compostos orgânicos das frações mais pesadas do petróleo, as resinas e os asfaltenos. Dessa forma, óleos de maior densidade tentem a apresentar maior quantidade desses três elementos. O enxofre ocorre principalmente na forma de sulfetos, benzotiofenos, moléculas policíclicas e gás sulfídrico, de forma que raramente se encontra enxofre elementar no petróleo. O nitrogênio ocorre em piridinas, pirróis, porfirinas e em compostos policíclicos. A presença do nitrogênio é a responsável pela maior quantidade de água em emulsão, pela formação de gomas durante o refino e pela contaminação dos catalisadores. O oxigênio aparece na forma de ácidos

carboxílicos, fenóis, ésteres, cetonas, dentre outras. Este elemento também é responsável pela formação de gomas, além da corrosividade do óleo.

Metais também podem ser encontrados no petróleo bruto, em teores de 0,01% a 0,04%. Os mais encontrados são: vanádio, níquel, ferro, zinco, mercúrio, boro, sódio, potássio, cálcio e magnésio.

Compostos insaturados como as olefinas (alcenos) não estão presentes nos óleos brutos. Em geral esses compostos se formam durante o refino, em processos térmicos ou catalíticos. São compostos mais reativos do que os saturados e aromáticos, o que explica sua ausência no petróleo.

Outra forma de classificação dos petróleos leva em conta a sua densidade. Esta densidade é expressa na forma de um parâmetro adimensional denominado *Grau API*. Este parâmetro é calculado a partir da seguinte equação (11):

$$API = \frac{141,5}{dr(60/60)} - 131,5 \quad (1)$$

onde  $dr(60/60)$  é a densidade do óleo a 60°F, relativa à densidade da água a 60°F. Pode-se perceber que, quanto maior a densidade de uma amostra, menor é o seu valor API. Como já visto, o teor de impurezas como enxofre, nitrogênio e oxigênio depende da densidade do óleo considerado. Assim como estes teores, outras características dos óleos estão diretamente ligadas à densidade, e portanto, ligadas ao Grau API. A Tabela 2 mostra a classificação dos óleos segundo a American Petroleum Institute (11):

Tabela 2: Classificação dos petróleos segundo a API

<b>Classificação</b>	<b>°API</b>	
<b>Petróleo Leve</b>	>31,1	
<b>Petróleo Médio</b>	22,3	–
	31,1	
<b>Petróleo Pesado</b>	10,0	–
	22,3	
<b>Petróleo Extra-Pesado</b>	<10,0	

A definição do que são petróleos convencionais ou não-convencionais é dinâmica e depende da tecnologia de exploração e refino disponível. Assim, óleos convencionais são aqueles facilmente extraídos e refinados. Os óleos não-convencionais, ao contrário, apresentam especificidades que dificultam seu processamento. Apenas refinarias com tecnologia adequada conseguem processar estes óleos. Estas especificidades podem ser: o alto nível de acidez, a alta concentração de compostos nitrogenados, alta densidade (API<10), dentre outras.

Outra variável utilizada para caracterizar o petróleo é o Resíduo de Carbono Conradson (RCC). Este valor é obtido através da degradação térmica de uma amostra do óleo (aquecimento em temperaturas acima de 500°C) e verificação da porcentagem de massa residual dessa degradação. O resíduo de carbono está diretamente ligado com a quantidade de material asfáltico do óleo analisado (12).

## **2.2. Coqueamento Retardado**

O Coqueamento Retardado é uma técnica de processamento *downstream* (que altera a estrutura molecular dos hidrocarbonetos). Sua função é transformar os resíduos de destilação (principalmente da destilação à vácuo), constituídos de moléculas de elevado peso molecular, em compostos mais leves e de maior valor agregado. Por este motivo é um dos processos denominados de “fundo de barril”. É um processo térmico e não-catalítico de

craqueamento das moléculas (8). Através dele obtêm-se, gás, nafta, combustíveis (principalmente diesel), gasóleo leve e pesado, bem como o coque de petróleo (ou coque verde).

Dentre as vantagens deste processo em relação a outros processos de craqueamento, estão o seu menor custo operacional (8) e a diminuição de óleos residuais produzidos pela refinaria. É também bastante utilizado quando a carga contém índices altos de metais - indesejáveis em processos catalíticos – já que o coque produzido concentra os metais presentes na carga.

O coque verde produzido é formado por carbono sólido e impurezas, e pode ser destinado à queima para produção de energia, ou à fabricação de eletrodos para a indústria metalúrgica (8). O coque calcinado (oxidado em altas temperaturas) é utilizado principalmente na fabricação de alumínio, também como eletrodo (13). A qualidade e o tipo de coque produzido (shot, agulha ou esponja) dependem fortemente da qualidade da alimentação.

O “shot-coke” é rico em asfaltenos e forma-se a partir da aglomeração das moléculas em estruturas esféricas. Apresenta também elevados teores de metais e enxofre. O “coque esponja” é mais poroso e caracteriza-se por ser rico em resinas e asfaltenos. É um coque de maior qualidade e que pode ser calcinado para a fabricação dos anodos da indústria de alumínio, dependendo do seu teor de matéria volátil (VCM – Volatile Combustible Matter) (14) e do teor de enxofre (menor que 2,5% em peso) (11). O “coque agulha” é de altíssima qualidade. Apresenta baixa quantidade de resinas, asfaltenos e metais. Ao mesmo tempo possui grande quantidade de aromáticos. São utilizados principalmente na indústria aeroespacial e na produção de substâncias como fósforo e dióxido de titânio (11).

A primeira etapa consiste na passagem da corrente de alimentação, pré-aquecida, por uma torre fracionadora, onde as frações mais leves ainda presentes no resíduo de vácuo são retiradas e apenas as frações mais pesadas seguem no processo. Estes compostos mais pesados então seguem para uma fornalha de tubos, que deve ser projetada para que a formação de coque nos tubos seja mínima. As temperaturas neste forno são mantidas por volta de 500°C, e o tempo de residência neste é menor do que em outros

processos térmicos semelhantes, para que o coqueamento não seja iniciado ainda no forno. Em seguida, o material segue para as câmaras de coqueamento, isoladas termicamente, onde as reações (de craqueamento e de formação do coque) efetivamente acontecem (15,11). As frações mais leves geradas na câmara incorporam-se ao vapor que retorna à torre fracionadora. Este vapor também contém amônia e gás sulfídrico, e na torre passam por etapas de remoção.

Dessa forma, os parâmetros chave deste processo são o tempo de residência, os perfis de temperatura e pressão, a razão de reciclo e as características da carga. Em relação às características da carga, existem modelos para estimativa de rendimentos dos produtos baseados no Resíduo de Carbono (16). Outros modelos semelhantes utilizam os asfaltenos como base para essas estimativas (17).

O coqueamento retardado é o único processo semicontínuo encontrado numa refinaria. Isto porque o coque produzido é acumulado nas câmaras e deve ser retirado quando estas estão cheias. Para isso utilizam-se técnicas de perfuração do material com jatos d'água. O fluxo material é desviado para outra câmara de coqueamento em paralelo, vapor d'água é injetado na câmara cheia para a purga dos hidrocarbonetos gasosos restantes, e água é posteriormente injetada para resfriá-lo antes da aplicação dos jatos (13,11).

Da fracionadora, a nafta segue para o processo de hidrotreatamento, os gasóleos leves são tratados e misturados no "blend" de gasóleo, os gasóleos pesados seguem para um novo processo de craqueamento e se transformam em frações mais leves (11).

O projeto do reator de coque deve levar em conta que seu volume deve ser adequado para o acúmulo do coque durante o período de enchimento, e que a velocidade superficial deve ser a máxima possível. Também é importante considerar o nível máximo do leito e as características da espuma que se forma.

A quantidade de coque produzido está diretamente relacionado à quantidade de frações pesadas na carga. Assim, alimentações do processo com petróleos mais pesados levam à formação de maior quantidade de coque

nas câmaras (11). A Figura 1, retirada de (11), mostra um fluxograma de um processo de coqueamento retardado, no qual conseguimos observar as duas câmaras de coqueamento em paralelo, e a diferença entre a temperatura de entrada da carga na torre fracionadora e a temperatura da carga antes de entrar nas câmaras. Observa-se também as altas pressões do óleo saído da fornalha, entre 1,5 barg e 7,0 barg.

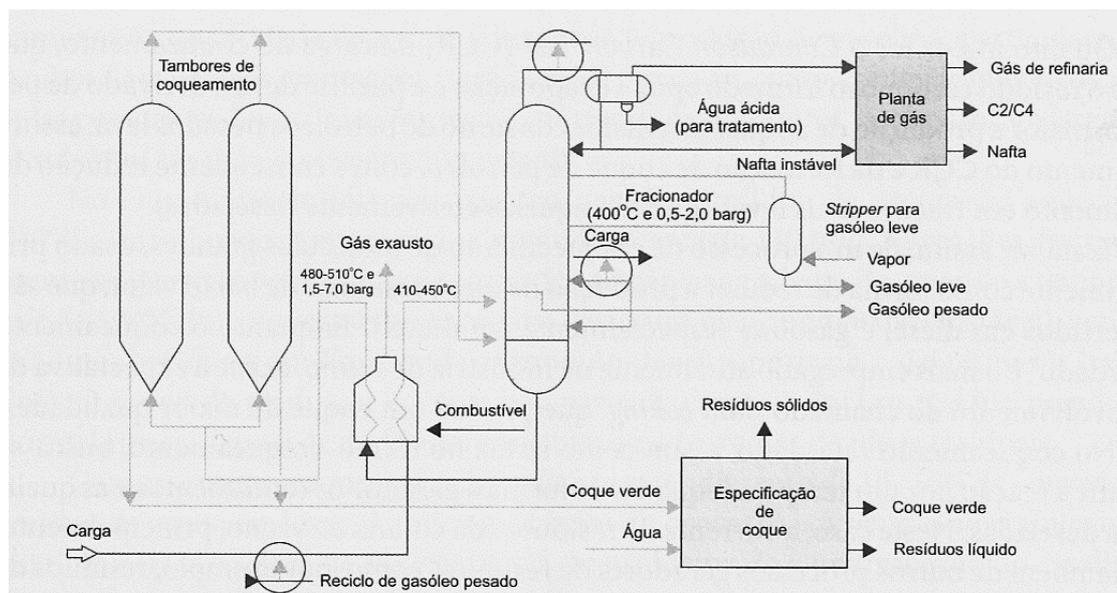


Figura 1: Fluxograma de processo de coqueamento retardado com as principais operações.

A corrente de entrada do forno é uma carga combinada de alimentação fresca (resíduo da destilação a vácuo) e corrente de reciclo. O reciclo é obtido na própria fracionadora, na condensação dos componentes de maior ponto de ebulição presentes no efluente do reator de coque. O reciclo possui valores altos de Resíduo de Carbono e de insolúveis em n-heptano. Altos valores para o reciclo resultam numa corrente de gasóleo pesado com menores valores de Resíduo de Carbono e teor de metais, pois estes componentes mais pesados (dentre eles também os metais) são convertidos em coque.

### 2.3. Cromatografia HPLC

A cromatografia é uma técnica de separação desenvolvida no início do século XX pelo botânico russo Mikhael Tswett. Este cientista desenvolveu um método que possibilitava a separação de pigmentos a partir de extratos de

folhas das plantas. O mais importante desenvolvimento desta técnica se deu a partir de 1970, com a “Cromatografia Líquida Moderna”, desenvolvida por Snyder e Kirkland. Nesta época é que a cromatografia passou a ter maior precisão, velocidade, e a possibilidade de realizar separações mais difíceis (18).

O princípio de funcionamento da cromatografia é a interação dos compostos em duas fases. Uma das fases é sólida, denominada *fase estacionária*, a outra fase é fluida, denominada *fase móvel*. A fase móvel é eluída através da fase estacionária, e os diferentes compostos das duas fases interagem em maior ou menor grau, causando a separação. Assim, aqueles compostos da fase móvel que apresentarem grande interação com a fase estacionária demoram um tempo maior para serem eluídos do que os compostos que interagem menos.

Existem vários mecanismos de separação utilizados: a *adsorção* do composto na superfície da fase estacionária; a *partição*, ou seja, a distribuição do analito entre as duas fases; a *troca iônica*, ou seja, a interação entre íons da fase móvel e cargas em excesso na fase estacionária, dentre outros. É comum o uso de fases estacionárias chamadas “fases ligadas”. Nestas fases estacionárias, utiliza-se um suporte (por exemplo, sílica), e acopla-se quimicamente uma molécula cuja interação com a fase móvel seja adequada .

As técnicas cromatográficas podem ser classificadas de diferentes formas, levando em consideração a geometria (em coluna ou planar), a fase móvel (líquida, gasosa ou supercrítica), ou mesmo a disposição da fase estacionária (empacotada ou tubular aberta) utilizadas.

A técnica a ser utilizada nas separações SARA é o HPLC (High Performance Liquid Chromatography), ou CLAE (Cromatografia Líquida de Alta Eficiência). É uma técnica de cromatografia moderna, e que se diferencia da cromatografia clássica pelo uso de elevadas pressões, que aumentam a eficiência de separação. Utiliza-se uma coluna em que está presente a fase estacionária, e por ela passa a fase móvel, que se movimenta devido a ação de uma bomba. Na cromatografia clássica, a coluna apresenta as extremidades abertas e, portanto, o movimento da fase móvel deve-se apenas à ação

gravitacional. A Figura 2, retirada de (19), ilustra esquematicamente o funcionamento do equipamento de HPLC:

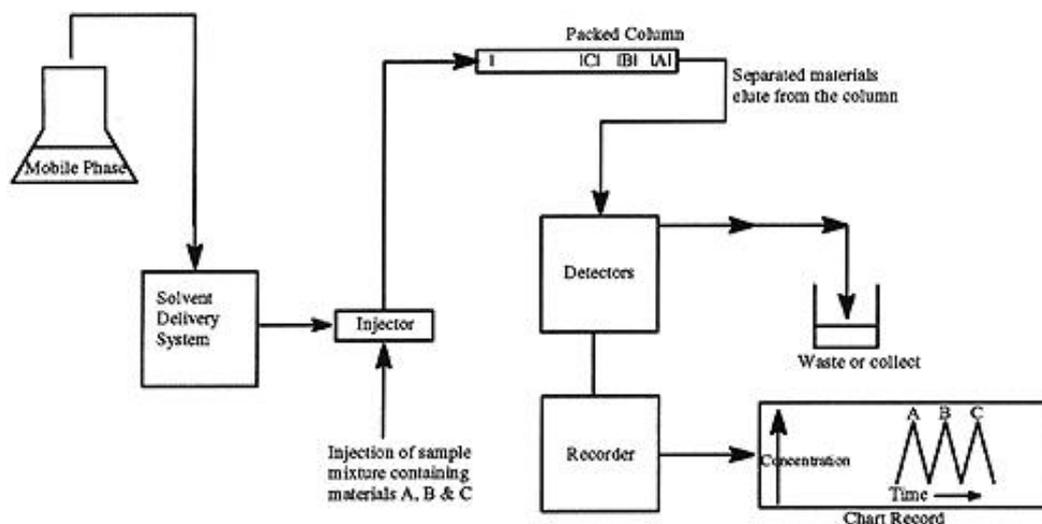


Figura 2: Esquema de funcionamento de um equipamento de HPLC.

Os compostos que saem da coluna passam por um detector, e o tempo que o composto demorou para atravessar a coluna é registrado. O composto, então, pode ser coletado após estas etapas. Ao fim da análise, obtém-se um gráfico denominado cromatograma, que mostra a concentração detectada em função do tempo.

A maior eficiência do HPLC em relação à cromatografia clássica também se deve ao uso de colunas com melhor qualidade, técnicas automatizadas de injeção de amostra no equipamento, controle da vazão de eluente, dentre outros fatores. Atualmente, pouco se utiliza a cromatografia clássica, e por isso, muitos pesquisadores já se referem ao HPLC simplesmente por LC (18).

#### 2.4. Espectrometria MALDI-TOF

Os asfaltenos são os compostos mais pesados e mais polares presentes nos óleos brutos e que se apresentam em estruturas coloidais. Embora seja importante o conhecimento das características dos asfaltenos para a sua adequada manipulação nos processos de refino, não se sabe o suficiente a respeito de suas estruturas moleculares, reatividades, dentre outras propriedades associadas. É possível encontrar na literatura diversos valores

para a massa molecular média desses compostos (20). A diversidade dos valores deve-se às especificidades de cada fonte de petróleo e também às diferenças das técnicas analíticas usadas. Os valores encontrados variam de 400uma a 10000uma (20).

Para alcançar este objetivo, devem ser utilizadas mais de uma técnica de determinação de massas moleculares, de forma a comparar os resultados de cada uma. Encontra-se na literatura a análise de asfaltenos por GPC – Gel Permeation Chromatography; Espectroscopia por Ressonância Magnética Nuclear; Espectroscopia de Massas por Dessorção a Laser; dentre outras técnicas.

A espectrometria de massas é considerada como um dos melhores métodos de determinação da distribuição da massa molecular dos asfaltenos, pois consegue detectar uma ampla faixa de massas. Diversos estudos utilizando a espectrometria de massas MALDI-TOF (Matrix-assisted laser desorption/ionization – Time of flight) para análise dos asfaltenos podem ser encontrados (20,21). A análise por MALDI-TOF é bastante comum para a identificação de macromoléculas biológicas, polímeros e frações de petróleo.

Esta técnica, introduzida em 1988 (10,21), baseia-se na ionização da amostra através da incidência de laser sobre elas. Encontra-se misturada à amostra uma substância denominada matriz, cujo objetivo é absorver a energia do laser e dispersá-la para moléculas da amostra, ionizando-as sem fragmentá-las. Dessa forma, a matriz melhora a detecção do sinal de amostras que absorveriam pouca energia diretamente do laser, e dificulta a detecção de massas menores que as reais devido à fragmentação.

Após a ionização, as moléculas carregadas eletricamente são dessorvidas, aceleradas através de um campo elétrico, e passam por uma coluna em que se mantém vácuo. O detector encontra-se na outra extremidade desta coluna. A relação massa/carga ( $m/z$ ) é determinada pelo detector TOF com base no tempo de vôo das moléculas através desta coluna. Os valores da intensidade do sinal detectado em função do valor  $m/z$  são registrados em um espectro.

É possível admitir que o espectro obtido corresponda à distribuição de massas moleculares da amostra analisada. Para isso, as condições de

operação do equipamento devem ajustadas de modo que se possa considerar uma mesma eficiência de ionização para todas as moléculas presentes na amostra e, além disso, que a ionização ocorra pelo recebimento de apenas um próton ( $H^+$ ) por molécula. Nessas condições, a ordenada do gráfico (intensidade de sinal) para um dado  $m/z$  é proporcional à quantidade do íon presente na amostra.

Existem dois modos de operação do equipamento MALDI-TOF, o linear e o reflectron. Enquanto, no modo linear, as moléculas ionizadas são detectadas ao fim da coluna, no modo reflectron, espelhos magnéticos são ativados e as moléculas são desviadas para outro detector. A vantagem do modo reflectron é o aumento da resolução dos sinais, isto é, consegue-se distinguir sinais de relação  $m/z$  mais próxima do que no modo linear.

Alguns parâmetros da análise por MALDI-TOF devem ser bem escolhidos, de forma a minimizar problemas de fragmentação ou polimerização das moléculas e melhorar a eficiência de ionização. Assim, é preciso escolher adequadamente a matriz a ser utilizada, a potência do laser, e o modo de operação. Em geral, costuma-se escolher matrizes, potências de laser e modos de operação já utilizados anteriormente em outros estudos com o mesmo composto, ou escolher dentre algumas possibilidades testadas. É possível que esse mesmo equipamento seja utilizado para analisar amostras que não sejam misturadas com uma matriz. Nestes casos, podemos denominar a técnica como LDI-TOF.

## **2.5. Espectrometria de Fluorescência**

Outra técnica analítica muito utilizada para caracterização de petróleo é a Espectrometria de Fluorescência. É uma técnica baseada na capacidade de um composto, denominado fluoróforo, em emitir radiação quando excitado por uma fonte (UV, raios X, etc.), em intervalo de tempo da ordem de  $10^{-9}$  s. Ao receber tal radiação elétrons passam do estado fundamental num átomo para o estado excitado (atingem uma camada eletrônica mais externa). Ao retornar ao estado fundamental, a energia inicialmente absorvida é liberada também sob a forma de radiação (2, 22).

De maneira simplificada, a fonte, que permanece em funcionamento durante a análise, é ajustada para um determinado comprimento de onda de excitação ( $\lambda_{exc}$ ), e o detector analisa a quantidade de fótons emitidos pela amostra excitada para uma determinada faixa de comprimentos de onda (espectro em estado estacionário). Se este procedimento for repetido para diferentes valores de  $\lambda_{exc}$ , obtém-se o Espectro de Fluorescência Total (EFT).

Também é bastante comum utilizar uma excitação em pulso com o  $\lambda_{exc}$ , e obter como resposta do detector a quantidade de fótons emitidos em função no tempo. Esta técnica corresponde à Fluorescência de Resolução Temporal (EFRT).

A fluorescência observada no petróleo se deve principalmente à presença de compostos aromáticos, e é muito influenciada pelas propriedades físicas dos óleos, como densidade, viscosidade e densidade óptica. Verifica-se que hidrocarbonetos saturados não apresentam significativa emissão de fluorescência, e que, dentre os aromáticos, os compostos menores e pouco substituídos apresentam um maior rendimento de emissão. Isto ocorre pois a fluorescência está relacionada com os elétrons de orbitais  $\pi$  em ligações duplas entre carbonos (22).

A diversidade da composição química do petróleo aumenta a complexidade das análises por fluorescência. Apesar disso, é uma análise comum na indústria do petróleo, utilizada para classificação de diferentes amostras.

Encontram-se na literatura estudos que utilizam a fluorescência para a caracterização de petróleo com comprimentos de onda de excitação na faixa de 300nm a 500nm, com obtenção dos espectros de emissão entre 300 nm e 700 nm (23).

## 2.6. Regressão em Componentes Principais (PCR)

A Regressão em Componentes Principais (PCR – Principal Component Regression) é uma técnica estatística multivariada baseada na Análise de Componentes Principais (PCA – Principal Component Analysis).

Dado um conjunto de  $p$  variáveis, o PCA gera combinações lineares dessas  $p$  variáveis com o objetivo de reduzir a dimensão desse conjunto (diminuir o número de variáveis estatisticamente significantes). A variabilidade total do conjunto de dados só pode ser totalmente descrita através de todos os  $p$  componentes gerados. Apesar disso, a maior parte dessa variabilidade é descrita por um pequeno número de componentes, os componentes principais. Escolhidos  $k$  componentes principais que representem uma dada porcentagem da variabilidade total do conjunto, temos a redução da dimensão dos dados de uma matriz de  $n$  observações por  $p$  variáveis, para uma nova matriz de  $n$  observações por  $k$  variáveis (com  $k < p$ ) (24).

O PCR envolve, após a determinação dos  $k$  componentes principais, uma regressão multilinear utilizando-se os valores calculados dos  $k$  componentes para cada uma das  $n$  observações (valores denominados *scores*) para obter uma equação de predição de uma variável  $Y$  que se deseja correlacionar (com valores também conhecidos para as  $n$  observações). Os cálculos envolvidos no PCR são apresentados resumidamente a seguir.

### 2.6.1. Cálculo dos Componentes Principais

Seja  $X_1, X_2, X_3, \dots, X_p$  um conjunto de  $p$  variáveis. Os componentes, como dito anteriormente, são  $p$  novas variáveis obtidas através de combinações lineares das variáveis originais. Geometricamente, considerando as  $p$  variáveis originais como eixos de um sistema de coordenadas, os  $p$  componentes consistem numa rotação desse sistema de coordenadas (24). Os novos eixos obtidos representam direções com máxima variabilidade. Será utilizado nos cálculos o vetor  $\mathbf{X}' = [X_1, X_2, X_3, \dots, X_p]$ , cuja matriz de covariância é  $\Sigma$  e cujos auto-valores são  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Temos as seguintes combinações lineares (24):

$$\begin{aligned}
C_1 &= \mathbf{a}'_1 \mathbf{X} = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\
C_2 &= \mathbf{a}'_2 \mathbf{X} = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\
&\vdots \\
C_p &= \mathbf{a}'_p \mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p
\end{aligned} \tag{2}$$

$$\text{Var}(C_i) = \mathbf{a}'_i \boldsymbol{\Sigma} \mathbf{a}_i \quad i = 1, 2, \dots, p \tag{3}$$

$$\text{Cov}(C_i, C_k) = \mathbf{a}'_i \boldsymbol{\Sigma} \mathbf{a}_k \quad i, k = 1, 2, \dots, p \tag{4}$$

Os componentes principais são obtidos quando essas combinações lineares  $C_1, C_2, \dots, C_p$  apresentam variâncias (eq.(3)) que sejam as máximas possíveis. Mostra-se que, sendo  $\boldsymbol{\Sigma}$  a matriz de covariância de  $\mathbf{X}$  com os pares de auto-valores e auto-vetores  $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ , o componente principal de ordem  $i$  é dado por:

$$C_i = \mathbf{e}'_i \mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X_p \quad i = 1, 2, \dots, p \tag{5}$$

$$\text{Var}(C_i) = \mathbf{e}'_i \boldsymbol{\Sigma} \mathbf{e}_i = \lambda_i \quad i = 1, 2, \dots, p \tag{6}$$

$$\text{Cov}(C_i, C_k) = \mathbf{e}'_i \boldsymbol{\Sigma} \mathbf{e}_k = 0 \quad i \neq k \tag{7}$$

Dessa forma, o algoritmo do PCA calcula os auto-valores e auto-vetores da matriz de covariância  $\boldsymbol{\Sigma}$ . Os valores de  $C_i$  calculados através da eq.(5) para cada uma das  $n$  observações são conhecidos como *scores*. Por sua vez, os coeficientes  $e_{ij}$  ( $j = 1, 2, \dots, p$ ) para o cálculo do  $i$ -ésimo componente são denominados *loadings*

Em lugar da matriz de covariância, a matriz de correlação  $\boldsymbol{\rho}$  pode ser utilizada e, neste caso, os componentes principais obtidos estão relacionados com as variáveis padronizadas (variáveis  $Z$ ):

$$Z_i = \frac{(X_i - \mu_i)}{\sqrt{\sigma_{ii}}} \quad i = 1, 2, \dots, p \quad (8)$$

$$\mathbf{Z} = \left( \mathbf{V}^{\frac{1}{2}} \right)^{-1} (\mathbf{X} - \boldsymbol{\mu})$$

$$V^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{\sigma_{11}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\sigma_{11}} \end{bmatrix} \quad (9)$$

### 2.6.2. Regressão multilinear

A análise de regressão é um método estatístico utilizado para prever valores de uma ou mais variáveis de *resposta* (variáveis dependentes), através de um conjunto de variáveis de *predição* (independentes) (26). Sendo Y uma variável de resposta e  $z_1, z_2, \dots, z_r$  as r variáveis independentes, um modelo de regressão linear tem a seguinte forma (24, 25):

$$Y = \beta_0 + \beta_1 z_1 + \dots + \beta_r z_r + \varepsilon \quad (10)$$

A eq.(10) indica que o valor de Y é uma combinação linear dos parâmetros  $\beta$ , cujos coeficientes são as variáveis de predição, além de um erro aleatório  $\varepsilon$ . Conhecidos os valores de  $z_i$  e Y para n observações, obtemos a seguinte equação, na forma matricial (24, 25):

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & z_{12} & \dots & z_{1r} \\ 1 & z_{21} & z_{22} & \dots & z_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & z_{n2} & \dots & z_{nr} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (11)$$

Ou:

$$\mathbf{Y} = \mathbf{Z} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (12)$$

Os erros  $\varepsilon_j$  devem ser tal que seus valores esperados sejam zero, que a a variância  $\sigma^2$  seja a mesma para todos, e que os erros para as observações não sejam correlacionados. Na forma de matrizes, estas restrições se apresentam na eq.(13):

$$\begin{aligned} E(\boldsymbol{\varepsilon}) &= \mathbf{0}; \\ \text{Cov}(\boldsymbol{\varepsilon}) &= E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\mathbf{I} \end{aligned} \quad (13)$$

Os parâmetros  $\beta$  devem ser estimados para que um modelo seja ajustado aos dados de  $Y$  e  $z$  existentes. Para isso, a técnica de estimativa por *mínimos quadrados* é bastante comum. Esta técnica se baseia na minimização de uma função  $S$  que compreende a soma dos quadrados dos erros em cada predição. O erro pode ser calculado da seguinte forma:

$$\varepsilon = Y - (\beta_0 + \beta_1 z_1 + \dots + \beta_r z_r) \quad (14)$$

Dada uma matriz  $\mathbf{b}$  de candidatos a coeficientes  $\beta_{ij}$ , temos que a seguinte função (erro quadrático) deve ser minimizada (24):

$$S(\mathbf{b}) = \sum_{j=1}^n (Y_j - (b_0 + b_1 z_{j1} + \dots + b_r z_{jr}))^2 = (\mathbf{Y} - \mathbf{Z}\mathbf{b})'(\mathbf{Y} - \mathbf{Z}\mathbf{b}) \quad (15)$$

Os valores de  $\mathbf{b}$  que minimizam  $S(\mathbf{b})$  são os valores estimados adotados, são denotados por  $\hat{\beta}$ , e podem ser calculados pela seguinte operação (24, 25):

$$\hat{\beta} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \quad (16)$$

Pode-se também realizar a regressão multilinear envolvendo mais de uma variável de resposta,  $Y_1, Y_2, \dots, Y_m$ . Neste caso, assume-se que cada resposta possui seu próprio modelo de regressão linear. A álgebra envolvida é completamente análoga à desenvolvida para apenas uma variável dependente. A diferença estará na dimensão da matriz das respostas  $\mathbf{Y}$  e dos erros  $\boldsymbol{\varepsilon}$ . Ambas possuíam dimensão  $n \times 1$  e passam a ter dimensão  $n \times m$ :

$$\mathbf{Y} = \begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1m} \\ Y_{21} & Y_{22} & \dots & Y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \dots & Y_{nm} \end{bmatrix}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \dots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22} & \dots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \dots & \varepsilon_{nm} \end{bmatrix} \quad (17)$$

Aplicando-se o conceito aqui desenvolvido sobre regressão multilinear ao PCA, tem-se que a matriz de preditores  $\mathbf{Z}$  corresponde à matriz  $\mathbf{C}$  (de dimensão  $n \times k$ ) com os scores dos  $k$  componentes principais calculados para as  $n$  observações (25).

## 2.7. Regressão por Mínimos Quadrados Parciais (PLSR)

A Regressão por Mínimos Quadrados Parciais (PLSR – Partial Least Squares Regression) é uma técnica de predição já bastante estudada para dados multivariáveis. É uma ferramenta especialmente importante quando se tem apenas um conhecimento parcial do funcionamento do conjunto de dados (por exemplo, na predição de certas propriedades em função de dados de espectrometria ou cromatografia) (26).

A regressão através do PLS costuma gerar resultados mais bem correlacionados do que através do PCA, apesar de seguir um princípio algébrico muito semelhante. Isto ocorre porque o cálculo dos *scores* e dos *loadings* das novas variáveis obtidas leva em conta não somente as variáveis independentes  $\mathbf{X}$ , mas também as variáveis dependentes  $\mathbf{Y}$ .

Diversos algoritmos para os cálculos do PLS são conhecidos e podem ser implementados. Todos se baseiam em modelos com duas relações *externas* ( $\mathbf{X}$  como combinação linear dos scores das novas variáveis e  $\mathbf{Y}$  como outra combinação linear semelhante) e uma relação *interna* (relação entre os scores das duas relações externas). As equações deste modelo, caso seja escolhido um número de componentes  $a \leq m$  (com  $m$  sendo o número de variáveis independentes), são as seguintes (27):

$$X = TP' + E = \sum t_h p_h + E \quad h = 1, 2, \dots, a \quad (18)$$

$$Y = UQ' + F^* = \sum u_h q_h + F^* \quad h = 1, 2, \dots, a \quad (19)$$

Onde **T** e **U** são as matrizes de scores para cada transformação linear, **P'** e **Q'** são os coeficientes de regressão, e **E** e **F\*** são os resíduos de cada equação. Esses resíduos devem ser minimizados, em especial o resíduo **F\***, uma vez que o objetivo da técnica é descrever **Y** o melhor possível.

Quanto às dimensões das matrizes na eq. (18), temos que **X** é de ordem  $n \times m$  ( $n$  observações e  $m$  variáveis originais); **T** é de ordem  $n \times a$  ( $n$  observações e  $a$  componentes escolhidos); **P'** é de ordem  $a \times m$ , e **E** também é de ordem  $n \times m$ . Na eq. (19) verificam-se as mesmas dimensões, exceto pelo fato de que **Y** não possui  $m$  variáveis mas  $p$ . Em geral,  $p \leq m$ .

A relação interna é frequentemente uma equação que envolve **U** e **T**, através de seus vetores coluna. A relação mais simples que pode ser utilizada é apresentada na eq. (20) e mostra que o vetor vertical  $u_h$ , uma coluna da matriz de scores **U**, é estimado através da multiplicação de um vetor de parâmetros  $b_h$  pelo vetor vertical  $t_h$  (coluna da matriz de scores **T**):

$$\hat{u}_h = b_h t_h \quad h = 1, 2, \dots, a \quad (20)$$

Nos algoritmos outras matrizes são também utilizadas, de forma a estimar os loadings e os scores de um determinado número  $a$  de componentes. O principal algoritmo é denominado NIPALS-PLS (25, 27). Uma alternativa mais robusta baseada no próprio NIPALS-PLS é proposta pelo algoritmo SIMPLS (27). O algoritmo SIMPLS já se apresenta implementado em uma função do software Matlab, encontrada no pacote *Statistics Toolbox*.

Para os cálculos efetuados pelo SIMPLS, não somente as eqs. (18) a (20) são utilizadas, mas também outras equações envolvendo vetores auxiliares e vetores de pesos. Apesar disso, uma matriz de coeficientes **B<sub>PLS</sub>** é calculada de

forma que a predição dos valores de  $\mathbf{Y}$  pode ser realizada pela simples operação matricial (27):

$$\hat{\mathbf{Y}}_0 = \mathbf{X}_0 \mathbf{B}_{PLS} \quad (21)$$

Onde  $\mathbf{X}_0$  e  $\hat{\mathbf{Y}}_0$  correspondem às matrizes  $\mathbf{X}$  e  $\mathbf{Y}$ , porém centradas na média. O último passo da predição é retornar a matriz  $\hat{\mathbf{Y}}_0$  para a sua média original, e realizar a operação inversa à normalização (caso  $\mathbf{X}$  e  $\mathbf{Y}$  tenham sido normalizadas antes da regressão).

Tão importante quanto o próprio cálculo do modelo é a validação do mesmo. Validar o modelo leva em conta três aspectos principais (26): escolha do número de componentes adequado para caracterizar o conjunto de dados; a adequação dos valores preditos pelo modelo e a representatividade dos dados coletados e utilizados na modelagem.

A principal forma de validação de modelo consiste na separação de um conjunto de observações das variáveis, de forma que esse conjunto não participa do ajuste do modelo pelo PLSR, mas participa de sua validação (test-set). O conjunto de observações utilizado no ajuste é denominado “learning-set”. Na validação, as variáveis  $\mathbf{Y}$  são calculadas para as observações do test-set, e os erros (desvios em relação ao valor experimental) médios quadráticos (MSE – Mean Square Error) são obtidos. Tais erros devem ser mínimos.

Outra maneira de se validar o modelo e obter um erro mínimo de predição é através da validação cruzada (CV – Cross-Validation). Dado um valor  $k$ , as observações são separadas em  $k$  grupos de maneira aleatória. Define-se o modelo utilizando  $k-1$  grupos de observações, e utiliza-se o último grupo para o cálculo dos erros quadráticos de predição. Um comportamento muito comum e interessante da validação cruzada pode ser observado: conforme aumenta o número de componentes na regressão, o erro quadrático médio cai (nos primeiros componentes), passa por um mínimo e cresce novamente até atingir um valor quase constante para um grande número de componentes. Isto permite escolher o número de componentes a ser utilizado como aquele que

apresentou o menor valor de MSE. Utilizando-se um grande número de componentes, verifica-se que o ajuste do modelo melhora muito, gerando coeficientes de determinação ( $R^2$ ) muitíssimo próximos de 1,0. Apesar disso, não são bons modelos, porque levam em conta praticamente toda a variabilidade do sistema, inclusive ruídos (em espectros ou cromatogramas) e erros experimentais.

### 3. Procedimento Experimental

As amostras preparadas e analisadas foram retiradas de sete pontos da refinaria (RPBC – Refinaria Presidente Bernardes – Petrobras), como indica a Figura 3, que ilustra o fluxograma do processo:

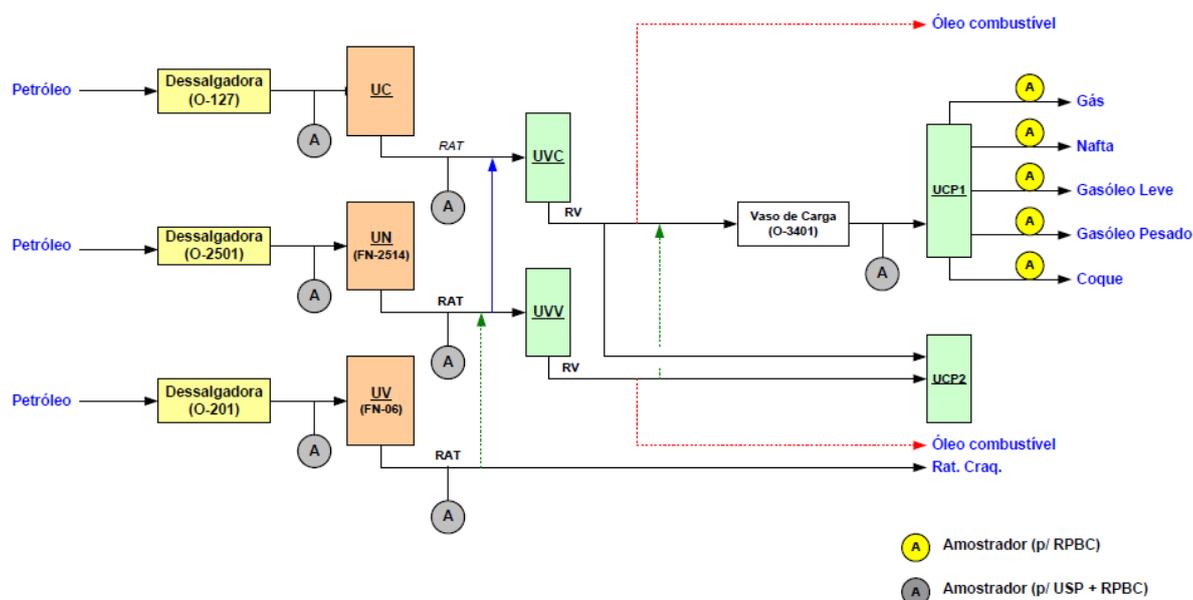


Figura 3: Fluxograma do processo de Destilação e Coqueamento Retardado da RPBC

- 1) Petróleo: óleo bruto alimentado nas torres de destilação, após passagem por dessalgadora;
- 2) Resíduo Atmosférico (RAT): fração do petróleo obtida no fundo da primeira torre de destilação, à pressão atmosférica;
- 3) Resíduo de Vácuo (RV): fração obtida no fundo da segunda torre de destilação, à vácuo. Esta segunda torre tem o RAT como alimentação. O RV é a alimentação do Coqueamento Retardado.

O anexo 1 mostra quais são as 67 amostras coletadas na refinaria e utilizadas na análise de dados, bem como a data de coleta, a origem na refinaria e os valores das propriedades “bulk” (grau API, teor de asfaltenos e Resíduo de Carbono) que foram fornecidos pela Petrobras.

### **3.1. Separação dos Óleos em Frações (SARA)**

A metodologia denominada SARA utiliza a cromatografia líquida para a separação das frações de petróleo que dão o nome ao método (saturados, aromáticos, resinas e asfaltenos). Este fracionamento e a posterior análise das frações SARA são de grande importância na caracterização do petróleo, sendo que a metodologia empregada atualmente, a ASTM D2007-93, emprega o procedimento de cromatografia clássica de coluna aberta, tornando a análise irreproduzível e dispendiosa.

A cromatografia líquida de alta eficiência (HPLC) pode ser empregada com o intuito de resolver os problemas que ocorrem na análise SARA clássica, pois possui alto grau de reprodutibilidade, suas análises são rápidas, automatizadas e a eficiência de separação é muito maior que a da cromatografia clássica. Até o presente momento não se encontra disponível nenhuma metodologia ASTM para análise das frações do petróleo através de HPLC.

A primeira etapa da separação das frações SARA é a precipitação dos asfaltenos. Os asfaltenos pertencem a uma classe de compostos definida pela sua insolubilidade em solventes parafínicos como o n-pentano, n-hexano ou o n-heptano. Quando n-pentano é utilizado para separar os asfaltenos, uma maior quantidade de precipitado é obtida em comparação com a separação utilizando n-heptano. Isto ocorre porque no primeiro caso, não são somente os asfaltenos que se precipitam, mas também as resinas. A precipitação dos asfaltenos é feita mais comumente utilizando-se o n-heptano, já que solventes de cadeias maiores não geram melhoras significantes nos resultados (10). A Figura 4 esquematiza o processo de separação adotado:

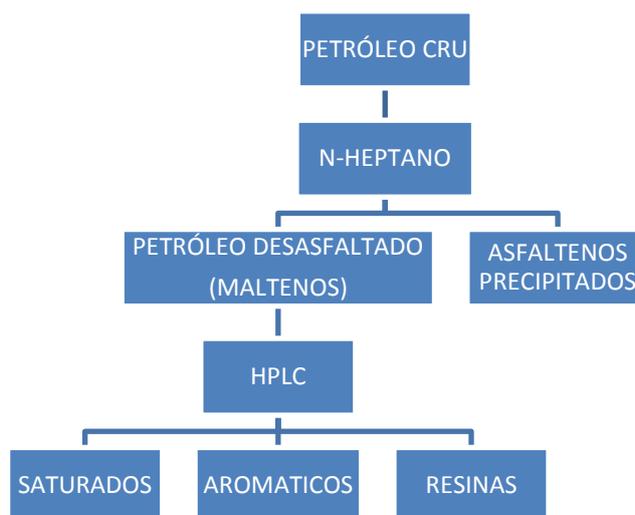


Figura 4: Esquema de separação Saturados, Aromáticos, Resinas e Asfaltenos (SARA)

Em algumas metodologias, a proporção de n-heptano em relação ao óleo amostrado é de 40:1 (10), embora também existam estudos em que a proporção é de 20:1. O tempo para a precipitação dos asfaltenos no solvente varia entre 16h e 24h (10). O precipitado então é separado do óleo desasfaltado por diversas formas, como a filtração ou a centrifugação, e em seguida passa por secagem. As frações SAR ainda presentes na solução óleo/n-heptano são separadas através do HPLC.

O foco deste trabalho é a precipitação e caracterização dos asfaltenos, de forma que não foram realizadas as separações dos maltenos com o HPLC.

### 3.1.1. Precipitação dos Asfaltenos

Foram pesados por volta de 1g de cada amostra que, em seguida, foram solubilizadas em aproximadamente 40 mL de n-heptano (Synth – P.A.). As amostras foram mantidas por, pelo menos, 24h em um local escuro, para evitar degradações das substâncias devido à energia luminosa. Após esse período de precipitação, os tubos contendo as suspensões foram centrifugados e a parte líquida (heptano e frações SAR) foi reservada em tubos novos. Aos tubos contendo precipitados e resíduos das frações líquidas, foi adicionado n-heptano novamente para lavar a parte sólida. Deixou-se esta nova suspensão em repouso por mais 30 min, os tubos foram novamente centrifugados, e o n-

heptano resultante da lavagem foi descartado. Os tubos com material sólido e resíduos de n-heptano foram secos por 24h em uma estufa mantida a 70°C.

### 3.2. Análise por Espectrometria MALDI-TOF

Os asfaltenos precipitados pela metodologia descrita foram caracterizados pelo espectrômetro de massas MALDI-TOF *Shimadzu Axima Confidence*.

Antes da obtenção dos espectros, foi necessário determinar a matriz a ser utilizada na análise. Outros trabalhos em que asfaltenos foram caracterizados por MALDI-TOF utilizam diversas matrizes como o ácido sinapínico (21). Como não há uma matriz amplamente utilizada para asfaltenos, um teste foi realizado para a escolha da matriz. As quatro primeiras amostras de asfaltenos secas, precipitadas de petróleos brutos, foram dissolvidas em tolueno e em seguida dispostas sobre a placa de amostras do equipamento. Cada amostra foi colocada em 12 *spots* da placa, 0,5 µL em cada um. 4 *spots* permaneceram sem matriz, a 4 deles foi adicionado antraceno como matriz, e aos outros 4 foi adicionado tri-hidroxiantraceno. As análises foram feitas após a evaporação do solvente dos *spots*. Os melhores espectros foram obtidos nos *spots* com amostras sem matriz, confirmando o que outro estudo já havia afirmado (28). A partir deste teste, todos os espectros foram obtidos sem o uso de matriz. Sendo assim, a técnica pode ser também denominada LDI-TOF, uma vez que não é mais utilizada uma matriz para auxiliar na ionização e dessorção do analito.

Nestas análises preliminares também foram testadas diferentes potências de laser e o modo de operação. Os melhores resultados foram obtidos com potência entre 120 e 125 dentro da escala do equipamento que vai de 0 a 180. O modo de operação escolhido foi o linear.

No preparo das soluções pesou-se por volta de 1,0 mg de sólido seco, e dissolveu-se em 1 mL de tolueno. Alíquotas de 1,0 µL de cada solução foram depositadas sobre os *spots* da placa do MALDI-TOF. A última etapa é a introdução da placa no equipamento e a determinação dos parâmetros adequados no software que o controla. Para cada amostra, 100 perfis (gráficos de intensidade relativa do sinal em função da relação m/z) foram extraídos. A

mesma potência foi utilizada para todos os espectros. Estes 100 perfis são processados pelo software e as figuras aqui apresentadas apresentam o espectro após a média desses 100 perfis.

No Anexo 2 é possível observar alguns dos 67 espectros de massa obtidos.

### **3.3. Caracterização por Espectrometria de Fluorescência**

Os espectros de emissão de fluorescência no estado estacionário, foram obtidos empregando Espectrômetro para Estado Estacionário e Tempo de Vida de Fluorescência modelo FLS920 (Edinburgh Instruments), que é um sistema modular para a obtenção de espectros no estado estacionário na região espectral do UV-visível, juntamente com decaimentos de fluorescência no intervalo de tempo entre 100 ps e 50  $\mu$ s, empregando o método de Contagem de Monofótons com Correlação Temporal. Para a obtenção de espectros no estado estacionário o equipamento emprega lâmpada nF900 controlada por computador, com fonte e sincronização com a fotomultiplicadora, com a excitação na região do UV-visível sendo feita com gás hidrogênio (largura do pulso ótico: abaixo de 0.8 ns), enquanto que a linha de excitação no UV é feita com gás nitrogênio (largura do pulso ótico:  $\sim$ 1.1 ns).

Neste trabalho foram extraídos apenas os espectros de fluorescência das amostras de petróleo bruto, sem a etapa de separação dos asfaltenos. Não foram analisadas as amostras de Resíduo Atmosférico nem de Resíduo de Vácuo.

## **4. Análise de Dados**

### **4.1. Preparação de Dados e Seleção de Variáveis**

A primeira etapa da análise de dados consiste na preparação e limpeza dos dados utilizados. Foram construídas matrizes de dados das variáveis independentes em que cada linha corresponde a uma amostra e cada coluna a uma variável (valores  $m/z$  para o LDI-TOF ou  $\lambda$  para a Fluorescência). Algumas amostras também foram excluídas das matrizes de dados por apresentarem inconsistências quanto aos valores das propriedades fornecidos pela refinaria.

#### 4.1.1. Dados dos Espectros LDI-TOF

Os espectros do LDI-TOF foram extraídos numa faixa de valores  $m/z$  de 100 a 3000. O software do equipamento fornece em arquivo de texto as intensidades em função dos valores  $m/z$  (as variáveis do conjunto de dados). A matriz de dados foi então estruturada de modo que cada amostra possuísse inicialmente 2901 variáveis, ou seja, valores para as variáveis  $m/z=100, 101, 102, \dots, 3000$ . Como os valores  $m/z$  fornecidos no arquivo de texto possuem precisão na segunda casa decimal, foram acumuladas na variável  $m/z=100$  as intensidades fornecidas no arquivo para  $m/z$  entre 100,00 e 100,99, acumuladas em  $m/z=101$  as intensidades para  $m/z$  entre 101,00 e 101,99, e assim sucessivamente.

Em seguida, uma redução do número de variáveis foi realizada, eliminando as colunas da matriz cujas intensidades eram todas nulas e que, portanto, não influem na variabilidade dos dados. Com esta limpeza, restaram 1798 variáveis  $m/z$ . Dessa forma, a matriz de dados do LDI-TOF final possui dimensão 67 x 1798.

Estas duas etapas de limpeza foram realizadas com o auxílio do Microsoft Excel.

#### 4.1.2. Dados de Fluorescência em Estado Estacionário

Os espectros de fluorescência foram extraídos na faixa de comprimentos de onda de 350 nm a 650 nm. O arquivo de texto obtido apresenta a contagem de fótons pelo detector em função do comprimento de onda  $\lambda$ , com um passo unitário entre os comprimentos de onda. Dessa forma, a matriz de dados obtida apresentou inicialmente 301 colunas e 31 linhas (31 amostras de petróleo bruto).

Alguns dos espectros de fluorescência obtidos são apresentados na Figura 5:

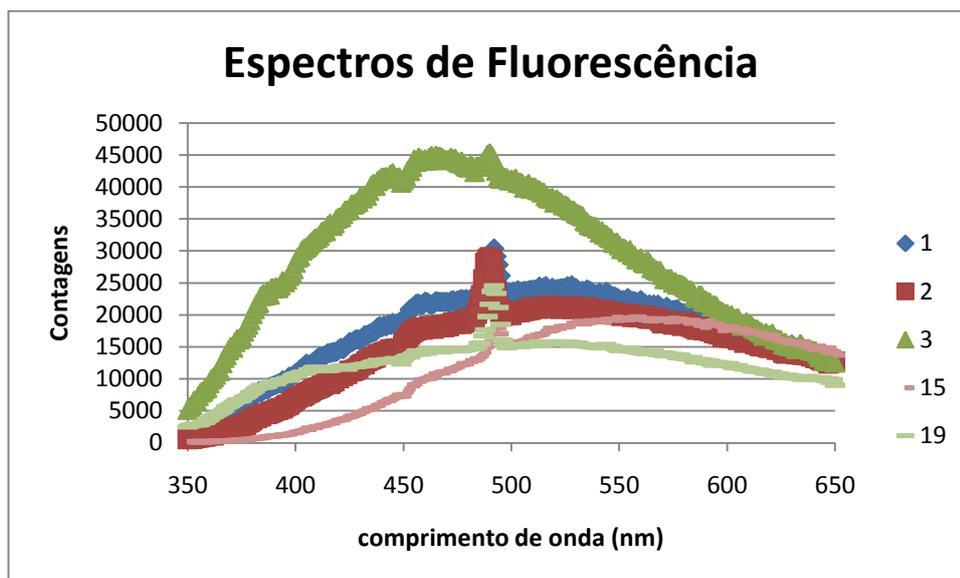


Figura 5: Espectros de Fluorescência em Estado Estacionário para amostras 1, 2, 3, 15 e 19.

Da figura, se observa que em comprimentos de onda por volta de 500 nm existe uma perturbação no espectro que, a princípio, não deveria ocorrer. Investigando melhor esses valores, observou-se que tal perturbação ocorre, para todas as amostras, entre 482 nm e 497 nm. Para efeitos da análise multivariada, decidiu-se excluir estas variáveis, resultando numa matriz final de dados de ordem 31 x 285.

#### 4.2. Regressão por Componentes Principais (PCR)

A regressão pela técnica do PCR foi realizada com o auxílio do software Minitab para predição dos dados bulk através dos espectros de fluorescência. Inicialmente foi realizado o PCA com a matriz dos dados de fluorescência para que fossem calculados os seus auto-vetores e auto-valores (respectivamente os componentes e suas variâncias). O número de componentes principais escolhido baseou-se na variância acumulada desses componentes. Escolhido o número de componentes e calculados os scores de cada amostra para os componentes (matriz de scores de dimensão: N° de amostras x N° de componentes), a regressão foi então realizada, bem como a análise do coeficiente de determinação ( $R^2$ ) calculado e dos resíduos ( $y_{\text{calc}} - y_{\text{exp}}$ ).

Não foi realizado PCR para os dados do LDI-TOF.

### 4.3. Regressão por Mínimos Quadrados Parciais (PLSR)

A regressão pelo PLS foi realizada tanto para os dados de Fluorescência quanto para os dados do LDI-TOF, através do software Matlab. Foi realizada a escolha do número de componentes através da validação cruzada com 10 partições (10-fold cross-validation - CV) e da análise do erro quadrático médio na predição (MSEP) calculado através do CV.

Um script no Matlab foi escrito para realizar a análise de forma prática e obter as matrizes de coeficientes ( $\mathbf{B}_{PLS}$ ), gráficos de variância acumulada de  $\mathbf{X}$  e de  $\mathbf{Y}$  em função do número de componentes, o gráfico de MSEP em função também do número de componentes, os três gráficos com os valores preditos em função dos experimentais (para grau API, teor de asfaltenos e Resíduo de Carbono), os histogramas com as freqüências de resíduos para cada um dos três ajustes, os gráficos de dispersão de resíduos em função do valor ajustado pelos modelos, e os fatores  $R^2$  obtidos nos ajustes.

## 5. Resultados e Discussão

### 5.1. Espectros LDI-TOF

Para uma primeira comparação dos espectros de LDI-TOF, foram escolhidos, dos 83 espectros, 21 espectros que correspondem às seguintes amostras (as numerações das amostras podem ser verificadas no Anexo 1):

Tabela 3: Mapeamento dos espectros a serem comparados.

Data	Petróleo UC	RAT UC	Petróleo UN	RAT UN	Petróleo UV	RAT UV	RV
25/02/11	46	26	19	23	25	21	22
02/03/11	47	44	48	69	38	73	70
29/03/11	60	30	56	32	53	34	29

Estes espectros permitem que se compare a distribuição de massas moleculares da carga das três torres de destilação atmosférica (UC, UN e UV) com o resíduo das mesmas e com o resíduo de destilação a vácuo. Os diagramas obtidos estão presentes no Anexo 2 (espectros de RV foram

repetidos no final de cada página para efeito de comparação destes com os correspondentes espectros de Petróleo e RAT de uma certa origem, para amostras coletadas na mesma data).

Os espectros relacionam a intensidade relativa do sinal obtido (em unidades arbitrárias) com a relação massa/carga detectada. Admite-se que a carga dos íons seja sempre unitária, devido à retirada de um elétron da molécula ou à adição de  $H^+$  devido à incidência do laser de  $N_2$ . Assim, os valores da abscissa são aqui tratados diretamente como massa molecular, cuja unidade é o dalton, Da.

Não existe, a princípio, uma faixa de massas determinada para os asfaltenos. Enquanto alguns estudos apresentam espectros LDI-TOF com distribuição de massas por volta de 2000 Da (21), outros apresentam resultados semelhantes aos obtidos aqui, com distribuições de massas entre 200 Da e 800 Da (29). Uma hipótese bastante provável é a de que a massa molecular dos compostos asfálticos variam fortemente com o local de onde são extraídos.

É possível perceber que, para uma mesma data e uma mesma torre, os espectros de petróleo e de RAT são muito semelhantes entre si, e diferentes do espectro de RV correspondente. Este fato reforça a afirmação de que os asfaltenos, que são compostos mais pesados presentes no petróleo, são obtidos após a destilação no resíduo, sem sofrer transformações químicas que possivelmente alterariam sua distribuição de massa molecular. Pode-se observar também que as amostras de RV mostram distribuições de massa bastante uniformes, que pelo seu formato apresentam semelhança com a distribuição normal. A faixa de massas nos espectros RV cuja intensidade de sinal foi considerável é mais estreita do que as faixas observadas nos petróleos e nas amostras de RAT e, além disso, estas faixas estão localizadas em regiões com maior massa. Isso indica que, de uma gama maior de compostos presentes nos asfaltenos de Petróleos e de RAT, somente permanecem no RV os compostos mais pesados, de maneira que os mais leves são separados na destilação a vácuo.

Os petróleos alimentados na refinaria são formados de misturas (“blends”) de óleos brutos provenientes de diferentes reservatórios e, portanto, de

diferentes poços. As características de cada um destes óleos, bem como a porcentagem de cada um deles na mistura, vai influenciar decisivamente na distribuição de massas obtida pelo LDI-TOF. Dessa forma, os espectros cuja distribuição se afasta muito da normal, ou que aparentam possuir duas regiões com distribuição normal demonstram a influência dos teores de cada petróleo no blend que é alimentado nas torres de destilação.

## 5.2. Regressões e Correlações

### 5.2.1. PCR para os dados de Fluorescência

O PCA foi realizado através do Minitab, utilizando-se a matriz de correlação ( $\rho$ ) para o cálculo dos auto-valores e auto-vetores. A Tabela 4 mostra os resultados obtidos para os auto-valores (variâncias), bem como as contribuições percentuais de cada um deles para a variância total e a variância acumulada:

Tabela 4: Auto-valores calculados para os dados de fluorescência.

Componente	Auto-valor	Proporção	Proporção acumulada
PC1	190,97	0,67	0,67
PC2	87,6	0,307	0,977
PC3	5,07	0,018	0,995
PC4	1,11	0,004	0,999
PC5	0,12	0	1
PC6	0,02	0	1
PC7	0,01	0	1
PC8	0,01	0	1

Com esses dados, percebe-se que com o uso apenas dois dos primeiros componentes principais, 97,7% da variância dos dados de  $X$  é considerada. Por esse motivo, os componentes principais PC1 e PC2 foram escolhidos para a regressão multilinear.

A etapa de regressão, como já visto anteriormente, necessita dos scores dos dois componentes calculados para todas as 31 amostras. A matriz de loadings dos componentes (de ordem  $285 \times 2$ ) é fornecida pelo Minitab, mas não

será aqui apresentada. Por sua vez, a matriz dos scores é apresentada na Tabela 5:

Tabela 5: Matriz dos scores calculados dos componentes do PCR para as 31 amostras.

<b>amostra</b>	<b>score PC1</b>	<b>score PC2</b>
1	2,0718	-3,4648
2	-3,40161	0,25218
3	25,52803	-22,8193
4	16,53372	-12,0408
8	12,81618	12,72228
13	1,474811	-0,96185
14	10,50079	-1,02174
15	-8,4075	6,448296
18	6,493146	12,95238
19	-7,68345	-8,38023
20	-2,18468	9,082718
37	5,986472	-6,06687
38	-13,0261	9,52577
39	23,74339	18,59862
40	-9,40037	4,159697
41	-3,8894	-0,1013
42	-9,17356	11,23541
45	5,558945	3,5796
46	2,982929	2,373771
47	-20,7018	3,908385
48	-25,5721	0,049495
49	-18,7101	5,45671
50	11,40213	12,25346
52	0,153443	-5,10507
56	-8,67458	-8,48354
59	-10,8768	-5,54833
60	-14,4186	-11,8253
63	-6,71548	-14,2572
67	-10,0973	-8,18681
75	30,54751	2,589208
76	17,14	-6,92468

A regressão multilinear realizada em seguida, também pelo software Minitab calculou os coeficientes da matriz  $\beta$  da eq.(12). A saída completa do programa é apresentada no Anexo 3. A equação de regressão proposta (com o coeficiente de determinação já calculado) é:

$$^{\circ}API = 26,3494 + 0,0837 * PC1 - 0,0343 * PC2$$

$$R^2 = 15,07\% \quad (22)$$

O  $R^2$  calculado mostra que os valores previstos pelo modelo estão muito pouco correlacionados com os valores experimentais. Dessa forma, essa regressão não é adequada para fazer a predição do grau API através dos dados da espectrometria de fluorescência. Outro indício de que a correlação da eq.(22) não é adequada é o histograma dos resíduos apresentado na Figura 6:

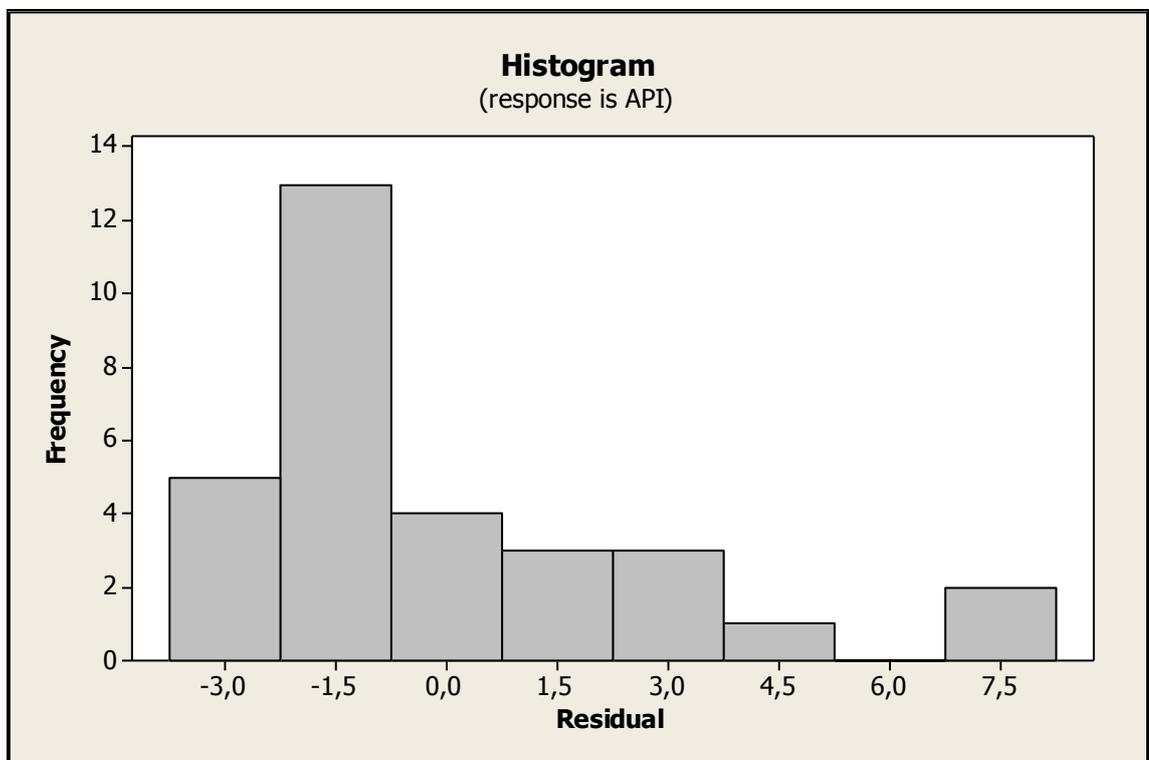


Figura 6: Histograma de distribuição de resíduos para o PCR em relação ao grau API.

Os resíduos deveriam estar centrados na média 0 e possuir distribuição próxima à normal, o que aparentemente não ocorre na Figura 6. Os dois resíduos encontrados na faixa de 7,5 estão um pouco afastados dos restantes. Utilizando-se o critério dos 6 sigma para detecção de possíveis pontos anômalos (outliers), ainda admitindo distribuição normal dos valores, temos os dados da Tabela 6:

Tabela 6: Detecção de outliers no PCR para predição do grau API.

$\bar{x}$	0
$\sigma$	2,85
$\bar{x} - 3\sigma$	-8,55
$\bar{x} + 3\sigma$	8,55

Estes parâmetros indicam que, se a distribuição dos resíduos fosse normal, nenhum deles seria um outlier, uma vez que todos estão na faixa de  $\bar{x} \pm 3\sigma$ .

O Minitab ainda calcula, na regressão, quais são os pontos que apresentaram comportamento possivelmente anômalo (neste caso, as observações correspondentes às amostras de petróleo numeradas 3, 4 e 8). Realizando novamente a regressão sem essas observações, o  $R^2$  apresenta um valor ainda mais baixo, e novos pontos potencialmente anômalos são calculados pelo Minitab. Isso indica uma dispersão não desejada dos valores.

Outras tentativas de realizar o PCR foram realizadas. Em todas elas foram utilizados os mesmos componentes principais PC1 e PC2 e, portanto, a mesma matriz com os scores dos componentes foi utilizada. A regressão com o maior valor de  $R^2$  foi obtida para a predição do Resíduo de Carbono (RC) utilizando-se cinco amostras a menos do que as originais. De qualquer forma, o valor do coeficiente foi  $R^2=23,90\%$ , um valor ainda muito baixo de correlação entre dados preditos e experimentais. Por esses motivos, tais correlações não são aqui detalhadas.

Levando em consideração esses resultados, a técnica do PCR foi abandonada para a investigação das correlações.

### 5.2.2. PLSR através do Matlab

Os cálculos do PLSR foram efetuados com o Matlab. O script utilizado para o PLSR com os dados do LDI-TOF está transcrito no Anexo 4 e é basicamente o mesmo script utilizado para os dados de fluorescência, com pouquíssimas alterações.

O script carrega um conjunto de matrizes para a área de trabalho do Matlab. A Tabela 7 fornece os nomes e as ordens das matrizes de dados  $X$  e  $Y$  utilizadas no PLSR

Tabela 7: Matrizes de variáveis preditoras e de resposta carregadas no Matlab

	$X$		$Y$	
<b>PLSR para Fluorescência</b>	<i>fluorescencia2</i>	31x285	<i>BulkExp</i>	31x3
<b>PLSR para LDI-TOF</b>	<i>maldi</i>	67x1798	<i>BulkExpMaldi</i>	67x3

As duas matrizes de dados são então padronizadas através da função “*zscore*”, pois as variáveis possuem diferentes ordens de grandeza. Essa padronização é a mesma descrita pela eq.(8).

A função “*plsregress*” utilizada é parte integrante do Statistics Toolbox do Matlab. Como entradas dessa função, temos as matrizes de variáveis padronizadas  $X$  e  $Y$ , o número de componentes a ser computado, o parâmetro ‘*cv*’ que indica a regressão com cross-validation e o parâmetro  $k$  que indica o número de partições dos dados a serem utilizados no cross-validation. Em todas as análises desse trabalho, o valor  $k=10$  foi fixado. O objetivo, então, foi obter um valor adequado para o número de componentes, para que o ajuste seja satisfatório.

A função retorna as matrizes de scores e loadings para ambas as matrizes  $X_0$  e  $Y_0$ , além da matriz de coeficientes  $\beta$  (“*Beta*”, no script), a matriz “*PCTVAR*”, com as porcentagens da variância em  $X$  e em  $Y$  por componente, “*MSE*”, com os erros quadráticos médios de predição calculados pelo CV para  $X$  e  $Y$ , e “*Stats*”, uma estrutura que retorna, dentre outros, os resíduos dos preditores  $X$  e das respostas  $Y$ .

Além das matrizes do modelo, o script foi escrito de forma a fornecer os gráficos importantes para a análise dos ajustes (variâncias acumuladas, MSEP, gráficos de valores ajustados e de resíduos).

### 5.2.3. PLSR para os dados de Fluorescência

Para a escolha do número de componentes a ser utilizado neste PLSR, inicialmente um número grande foi utilizado ( $n\text{-comp}=20$ ), e o MSEP foi analisado. A Figura 7 foi então obtida para o MSEP:

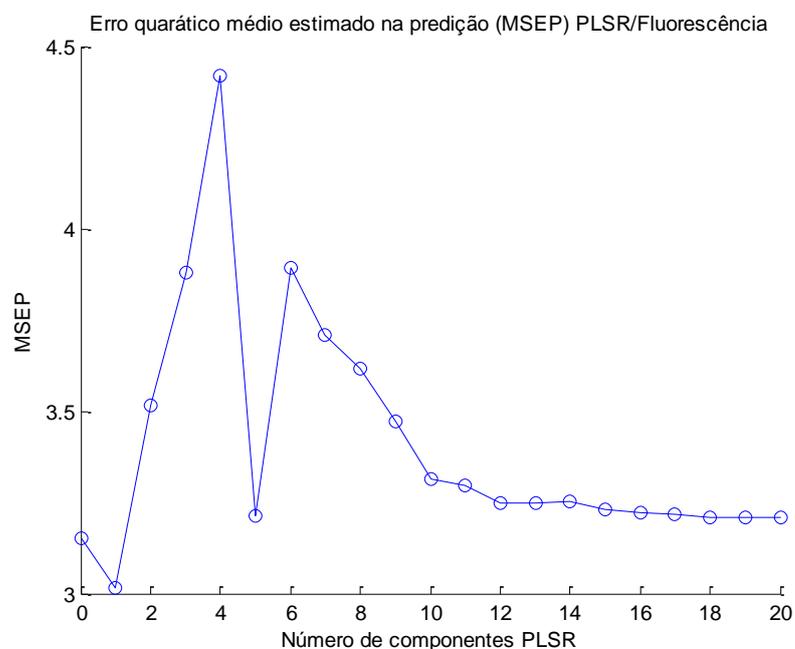


Figura 7: MSEP em função do número de componentes (PLSR/Fluorescência)

O que se observa dessa figura é que o MSEP para por valores mínimos quando são usados dois ou cinco componentes. Com o uso de dez ou mais componentes, o MSEP é aparentemente menor do que para cinco componentes, e este erro cai continuamente com o aumento do número de componentes. Apesar dessa queda do MSEP, não é interessante utilizar um grande número de componentes na modelagem, pois pode ocorrer o fenômeno de sobreajuste. Apesar disso, o uso de cinco componentes leva em consideração uma porcentagem total de menos de 50% da variância das respostas em  $Y$ , como pode-se observar na Figura 8:

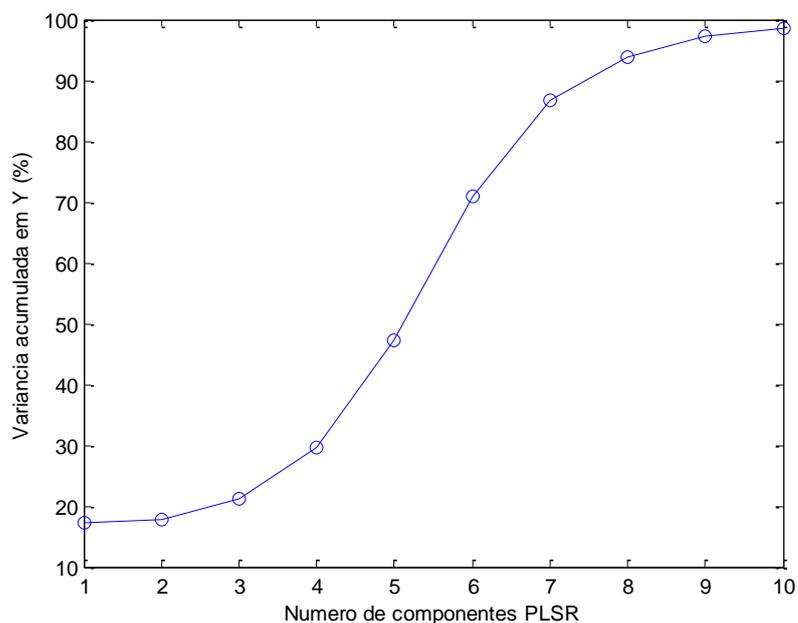


Figura 8: Variância acumulada em Y em função do número de componentes (PLSR/Fluorescência)

Dessa forma, como esperado, os valores ajustados para o PLSR com 5 componentes não obtiveram boas correlações com os valores experimentais. Os  $R^2$  calculados estão na Tabela 8:

Tabela 8: Coeficientes de determinação ( $R^2$ ) para PLSR/Fluorescência e 5 componentes

	$R^2$
<b>API</b>	0,4233
<b>% Asf</b>	0,4703
<b>RC</b>	0,5267

Novas tentativas foram realizadas com diferentes números de componentes. Resultados muito satisfatórios foram obtidos para o PLSR com 10 componentes. Os resultados gráficos são apresentados no Anexo 5.

A partir das figuras do Anexo 5, é possível observar que, com 10 componentes, quase 100% da variância total de ambas as matrizes  $X$  e  $Y$  é considerada. Percebe-se também que os valores calculados pelo modelo para as três propriedades estão bem correlacionados com os valores experimentais (os gráficos de  $y_{calc}$  em função de  $y_{exp}$  apresentam a tendência de retas muito próximas à bissetriz do primeiro quadrante) e que os resíduos apresentam uma

distribuição que se assemelha á normal (principalmente para os ajustes de teor de Asfaltenos e Resíduo de Carbono). Os coeficientes de determinação obtidos foram:

Tabela 9: Coeficientes de determinação ( $R^2$ ) para PLSR/Fluorescência e 10 componentes

	$R^2$
<b>API</b>	0.9837
<b>% Asf</b>	0.9849
<b>RC</b>	0.9891

Regressões com maior número de componentes não foram realizadas, uma vez que os resultados obtidos com 10 componentes foram satisfatórios e que o acréscimo de novos componentes não reduz significativamente o erro quadrático de predição (como pode-se observar na Figura 7).

#### **5.2.4. PLSR para os dados do LDI-TOF**

De maneira análoga à seção anterior, um estudo preliminar dos valores de MSEF para um grande número de componentes foi realizado para que se possa estimar a quantidade de componentes que minimizará o erro quadrático de predição e, ao mesmo tempo, fornecerá um modelo com ajuste satisfatório. A Figura 9, gerada através de um PLSR com 30 componentes, mostra como se comporta o erro de predição:

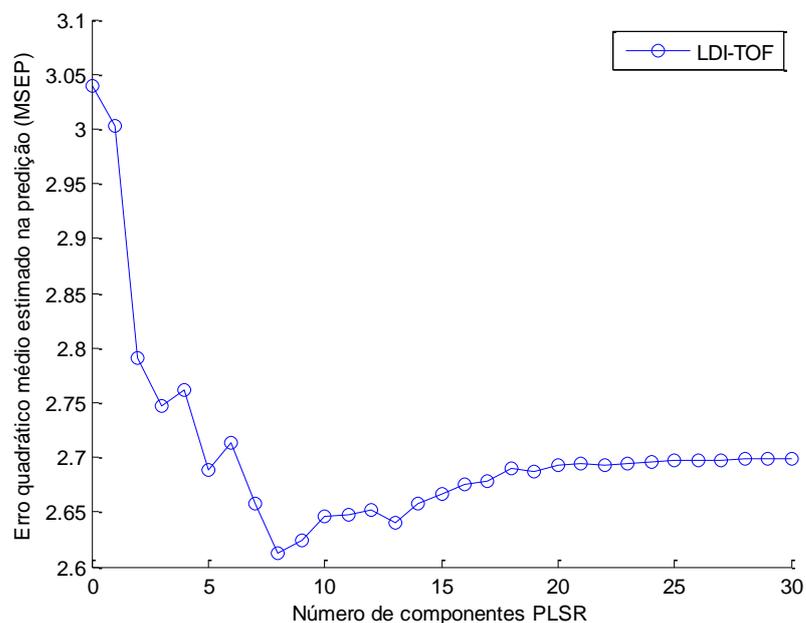


Figura 9: MSEP em função do número de componentes (PLSR/LDI-TOF)

Dessa forma, as regressões com 5, 8 ou 13 componentes apresentam valores MSEP mínimos “locais” e precisam de melhor investigação. Com esse objetivo, o PLSR foi realizado para estes números de componentes. Os resultados dos coeficientes de determinação se apresentam na Tabela 10:

Tabela 10: Coeficientes de determinação ( $R^2$ ) para PLSR/LDI-TOF com 5, 8 e 13 componentes.

$R^2$	5 comp	8 comp	13 comp
<b>API</b>	0.8723	0.9560	0.9895
<b>% Asf</b>	0.8804	0.9522	0.9930
<b>RC</b>	0.8991	0.9555	0.9903

Os valores de  $R^2$  para o PLSR com 13 componentes mostram uma boa correlação entre dados calculados e experimentais. É interessante notar que para regressões com um número de componentes maior que 13, o erro quadrático passa a aumentar e seu valor estabiliza em torno de 2,7.

O Anexo 6 mostra todos os gráficos de variâncias, valores ajustados e resíduos obtidos para o PLSR com 13 componentes. Através deles observa-se que a variância acumulada em  $Y$  para esses 13 componentes corresponde a quase 100% da variância total. O mesmo não ocorre para a variância

acumulada em  $\mathbf{X}$ , que corresponde a pouco mais de 50% da variância total dessa matriz. Também por esse motivo, o PLSR com 13 componentes é preferível, uma vez que para 5 e 8 componentes uma representatividade ainda menor da variância da matriz  $\mathbf{X}$  é atingida. Apesar disso, mesmo com apenas 5 componentes, quase 90% da variância total da matriz  $\mathbf{Y}$  é alcançada. Esse é um dos prováveis motivos pelo qual os valores de  $R^2$  para 5 e 8 componentes ainda são bastante altos, mesmo com a baixa representatividade de  $\mathbf{X}$ .

Também se pode avaliar a qualidade dos ajustes, através dos três gráficos das propriedades calculadas em função dos valores experimentais, e as distribuições dos resíduos. Tais distribuições também se assemelham a distribuições normais, corroborando para a validação do modelo.

## 6. Conclusões

Os espectros de massa, em uma análise qualitativa e preliminar, mostram que as distribuições de massa dos asfaltenos oriundos da alimentação de uma coluna de destilação atmosférica e de seu resíduo são bastante idênticas, considerando-se um mesmo instante de amostragem. Apesar disso, o resíduo da destilação a vácuo correspondente apresenta uma distribuição mais bem definida de massa, em alguns casos aparentando ser uma distribuição normal. Outra característica importante dos espectros de RV é a concentração da curva faixas de massa maiores. Algumas amostras apresentaram espectros indicando uma distribuição não gaussiana de massas e outras apresentaram duas regiões distintas no espectro, demonstrando uma influência das porcentagens dos blends de petróleo alimentados ao processo.

A técnica multivariada do PCR aplicada a dados de Espectrometria de Fluorescência para a predição do grau API, teor de Asfaltenos e Resíduo de Carbono não resultou em modelos com bom ajuste, de maneira que o  $R^2$  máximo obtido foi de 23,9%. Por esse motivo, esta técnica não foi utilizada para a análise dos dados de Espectrometria LDI-TOF.

Por sua vez, a técnica do PLSR resultou em modelos cujo  $R^2$  mínimo esteve acima de 0,98, para a predição das propriedades a partir de ambas as técnicas espectrométricas. O modelo aqui proposto pode ser facilmente calculado através do script do Matlab transcrito no Anexo 4. Os resultados mais satisfatórios foram obtidos através do PLSR com 10 componentes para os dados de Fluorescência e com 13 componentes para os dados do LDI-TOF. Uma validação dos modelos já foi realizada em paralelo com a escolha do número de componentes, através do cálculo dos erros quadráticos médios de predição após a validação cruzada (10-fold cross-validation).

Dessa forma, as matrizes de coeficientes  $\beta$  calculadas pelo Matlab através destes modelos podem ser utilizadas para prever as propriedades aqui ajustadas através da eq.(21).

## 7. Referências Bibliográficas

- (1) WOODS, J. et al. **Canadian Crudes: A Comparative Study of SARA Fractions from a Modified HPLC Separation Technique**. Oil & Gas Science and Technology – Rev. IFP., Vol. 63, v. 1 p. 151-163. 2008.
- (2) PANTOJA, P.A. **Caracterização de petróleo por técnicas espectroscópicas e sua utilização na avaliação e desenvolvimento de métodos de inferência em processo de refino**. 2010. 143 p. Tese (Doutorado) – Escola Politécnica, Universidade de São Paulo. 2010.
- (3) KHARRAT, A.M. et al. **Issues with comparing SARA methodologies**. Energy & Fuels v. 21, p. 3618-3621. 2007.
- (4) VAZQUEZ, D.; MANSOORI, G.A.. **Identification and measurement of petroleum precipitates**. Journal of Petroleum Science and Engineering v. 26, n. 1-4, p. 49-55. 2000.
- (5) ASKE, N.; KALLEVIK. H.; SJOBLOM, J. **Determination of Saturate, Aromatic, Resin and Asphaltenic (SARA) Components in Crude Oils by Means of Infrared and Near-Infrared Spectroscopy**. Energy & Fuels, v. 15, p.1304 –1312. 2001.
- (6) CASTRO, L.V.; VAZQUEZ, F. **Fractionation and Characterization of Mexican Crude Oils**. Energy & Fuels , v. 23, p. 1603-1609. 2009.
- (7) THOMAS, J.E. **Fundamentos de Engenharia de Petróleo**. 2ª ed. Rio de Janeiro.: Ed. Interciência. 2004.
- (8) SZKLO, A.; ULLER, V.C. **Fundamentos do Refino de Petróleo: tecnologia e economia**. Ed. 2. Editora Interciência: Rio de Janeiro, 2008.
- (9) SIMANZHENKOV, V.; IDEM, R. **Crude Oil Chemistry**. New York. Marcel Dekker Inc.. 2003.
- (10) ANCHEYTA, J.; TREJO, F.; RANA, M.S. **Asphaltenes: chemical transformation during hydroprocessing of heavy oils**. New York. Taylor & Francis Group. 2009.
- (11) SZKLO, A.; ULLER, V.C. **Fundamentos do Refino de Petróleo: tecnologia e economia**. Ed. 2. Editora Interciência: Rio de Janeiro, 2008.
- (12) NOEL, F. **Alternative to the Conradson Carbon Residue**. Fuel. v. 63, p. 931 – 934, 1984.
- (13) MARIANO, J.B. **Impactos Ambientais do Refino de Petróleo**. Rio de Janeiro. 2001. 216 p. Dissertação ( Mestrado) Universidade Federal do Rio de Janeiro. 2001.
- (14) SUBERO, J.; LYON, L.; BERTANCOURT, R.. **Evaluación de la influencia microestructural del coque de petróleo calcinado sobre el choque térmico del ánodo utilizado en el proceso de reducción electrolítica del aluminio**. uct, Puerto Ordaz, v. 10, n. 39, jul. 2006 . Disponible em

<[http://www.scielo.org/ve/scielo.php?script=sci\\_arttext&pid=S1316-48212006000300013&lng=es&nrm=iso](http://www.scielo.org/ve/scielo.php?script=sci_arttext&pid=S1316-48212006000300013&lng=es&nrm=iso)>. Adicionado em 07 agosto 2012.

- (15) BLAND, W.F.; DAVIDSON, R.L. **Petroleum Processing Handbook**. New York. McGraw-Hill. 1967.
- (16) NELSON, W.L. **Thermal cracking and decomposition process**. Petroleum Refining Engineering. 4ªed. New York. McGraw Hill. 1958.
- (17) SPEIGHT, J.G. **An Evaluation of the delayed coking product yield of heavy feedstocks using asphaltene content and carbon residue**. Revue de L'Institut Français du Pétrole. v. 52, n. 1, 1997.
- (18) LANÇAS, F.M. **Cromatografia Líquida Moderna HPLC / CLAE**. 1ed. Campinas: Editora Atomo. 2009.
- (19) **High Performance Liquid Chromatography (HPLC)**. Iowa. Office of Biotechnolgy – Iowa State University. Disponível em: <http://www.protein.iastate.edu/hplc.html> - acesso em 21/11/2011
- (20) TANAKA, R. et al. **Analysis of the Molecular Weight Distribution of Petroleum Asphaltenes Using Laser Desorption – Mass Spectrometry**. Energy & Fuels. v. 18, p. 1405-1413.2004.
- (21) TREJO, F. et al. **Characterization of Asphaltenes from Hydrotreated Products by SEC, LDMS, MALDI, NMR and XRD**. Energy & Fuels. v. 21, p. 2121-2128. 2007
- (22) PRADIER, B. et al. **Chemical basis of fluorescence alteration of crude oils and kerogens – I. Microfluorimetry of an oil and its isolated fractions; relationships with chemical structure**. Advances in Organic Geochemistry. v. 16, n. 1-3, p. 451-460. 1990.
- (23) ALBUQUERQUE; F.C., NICODEM; D.E., RAJAGOPAL, K. **Estudo da agregação de asfaltenos por espectroscopia de fluorescência**. In: 3º Congresso Brasileiro de P&D em Petróleo e Gás. Rio e Janeiro. 2004.
- (24) JOHNSON, R.A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis**. 5ª ed. Prentice Hall. 2002.
- (25) GELADI, P.; KOWALSKI, B.R. **Partial Least-Squares Regression: A Tutorial**. Analytica Chimica Acta. v. 185, p. 1-17. 1986.
- (26) BRERETON, R.G. **Chemometrics: Data Analysis for the Laboratory and Chemical Plant**. John Wiley & Sons. 2003.
- (27) DE JONG, S. **SIMPLS: an alternative approach to partial least squares regression**. Chemometrics and Intelligent Laboratory Systems, v. 18, p. 251-263. 1993.
- (28) POMERANTZ, A.E. et al. **Two-step Laser Mass Spectrometry of Asphaltenes**. J. Am. Chem. Soc. v. 130, pp. 7216–7217. 2008.

- (29) SMANIOTTO, A.. **Laser desorption/ionization mass spectrometric study on asphaltenes from different heavy crude oils and products originating from their thermal treatment.** Rapid Communications in Mass Spectrometry v. 23, p. 725-728. 2009.

## 8. Anexos

### 8.1. Anexo 1: Amostras utilizadas

Tabela 11: Amostras utilizadas nas modelagens

Tubo	Amostra	Unidade	Data	°API	% Asfaltenos	RC
13	Petróleo	UC-AA	07/02/11	24,2391	1,4836	5,14
15	Petróleo	UC-AA	08/02/11	23,5774	1,4437	5,27
37	Petróleo	UC-AA	23/03/11	25,3026	1,8374	5,15
42	Petróleo	UC-AA	06/04/11	28,6864	1,6621	4,60
46	Petróleo	UC-AA	25/02/11	27,1563	1,5566	5,29
47	Petróleo	UC-AA	02/03/11	25,9603	1,8162	5,64
49	Petróleo	UC-AA	01/03/11	23,3952	0,9730	5,13
60	Petróleo	UC-AA	29/03/11	28,0473	1,9513	5,00
63	Petróleo	UC-AA	15/03/11	24,5243	2,0025	5,89
75	Petróleo	UC-AA	11/07/11	25,4199	0,5934	4,83
3	Petróleo	UN-AA	22/10/10	31,8094	0,5024	2,23
4	Petróleo	UN-AA	23/10/10	35,3163	0,3733	1,98
8	Petróleo	UN-AA	28/10/10	35,1767	0,4638	1,55
14	Petróleo	UN-AA	07/02/11	27,9115	1,9379	5,16
18	Petróleo	UN-AA	08/02/11	28,5627	1,8159	4,68
19	Petróleo	UN-AA	25/02/11	25,0325	1,8305	5,75
40	Petróleo	UN-AA	06/04/11	25,3457	2,0476	5,50
41	Petróleo	UN-AA	23/03/11	24,3450	2,0982	5,73
48	Petróleo	UN-AA	02/03/11	25,6222	1,5089	5,26
50	Petróleo	UN-AA	01/03/11	25,3554	1,5214	5,38
52	Petróleo	UN-AA	24/03/11	24,3877	2,0646	5,73
56	Petróleo	UN-AA	29/03/11	23,7283	0,6546	4,95
67	Petróleo	UN-AA	15/03/11	23,7788	2,3592	6,26
1	Petróleo	UV-AA	22/10/10	23,7420	1,1906	5,28
2	Petróleo	UV-AA	28/10/10	24,7543	2,0052	5,17
20	Petróleo	UV-AA	08/02/11	23,5772	1,4437	5,27
38	Petróleo	UV-AA	02/03/11	25,5355	0,4675	4,16
39	Petróleo	UV-AA	06/04/11	25,7454	1,2005	4,90
45	Petróleo	UV-AA	01/03/11	25,7315	0,4600	4,14
59	Petróleo	UV-AA	15/03/11	29,6441	1,1704	3,40
76	Petróleo	UV-AA	11/07/11	25,4198	0,5934	4,83
26	RAT	UC-AA	25/02/11	13,5667	2,5532	9,46
27	RAT	UC-AA	08/02/11	13,4197	2,2076	8,69
30	RAT	UC-AA	29/03/11	13,0930	3,9083	11,16
44	RAT	UC-AA	02/03/11	13,3408	2,8103	9,86
79	RAT	UC-AA	11/07/11	14,9874	1,0853	7,48

Continua

Tabela 11: continuação

<b>Tubo</b>	<b>Amostra</b>	<b>Unidade</b>	<b>Data</b>	<b>°API</b>	<b>% Asfaltenos</b>	<b>RC</b>
82	RAT	UC-AA	06/04/11	13,8784	2,9243	9,03
9	RAT	UN-AA	28/10/10	18,9579	1,7032	7,57
16	RAT	UN-AA	22/10/10	19,8350	0,7491	5,50
23	RAT	UN-AA	25/02/11	13,1730	2,7882	9,87
28	RAT	UN-AA	08/02/11	14,4124	3,1868	9,17
32	RAT	UN-AA	29/03/11	14,0783	1,1793	7,59
35	RAT	UN-AA	23/03/11	12,7806	3,2602	10,63
61	RAT	UN-AA	11/07/11	12,8806	3,9051	10,95
69	RAT	UN-AA	02/03/11	12,9325	3,0205	9,50
71	RAT	UN-AA	01/03/11	13,0939	2,6507	9,37
81	RAT	UN-AA	06/04/11	12,5204	4,2819	10,68
5	RAT	UV-AA	28/10/10	13,0064	3,4409	9,79
6	RAT	UV-AA	22/10/10	13,5992	2,0011	8,94
7	RAT	UV-AA	23/10/10	13,5014	2,1770	9,20
21	RAT	UV-AA	25/02/11	19,6156	0,5624	5,20
24	RAT	UV-AA	08/02/11	13,4196	2,2075	8,69
34	RAT	UV-AA	29/03/11	12,6937	3,9785	11,44
36	RAT	UV-AA	15/03/11	15,0547	2,6914	8,53
73	RAT	UV-AA	02/03/11	16,0807	0,8733	7,10
74	RAT	UV-AA	01/03/11	16,2891	0,8619	7,08
77	RAT	UV-AA	11/07/11	14,9872	1,0853	7,48
10	RV	UCP-I	28/10/10	10,1906	5,7189	17,38
11	RV	UCP-I	23/10/10	9,8544	4,4675	16,40
12	RV	UCP-I	22/10/10	11,5793	3,8048	15,40
17	RV	UCP-I	07/02/11	8,7087	6,2661	18,36
22	RV	UCP-I	25/02/11	8,8321	6,7167	18,31
29	RV	UCP-I	29/03/11	8,9224	6,0004	18,13
31	RV	UCP-I	24/03/11	8,0965	6,8132	19,12
33	RV	UCP-I	23/03/11	8,0431	6,9757	18,95
70	RV	UCP-I	02/03/11	8,6942	6,7567	18,45
72	RV	UCP-I	01/03/11	8,9385	5,9962	18,03

Tabela 12: Amostras utilizadas apenas para comparação dos espectros LDI-TOF.

<b>Tubo</b>	<b>Amostra</b>	<b>Unidade</b>	<b>Data</b>	<b>°API</b>	<b>% Asfaltenos</b>	<b>RC</b>
25	Petróleo	UV-AA	25/02/11	26,7274	0,5062	2,67
53	Petróleo	UV-AA	29/03/11	23,9327	2,5012	6,21

## 8.2. Anexo 2: Espectros de MALDI-TOF das amostras de Petr leo, RAT e RV de 25/02/11, 02/03/11, 29/03/11 e amostras de RV de

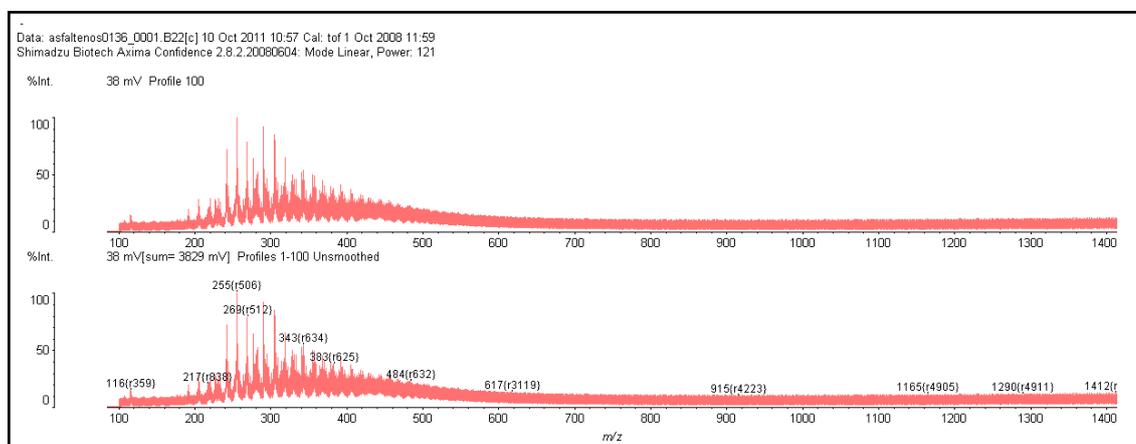


Figura 10: 25/02/11 – Petr leo UC (Amostra 46).

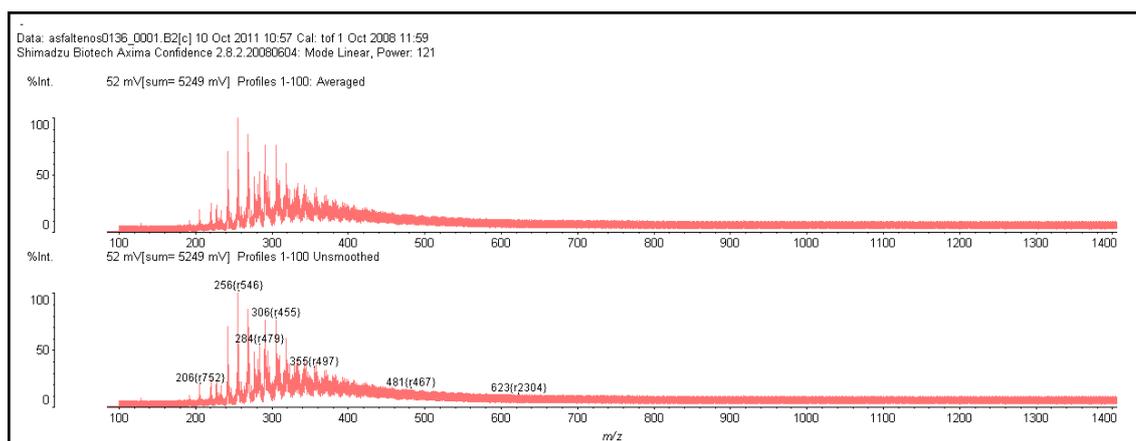


Figura 11: 25/02/11 – RAT UC (Amostra 26)

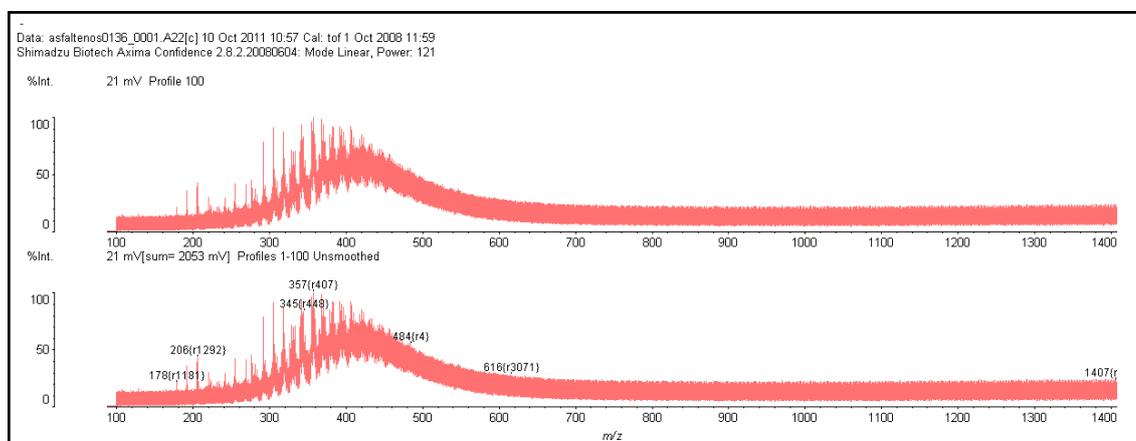


Figura 12: 25/02/11 – RV (Amostra 22)

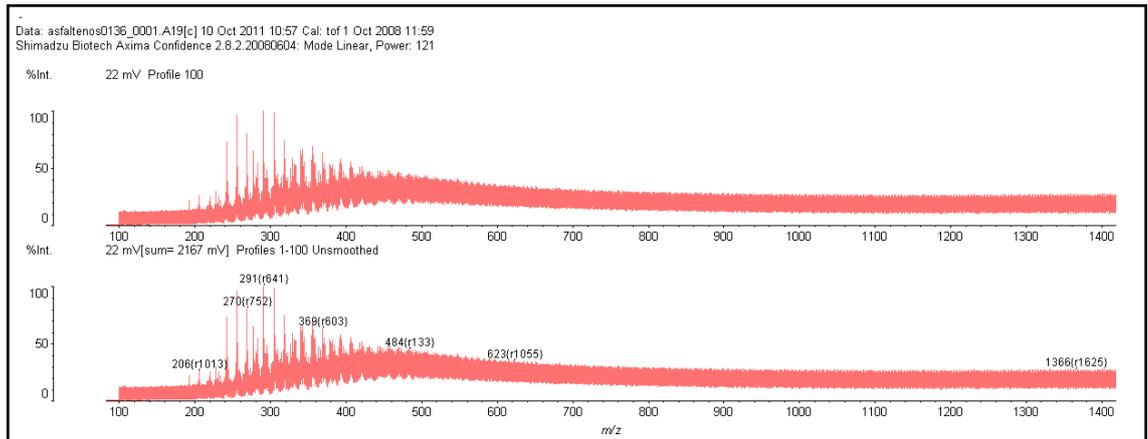


Figura 13: 25/02/11 – Petróleo UN (Amostra 19)

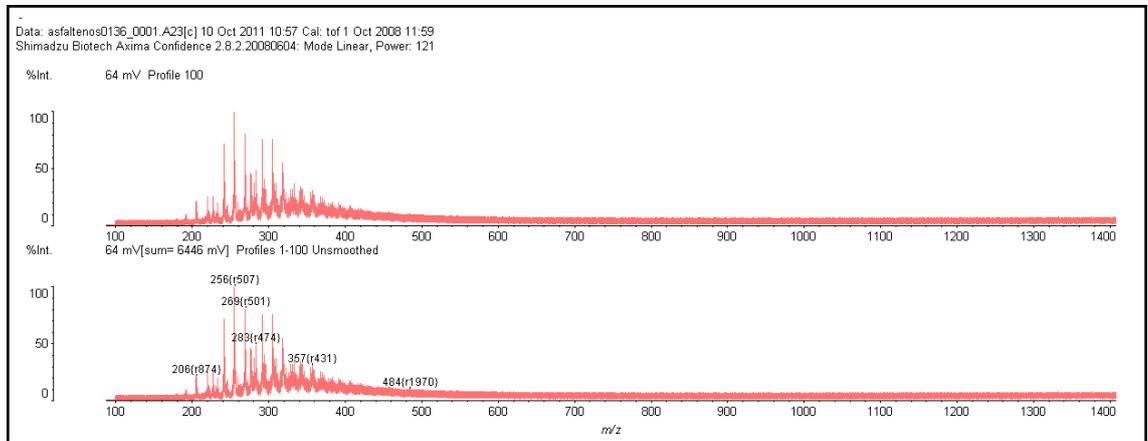


Figura 14: 25/02/11 – RAT UN (Amostra 23)

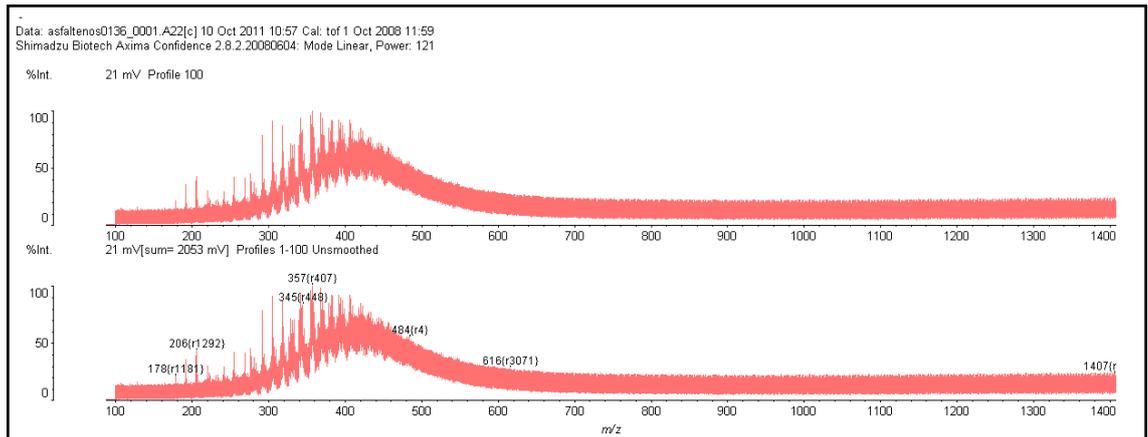


Figura 15: 25/02/11 – RV (Amostra 22)

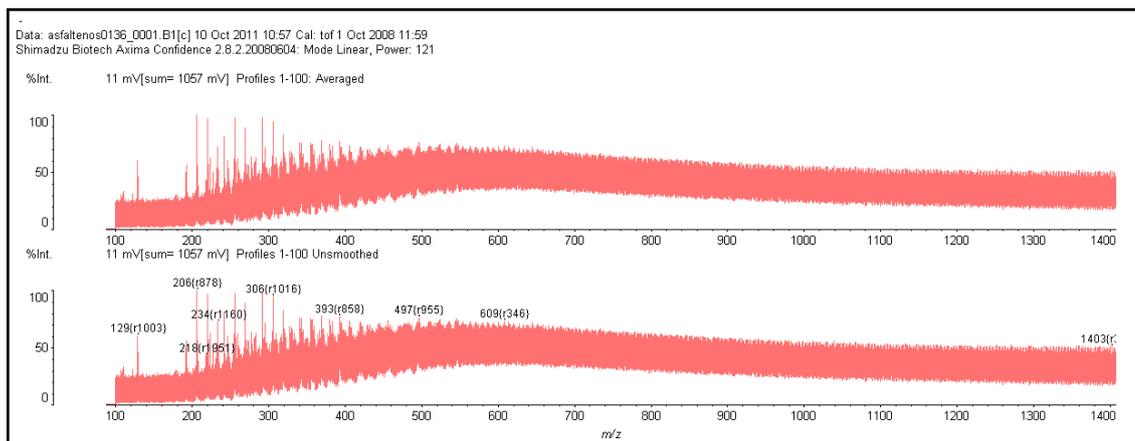


Figura 16: 25/02/11 – Petróleo UV (Amostra 25)

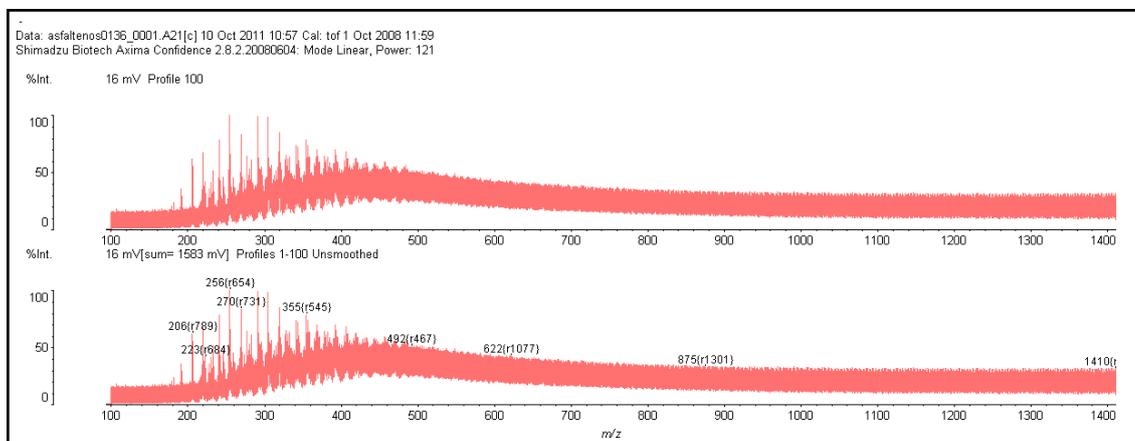


Figura 17: 25/02/11 – RAT UV (Amostra 21)

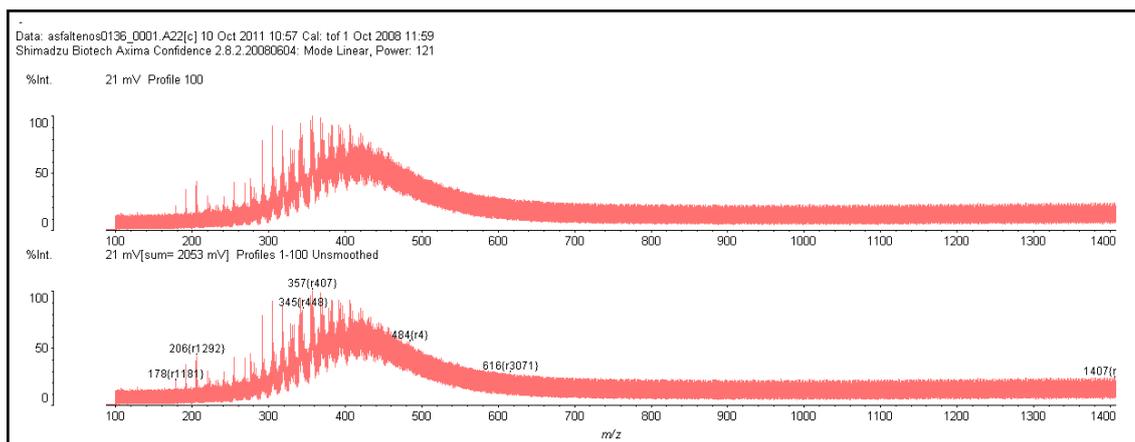


Figura 18: 25/02/11 – RV (Amostra 22)

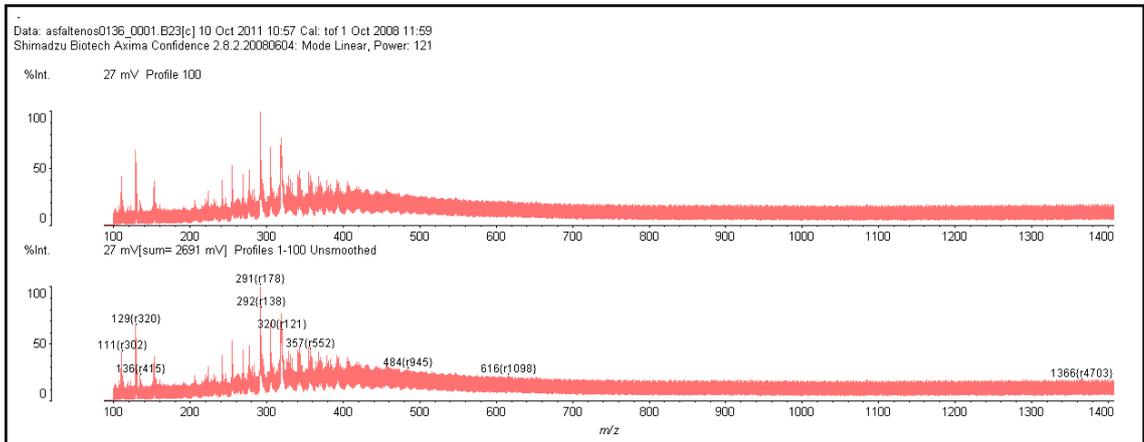


Figura 19: 02/03/11 Petróleo UC (Amostra 47)

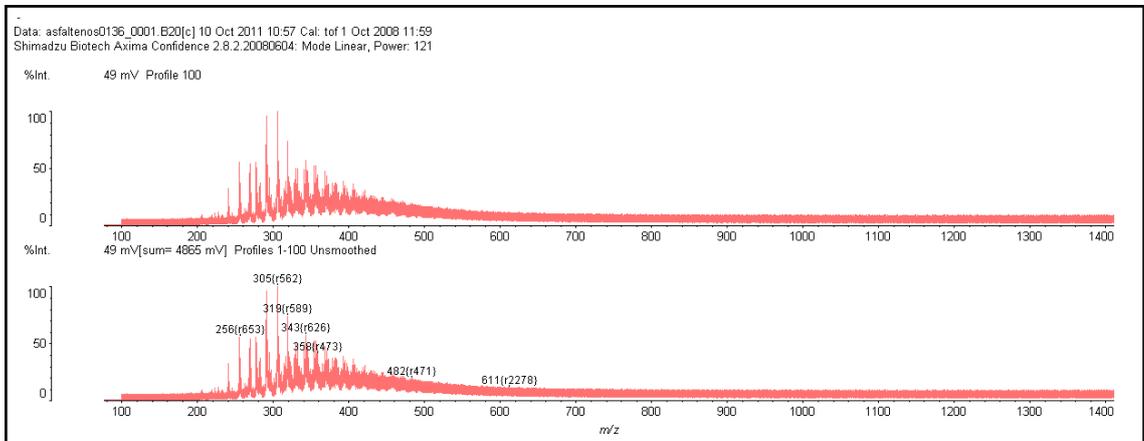


Figura 20: 02/03/11 RAT UC (Amostra 44)

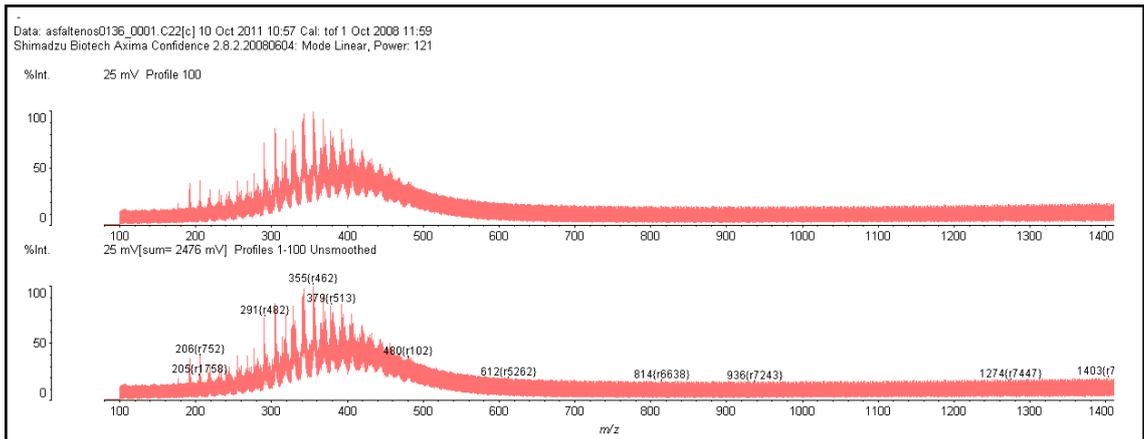


Figura 21: 02/03/11 RV (Amostra 70)

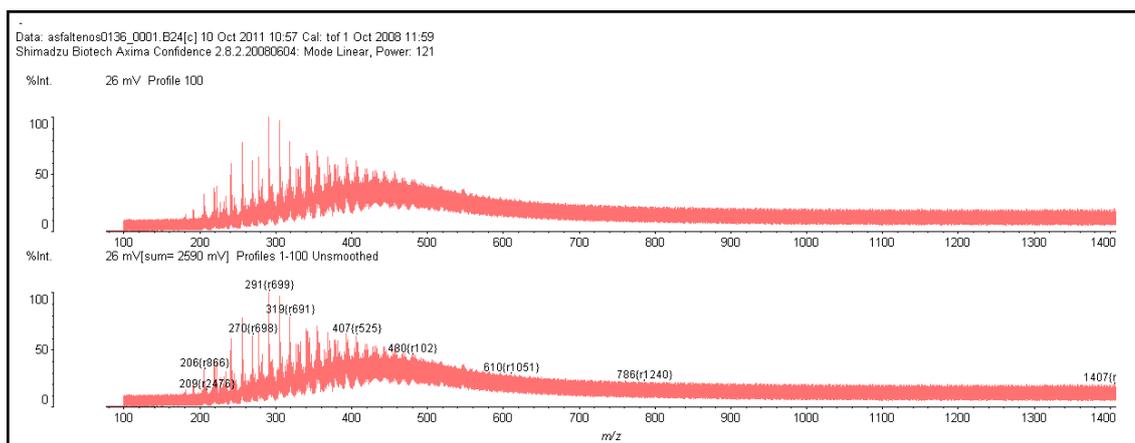


Figura 22: 02/03/11 Petr6leo UN (Amostra 48)

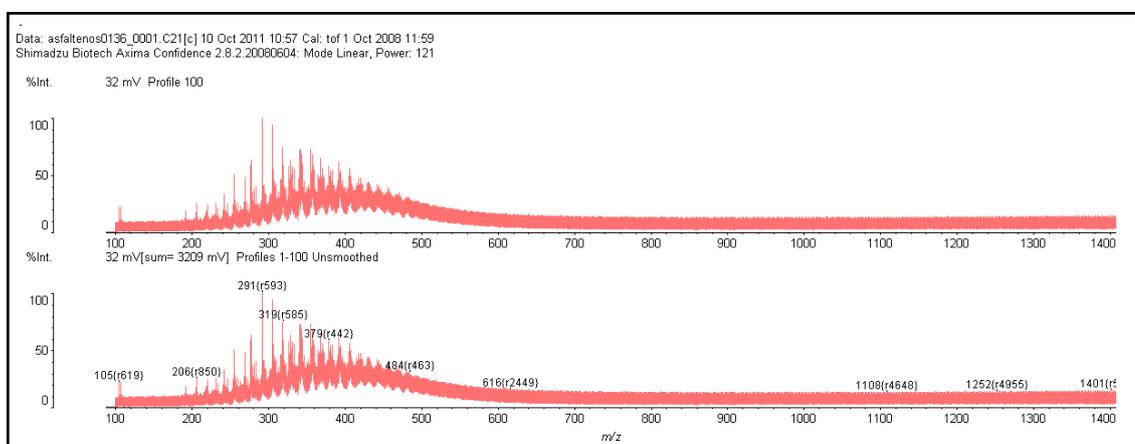


Figura 23: 02/03/11 RAT UN (Amostra 69)

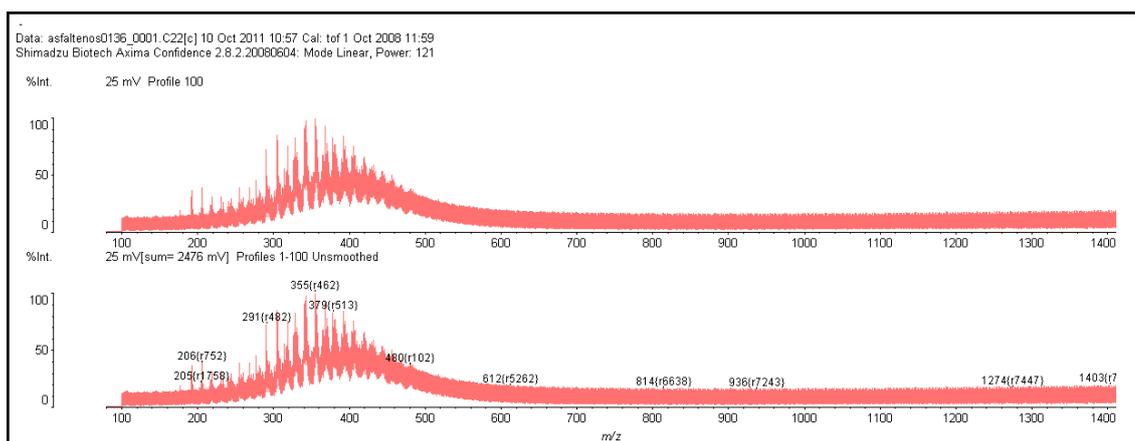


Figura 24: 02/03/11 RV (Amostra 70)

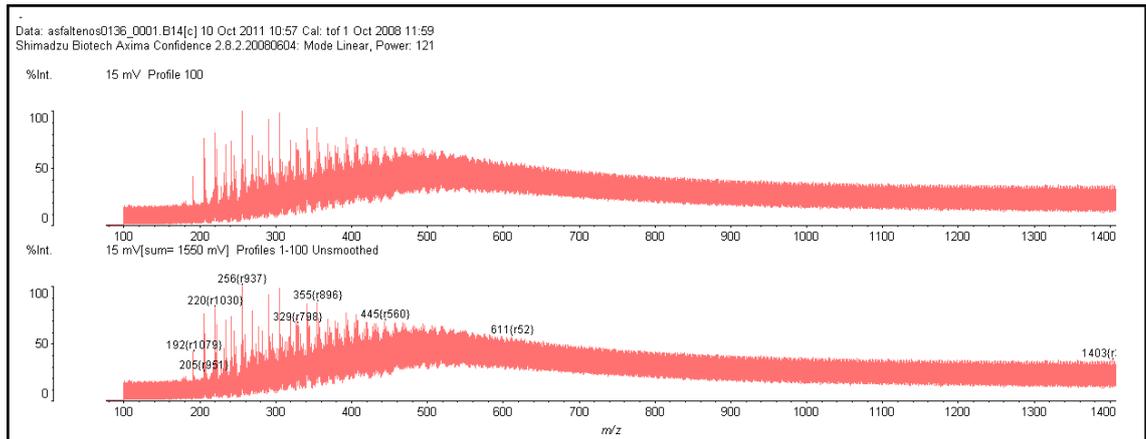


Figura 25: 02/03/11 Petr6leo UV (Amostra 38)

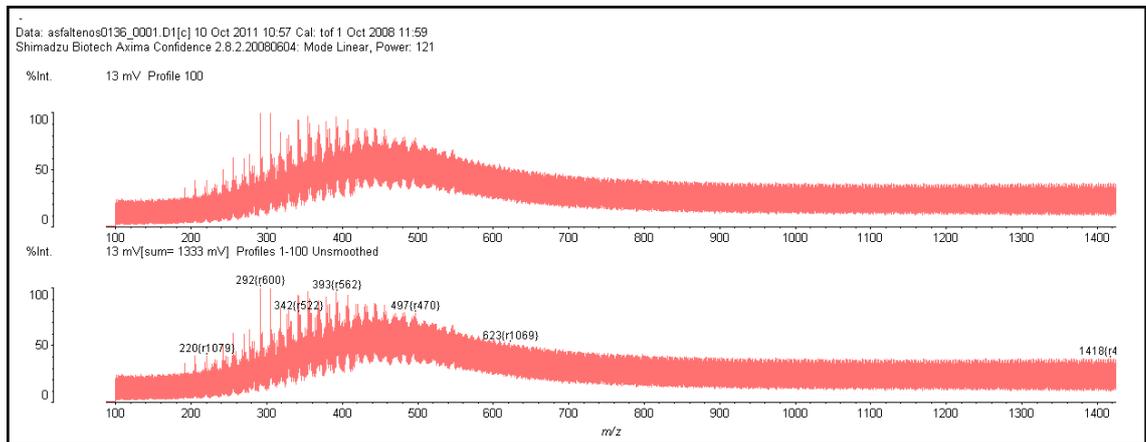


Figura 26: 02/03/11 RAT UV (Amostra 73)

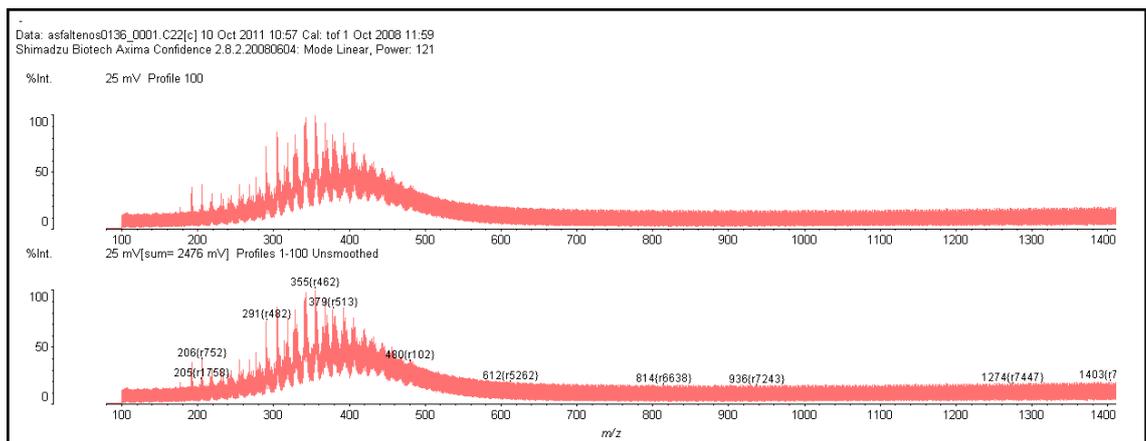


Figura 27: 02/03/11 RV (Amostra 70)

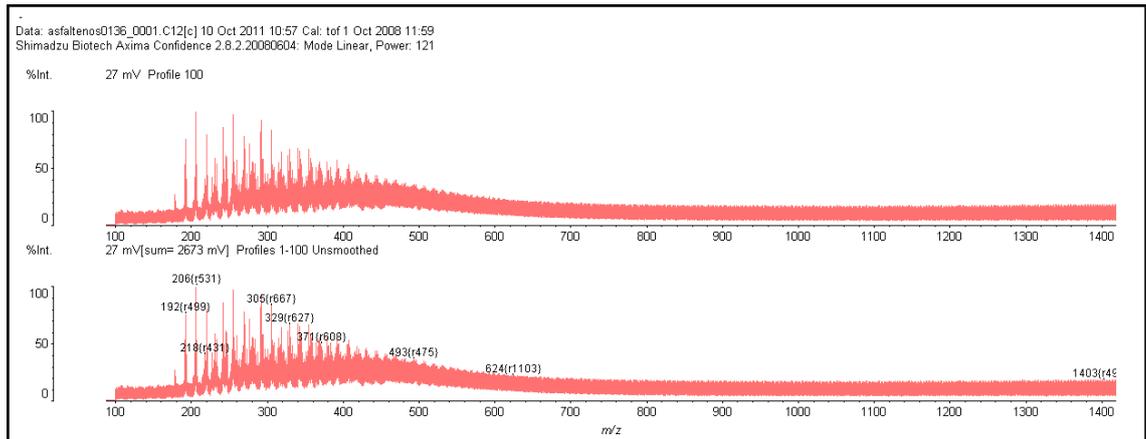


Figura 28: 29/03/11 Petr6leo UC (Amostra 60)

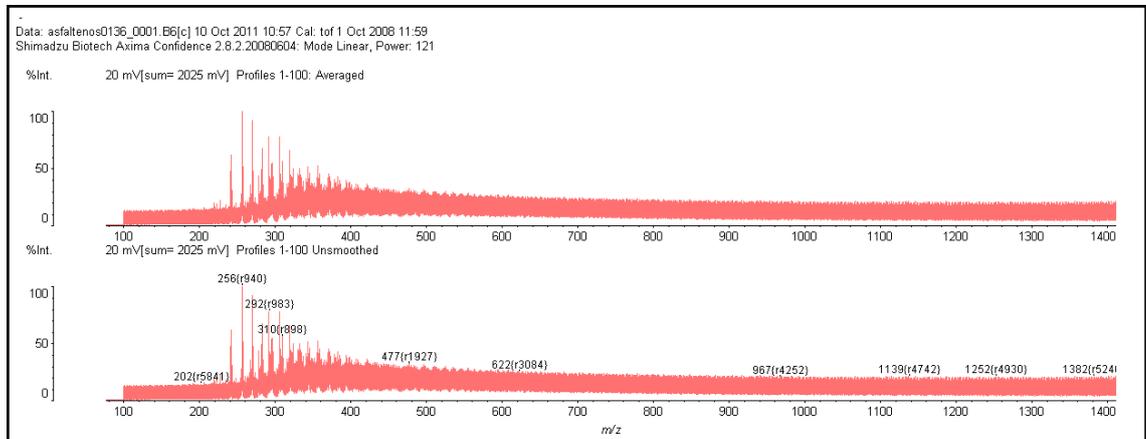


Figura 29: 29/03/11 RAT UC (Amostra 30)

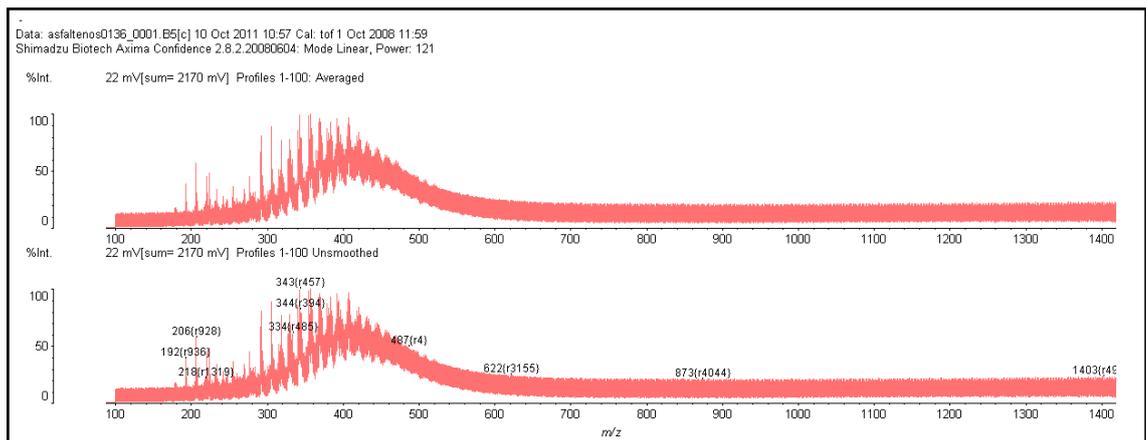


Figura 30: 29/03/11 RV (Amostra 29)

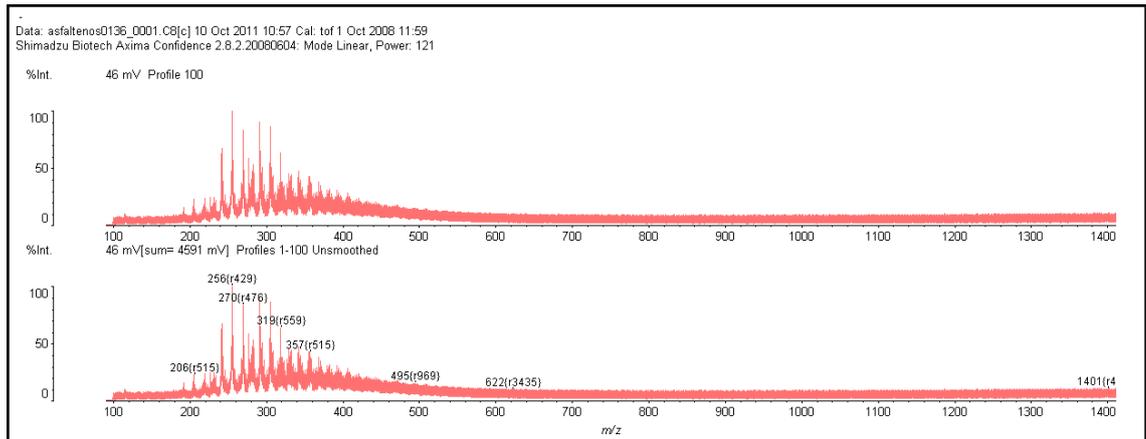


Figura 31: 29/03/11 Petróleo UN (Amostra 56)

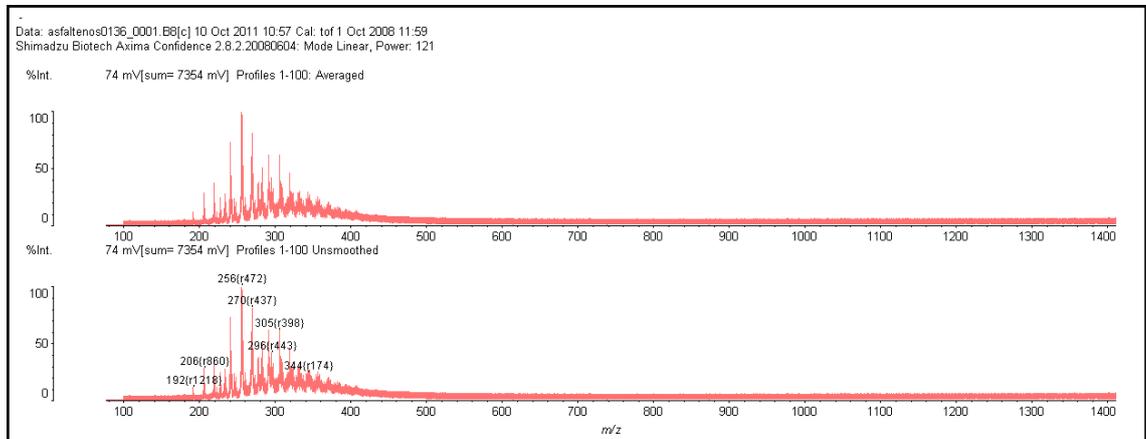


Figura 32: 29/03/11 RAT UN (Amostra 32)

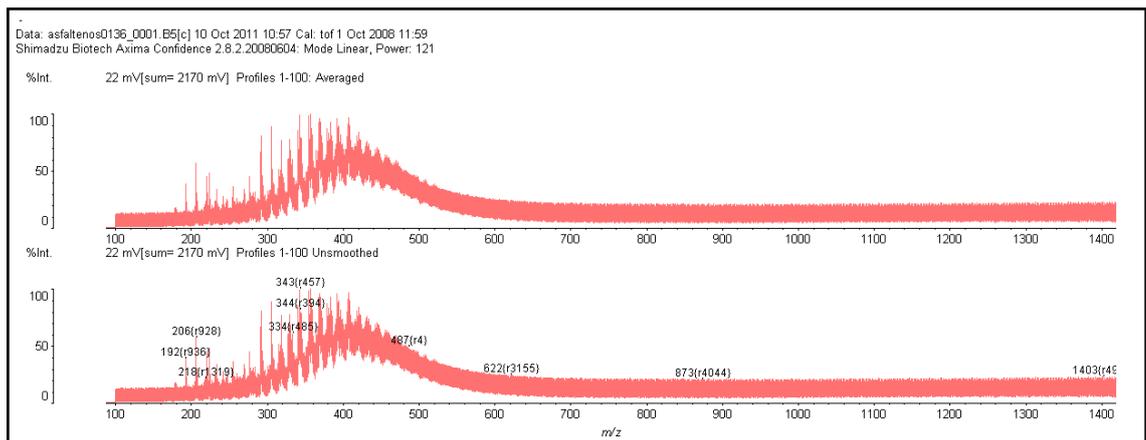


Figura 33: 29/03/11 RV (Amostra 29)

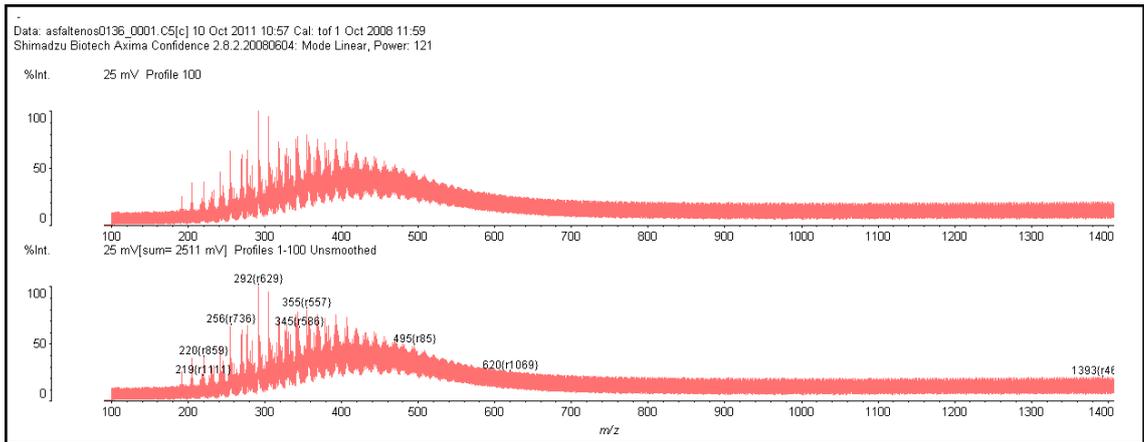


Figura 34: 29/03/11 Petróleo UV (Amostra 53)

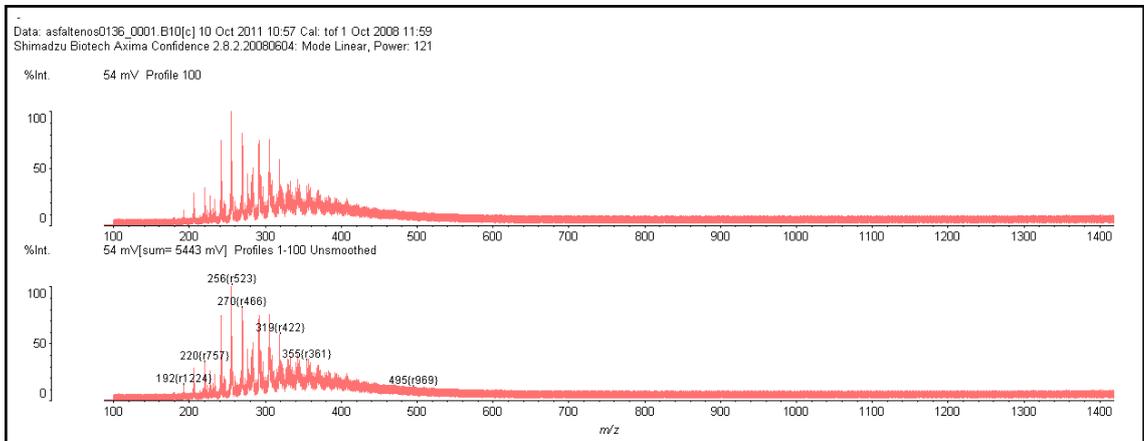


Figura 35: 29/03/11 RAT UV (Amostra 34)

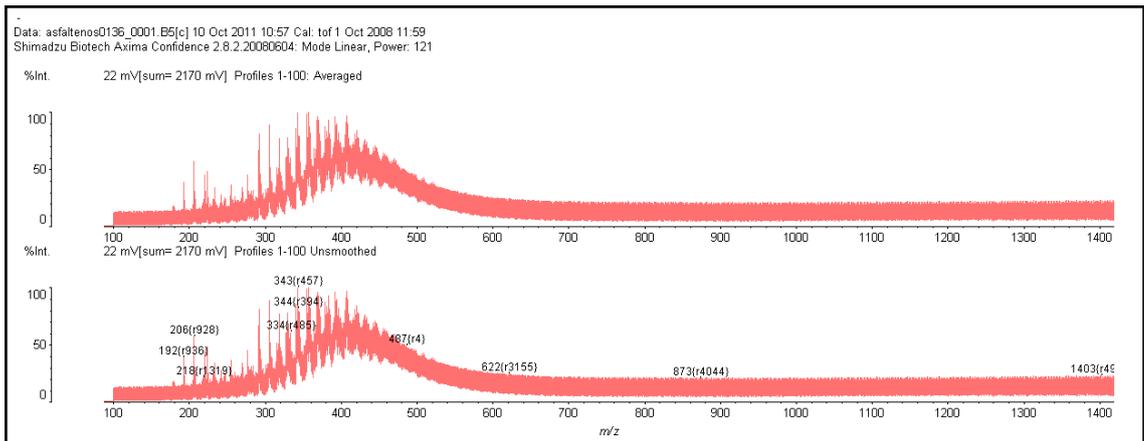


Figura 36: 29/03/11 RV (Amostra 29)

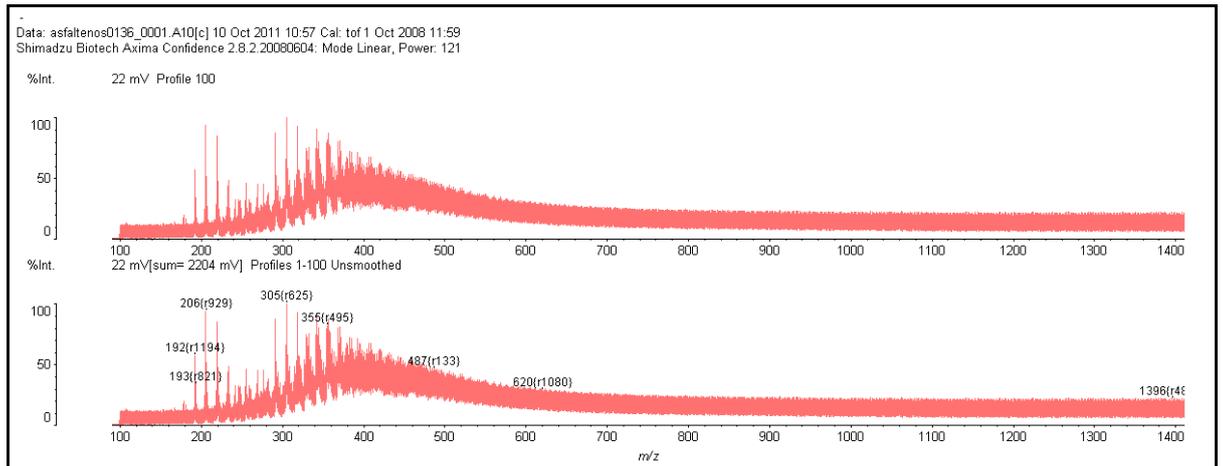


Figura 37: 28/10/10 RV (Amostra 10)

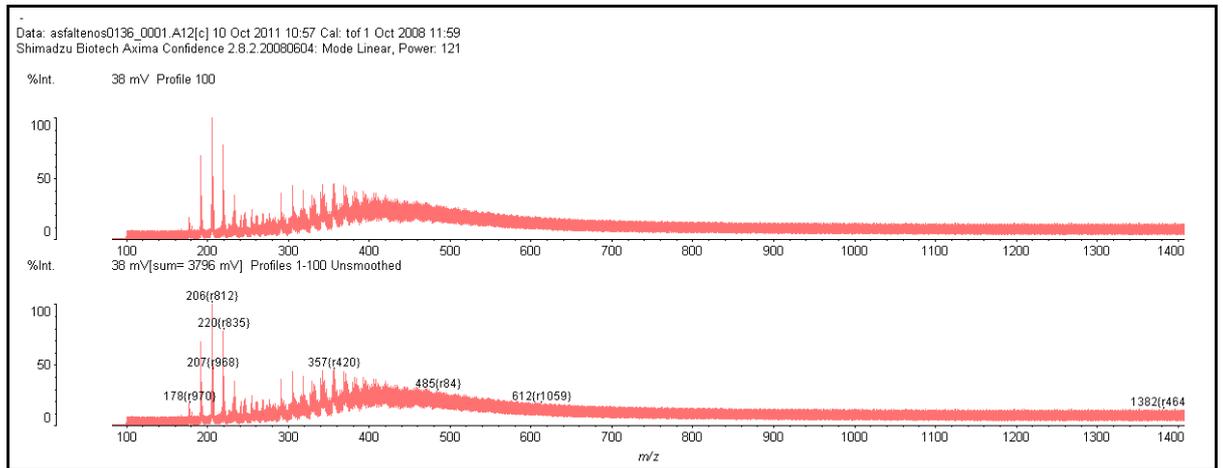


Figura 38: 22/10/10 RV (Amostra 12)

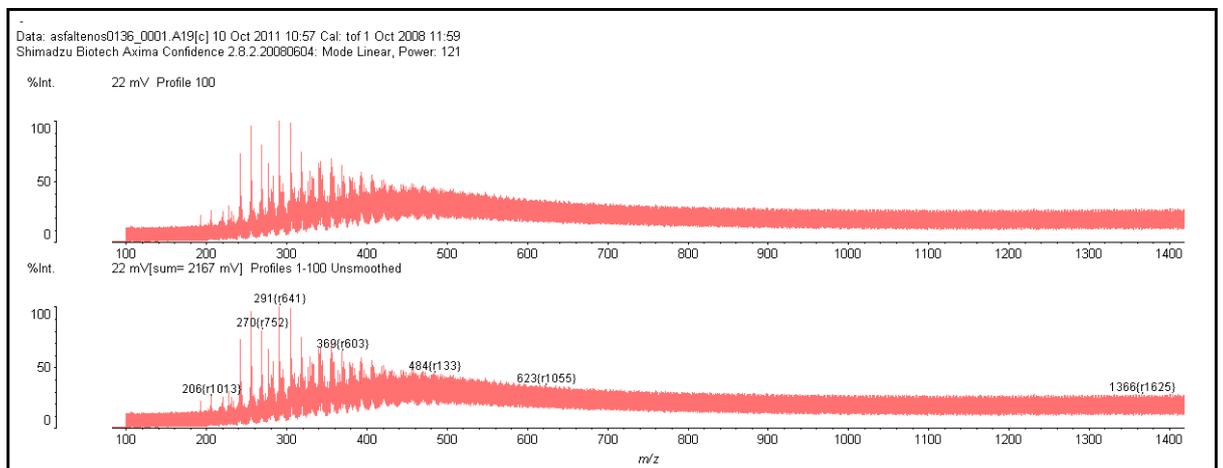


Figura 39: 25/02/11 Petróleo (Amostra 19)

### 8.3. Anexo 3: Resultados do Minitab para regressão do PC1 e PC2 (Fluorescência) e predição do °API

#### Regression Analysis: API versus score PC1; score PC2

The regression equation is

$$\text{API} = 26,3 + 0,0837 \text{ score PC1} - 0,0343 \text{ score PC2}$$

Predictor	Coef	SE Coef	T	P
Constant	26,3494	0,5295	49,76	0,000
score PC1	0,08366	0,03895	2,15	0,041
score PC2	-0,03427	0,05751	-0,60	0,556

S = 2,94809    R-Sq = 15,1%    R-Sq(adj) = 9,0%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	43,184	21,592	2,48	0,102
Residual Error	28	243,354	8,691		
Total	30	286,538			

Source	DF	Seq SS
score PC1	1	40,097
score PC2	1	3,086

#### Unusual Observations

Obs	score PC1	API	Fit	SE Fit	Residual	St Resid
3	25,5	31,809	29,267	1,729	2,542	1,06 X
4	16,5	35,316	28,145	1,084	7,171	2,62R
5	12,8	35,177	26,986	1,032	8,191	2,97R

R denotes an observation with a large standardized residual.

X denotes an observation whose X value gives it large leverage.

## 8.4. Anexo 4: Script para cálculo do PLSR através do Matlab

```

close all
clear all
load 'maldi2'
X = zscore(maldi);
Y = zscore(BulkExp);
[n,p] = size(X);
[XL,YL,XS,YS,Beta,PCTVAR,MSE,Stats] = plsregress(X,Y,13,'CV',10);
figure(1),hold on,plot(1:13,cumsum(100*PCTVAR(2,:)),'-bo');
title('Scree Plot PLSR/LDI-TOF')
xlabel('Número de componentes PLSR');
ylabel('Variância acumulada em Y (%)');
figure(2),hold on,plot(1:13,cumsum(100*PCTVAR(1,:)),'-ro');
title('Scree Plot PLSR/LDI-TOF')
xlabel('Número de componentes PLSR');
ylabel('Variância acumulada em X (%)');
figure(3),hold on,plot(0:13,MSE(2,:),'b-o');
title('Erro quarático médio estimado na predição (MSEP) PLSR/LDI-TOF')
xlabel('Número de componentes PLSR');
ylabel('MSEP');

Yfit = [ones(n,1) X]*Beta;

figure(4),hold on,plot(Y(:,1),Yfit(:,1),'bo');
title('Ajuste para API - PLSR/LDI-TOF');
xlabel('Valor experimental');
ylabel('Valor ajustado pelo modelo');
legend({'13 componentes'},'location','NW');
figure(5),hold on,plot(Y(:,2),Yfit(:,2),'r^');
title('Ajuste para % Asfaltenos - PLSR/LDI-TOF')
xlabel('Valor experimental');
ylabel('Valor ajustado pelo modelo');
legend({'13 componentes'},'location','NW');
figure(6),hold on,plot(Y(:,3),Yfit(:,3),'k+');
title('Ajuste para RC - PLSR/LDI-TOF')
xlabel('Valor experimental');
ylabel('Valor ajustado pelo modelo');
legend({'13 componentes'},'location','NW');

TSSapi = sum((Y(:,1)-mean(Y(:,1))).^2);
RSSapi = sum((Y(:,1)-Yfit(:,1)).^2);
rsquared_api = 1 - RSSapi/TSSapi
TSSasf = sum((Y(:,2)-mean(Y(:,2))).^2);
RSSasf = sum((Y(:,2)-Yfit(:,2)).^2);
rsquared_asf= 1 - RSSasf/TSSasf
TSS_RC = sum((Y(:,3)-mean(Y(:,3))).^2);
RSS_RC = sum((Y(:,3)-Yfit(:,3)).^2);
rsquared_RC = 1 - RSS_RC/TSS_RC

residuals = Y - Yfit;

figure(7),hold on,hist(residuals(:,1))
title('Distribuição de Resíduos para API (PLSR/LDI-TOF)');
xlabel('Resíduo');
ylabel('Frequência');
legend({'13 componentes'},'location','NW');
figure(8),hold on,hist(residuals(:,2))

```

```
title('Distribuição de Resíduos para % Asfaltenos (PLSR/LDI-TOF)');
xlabel('Resíduo');
ylabel('Frequência');
legend({'13 componentes'}, 'location', 'NW');
figure(9), hold on, hist(residuals(:,3))
title('Distribuição de Resíduos para RC (PLSR/LDI-TOF)');
xlabel('Resíduo');
ylabel('Frequência');
legend({'13 componentes'}, 'location', 'NW');

figure(10), hold on, plot(Yfit(:,1), residuals(:,1), 'mo')
title('Resíduos por Observação para API (PLSR/LDI-TOF)');
xlabel('Valor ajustado pelo modelo');
ylabel('Resíduo');
legend({'13 componentes'}, 'location', 'NW');
figure(11), hold on, plot(Yfit(:,2), residuals(:,2), 'm+')
title('Resíduos por Observação para % Asfaltenos (PLSR/LDI-TOF)');
xlabel('Valor ajustado pelo modelo');
ylabel('Resíduo');
legend({'13 componentes'}, 'location', 'NW');
figure(12), hold on, plot(Yfit(:,3), residuals(:,3), 'm^')
title('Resíduos por Observação para RC (PLSR/LDI-TOF)');
xlabel('Valor ajustado pelo modelo');
ylabel('Resíduo');
legend({'13 componentes'}, 'location', 'NW');
```

### 8.5. Anexo 5: Resultados do PLSR/Fluorescência com o uso de 10 componentes

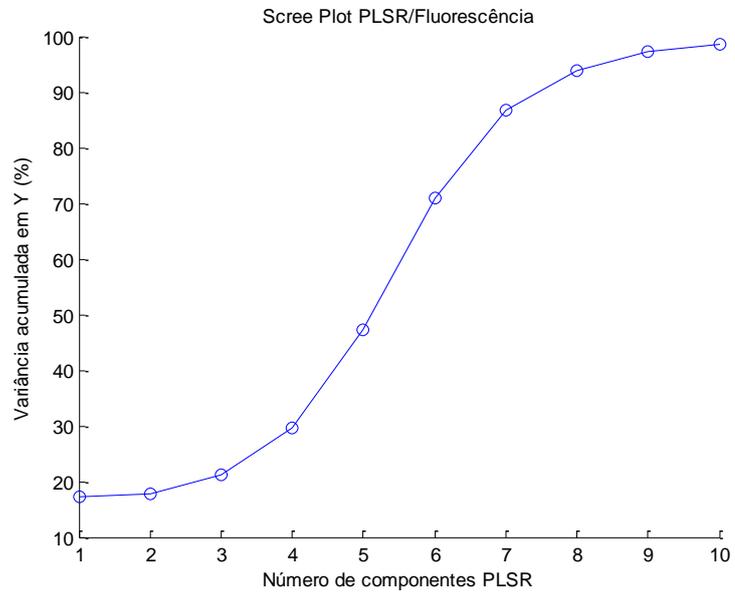


Figura 40: Variância acumulada na matriz Y (PLSR/Fluorescência)

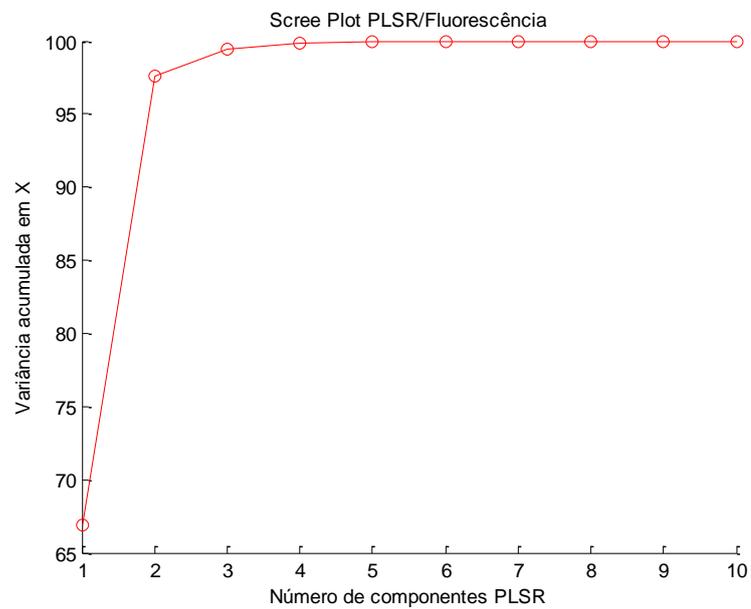


Figura 41: Variância acumulada na matriz X (PLSR/Fluorescência)

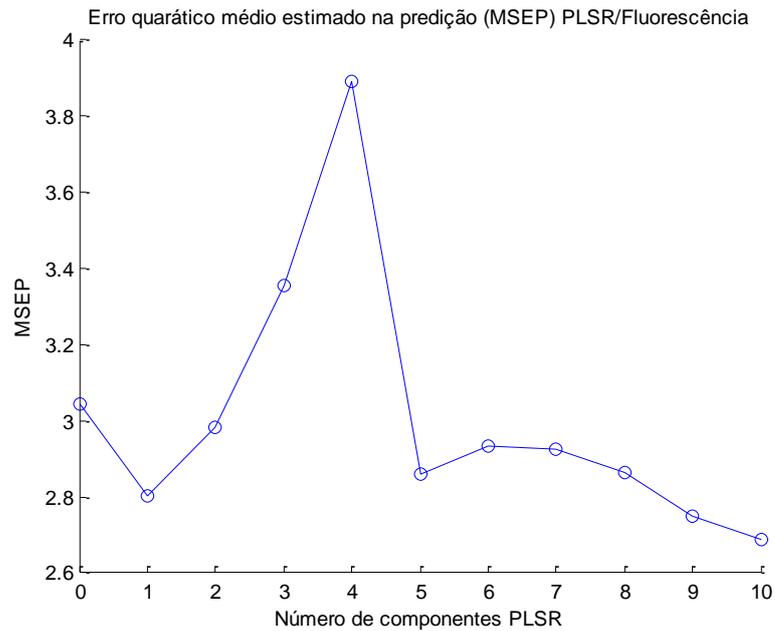


Figura 42: MSEP em função do número de componentes (PLSR/Fluorescência)

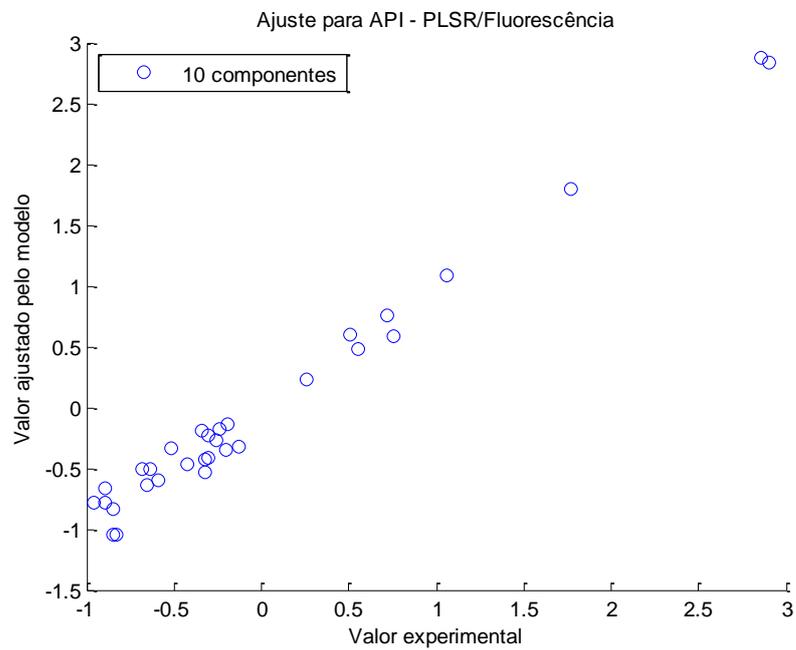


Figura 43: API calculado em função do API experimental (PLSR/Fluorescência) com 10 componentes

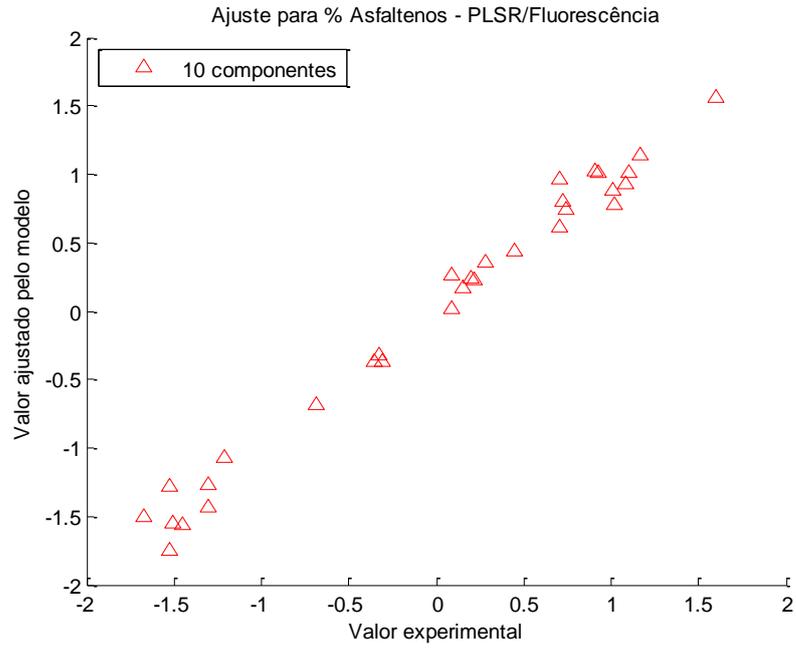


Figura 44: Teor de asfaltenos calculado em função do experimental (PLSR/Fluorescência) com 10 componentes

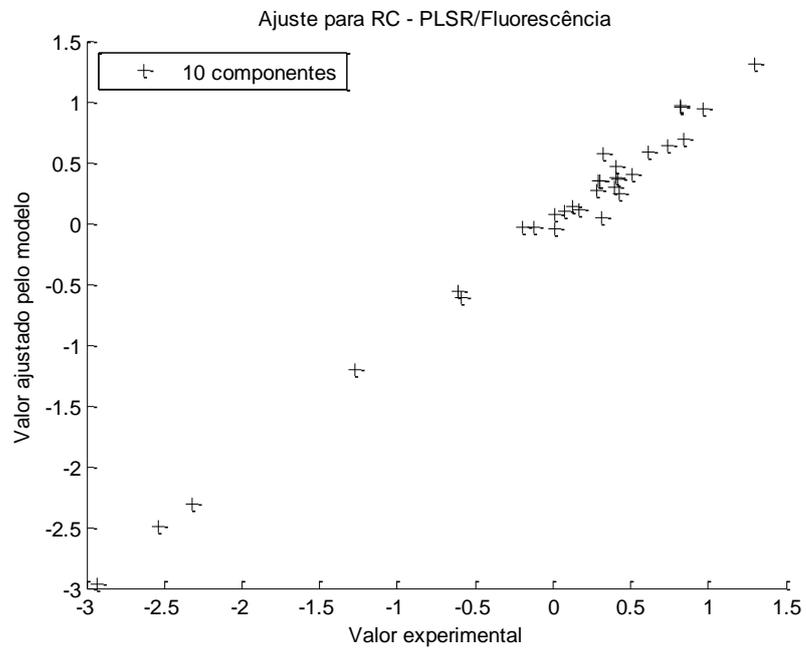


Figura 45: RC calculado em função do RC experimental (PLSR/Fluorescência) com 10 componentes

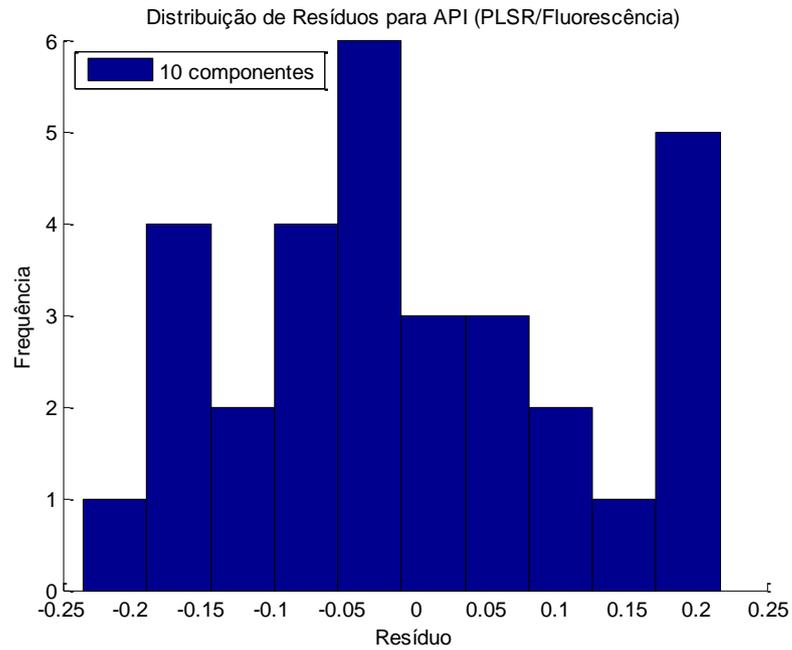


Figura 46: Distribuição de Resíduos para o modelo do API (PLSR/Fluorescência) com 10 componentes

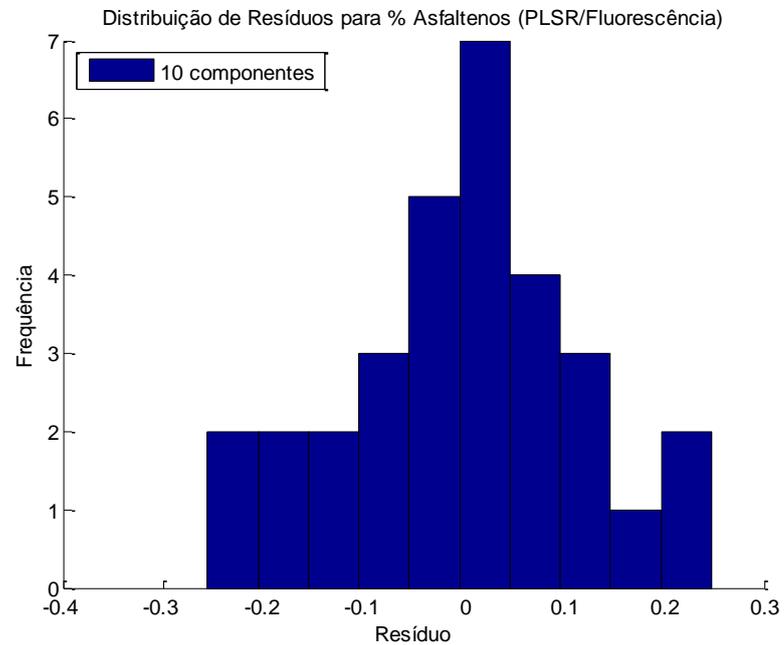


Figura 47: Distribuição de Resíduos para o modelo do teor de asfaltenos (PLSR/Fluorescência) com 10 componentes

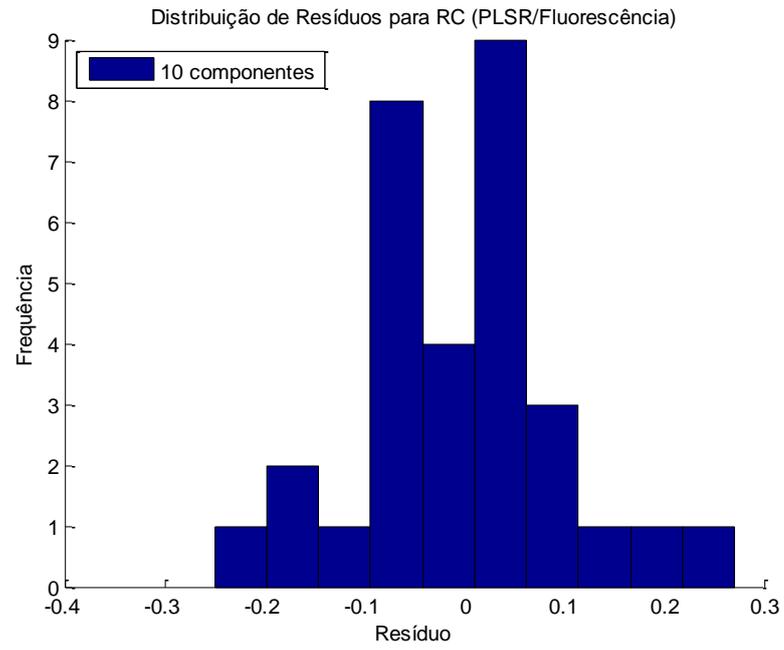


Figura 48: Distribuição de Resíduos para o modelo do RC (PLSR/Fluorescência) com 10 componentes

## 8.6. Anexo 6: Resultados do PLSR/LDI-TOF com o uso de 13 componentes

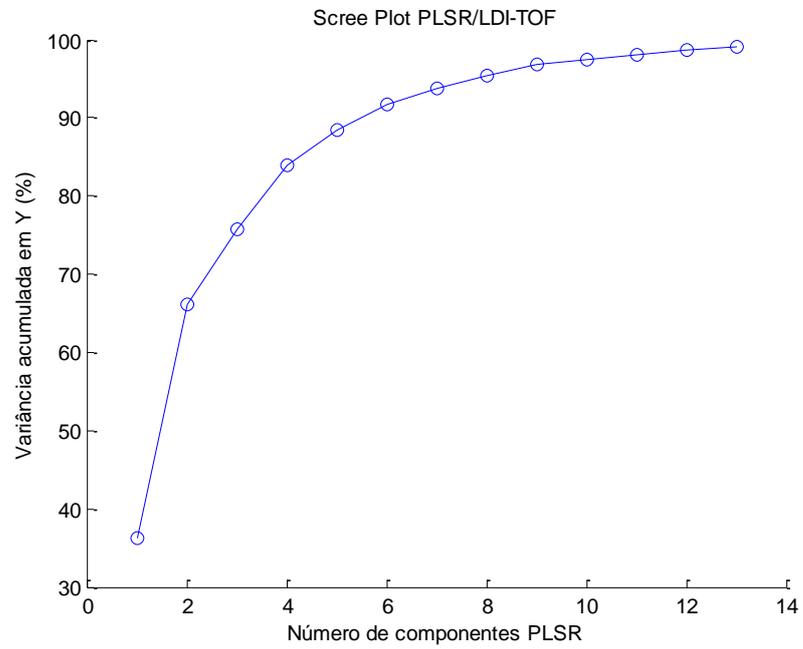


Figura 49: Variância acumulada na matriz Y (PLSR/LDI-TOF)

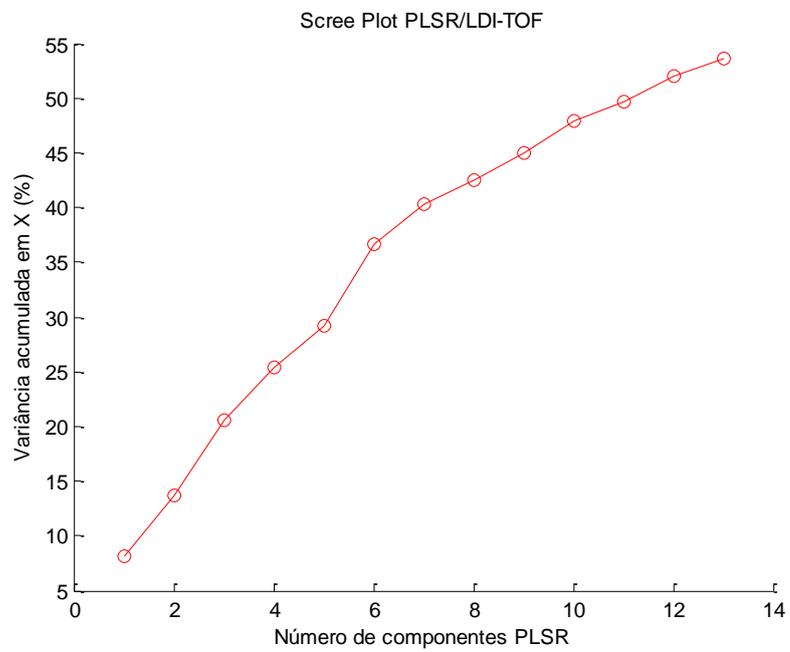


Figura 50: Variância acumulada na matriz X (PLSR/LDI-TOF)

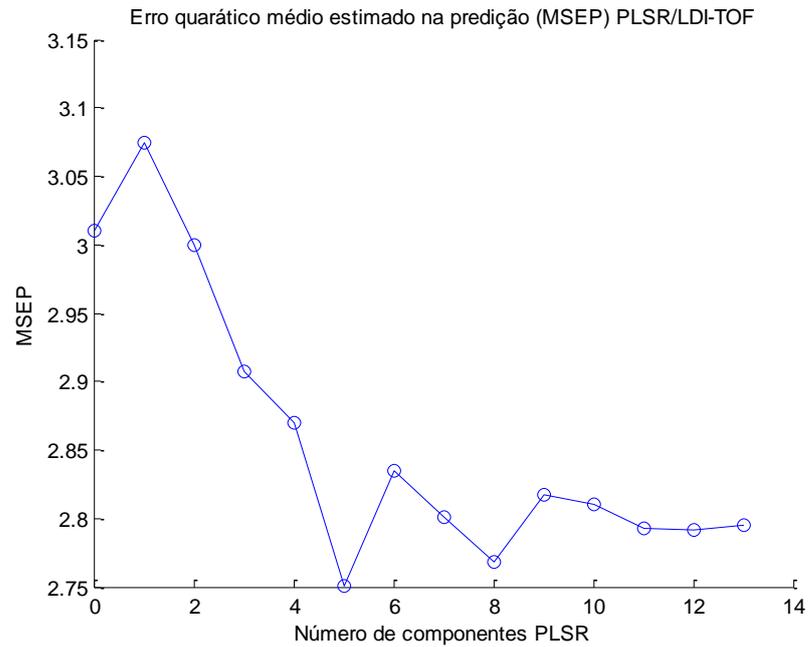


Figura 51: MSEP em função do número de componentes (PLSR/LDI-TOF)

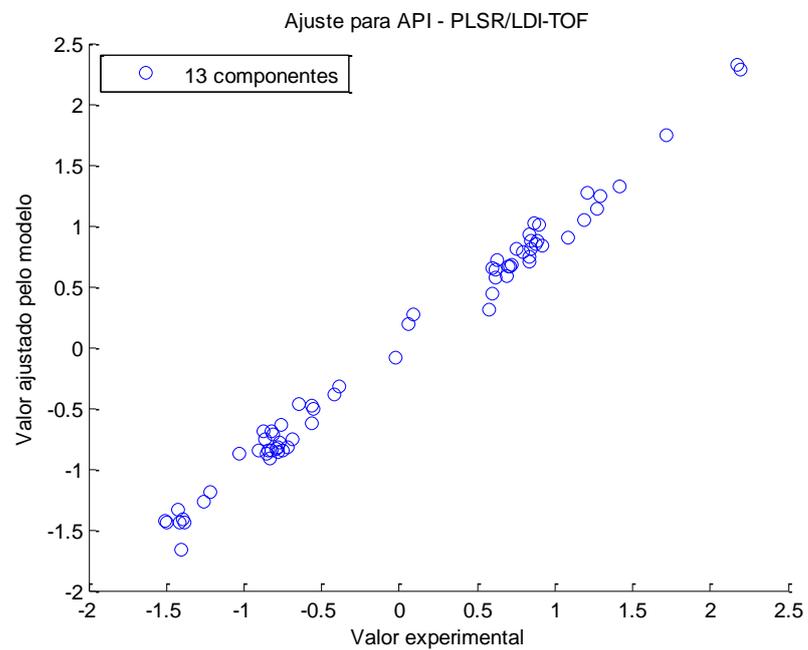


Figura 52: API calculado em função do API experimental (PLSR/LDI-TOF) com 13 componentes

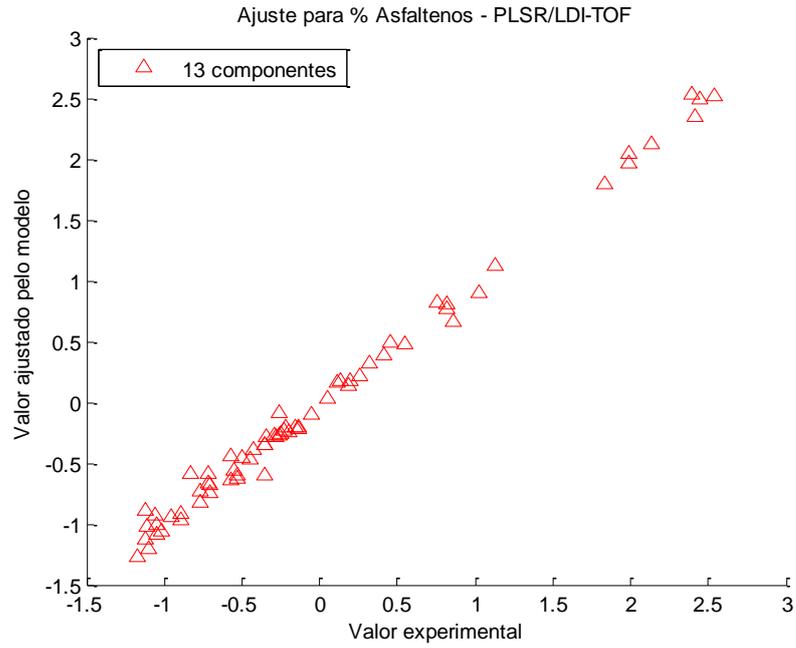


Figura 53: Teor de asfaltenos calculado em função do experimental (PLSR/LDI-TOF) com 13 componentes

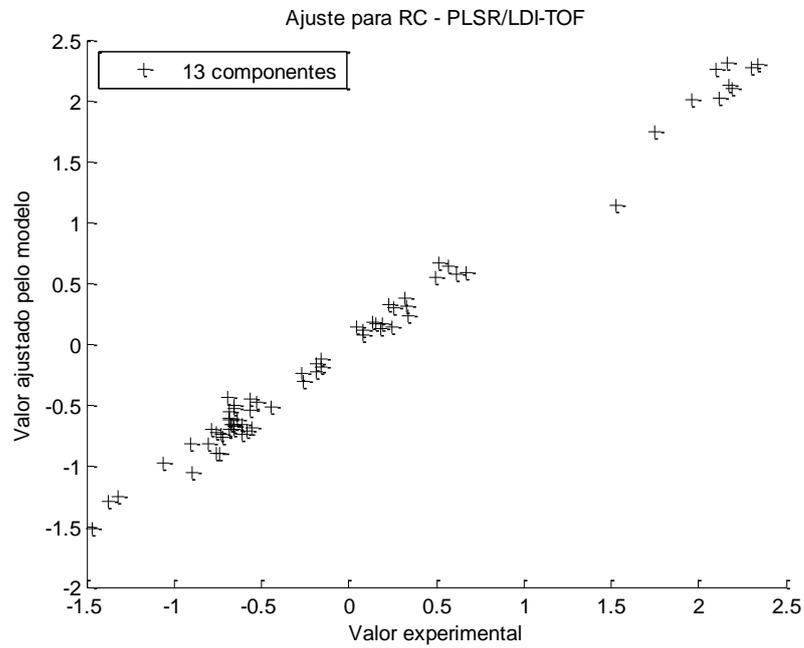


Figura 54: RC calculado em função do RC experimental (PLSR/LDI-TOF) com 13 componentes

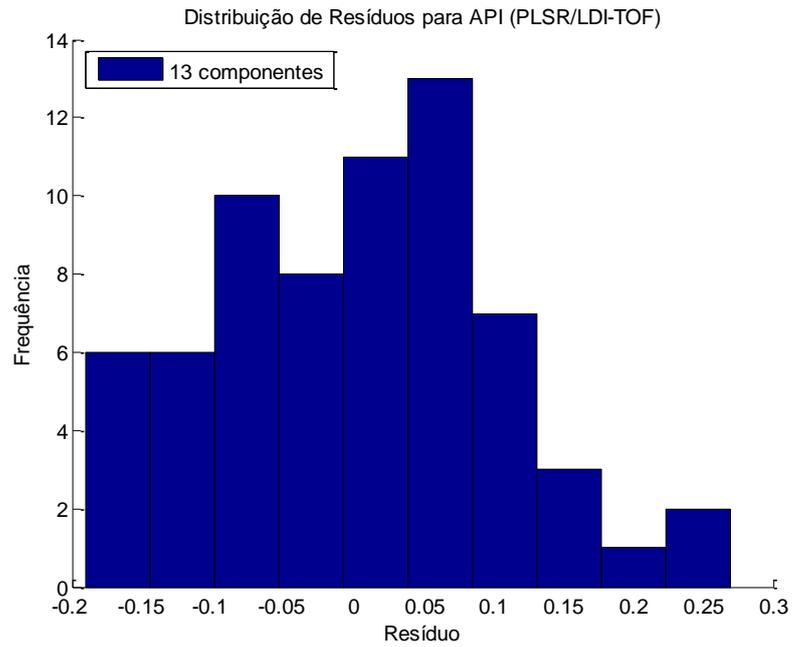


Figura 55: Distribuição de Resíduos para o modelo do API (PLSR/LDI-TOF) com 13 componentes

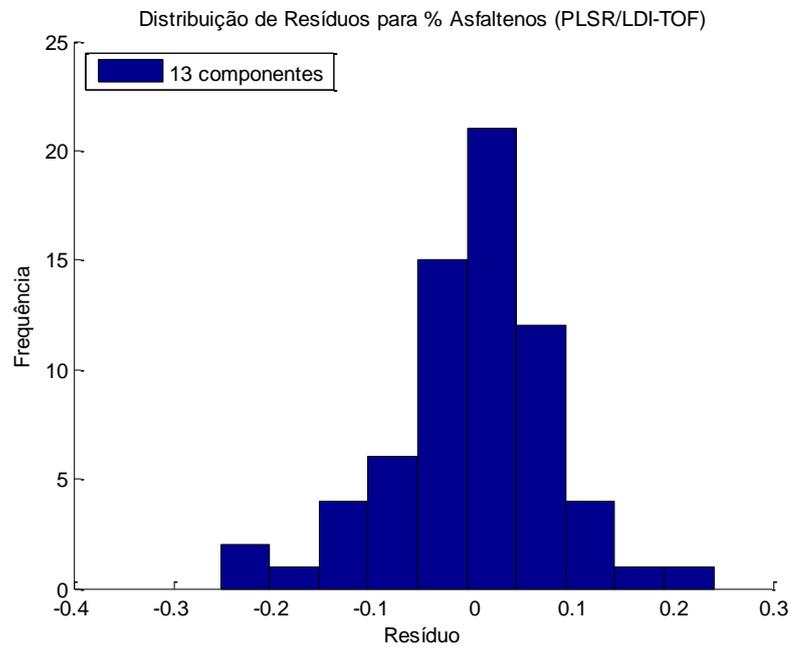


Figura 56: Distribuição de Resíduos para o modelo do teor de asfaltenos (PLSR/LDI-TOF) com 13 componentes

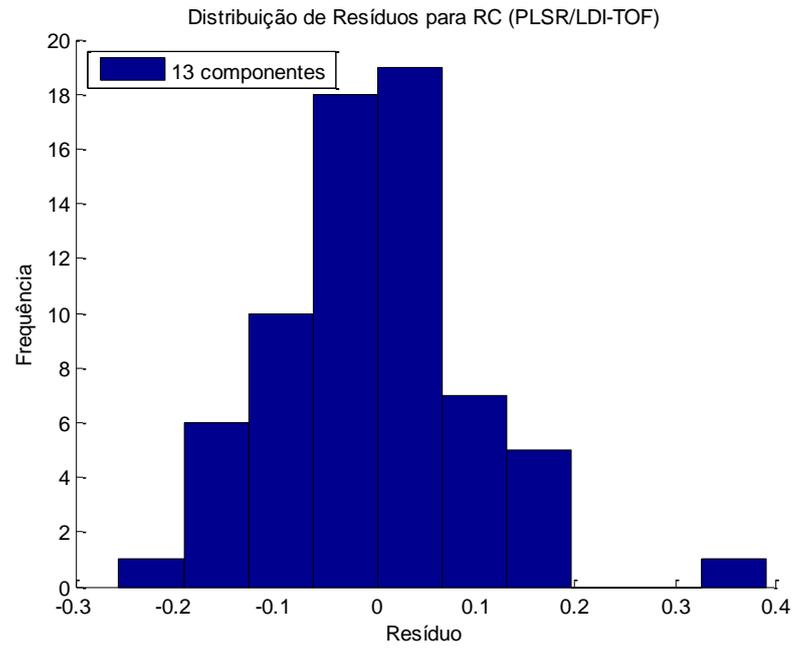


Figura 57: Distribuição de Resíduos para o modelo do RC (PLSR/LDI-TOF) com 13 componentes