



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

MILLA BEATRICE ENGELMANN DE OLIVEIRA GARCIA

**PREDIÇÃO DE CURVA DE DESTILAÇÃO DE GASOLINA DE
PIRÓLISE A PARTIR DE RESULTADOS DE CROMATOGRAFIA
EM FASE GASOSA E CALIBRAÇÃO MULTIVARIADA**

DISSERTAÇÃO DE MESTRADO

Porto Alegre - RS
Dezembro de 2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

MILLA BEATRICE ENGELMANN DE OLIVEIRA GARCIA

**PREDIÇÃO DE CURVA DE DESTILAÇÃO DE GASOLINA DE
PIRÓLISE A PARTIR DE RESULTADOS DE CROMATOGRAFIA
EM FASE GASOSA E CALIBRAÇÃO MULTIVARIADA**

Dissertação apresentada como requisito
parcial para a obtenção do grau de Mestre
em Química

Prof. Dr. Adriano de Araújo Gomes
Orientador

Porto Alegre-RS,
Dezembro de 2021.

*“Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar.
Mas o mar seria menor se lhe faltasse uma gota.”*

Agradecimentos

À Deus, pelo dom da vida e por tantas graças concedidas.

Ao meu orientador, Adriano Araújo Gomes, por ter me acolhido com tanta atenção depois de tantos anos fora da sala de aula, pela paciência com minha realidade e com minhas dificuldades, e pelos ensinamentos em quimiometria.

À UFRGS pela oportunidade de realizar esse Mestrado que para mim é tão valioso.

À Braskem pela oportunidade dada de usar seus recursos. Em especial ao meu coordenador Ricardo Casal por ter me oportunizado aquilo que me é tão caro: o tempo. Aos meus colegas de dia-a-dia, que entenderam minhas ausências e me deram o apoio necessária para o meu desenvolvimento. E aqueles colegas que toparam o desafio de aprender junto comigo e entregar ainda mais valor à Braskem: Bruna Dias, Carlos Becker, Gabriel Hanauer, Julie Skalee, Simone Ziles e Tainara Strassburger.

Aos meus colegas da Braskem da Bahia, que anos atrás, ao me convidarem para participar de um curso, me mostraram a Quimiometria, pela qual imediatamente me encantei. Em especial à Ana Rosa, pelas dúvidas dirimidas e sugestões ao longo dessa caminhada.

À minha mãe pelo entendimento, pela força, pela fortaleza. Aos meus queridos irmãos, cunhados e sogra, por serem tão queridos comigo e com meus filhos e pelas dicas para seguir adiante.

Ao trio maravilhoso do meu dia-a-dia: meu marido Éder e meus filhos Ivan e Lara. Vocês são a luz diária que Deus enviou para fazer brilhar meus dias! Obrigada pela paciência e desculpem-me por necessidades de ausência, por trocar horas de brincadeiras por estudo e escrita.

Muito obrigada a todos! Sem vocês, essa conquista não seria possível!

Sumário

1 Introdução	13
1.1 Objetivos	14
<i>1.1.1 Objetivos Específicos</i>	<i>14</i>
2 Revisão Bibliográfica	16
2.1 Gasolina de Pirólise: definição e processo de obtenção	16
2.2 Cromatografia em fase gasosa	20
<i>2.2.1 Injetores</i>	<i>21</i>
<i>2.2.2 Colunas</i>	<i>21</i>
<i>2.2.3 Detectores</i>	<i>22</i>
<i>2.2.4 Calibração por Número de Carbono Efetivo</i>	<i>24</i>
<i>2.2.5 Destilação Simulada</i>	<i>25</i>
2.4 Métodos Quimiométricos	28
<i>2.4.1 Análise Exploratória de Dados</i>	<i>29</i>
<i>2.4.2 Pré-Tratamento dos dados</i>	<i>31</i>
<i>2.4.3 Separação das amostras</i>	<i>31</i>
<i>2.4.4 Calibração Multivariada</i>	<i>32</i>
<i>2.4.5 Regressão Linear Múltipla</i>	<i>33</i>
<i>2.4.6 Regressão pelo Método dos Quadrados Mínimos Parciais</i>	<i>33</i>
<i>2.4.7 Avaliação dos Modelos</i>	<i>35</i>
<i>2.4.8 Validação Cruzada</i>	<i>35</i>
<i>2.4.9 Métricas de Desempenho</i>	<i>36</i>
2.5 Repetibilidade e Reprodutibilidade	38
3 PARTE EXPERIMENTAL	40
3.1 Amostras	40
3.2 Cromatografia em fase gasosa	41
3.3 Destilação	41
<i>3.3.1 Repetibilidade e Reprodutibilidade do ASTM D86</i>	<i>41</i>
3.4 Processamento de Dados	43
4 RESULTADOS E DISCUSSÕES	45
4.1 Estatística descritiva do conjunto de dados	45
4.2 Pré processamento do conjunto de dados - autoescalamento	51
4.3 Separação das Amostras	51

4.4 Análise de Componentes Principais	53
4.5 MLR.....	54
4.6 PLS.....	59
4.7 Comparação entre os modelos MLR e PLS	63
4.8 Comparação dos modelos com a Referência ASTM D86 e aplicação na indústria.....	64
5 CONCLUSÃO	70
REFERÊNCIAS BIBLIOGRÁFICAS.....	72
APÊNDICE A – PUBLICAÇÃO CIENTÍFICA.....	77

Lista de Figuras

Figura 1: Fluxo de produção das várias frações que compõem a gasolina de pirólise, desde o craqueamento à vapor de nafta até seu processamento. Fonte própria.....	19
Figura 2: Sistema cromatográfico simplificado. Fonte: modificado do site da DCtech ¹⁷	20
Figura 3: Esquema de um GC-MS e algumas de suas possibilidades de configuração. Fonte: modificado do site da Revista Analytica ^{23, 24}	24
Figura 4: Destilador automático e indicações dos principais componentes. Fonte própria.....	26
Figura 5: Esquema simplificado de aplicação de Quimiometria no desenvolvimento de preditor. Fonte própria.....	29
Figura 6: Disponibilização dos dados da amostra de <i>PYGAS</i> Pesada, conforme disponibilizado para produção e engenharia.	40
Figura 7: Organização dos dados na matriz X e vetores Y. Fonte própria.....	44
Figura 8: gráficos de distribuição de cada variável x em carta controle.....	46
Figura 9: Gráficos de distribuição de cada resultado y de cada parâmetro da destilaçãoem carta controle.	49
Figura 10: Cromatograma de uma amostra de <i>PYGAS</i> pesada, representado pelo gráfico e pela lista de componentes com as respectivas concentrações.....	50
Figura 11: Gráfico de Destilação da mesma amostra de <i>PYGAS</i> pesada.....	50
Figura 12: Distribuição dos dados da matriz X , graficamente. No Quadro 1, os valores originais de cada variável x . No quadro 2, os dados autoescalados de cada variável x ..	51
Figura 13: Resultados da análise de componentes principais sobre os blocos de dados X (a: escores, b: pesos e c: Resíduos Q versus T^2) e Y (d: escores, e: pesos e f: Resíduos Q versus T^2).....	53
Figura 14: Gráficos de cada parâmetro dos valores atuais (referência) versus preditos pela equação obtida por MLR, para as 20 amostras do conjunto de predição.	55
Figura 15: O peso de cada variável x definida para o modelo MLR de cada ponto da destilação	57

Figura 16: O peso de cada variável x para o modelo PLS de cada ponto da destilação.	60
Figura 17: Gráfico dos resultados de destilação preditos versus referência para cada parâmetro de destilação para os modelos de predição obtidos pelo PLS.	62
Figura 18: Figuras de mérito plotadas para MLR e PLS. Quadro A com os valores de RMSEP. Quadro B com os valores de <i>bias</i> . Quadro C com os valores de r	64
Figura 19: Resíduos do modelo MLR para o conjunto de predição, em comparação com a repetibilidade e reprodutibilidade do ASTM D86.	66
Figura 20: Resíduos do modelo PLS para o conjunto de predição, em comparação com a repetibilidade e reprodutibilidade do ASTM D86.	67
Figura 21: Exemplos de resultados disponibilizados para Engenharia e Produção dos dados da amostra de <i>PYGAS</i> Pesada, com valores práticos realizados conforme ASTM D86 e valores preditos pela equação.	69

Lista de Tabelas

Tabela 1: Valores e equações de repetibilidade e reprodutibilidade conforme ASTM D86-18 para cada ponto da destilação.....	42
Tabela 2: Apresentação resumida dos resultados cromatográficos das 68 amostras, em % massa.....	45
Tabela 3: Correlação de Pearson entre todas as variáveis.	47
Tabela 4: Estatística básica dos resultados de destilação das 68 amostras.....	48
Tabela 5: Amostras do conjunto de predição e a respectiva posição na distribuição.	52
Tabela 6: Amostras do conjunto de calibração/validação e a respectiva posição na distribuição.	52
Tabela 7: Apresentação dos resultados obtidos de cada parâmetro aplicando MLR.....	54
Tabela 8: Valores de b_0 de cada equação de destilação obtida por MLR.	57
Tabela 9: Resumo estatístico de calibração e validação cruzada da aplicação do PLS..	59
Tabela 10: Resumo estatístico da predição na aplicação do PLS.....	62
Tabela 11: Quadro demonstrativo entre valores de repetibilidade e reprodutibilidade do ASTM D86-18 ¹ e RMSEP por MLR e PLS.....	65

Lista de abreviaturas

- ASTM – órgão norte-americano de normalização, que desenvolve e publica normas técnicas para uma ampla gama de materiais, produtos, sistemas e serviços (do inglês “*American Society for Testing and Materials*”)
- BTX – benzeno, tolueno e xilenos
- ECN – número de carbono efetivo (do inglês “Effective Carbon Number”)
- FID – detector de ionização por chama (do inglês “Flame Ionization Detector”) HC(s) – hidrocarboneto(s)
- MLR – Regressão linear múltipla (do inglês “Multiple Linear Regression”)
- PC – componente principal (do inglês “Principal Component”)
- PCA – análise de componentes principais (do inglês “*Principal Component Analysis*”)
- PIE – ponto inicial de ebulição PFE – ponto final de ebulição
- PLS – regressão pelo método dos quadrados mínimos parciais (do inglês “*Partial least-squares*”)
- PRESS – soma dos quadrados dos erros de previsão (do inglês “*Prediction Error Sum of Squares*”)
- PYGAS – gasolina de pirólise (do inglês “*Pyrolysis Gasoline*”)
- RMSEC – erro médio quadrático da calibração (do inglês “*Root Mean Square Error of Calibration*”)
- RMSECV – erro médio quadrático da validação (do inglês “*Root Mean Square Error of Validation*”)
- RMSEP – erro médio quadrático da predição (do inglês “*Root Mean Square Error of Prediction*”)
- RES – resíduo de destilação
- SDV – desvio padrão dos erros de validação (do inglês “*Standard Deviation of Validation*”)
- SPXY – separação de amostras baseado na distância x-y (do inglês “*Sample set Partitioning based on joint x-y distances*”)
- VL – variável latente

Resumo

A gasolina de pirólise pesada, também conhecida por *PYGAS* pesada (do acrônimo em inglês “pyrolysis gasoline”), é uma mistura de hidrocarbonetos majoritariamente maiores que 8 carbonos, obtida a partir do fracionamento dos produtos obtidos a partir do craqueamento da nafta. *PYGAS* pesada tem seu controle no processo produtivo realizado por análises de destilação, segundo o ASTM D86, e composição por cromatografia em fase gasosa com uso de detector de ionização por chama. Uma vez que o volume de amostra para a análise de destilação é muito superior ao de cromatografia (100:1) e existe o risco de princípio de incêndio em destilações se houver vidraria trincada ou montagem inadequada do sistema, foi desenvolvido o trabalho de prever os valores de destilação de *PYGAS* pesada a partir dos resultados de cromatografia em fase gasosa, com aplicação de calibração multivariada. Neste trabalho foram aplicadas Regressão Linear Múltipla (MLR) e Regressão pelo Método dos Quadrados Mínimos Parciais (PLS), comparando os resultados obtidos entre os dois métodos de regressão e com os valores de referência de repetibilidade e reprodutibilidade do ASTM D86 de 2018. Para os pontos de destilação 20%, 30%, 40%, 50%, 60%, 70%, 80% e 90%, tanto no MLR quanto no PLS, os valores de erro médio quadrático da predição (RMSEP) ficaram inferiores 1 °C e r maior que 0,94 e os valores de erro obtidos ficaram dentro do intervalo de reprodutibilidade para cada ponto. Para o Ponto Inicial de Ebulição, o RMSEP foi de 7,5 °C para o MLR e 9,9 °C para o PLS, enquanto os valores de r foram de 0,4 e -0,1, respectivamente. O Ponto Inicial de Ebulição predito obteve resultados insatisfatórios quando comparado aos valores de repetibilidade e reprodutibilidade da referência. Para destilação 5%, os RMSEPs foram de 2,4 °C e 3,1 °C e os valores de r foram de 0,78 e 0,73, respectivamente para MLR e PLS. O ponto de destilação 10% teve valores semelhantes aos pontos 20% e 30% no MLR, porém no PLS os resultados ficaram um pouco diferentes, com RMSEP de 1,5 °C e r de 0,8. No comparativo com o ASTM D86, com resultados foram satisfatórios. Para os pontos de destilação 95% e Ponto Final de Ebulição, os valores de RMSEP foram entre 1,5 °C e 2,5 °C e os valores de r foram superiores a 0,94 nos dois métodos, bem como os erros de cada ponto ficaram dentro do intervalo de reprodutibilidade do ASTM D86. O resíduo de destilação não tem valores de repetibilidade ou reprodutibilidade na referência, porém ficou evidente a impossibilidade do MLR e do PLS de predizerem valores menores 0,5% v/v, o que pode ser explicado pela etapa manual inerente a esta leitura. Comparativamente, os valores de MLR se apresentaram semelhantes ou melhores que PLS, porém nenhum dos dois métodos de calibração multivariada foi capaz de prever os valores de Ponto Inicial de Ebulição da destilação de *PYGAS* pesada.

Palavras-chave: Quimiometria, gasolina de pirólise, destilação, cromatografia em fase gasosa.

Abstract

The heavy pyrolysis gasoline, also known as heavy PYGAS, is a mixture of hydrocarbons mostly composed by more than 8 carbons, obtained through the fractionation of the products obtained in the fractionation of naphtha. The control of heavy PYGAS production process is developed through distillation analysis, according to ASTM D86, and chromatography composition through flame ionization detector usage. Since the sample volume is considerably higher compared to the chromatography volume (100:1), and there is a risk of fire onset in distillation if the glassware is cracked or if the system is improperly assembled, a work to predict the values of heavy PYGAS distillation was developed from the gas chromatography results, with multivariate calibration application. In this work, Multiple Linear Regression (MLR) and Partial Least Square (PLS) were applied to compare the results obtained between both methods and the reference values of ASTM D86 from 2018 repeatability and reproducibility. For the distillation values of 20%, 30%, 40%, 50%, 60%, 70%, 80% and 90%, both for MLR and PLS, the root mean square error of prediction (RMSEP) values were 1 °C inferior, presented a r higher than 0.94 and the error values obtained were in the reproducibility interval for each point. For Initial Boiling Point, the RMSEP value was 7.5 °C to the MLR and 9.9 °C to PLS, whilst the r values were 0.4 and -0.1, respectively. For a 5% distillation, the RMSEPs were 2.4 °C and 3.1 °C and the r value were 0.78 and 0.73, respectively for MLR and PLS. The 10% distillation point reached similar values to the 20% and 30% points in MLR, however, the PLS results were slightly different, with RMSEP of 1.5 °C and r value of 0.8. According to the results comparison with the ASTM D86, the results were satisfactory. For the 95% distillation points and Final Boiling Point, the RMSEP were 1.5 °C and 2.5 °C, the r values were higher than 0.94 for both methods, as well as the errors for each point, which were in the ASTM D86 repeatability range. The distillation residue did not present reproducibility or repeatability values in the reference, yet the impossibility of MLR and PLS predictability for values inferior than 0.5% v/v became evident, which can be explained by the inherent manual step of this reading. In comparison, the MLR values were presented in similarity, or better, with the PLS, but none of the multivariate calibration methods were capable to predict the Initial Boiling Point distillation values for the heavy PYGAS.

Keywords: Chemometrics; pyrolysis gasoline; distillation, gas chromatography.

1 Introdução

A indústria petroquímica utiliza nafta, uma fração do petróleo obtida pelas refinarias, para craqueamento nos fornos de pirólise, obtendo eteno, propeno, 1,3-butadieno, compostos aromáticos, entre outros. O processo de craqueamento gera o *mix* de produtos, que então precisam ser separados ¹⁰.

A mistura de compostos hidrocarbonetos de 5 a 12 carbonos obtida no craqueamento é denominada gasolina de pirólise ^{9, 11}. A gasolina de pirólise é uma mistura de três frações distintas obtidas ao longo do processo de separação: a fração leve, que origina *PYGAS* leve, oriunda da separação dos compostos de 5 carbonos que segue o fracionamento de gases para obtenção de eteno e propeno; a fração média, que origina *PYGAS* média, oriunda da separação dos compostos de 6 a 8 carbonos que segue o fracionamento de líquidos para obtenção de aromáticos e alifáticos; e a fração pesada, que origina *PYGAS* pesada, oriunda da separação dos compostos maiores que 8 carbonos que segue o fracionamento de óleo, separando os hidrocarbonetos recuperáveis da fração pesada que formam o óleo, que poderá ser usado como combustível ^{10, 12}.

Uma das formas de controle do processo produtivo é o acompanhamento analítico das diferentes frações de *PYGAS*. A gasolina de pirólise pesada é analisada periodicamente quanto à sua composição cromatográfica e o seu perfil de destilação. Essas duas análises geram dados para o processo produtivo para definir a melhor condição de operação, maximizando a obtenção de hidrocarbonetos que gerarão compostos aromáticos de 6 a 9 carbonos, sem contaminar a fração com óleo. O controle de qualidade da gasolina de pirólise pesada se dá pelas análises de composição, através da técnica de cromatografia em fase gasosa, e pela análise de destilação, conforme norma ASTM D86 ¹. O controle de qualidade se faz necessário para controle do processo produtivo, garantindo baixos teores de compostos de até 4 carbonos a fim de evitar a perda destes, que podem ser 1,3-butadieno, 1-buteno e outros de interesse na fração de gases, e o corte adequado da fração mais pesada, evitando o arraste de óleo para o sistema (compostos de 12 carbonos e maiores).

A cromatografia em fase gasosa é uma técnica consolidada na indústria petroquímica, sendo uma das análises mais utilizadas, que utiliza pouca quantidade de amostra (1 mL ou menos) e nenhum reagente ¹³. A análise de destilação também é

uma técnica muito utilizada na indústria petroquímica, sem necessidade de reagentes, porém utiliza grande quantidade de amostra (100 mL)¹ e a análise gera vapores dos produtos destilados, que no caso da gasolina de pirólise, são tóxicos aos seres humanos e ao meio ambiente. A análise de destilação também combina compostos orgânicos com calor, em atmosfera ambiente, ou seja, com oxigênio, podendo levar a princípio de incêndio no caso de a substância entrar em contato com as partes aquecidas do equipamento, comum de ocorrer quando há alguma fissura na vidraria. Outro benefício da utilização de modelo de predição para determinação dos valores de destilação a partir de cromatografia está associado ao custo: a hora-homem é um dos maiores custos num laboratório petroquímico e execução da análise de destilação requer o dobro do tempo de um analista em relação à execução da análise de composição. Avaliando os cenários de risco e econômico, o presente trabalho busca utilizar os resultados cromatográficos aliados às ferramentas de tratamento de dados multivariados para obtenção dos valores de destilação.

Diversos trabalhos na literatura já utilizam dados de análises por cromatografia em fase gasosa para determinar aspectos físico-químicos de gasolinas ²⁻⁶, porém todos esses são em gasolina automotiva. Existem outros trabalhos também que trazem aplicação de cromatografia em fase gasosa e calibração multivariada para obter aspectos físico-químicos de produtos derivados de petróleo ^{7, 8}, porém nenhum aplicado à gasolina de pirólise.

1.1 Objetivos

Propor métodos multivariados aplicados aos resultados da análise de cromatografia em fase gasosa, para obtenção dos resultados de destilação da gasolina de pirólise pesada.

1.1.1 Objetivos Específicos

- ✓ Identificar amostras anômalas via método de análise exploratória não supervisionado, a análise de componentes principais – PCA.

- ✓ Aplicar MLR para obtenção da predição dos resultados de destilação e avaliar os resultados obtidos.
- ✓ Aplicar PLS para obtenção da predição dos resultados de destilação e avaliar os resultados obtidos.
- ✓ Comparar os métodos MLR e PLS para as amostras em questão.
- ✓ Avaliar aplicação dos modelos obtidos na indústria.

2 Fundamentação Teórica

A determinação dos valores de destilação de derivados de petróleo a partir de análise por cromatografia em fase gasosa com a aplicação de métodos de regressão já foi apresentada por Fakayode et al ⁷, bem como a obtenção de outros parâmetros físico-químicos a partir da cromatografia foi demonstrada por Rodrigues et al ⁸.

É possível identificar trabalhos publicados que preveem os parâmetros físico-químicos da gasolina automotiva a partir da composição cromatográfica, como publicado por Flumignan et al ³, de Godoy et al ⁴, van Leeuwen et al ² e Mendes et al ⁶, e publicação para prever rapidamente a presença de adulterantes em gasolina automotiva, como publicado por Nespeca et al ⁵. No entanto, a aplicação de métodos quimiométricos para prever parâmetros ou até mesmo classificar gasolina de pirólise não foram encontrados.

A fundamentação teórica deste trabalho primeiro pretende explicar o que é gasolina de pirólise e como é obtida. Posteriormente serão abordadas as análises aplicadas neste trabalho: cromatografia em fase gasosa e destilação. Na sequência, os métodos quimiométricos serão apresentados e um esquema de como foi utilizado neste trabalho. As definições de repetibilidade e reprodutibilidades a partir da norma de referência, o ASTM D86, finalizarão este capítulo.

2.1 Gasolina de Pirólise: definição e processo de obtenção

Gasolina de pirólise é um subproduto do craqueamento à vapor da nafta, composta por hidrocarbonetos na faixa de 5 a 12 carbonos e alto conteúdo alquilaromático, processo este muito comum na indústria petroquímica. Além dos aromáticos, *PYGAS* (acrônimo do inglês, “Pyrolysis gasoline”) tem uma quantidade elevada de alcenos e alcadienos, como ciclopenteno, hexeno-1, ciclopentadieno, isopreno e estireno e, por causa da presença de dienos conjugados, é muito instável e reativa ⁹.

A Figura 1 apresenta a produção de *PYGAS* a partir do craqueamento da nafta. A nafta é quebrada em moléculas menores por um mecanismo via radical livre na ausência de catalisadores, no forno de pirólise. Ao sair do forno, a mistura de gás quente

é resfriada rapidamente nos trocadores de calor da linha de transferência, produzindo gás craqueado e líquido. O líquido é água, óleo e *PYGAS* pesada (baixo teor de benzeno, tolueno e outros C7-, mas muito C8+, incluindo naftaleno), que são condensados e fracionados. Essa fração de *PYGAS* pesada é o objeto de estudo deste trabalho. O óleo é enviado para a caldeira como combustível. A fração gasosa passa por vários estágios de compressão do gás, é resfriada e finalmente enviada ao sistema de lavagem cáustica para remover gases ácidos, dióxido de carbono e água. Um dos produtos desta seção é a gasolina de pirólise média, que contém grande quantidade de benzeno, tolueno e aromáticos de 8 carbonos (grupo denominado BTX, de benzeno, tolueno e xilenos) ¹⁰. Os gases seguem para as próximas etapas, que são essencialmente um processo de separação por destilação, refrigeração e extração.

Os produtos da compressão obtidos são secos, em seguida pré-resfriados e então refrigerados. A desmetanizadora remove o metano e outros gases leves, como hidrogênio e monóxido de carbono, sendo o metano enviado como gás combustível. Para obter eteno, etano e mais pesados, a corrente é separada por destilação e o acetileno é convertido à eteno por hidrogenação catalítica. Para a obtenção de propeno, propano e mais pesados, a corrente é separada por destilação e o metilacetileno e propadieno são convertidos à propeno por hidrogenação catalítica. Os compostos C4 (4 carbonos) são separados do *PYGAS* leve (C5s) e produzem especialidades como 1,3-butadieno, 1-buteno e MTBE (ou ETBE, dependendo se a petroquímica usa metanol ou etanol para produzir éter, respectivamente); MTBE é metil-tert-butil éter e ETBE é etil-tert-butil éter. A soma de gasolina de pirólise pesada, média e leve será o *PYGAS* usado para extrair os compostos aromáticos, separando-os dos não aromáticos.

Como *PYGAS* é muito instável, deve ser hidrogenada antes do processamento. A hidrogenação catalítica heterogênea de dois estágios é usada para estabilizar a gasolina de pirólise. O primeiro estágio de hidrogenação é mais brando, hidrogenando espécies muito reativas (como estireno e dienos conjugados). O segundo estágio é mais severo, onde as olefinas restantes são hidrogenadas e os compostos sulfurados são convertidos em gases leves (principalmente H₂S) ^{11, 12}. Entre o primeiro e o segundo estágios de hidrogenação, a despentanizadora (removedora de C5) e a desoctanizadora (removedora da fração C9+) são usadas para remover as frações sem aromáticos ou aromáticos pesados. O produto final do segundo estágio de hidrogenação é uma corrente rica em compostos aromáticos de 6 a 8 carbonos, e hidrocarbonetos não aromáticos saturados,

como alcanos e cicloalcanos. Este produto pode seguir processos de separação de compostos aromáticos com uso de solventes extratores e posterior destilação para separação dos produtos de interesse ¹³.

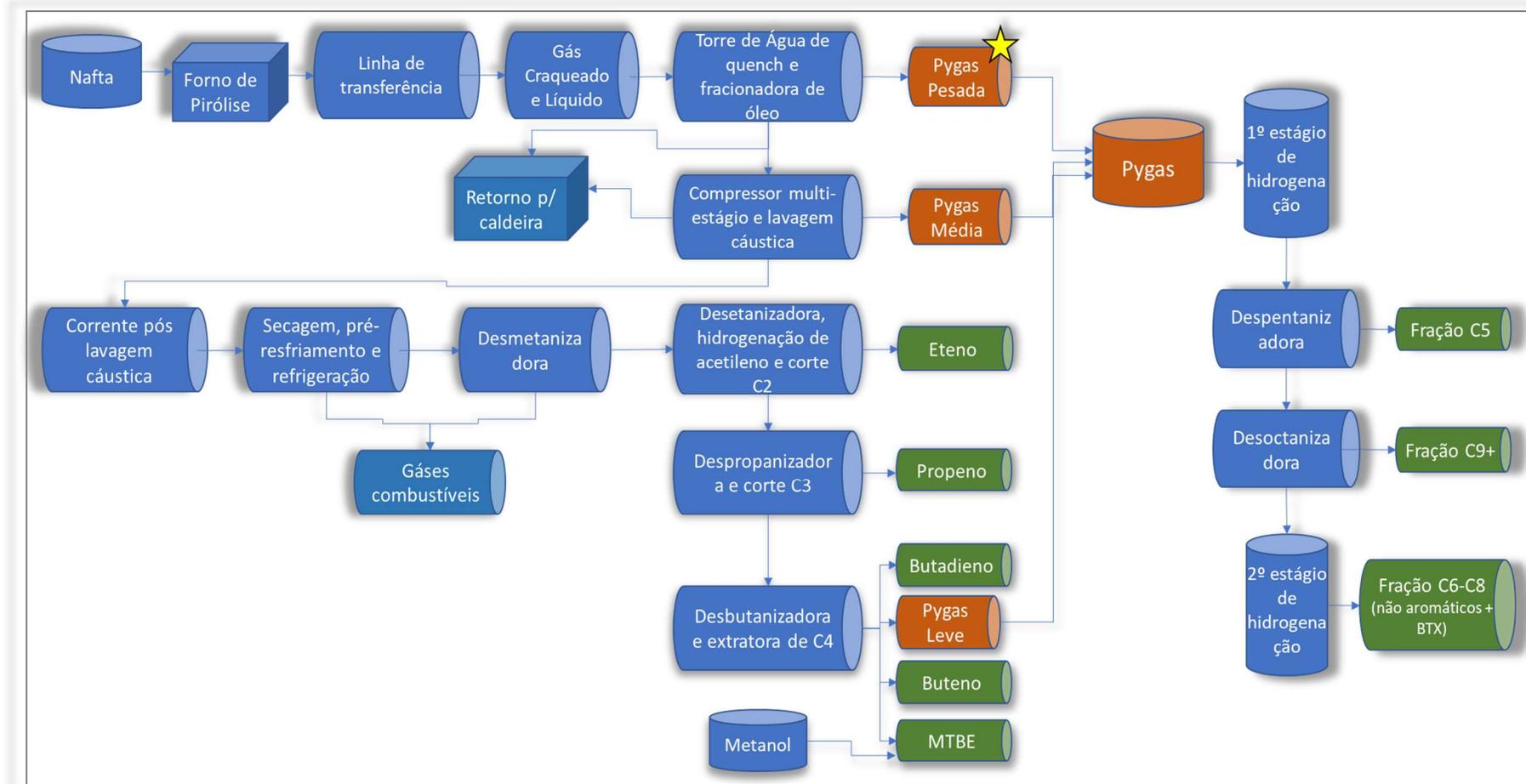


Figura 1: Fluxo de produção das várias frações que compõem a gasolina de pirólise, desde o craqueamento à vapor de nafta até seu processamento. Fonte própria.

2.2 Cromatografia em fase gasosa

A cromatografia em fase gasosa surgiu em 1952 com Martin e James ¹⁴ e rapidamente se tornou uma técnica muito aplicada na indústria de petróleo ¹⁵. Por definição, cromatografia em fase gasosa é uma técnica de separação de substâncias voláteis percolando um gás através de uma fase estacionária. A fase estacionária, que é o componente da coluna cromatográfica, pode ser um sólido, então denominada de cromatografia gás- sólida, ou um líquido filmico de alto peso molecular e baixa volatilidade, então denominada cromatografia gás-líquida. É uma técnica tanto qualitativa quanto quantitativa. Qualitativamente, os componentes são reconhecidos pelo seu tempo de retenção, após separados adequadamente. E, quantitativamente, a concentração de cada componente é proporcional a sua área para o mesmo tempo de retenção ¹⁶.

Um sistema cromatográfico é, resumidamente, composto de injetor, coluna e detector ^{16, 18}, necessitando de gás(es) para arraste e funcionamento do detector, conforme representação na Figura 2.

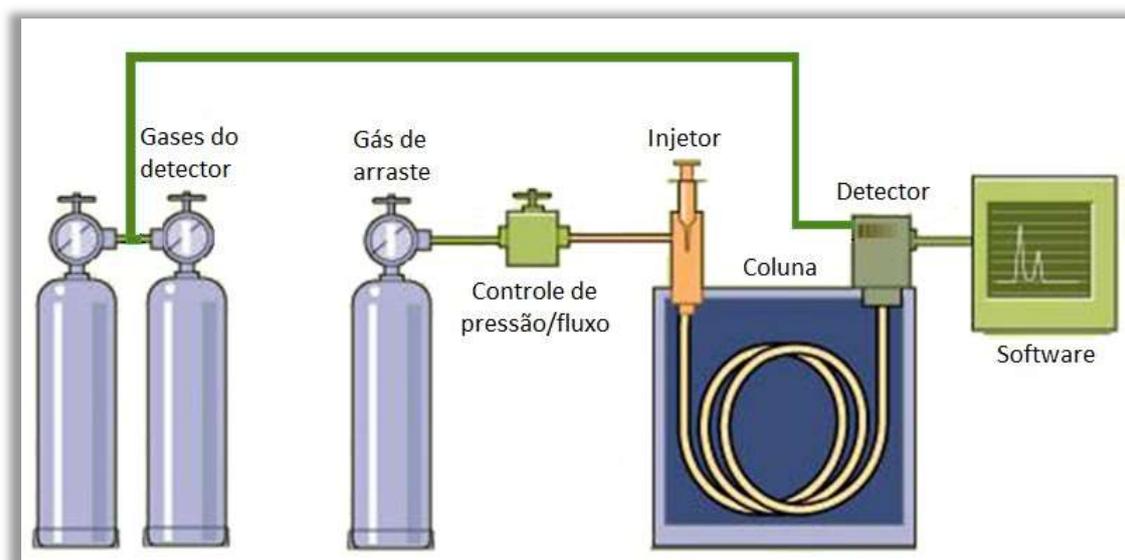


Figura 2: Sistema cromatográfico simplificado. Fonte: modificado do site da DCtech ¹⁷.

2.2.1 Injetores

A introdução da amostra na coluna cromatográfica é realizada através do injetor. Este precisa ser compatível com a coluna, com dimensões adequadas ao tipo de amostra e à temperatura de trabalho, garantindo a vaporização total da amostra.

Os injetores mais utilizados atualmente são:

- *Split/Splitless*¹⁸ – do inglês divisão/sem divisão. Utilizado para dividir a amostra em partes desiguais, necessitando do uso de *liner*, que tem a função de homogeneizar a amostra. Quando no modo *split*, encontra-se em condições de método informado como 1:80, por exemplo, e isso significa que de cada 80 partes de amostra introduzida no injetor, apenas 1 foi para a coluna, e as outras 79 para *vent* (descarte). Muito utilizado para colunas capilares. Quando está no modo *splitless*, é porque a função de divisão de amostra foi desabilitada, sendo introduzida toda amostra, aumentando o sinal gerado para os componentes de interesse por introduzir mais amostra.
- *On-column* – a amostra é injetada à frio diretamente na coluna, que precisa estar em temperatura próxima à de vaporização do solvente¹⁸. Usado para situação em que precisa aumentar a quantidade de amostra injetada.
- PTV – sistema de temperatura programada de vaporização, no qual separa o vapor do solvente dos analitos no *liner*, reduzindo a discriminação de alguns analitos como pode ocorrer em injeção *split/splitless*¹⁹. É semelhante a um injetor *split/splitless* convencional, no entanto, é equipado com um sistema sofisticado de controle de temperatura, que pode ser rapidamente aquecido ou resfriado durante a injeção.

2.2.2 Colunas

A coluna cromatográfica é por vezes denominada como o “coração” do sistema cromatográfico¹⁸, essa expressão se deve ao fato de nela ocorrer a separação dos compostos presentes na amostra. Além da diferenciação entre gás-sólido e gás-líquido, as colunas também são diferenciadas por serem empacotadas ou capilares.

As colunas empacotadas normalmente possuem diâmetro interno variando de 1/8” a

1/4” e de até 10 m de comprimento, em aço inoxidável ¹⁸, e a fase estacionária pode ser um sólido (exemplo: peneira molecular) ou um líquido distribuído sobre um suporte sólido (exemplo: carbowax 20M).

As colunas capilares possuem diâmetro interno variando de 0,05 mm à 0,53 mm, e até 150 m de comprimento. Normalmente o suporte é de sílica fundida. A fase estacionária pode ser sólida (exemplo: Al₂O₃/KCl, bastante utilizada para separação de hidrocarbonetos leves) ou líquida (exemplo: polimetilsiloxano, bastante utilizada para separação de hidrocarbonetos a partir de 5 carbonos).

2.2.3 Detectores

Existem hoje diversos detectores para cromatografia em fase gasosa. Os mais conhecidos são:

- TCD – do inglês, detector de condutividade térmica. Considerado um detector universal, pois gera sinal para qualquer composto que modifique a resistência do filamento de tungstênio, que é um constituinte da ponte de Wheatstone, ou seja, qualquer composto que não o gás de arraste utilizado ¹⁸. Bastante utilizado para determinações em nível percentual.
- ECD – do inglês, detector de captura de elétrons. É um detector de ionização que utiliza uma fonte radioativa para gerar elétrons livres (³H ou ⁶³Ni), do tipo seletivo, pois só gera sinal para compostos eletronegativos, como haletos orgânicos, nitratos, etc ¹⁸. Apesar de conhecido, não é muito utilizado devido às suas características radioativas e controles necessários devido a essa característica.
- FID – do inglês, detector de ionização por chama. É um detector usado tanto para detecção de traços quanto de percentual, sensível a compostos orgânicos, porém não detecta água e compostos inorgânicos. A sua seletividade se deve ao seu mecanismo de geração de sinal: O FID é composto por uma chama (mistura de H₂, ar e N₂) e ao passar por ela, os compostos queimam, sendo ionizados, e geram uma diferença de potencial entre os filamentos ¹⁶, ¹⁸. O FID foi introduzido em 1958 por McWilliam e Dewar ²⁰ e é o detector mais utilizado na

indústria de petróleo e petroquímica no controle de qualidade.

- SCD - do inglês, detector de queluminescência para enxofre. É detector específico para compostos sulfurados. Os compostos sulfurados reagem em uma câmara à alta temperatura e na presença de H_2 , formando SO como produto de interesse. O SO então reage com excesso de O_3 à baixa pressão, formando SO_2 , O_2 e $h\nu$, na região do UV-azul, sendo esta última então detectada, proporcional equimolarmente à presença de enxofre. Existe também o NCD, com o mesmo princípio, aplicável para detecção de compostos nitrogenados ²¹. Para análise de compostos sulfurados tem-se ainda outros detectores, como FPD ¹⁵, do inglês, detector fotométrico de chama e PFPD, do inglês, detector fotométrico de chama pulsado.
- PDHID - do inglês, detector de ionização de hélio por descarga pulsada. Um detector do tipo universal, com boa linearidade mesmo em baixas concentrações. Funciona ionizando o hélio pela descarga pulsada de alta tensão entre dois eletrodos de platina. Essa descarga proporciona a excitação do hélio e a obtenção do hélio metaestável capaz de ionizar todos os compostos, com exceção do neônio ²². Existe, também o PID, do inglês detector de ionização pulsada e o HID, do inglês detector de ionização de hélio, também do tipo universais ^{15,22}.
- MS – do inglês, espectrômetro de massas. Usado não só para quantificar, mas também para qualificar moléculas. De maneira resumida, é composto de uma fonte ionizadora para transformar as moléculas em íons, que então seguem para um analisador de massas sendo separadas de acordo a razão m/z e finalmente para o detector, gerando sinal proporcional à concentração ¹⁵. Para cada parte do sistema existem vários tipos, que podem ser visualizados na Figura 3, que resume as possibilidades de um sistema GC-MS.

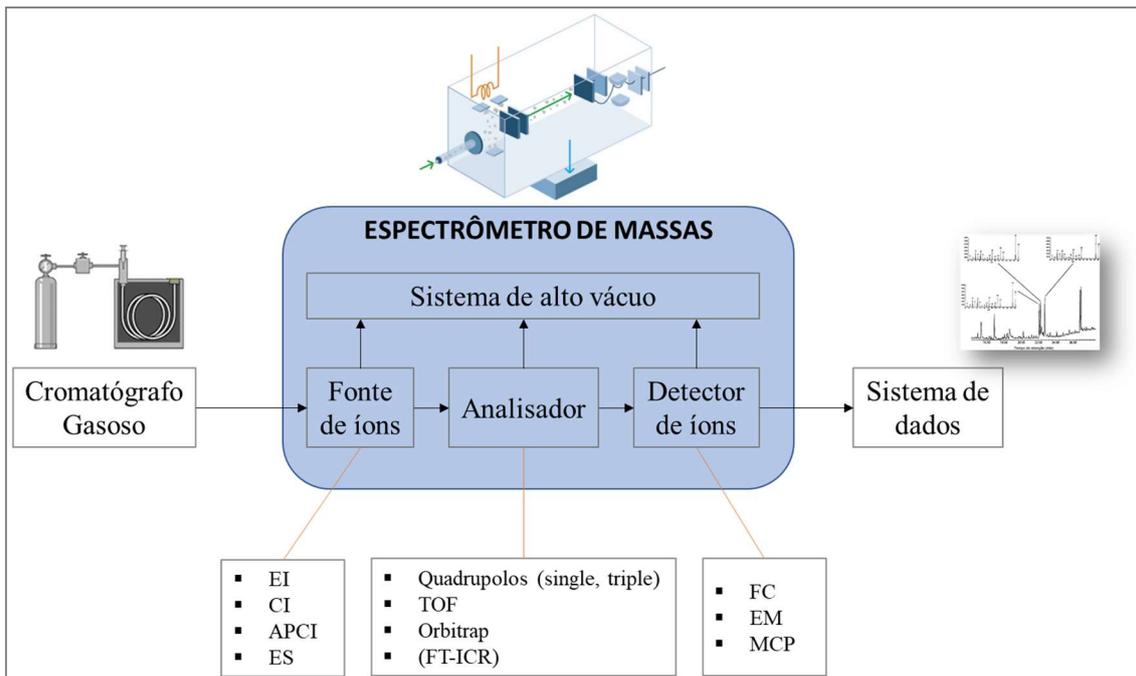


Figura 3: Esquema de um GC-MS e algumas de suas possibilidades de configuração. Fonte: modificado do site da Revista Analytica ^{23, 24}

2.2.4 Calibração por Número de Carbono Efetivo

O método mais comum de calibração é o ESTD (padronização externa), que usa um padrão de concentração conhecida do composto de interesse para obter um fator de resposta para o componente em questão, sendo este então aplicado a área obtida em cada análise realizada.

Como a cromatografia em fase gasosa possui detectores capazes de identificar a totalidade ou quase totalidade dos compostos presentes (uso do TCD e FID) é por vezes aplicada a normalização dos resultados obtidos a 100%, ou um valor que considerada a concentração de um composto de concentração conhecida (por exemplo, água analisada por outra técnica) de modo a obter o resultado total da composição da amostra, como exemplo, uma mistura que contenha 5% de água determinada e o resultado da cromatografia então normalizado a 95%.

Na indústria ainda se utiliza outro recurso: fator ECN (do inglês, número de carbono efetivo) acompanhado de normalização. É um método aplicável em muitas análises ¹⁴⁻²⁸, muito comum na indústria de petróleo e petroquímica, na qual existem muitas correntes com grande número de compostos em cada corrida cromatográfica.

ECN é um método capaz de propor um fator de resposta relativo para qualquer composto em um FID, conhecendo sua estrutura ^{22, 29}.

2.2.5 Destilação Simulada

Após o desenvolvimento das colunas de polidimetilsiloxano, totalmente não polares, os químicos da área do petróleo evidenciaram a eluição dos compostos por ponto de ebulição, desenvolvendo, a partir de 1960, a destilação simulada por cromatografia em fase gasosa – SimDist ¹⁵. A partir de padrões e amostras conhecidas, é possível obter o % recuperado a qualquer temperatura.

Algumas empresas vendem aplicações prontas para obter os resultados de destilação simulando o ASTM D86 a partir de uma análise cromatográfica. Como exemplo tem-se a aplicação AC8612 ³⁰, da Pac, e o Scion 436 SimDist Analyzer ³¹, da Scion. Se existem aplicações prontas, então por que utilizar quimiometria? Pelo custo de adquirir uma aplicação como esta, que atualmente é superior a R\$ 600.000. Além disto, não pode substituir os resultados em certificados de produtos, que possuem acordos e normas aplicadas, não podendo ser substituídos unilateralmente.

2.3 Destilação

As indústrias de petróleo e petroquímica utilizam amplamente a análise de destilação para obter dados referentes a seus processos produtivos e, também, dados sobre as características dos produtos obtidos. É uma técnica utilizada tanto para caracterizar produtos puros, como tolueno seguindo o ASTM D850 ³², por exemplo, quanto misturas complexas seguindo o ASTM D86 ¹.

A destilação consiste numa curva, que é a representação da temperatura de ebulição da mistura líquida versus o volume de destilado acumulado, a uma determinada pressão. De acordo com a ASTM, o método D86 padroniza a destilação de derivados de petróleo do tipo leves e médios, à pressão atmosférica sem o uso de refluxo intencional, em uma única etapa teórica, usando um equipamento de bancada para determinar a faixa de destilação.

A norma ASTM D86 define padrões de condições para a destilação, mas não gera resultados para definir o ponto de ebulição verdadeiro de substâncias (conhecido como TBP, do inglês “*true boiling point*”) ^{1,33}.

Os componentes básicos de um destilador são: balão de destilação, condensador associado a um banho de resfriamento, placa de suporte para o balão, fonte de aquecimento, suporte para o frasco coletor (proveta graduada), medidor de temperatura (termômetro ou termopar) e proveta graduada para recolher destilado (também denominado cilindro coletor). A Figura 4 é a foto de um destilador automático com a indicação das principais partes, dotado de sensores e controladores.



Figura 4: Destilador automático e indicações dos principais componentes. Fonte própria.

A norma também define condições de uso quanto à classificação da amostra, denominado grupo. A norma diferencia em grupo 1, 2, 3 e 4, que leva em consideração a pressão de vapor *Reid* da amostra, a temperatura do ponto inicial de destilação e a temperatura do ponto final de destilação para definir o grupo ao qual a amostra pertence. Por exemplo, para o grupo 2, grupo da *PYGAS* pesada, a pressão de vapor é < 65,5 kPa, sem critério para o ponto inicial de destilação e $\leq 250^{\circ}\text{C}$ para o ponto

final de destilação ¹. Uma vez definido o grupo, os arranjos da aparelhagem, a temperatura do condensador e outras variáveis operacionais são definidas pelo grupo ao qual a amostra pertence.

Uma vez classificada como grupo 2, as seguintes condições precisam ser atendidas para a análise de destilação:

- Balão de destilação de 125 mL.
- A utilização de um termômetro tipo 7C (ou 7F).
- Diâmetro do buraco da placa de suporte: 38 mm.
- Temperatura do balão no início do teste: entre 13 °C e 18 °C.
- Temperatura da proveta coletora no início do teste e nos 100 mL medidos: entre 13 °C e 18 °C.
- Temperatura da amostra estocada: < 10 °C.
- Temperatura do banho de resfriamento: entre 0 °C e 5 °C.
- Tempo da primeira aplicação de calor para o PIE: de 5 min à 10 min.
- Tempo do PIE para os 5% recuperados: de 60 s à 100 s.
- Taxa de condensação dos 5% recuperados até os 5 mL restantes no balão: 4 a 5 mL/min.
- Tempo de recuperação dos 5 mL restantes até o ponto final: máx. 5 min.

O procedimento experimental básico consiste em transferir 100 mL da amostra refrigerada para o balão de destilação de 125 mL e submetê-los a uma rampa de aquecimento suficiente para manter a taxa de destilação em 4 a 5 mL/min. O vapor destilado é condensado e recolhido na proveta refrigerada. O volume não evaporado no balão é o resíduo de destilação, que é medido vertendo o resíduo para uma proveta graduada. Cada temperatura é corrigida para a pressão de 760 mmHg, de modo a ter-se resultados comparáveis em qualquer condição. Para tal, aplica-se a equação de Sydney Young (Equação 1) ¹:

$$C_c = 0,00012 (760 - P) (273 + t_c) \quad \text{Equação 1}$$

Onde,

C_c = Correção algébrica da temperatura lida

P = pressão barométrica lida, em mmHg

t_c = temperatura lida, em °C

Além da correção da pressão atmosférica, também é necessária a correção do volume da perda, que pode ser feita aritmeticamente ou graficamente. Os equipamentos automáticos, como o aplicado neste trabalho, aplicam as correções de pressão e de perda por evaporação automaticamente.

Se o encaixe de cada parte do sistema de análise de destilação apresentar folga ou se houver trinca na vidraria utilizada, pode levar à evaporação ou derramamento de solvente para superfície aquecida do destilador, provocando princípio de incêndio.

2.4 Métodos Quimiométricos

A quimiometria consiste na aplicação de ferramentas matemática e estatística em dados com informação química, dados estes que representam um grande volume de informação e que, nem sempre, podem ser facilmente tratados pelas estratégias de análise univariada convencionais, ou, em último caso, fornecem informações incompletas. A quimiometria tornou-se uma ferramenta indispensável para mineração e análise de dados químicos, pois cada vez os instrumentos analíticos geram mais dados em menor tempo e há grande capacidade de processamento dos computadores atuais ^{34, 35}.

É possível dividir a aplicação da quimiometria em 4 diferentes necessidades:

- Planejamento e otimização de experimentos;
- Análise exploratória de dados;
- Classificação em diferentes grupos ou reconhecimento de padrões;
- Criação de modelos de calibração multivariada. Hoje também é utilizada na indústria a terminologia de “construção de preditores”.

Planejamento e otimização de experimentos, bem como classificação em diferentes grupos ou reconhecimentos de padrões, não são objeto desse estudo.

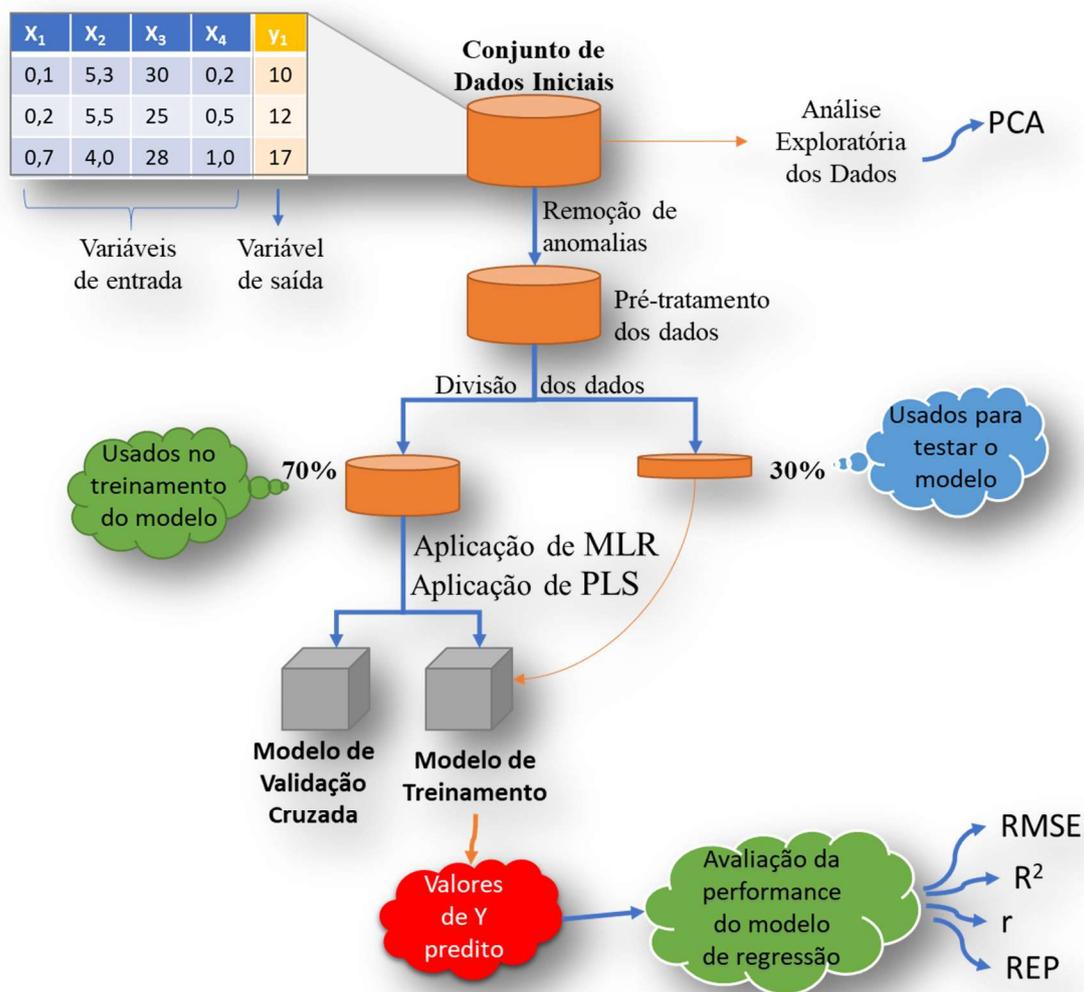


Figura 5: Esquema simplificado de aplicação de Quimiometria no desenvolvimento de preditor. Fonte própria.

A Figura 5 apresenta, de maneira esquemática, o caminho a ser seguido na construção e validação de modelos de calibração multivariada. Cada etapa presente na Figura 5 será detalhada nos subitens a seguir.

2.4.1 Análise Exploratória de Dados

Análise exploratória de dados é um tipo de método não supervisionado, ou seja, não se conhece a que classes pertencem os objetos ou essa informação não é usada na construção do modelo. Tipos de métodos como análise de componentes principais, mais conhecida por PCA, e agrupamento hierárquico, mais conhecida por HCA, são os mais

conhecidos e difundidos na literatura. No caso específico da PCA é importante discutir em mais detalhes, pois esta técnica contempla as bases de calibração do PLS, método aplicado nesta dissertação, que significa regressão pelo método dos quadrados mínimos parciais.

A etapa inicial na modelagem de dados consiste na disposição dos mesmos em estrutura matricial, de modo que cada coluna é uma variável e as amostras ficam dispostas nas linhas da matriz aqui definida como \mathbf{X} .

Estatisticamente, a PCA é uma técnica de análise multivariada relacionada com análise de autovetores/autovalores e a decomposição em valores singulares (acrônimo do inglês, SDV). Em termos matriciais, PCA é um método para decompor \mathbf{X} em duas matrizes menores \mathbf{T} e \mathbf{P} mais uma matriz de resíduos \mathbf{E} ³⁶:

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

Na equação 2, a matriz de escores \mathbf{T} corresponde às coordenadas das amostras em relação aos eixos das componentes principais (PCs) e possui dimensões $I \times K$. Já a matriz de *loadings* \mathbf{P} , de dimensão $K \times J$ representa a relação entre o eixo da PC e o eixo da variável original, sendo calculado pelo cosseno do ângulo entre esses eixos. Por fim, a variabilidade dos dados não explicada pelo modelo é expressa por uma matriz de resíduos \mathbf{E} , que apresenta a mesma dimensão da matriz \mathbf{X} , $I \times J$.

A PCA transforma grandes matrizes de dados em matrizes menores, as quais podem ser mais facilmente examinadas, plotadas e interpretadas. A PCA extrai os fatores mais importantes dos dados, as primeiras PCs (componentes principais), preservando a maior parte da variância. PCs são também denominadas autovetores ou fatores e cada uma delas é não correlacionada com a outra e são ortogonais entre si. Exemplificando, a PC1, que é a primeira delas, é definida pela direção dos dados originais que descreve a máxima variância. Cada um desses fatores descreve as interações multivariadas entre as variáveis medidas e revelam tendências subjacentes dos dados. Essa análise também nos permite identificar amostras que estejam com comportamento atípico, evidenciando-as graficamente ³⁷.

2.4.2 Pré-Tratamento dos dados

O pré-tratamento dos dados consiste em dar, matematicamente, condições de um melhor ajuste do modelo sobre os dados, ações como dar a mesma importância para todas as variáveis, correção de linha base, remoção de regiões não informativas, são exemplos da importância desta etapa.

Para as amostras, que são as linhas da matriz, pode-se utilizar correções dos tempos de retenção de um cromatograma ou de um espectro de RMN (ressonância magnética nuclear), com o método COW (*Correlation Optimized Warping*)^{37, 38}, ou o método DTW (*Dynamic Time Warping*)³⁸ ou o método PTW (*Parametric Time Warping*)³⁸; pode-se aplicar alisamento, reduzindo o ruído a partir da redução da componente aleatória dos dados frequentemente encontrado em cromatografia, podendo ser o alisamento pelo método da média, pelo método de Savitzky-Golay ou por filtro de Fourier³⁷, por exemplo. A depender da técnica a ser utilizada, busca-se a melhor transformação para o caso. Neste trabalho, como estão sendo utilizados os dados discretos disponíveis para o usuário final (equipes de produção e engenharia), essas técnicas não foram utilizadas por não ser utilizado sinal de equipamento.

Para as variáveis, que são cada uma das colunas da matriz de dados, os procedimentos utilizados podem ser: centragem dos dados na média, escalamento pela variância, escalamento pela amplitude e autoescalamento. O autoescalamento, método aplicado aos dados deste trabalho, consiste em subtrair de cada elemento de uma coluna da matriz o valor médio daquela coluna e dividir o resultado pelo desvio-padrão dessa coluna³⁷, conforme equação 3, tornando os dados adimensionais.

$$x_{ij \text{ autoescalado}} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad \text{Equação 3}$$

2.4.3 Separação das amostras

Na literatura sugere-se a divisão do conjunto de amostras em 70% para o conjunto de treinamento e 30% para o conjunto de predição, ou como 80% - 20%. O conjunto de predição é aquele em que as amostras não farão parte do conjunto de treinamento e/ou calibração, testando se o modelo obtido passa nos critérios de aceitação avaliando os resultados obtidos nas figuras de mérito. Além disso, o modelo

precisa ser validado, com um conjunto separado de amostras, que pode ser 20% das amostras de treinamento, ou do tipo validação cruzada, utilizada neste trabalho. A validação cruzada consiste em retirar uma amostra a cada modelo de treinamento, passando por todas e verificando os resultados obtidos em cada situação.

E como fazer de maneira adequada a separação das amostras? Utilizam-se algoritmos para que a separação seja adequada e obedeça a determinados critérios. Uma das formas seria a separação do conjunto de amostras de maneira randômica, obtendo a cada teste conjuntos diferentes, requerendo que todos os conjuntos sejam testados e verificados se os resultados seriam semelhantes.

Outra forma de separação do conjunto de amostras é a utilização do algoritmo SPXY ³⁹ (separação de amostras baseado na distância x-y), que é uma adaptação do Kernard-Stone, que considera a informação conjunta das variáveis independente (**X**) e dependente (**y**). Após a normalização de cada termo, a distância x-y atribui importâncias equivalentes na distribuição das amostras em ambos os espaços considerados. De maneira prática, também busca conter todos os valores extremos no conjunto de treinamento. Em métodos de regressão, o algoritmo SPXY é bastante utilizado ⁴⁰.

$$d_{xy}(p, q) = \frac{d_x(p, q)}{\max_{p, q \in [1, I]} d_x(p, q)} + \frac{\sqrt{(y_p - y_q)^2}}{\max_{p, q \in [1, I]} d_y(p, q)}; p, q \in [1, I] \quad \text{Equação 4}$$

Onde:

2.4.4 Calibração Multivariada

A calibração multivariada consiste em construir modelos empíricos de propriedades mensuráveis, a partir dos dados experimentais disponíveis, utilizando métodos matemáticos de regressão.

Os métodos mais populares são CLS (regressão pelo método dos quadrados mínimos clássico, do inglês “*Classical Least-Squares*”), MLR (Regressão linear múltipla, do inglês “*Multiple Linear Regression*”), PCR (regressão pelo método dos componentes principais, do inglês “*Principal Component Regression*”) e PLS ⁴¹. Serão detalhados os métodos MLR e PLS neste trabalho, que foram aplicados aos dados.

2.4.5 Regressão Linear Múltipla

O MLR pode ser caracterizado como uma técnica para resolver uma série de equações simultâneas. Em um sistema multicomponente que é determinado simultaneamente, a análise pode ser descrita medindo m variáveis x_j para a variável y , com o objetivo principal de criar uma relação linear entre elas. Isto pode ser representado matematicamente como:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m + e \quad \text{Equação 5}$$

Esta regressão permite estabelecer uma ligação entre um número reduzido de variáveis e uma propriedade das amostras. A previsão y_j da propriedade de interesse pode então ser descrita com a fórmula:

$$y_j = b_0 + \sum_{i=1}^k b_ix_i + e_{ij} \quad \text{Equação 6}$$

onde b_i é o coeficiente calculado, x_i o valor de cada variável considerada e e_{ij} é o erro.

Importante destacar que o uso do MLR requer que o número de amostras seja maior que o número de variáveis e que estas não sejam altamente correlacionáveis (não maior que 0,7 o coeficiente de correlação) ⁴¹.

2.4.6 Regressão pelo Método dos Quadrados Mínimos Parciais

O PLS tem como base o PCA e não possui as mesmas restrições de utilização que o MLR, sendo fortemente disseminado em quimiometria por sua vantagem de primeira ordem, ou seja, a capacidade de modelar e predizer amostras mesmo na presença de interferentes ³⁷.

Por esse método linear, a regressão considera a estrutura latente em ambos conjuntos de dados: nas variáveis (X) e propriedades (Y). Para predizer a propriedade de interesse (y), o alvo da redução de dimensionalidade dos dados em X tem como critério a preservação da explicação da variância em X com a previsão da propriedade de interesse, y.

Genericamente, PLS pode ser comparado à análise de componentes principais, porém uma versão supervisionada, cujos escores em U e T possuem uma relação

interna, em outras palavras, máxima covariância um com o outro ^{35, 42}.

$$X = T_A L_A^T + E = X_A + E \quad \text{Equação 7}$$

$$y = T_A q + e \quad \text{Equação 8}$$

A solução não se aplica a qualquer conjunto de pesos L , mas somente àqueles que satisfazem à restrição de maximizar covariância entre U e T ³⁷.

O processo é caracterizado pela possibilidade de extração de informação a partir de mais de um componente. Como na PCA, do resíduo não modelado pela primeira componente PLS, uma segunda componente pode ser extraída. Cada repetição desse processo resultará em um componente de regressão, que será utilizado na etapa de predição das amostras desconhecidas.

A decomposição da matriz X em escores e pesos pode ser feita por meio de diversos algoritmos, como NIPALS, utilizado nos estudos de regressão e reconhecimento de padrões, e bastante recomendado, uma vez que esse calcula as variáveis latentes (autovetores) um a um ^{37, 44}. O esquema a seguir, retirado de Márcia M. C. Ferreira (2015) ³⁷, descreve o algoritmo do método NIPALS.

1) Pré-processamento de X e y (autoescalamento, por exemplo)

2) Cálculo da primeira variável latente de cada bloco:

$$X = y w_1^T + E_x$$

$$\hat{w}_1 = X^T y (y^T y)^{-1} \quad \text{(fator - peso)}$$

$$\hat{w}_1 = \frac{\hat{w}_1}{\sqrt{\hat{w}_1^T \hat{w}_1}} \quad \text{(normalização)}$$

$$t_1 = X \hat{w}_1 \quad \text{(escores)}$$

$$X^T = l_1 t_1^T \Rightarrow \hat{l}_1 = X^T t_1 (t_1^T t_1)^{-1} \quad \text{(pesos)}$$

Y { Regressão do vetor y no vetor t_1

$$y = t_1 q_1 + f \Rightarrow \hat{q}_1 = (t_1^T t_1)^{-1}$$

3) Atualiza X e y {

$$E = X - t_1 l_1^T$$

$$f = y - t_1 q_1$$

$$X = E$$

$$y = f$$

4) Repete os cálculos para as próximas variáveis latentes

5) Calcula o vetor de regressão: $\hat{b} = W (L^T W)^{-1} q$

Para a determinação das VLs (variáveis latentes), foi utilizado o método desenvolvido por Haaland e Thomas ⁴⁴, que compara o PRESS obtido para cada VL testada, de modo a definir como o número ideal de VLs aquele que possuir o menor valor de PRESS, que é calculado como a soma dos quadrados dos erros de previsão, conforme equação ⁴¹:

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{Equação 9}$$

Onde n é o número de amostras; y_i e \hat{y}_i correspondem aos valores de referência e aos obtidos pelo modelo, respectivamente.

2.4.7 Avaliação dos Modelos

Crítérios para avaliação dos modelos obtidos foram desenvolvidos como os semelhantes aos processos de validação de metodologias analíticas, segundo o INMETRO ⁴⁵. São denominadas métricas de desempenho ou figuras de mérito e a seguir serão apresentados seus conceitos e equações gerais ⁴⁶ sendo também apresentado o conceito de validação cruzada, importante metodologia para validação do modelo de calibração.

2.4.8 Validação Cruzada

A validação do modelo de calibração pode ser realizada por um conjunto de amostras externo, que não foi utilizado no conjunto para a calibração, ou existe outra possibilidade que é a utilização da validação cruzada.

A validação cruzada é utilizada para a escolha do número de variáveis latentes, baseada na avaliação da magnitude dos erros de previsão. Nesta metodologia, um dos métodos mais usados é o ciclo interativo *leave-one-out*, que significa deixar uma amostra de fora do conjunto de calibração para desenvolvimento do modelo e então testar essa amostra no modelo, comparando o resultado obtido pela predição com o resultado real. O ciclo é realizado com todas as amostras do conjunto ⁴¹. A validação cruzada não requer um conjunto separado de amostras, o que contribui para conjuntos pequenos, em que é difícil a obtenção de mais amostras.

Um ponto de atenção para a utilização deste tipo de validação é com relação a presença de *outliers*, pois quando um *outlier* for a amostra retirada, não será possível

predizê-la ⁴¹.

Outro ponto de atenção da utilização desta metodologia é que em algum momento a amostra do extremo será a de validação, não tendo sido usado na calibração, sendo necessária a extrapolação do modelo para obter o resultado. Este fato explica por que os valores de validação são piores que o da calibração.

2.4.9 Métricas de Desempenho

A exatidão demonstra a concordância entre o valor verdadeiro e o valor medido (ou predito). Em quimiometria é representada por RMSEC a raiz quadrada do erro médio quadrático da calibração (e RMSECV para a validação cruzada e RMSEP para a predição). Incorpora tanto erros aleatórios quanto sistemáticos. O valor obtido está na mesma unidade de medida dos dados amostrais e quanto menor o valor, melhor o resultado. A equação para RMSEC é ⁴⁶:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n_v}} \quad \text{Equação 10}$$

Onde n_v é o número de amostras de validação; y_i e \hat{y}_i correspondem aos valores de referência e aos obtidos pelo modelo, respectivamente; n é o número de amostras utilizadas no cálculo da precisão.

BIAS

Bias, ou viés, é o valor calculado pela diferença entre a média da população e o valor verdadeiro, sendo este um erro sistemático. É dado pela equação ⁴⁶:

$$bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n_v} \quad \text{Equação 11}$$

Para avaliar se o *bias* é significativo, faz-se necessário o cálculo do SDV e o valor de t , do teste de significância de t student. Se o valor de t for maior que o t crítico calculado com base no número de amostras da validação (n_v-1 graus de liberdade), há indício de que há erro sistemático presente no modelo multivariado, e que este erro pode ser significativo. *Bias* é um vocabulário muito utilizado, mas em português a

tradução seria viés.

SDV

O desvio padrão dos erros de validação é dado por ⁴⁶:

$$SDV = \sqrt{\frac{\sum[(y_i - \hat{y}_i) - bias]^2}{n_v - 1}} \quad \text{Equação 12}$$

E o valor de t é dado por ⁴⁶:

$$t_{bias} = \frac{|bias|\sqrt{n_v}}{SDV} \quad \text{Equação 13}$$

Coefficiente de correlação de Pearson - r

O coeficiente de correlação de Pearson (r), também denominado coeficiente de correlação linear, é um grau de relação entre duas variáveis quantitativas e exprime o grau de correlação através de valores situados entre -1 e 1.

Quando o coeficiente de correlação se aproxima de 1, tem-se um aumento no valor de uma variável quando a outra também aumenta, ou seja, há uma relação linear positiva. Valor próximo a -1 significa que o aumento de uma variável acontece enquanto a outra diminui, ou seja, uma correlação linear negativa ou inversa. Valores próximos a 0 significam ausência de correlação. Neste trabalho é utilizado para comparar os valores de referência e os obtidos pelo modelo, conforme equação:

$$r = \frac{\sum_1^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_1^n (y_i - \bar{y})^2} \times \sqrt{\sum_1^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad \text{Equação 14}$$

Onde n_v é o número de amostras de validação; y_i e \hat{y}_i correspondem aos valores de referência e aos obtidos pelo modelo, respectivamente; n é o número de amostras utilizadas no cálculo da precisão.

R²

O coeficiente de determinação R^2 permite estimar a porcentagem da variação dos dados que é modelada adequadamente. Quanto mais próximo de 100, melhor o resultado. É dado por:

$$R^2 = \frac{\sum_1^n (\hat{y}_i - \bar{y})^2}{\sum_1^n (y_i - \bar{y})^2} \quad \text{Equação 15}$$

REP%

REP é o desvio padrão relativo, dado por ³⁷:

$$REP = 100 \frac{RMSECV}{\bar{y}} \quad \text{Equação 16}$$

2.5 Repetibilidade e Reprodutibilidade

No ASTM D86 ¹, a definição de repetibilidade e reprodutibilidade é a seguinte:

“Repetibilidade - A diferença entre os resultados de testes sucessivos, obtidos pelo mesmo operador usando o mesmo aparelho sob condições operacionais constantes em material de teste idêntico, iria, a longo prazo, na operação normal e correta deste método de teste, exceder os valores na Tabela 7, Tabela 8 e Tabela 9 apenas em um caso em vinte.” ASTM D86 ¹

“Reprodutibilidade - A diferença entre dois resultados de teste simples e independentes, obtidos por diferentes operadores trabalhando em laboratórios diferentes com material de teste idêntico, iria, a longo prazo, na operação normal e correta deste método de teste, exceder os valores na Tabela 7, Tabela 8 e Tabela 9 apenas em um caso em vinte.” ASTM D86 ¹

Em outras palavras, a repetibilidade ou repetitividade de um método é o grau de concordância entre os resultados de medições sucessivas de um mesmo mensurando, efetuadas sob as mesmas condições de medição (mesmo procedimento de medição, mesmo executante, mesmo instrumento de medição, utilizando nas mesmas condições, mesmo local, repetições em curto espaço de tempo). É dada pela equação:

$$r = 2,8 \times S_r \quad \text{Equação 17}$$

Onde r é a repetibilidade e S_r é o desvio padrão da repetibilidade, que vem da equação:

$$S_r = \sqrt{\frac{\sum_1^p S_i^2}{p}} \quad \text{Equação 18}$$

Onde:

p = Número de executantes

S_i = Desvio-padrão de um executante

A reprodutibilidade é o grau de concordância entre os resultados das medições de um mesmo mensurando, efetuadas em condições variadas de medição. As condições variadas podem incluir: princípio de medição, método de medição, executante, instrumento de medição, padrão de referência, local, condições de utilização e tempo. É dada pela equação:

$$R = 2,8 \times S_R \quad \text{Equação 19}$$

Onde **R** é a reprodutibilidade e **S_R** é o desvio padrão da reprodutibilidade, que vem da equação:

$$S_R = \sqrt{(S_{\bar{x}})^2 + \frac{(S_r)^2(n-1)}{n}} \quad \text{Equação 20}$$

Onde,

S_R = Desvio-padrão de reprodutibilidade

S _{\bar{x}} = Desvio-Padrão da média

S_r = Desvio-padrão de repetitividade

n = Número de repetições

Na prática, os valores de Repetibilidade e Reprodutibilidade deste método foram obtidos a partir de um programa Interlaboratorial de 2010, com a participação de 26 laboratórios que analisaram 21 conjuntos de amostras ¹, conforme descrito na versão de 2018 do ASTM D86.

3 PARTE EXPERIMENTAL

3.1 Amostras

As amostras e análises utilizadas neste trabalho são de acompanhamento de processo de uma indústria petroquímica, detalhadas a seguir. Neste trabalho foram utilizados dados de 68 amostras de *PYGAS* pesada, obtidas entre os anos de 2018 e 2019, num intervalo de tempo em que o mesmo cromatógrafo e o mesmo destilador foram utilizados para o conjunto de dados.

>>> Número da amostra					
>>> Ponto					
>>> Descrição		PYGAS PESADA	PYGAS PESADA	PYGAS PESADA	PYGAS PESADA
>>> Data - Hora		10/03/2020 06:00	05/03/2020 06:00	03/03/2020 06:00	27/02/2020 06:00
>>> Tipo de amostra		Programada	Programada	Programada	Programada
>>> Status		Encerrado	Encerrado	Aprovado	Encerrado
c4-	% massa	0.4686	0.4844	0.4801	0.4513
c5	% massa	0.5837	0.6203	0.6178	0.5695
benzeno	% massa	3.2057	3.2868	3.4313	3.3015
c6 nao arom 1	% massa	0.0336	0.036	0.0349	0.0337
c6 nao arom 2	% massa	0.772	0.8223	0.8493	0.7904
tolueno	% massa	9.5224	9.5081	10.119	9.9761
c7c8 nao arom 1	% massa	0.7388	0.7947	0.8835	0.7223
c7c8 nao arom 2	% massa	1.929	2.1426	2.3842	1.8074
etilbenzeno	% massa	4.31	4.5252	4.4038	4.2635
p+m-xileno	% massa	14.8448	14.9359	14.5771	13.9986
o-xileno	% massa	5.5131	5.8279	5.6808	5.3742
c9+ nao arom	% massa	1.3744	1.5698	1.8176	1.5839
estireno	% massa	16.8309	17.4537	16.4982	17.1522
alfa-metilestireno	% massa	1.2411	1.2615	1.2613	1.3373
c9a 1	% massa	0.1153	0.1182	0.1178	0.1137
c9a 2	% massa	1.1011	1.1249	1.0842	1.1646
c9a 3	% massa	0.9343	0.9247	0.9614	0.949
c9a 4	% massa	3.3804	3.2292	3.3526	3.3189
c9a 5	% massa	1.2883	1.2413	1.2798	1.2908
c9a 6	% massa	1.0301	0.881	0.8405	1.0373
c9a 7	% massa	1.1565	1.1751	1.2268	1.2334
c9a 8	% massa	0.7414	0.7451	0.7427	0.7833
c9a 9	% massa	6.525	6.0454	6.214	6.7068
c9a 10	% massa	3.4884	3.3955	3.4815	3.8317
c9a 11	% massa	3.2867	3.1089	3.114	3.4272
c9a 12	% massa	1.8328	1.7998	1.5888	1.6631
c9a 13	% massa	0.4417	0.4666	0.445	0.5304
demais c10+	% massa	4.8403	4.3315	4.5004	4.7069
indeno	% massa	8.1399	7.8751	7.6839	8.1734
naftaleno	% massa	0.3297	0.2686	0.3275	0.3076
DEST PIE	°C	61.5	63	80.6	84.5
DEST 05%	°C	126.8	125.8	126.4	126.8
DEST 10%	°C	130.9	130.6	130.9	131.6
DEST 20%	°C	136.9	135.9	136.9	136.9
DEST 30%	°C	140.9	140.1	141.1	141.2
DEST 40%	°C	145.1	144.2	145	145.1
DEST 50%	°C	148.6	147.1	148.6	149
DEST 60%	°C	151.8	150.8	152	152.2
DEST 70%	°C	155.7	154.7	156	156.4
DEST 80%	°C	161.1	159.8	161.5	161.5
DEST 90%	°C	168.7	167.7	169.8	168.5
DEST 95%	°C	174.8	173.6	177.3	174.3
DEST PFE	°C	188.2	187.5	192.1	186.8
DEST RESIDUO	% volume	1	1	1	1

Figura 6: Disponibilização dos dados da amostra de *PYGAS* Pesada, conforme disponibilizado para produção e engenharia.

Os resultados da composição cromatográfica estão em % massa de cada um dos 29 componentes, que pode ser um componente ou a soma de um grupo de componentes, e os resultados da destilação estão em °C para cada uma das 13 temperaturas e em % volume (% v/v) para o resíduo. A Figura 6 apresenta o formato de apresentação dos resultados para as áreas de produção e engenharia.

3.2 Cromatografia em fase gasosa

As análises cromatográficas foram realizadas em cromatógrafo gasoso modelo Agilent 7890A, com injetor tipo *split*, coluna de polimetilsiloxano de dimensões 60 m x 250 µm x 0,25 µm e detector FID. A razão de *split* é de 100:1 e a temperatura do injetor de 250 °C. As condições do forno são: 40 °C por 7 min, aumenta a uma taxa de 3 °C/min até 70 °C, depois aumenta a uma taxa de 8 °C/min até 100 °C, e nova taxa de aquecimento de 10 °C/min até 150 °C, ficando nesta por 25 min, e finalmente a última taxa de aquecimento de 20 °C/min até 230 °C, permanecendo nesta temperatura por 9 min. A análise ocorre à pressão constante de 18 psi. O injetor automático injeta 0,5 µL de amostra. A calibração dos componentes foi construída utilizando ECN.

A identificação de cada componente ou conjunto de componentes foi realizada com a aplicação da configuração de tipo de coluna e condições de análise em um GC-MS quadrupolo simples.

3.3 Destilação

As análises de destilação foram realizadas em destilador automático, modelo Herzog HDA 627, conforme norma ASTM D86 ¹, amostras do grupo 2. Para estas análises é usado o suporte de 38 mm de diâmetro, amostras com temperatura ≤ 10 °C, temperatura da proveta para recolhimento do destilado e do balão de destilação entre 13 °C e 18 °C, usando termômetro do tipo 7C.

3.3.1 Repetibilidade e Reprodutibilidade do ASTM D86

A norma ASTM D86 apresenta para o PIE o valor de 2,7 °C para a repetibilidade

do método para amostras do grupo 2 automático, e o valor 4,7 °C para a reprodutibilidade para as mesmas condições ¹. No caso do PFE, os valores de repetibilidade e reprodutibilidade são 3,3 °C e 7,1 °C, respectivamente. O ASTM D86 não apresenta valores de repetibilidade e reprodutibilidade para resíduo da destilação. Para as demais faixas de destilação, os valores são calculados a partir de equações dadas pela norma de referência. A Tabela 1 é foi retirada da norma ASTM D86 de 2018, com as equações de cada ponto da destilação.

Tabela 1: Valores e equações de repetibilidade e reprodutibilidade conforme ASTM D86 de 2018 para cada ponto da destilação.

Percentual evaporado	Repetibilidade °C	Reprodutibilidade °C
PIE	2,7	4,7
5	$1,4 + 2,8.(0,43.Sc + 0,24)$	$2,5 + 2,8.(0,43.Sc + 0,24)$
10	$0,9 + 2,8.(0,43.Sc + 0,24)$	$1,9 + 2,8.(0,43.Sc + 0,24)$
20	$0,9 + 2,8.(0,43.Sc + 0,24)$	$2,0 + 2,8.(0,43.Sc + 0,24)$
30	$0,8 + 2,8.(0,43.Sc + 0,24)$	$1,8 + 2,8.(0,43.Sc + 0,24)$
40	$0,9 + 2,8.(0,43.Sc + 0,24)$	$2,0 + 2,8.(0,43.Sc + 0,24)$
50	$1,0 + 2,8.(0,43.Sc + 0,24)$	$1,9 + 2,8.(0,43.Sc + 0,24)$
60	$1,1 + 2,8.(0,43.Sc + 0,24)$	$2,0 + 2,8.(0,43.Sc + 0,24)$
70	$1,5 + 2,8.(0,43.Sc + 0,24)$	$2,1 + 2,8.(0,43.Sc + 0,24)$
80	$1,1 + 2,8.(0,43.Sc + 0,24)$	$2,0 + 2,8.(0,43.Sc + 0,24)$
90	$1,8 + 2,8.(0,43.Sc + 0,24)$	$2,8 + 2,8.(0,43.Sc + 0,24)$
95	$2,0 + 2,8.(0,43.Sc + 0,24)$	$3,6 + 2,8.(0,43.Sc + 0,24)$
PFE	3,3	7,1

Como exemplo, para destilação de 50% volume, as equações da repetibilidade e reprodutibilidade são dadas respectivamente por:

$$\text{Repetibilidade 50\%, em } ^\circ\text{C} = 1,0 + 2,8 (0,43Sc + 0,24) \quad \text{Equação 21}$$

$$\text{Reprodutibilidade 50\%, em } ^\circ\text{C} = 1,9 + 2,8 (0,43Sc + 0,24) \quad \text{Equação 22}$$

Sc é o slope de temperatura em graus Celsius, dado por:

$$Sc = (T_U - T_L) / (V_U - V_L) \quad \text{Equação 23}$$

Onde,

T_U = temperatura da faixa acima, °C

T_L = temperatura da faixa abaixo, °C

V_U = o volume % recuperado ou evaporado correspondente a T_U

V_L = o volume % recuperado ou evaporado correspondente a T_L

Os valores de temperatura utilizados na equação foram os valores médios de cada ponto, a partir do conjunto das 68 amostras. Então para o exemplo do ponto da destilação 50%:

Repetibilidade 50%, em °C = $1,0 + 2,8 \{0,43 \times [(154,9 - 141,8) / (60 - 40)] + 0,24\} = 2,1$.

Reprodutibilidade 50%, em °C = $1,9 + 2,8 \{0,43 \times [(154,9 - 141,8) / (60 - 40)] + 0,24\} = 3,0$.

Para este conjunto de 68 amostras, os valores de repetibilidade e reprodutibilidade da destilação 50% são 2,1 °C e 3,0 °C, respectivamente.

As equações de cada ponto foram adquiridas da norma de referência e calculadas, para posterior comparação com as diferenças obtidas entre o valor medido (obtido na prática) e o valor calculado pelo modelo.

3.4 Processamento de Dados

Para os dados das 68 amostras disponibilizados conforme exemplificado na Figura 6 e resumidos nas Tabelas 2 e 4 do próximo capítulo, deve-se primeiramente organizar estes dados em matriz e vetores. A matriz X contém todas as variáveis de entrada, no caso, os resultados da análise cromatográfica. Cada coluna é uma variável e cada linha, uma amostra, e para cada valor da destilação é construído um vetor y , com cada amostra correspondendo a uma linha. A Figura 7 explicita a organização dos dados, que oriundos do exemplo da Figura 6, precisam ser transpostos para construção de X e y 's.

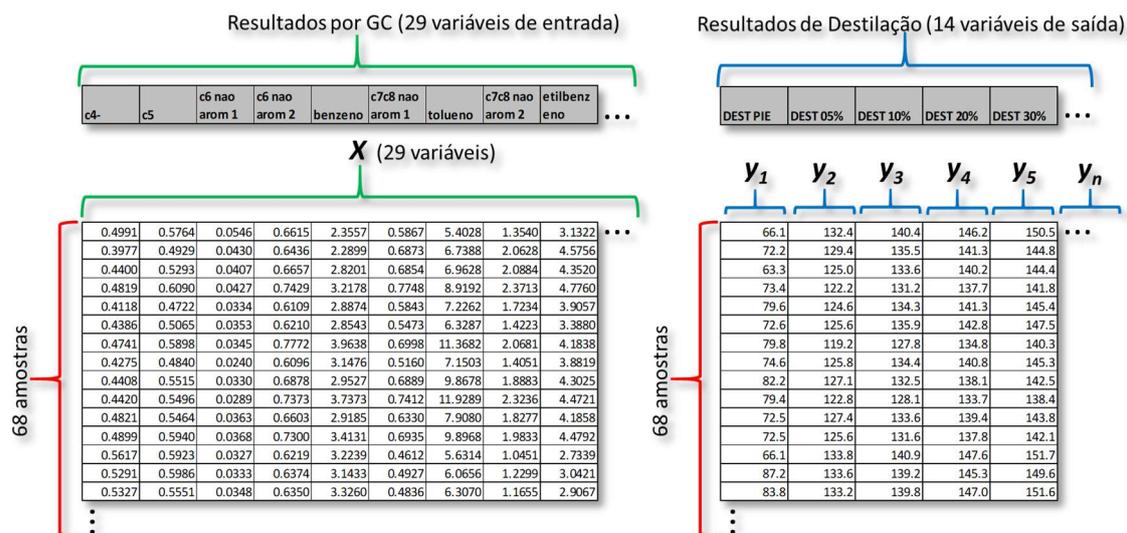


Figura 7: Organização dos dados na matriz X e vetores y . Fonte própria.

Após organização das matrizes, os dados foram pré-tratados, sendo estes autoescalados. As 68 amostras foram separadas em dois grupos: treinamento (calibração/validação) e predição. Foram separadas na razão aproximada 70/30, ficando 48 amostras no conjunto de treinamento e 20 amostras no conjunto de predição, utilizando o método SPXY para a separação dos conjuntos. No conjunto de treinamento, onde estão as amostras de calibração e validação, foi utilizada a validação cruzada.

O software MATLAB R2012b (The MathWorks, Inc., Natick, Massachusetts, EUA) e *toolboxes* de métodos que são: 1) não supervisionado – PCA – “pca_gui”⁴⁷; 2) supervisionado – MLR – “spa_gui”^{48, 49}; 3) supervisionado – PLS – “ispa_gui”^{40, 50}; 4) autoescalamento dos dados da matriz X e a separação das amostras usando MULTIVARIATE DATA HANDLING – “data_hand_gui”.

4 RESULTADOS E DISCUSSÕES

4.1 Estatística descritiva do conjunto de dados

A Tabela 2 contém a descrição de cada componente obtido pela composição cromatográfica junto à sua identificação, bem como valores da média, desvio padrão, valor mínimo e valor máximo de cada componente, que constituirão a matrix X, sendo cada componente (ou somatórios destes) uma variável x .

Tabela 2: Apresentação resumida dos resultados cromatográficos das 68 amostras, em % massa.

Componente	máximo	mínimo	média	desvio padrão
C4- (somatório de componentes com 4 carbonos ou menos)	0,98	0,34	0,531	0,136
C5 (somatório de componentes com 5 carbonos)	1,12	0,39	0,598	0,153
C6 naro 1 (componente não aromático de 6 carbonos)	0,08	0,02	0,036	0,011
C6 naro 2 (somatório de componentes não aromáticos com 6 carbonos)	1,37	0,47	0,682	0,167
BENZENO	6,29	2,29	3,311	0,779
C7C8 naro 1 (somatório de componentes não aromáticos com 7 e 8 carbonos)	1,24	0,37	0,589	0,174
TOLUENO	12,14	4,83	7,392	1,740
C7C8 naro 2 (somatório de componentes não aromáticos com 7 e 8 carbonos)	2,74	0,9	1,425	0,397
ETILBENZENO	5,44	2,73	3,602	0,616
P+M-XILENO (somatório de p-xileno e m-xileno)	14,85	7,01	10,888	1,929
O-XILENO	7,43	3,95	4,973	0,646
ESTIRENO	25,26	14,24	17,863	1,813
C9+ naro (somatório de componentes não aromáticos de 9 carbonos)	3,23	1,60	2,183	0,364
C9A 1 (composto aromático de 9 carbonos)	0,14	0,07	0,102	0,013
C9A 2 (composto aromático de 9 carbonos)	1,81	0,86	1,343	0,184
C9A 3 (composto aromático de 9 carbonos)	1,40	0,70	1,045	0,135
C9A 4 (composto aromático de 9 carbonos)	4,76	2,58	3,432	0,432
C9A 5 (composto aromático de 9 carbonos)	1,71	1,00	1,285	0,165
C9A 6 (composto aromático de 9 carbonos)	2,48	0,88	1,551	0,254
C9A 7 (composto aromático de 9 carbonos)	4,23	2,08	2,995	0,461
C9A 8 (composto aromático de 9 carbonos)	1,34	0,55	0,943	0,159
C9A 9 (composto aromático de 9 carbonos)	9,88	2,66	7,372	1,230
C9A 10 (composto aromático de 9 carbonos)	2,83	1,03	1,992	0,364
C9A 11 (composto aromático de 9 carbonos)	4,68	1,53	2,706	0,544
C9A 12 (composto aromático de 9 carbonos)	3,55	0,76	2,717	0,470
C9A 13 (composto aromático de 9 carbonos)	4,99	0,94	3,372	0,697
Outros C10+ (compostos com 10 carbonos ou mais)	10,75	0,54	6,031	2,267
INDENO	11,11	1,25	7,796	1,802
NAFTALENO	4,01	0,04	1,243	0,942

Como podem ser observadas, as amostras cobrem uma ampla faixa de concentração para todos os casos. Ilustrativamente, a Figura 8 apresenta os gráficos de distribuição de cada variável x , com a linha média e linha de máximo e mínimo de carta controle, identificando os pontos acima ou abaixo dos 3 desvios-padrão.



Figura 8: gráficos de distribuição de cada variável x .

Tabela 3: Correlação de Pearson entre todas as variáveis, que são os componentes obtidos pela composição cromatográfica.

	c4-	c5	c6 nao arom 1	c6 nao arom 2	benzeno	c7c8 nao arom 1	tolueno	c7c8 nao arom 2	etilbenze no	p+m-xileno	o-xileno	estireno	c9+ nao arom	c9a 1	c9a 2	c9a 3	c9a 4	c9a 5	c9a 6	c9a 7	c9a 8	c9a 9	c9a 10	c9a 11	c9a 12	c9a 13	demais c10+	indeno	naftaleno
c4-	1.0000	0.9771	0.7972	0.9165	0.8519	0.7180	0.5762	0.3910	0.2958	0.3104	0.2921	0.2190	-0.1422	-0.0536	-0.1578	-0.1727	-0.1095	-0.0792	-0.2190	-0.3561	-0.4043	-0.6006	-0.3621	-0.0540	-0.5077	-0.5485	-0.2927	-0.5657	-0.1149
c5	0.9771	1.0000	0.8413	0.9605	0.8517	0.8043	0.6500	0.5054	0.3851	0.3686	0.3137	0.2103	-0.0675	-0.0311	-0.2032	-0.1892	-0.1299	-0.1081	-0.1901	-0.3942	-0.4464	-0.6366	-0.4356	-0.0598	-0.5605	-0.5840	-0.3351	-0.6114	-0.1654
c6 nao arom 1	0.7972	0.8413	1.0000	0.8653	0.6553	0.8393	0.5096	0.5730	0.4264	0.2724	0.3617	0.1693	0.2011	0.1592	-0.1093	0.0167	0.0684	0.0580	-0.0608	-0.1629	-0.3409	-0.5237	-0.2979	0.0879	-0.4938	-0.4428	-0.3945	-0.5878	-0.2614
c6 nao arom 2	0.9165	0.9605	0.8653	1.0000	0.8356	0.8710	0.7496	0.6697	0.5427	0.4590	0.4196	0.2495	0.0358	0.0170	-0.2488	-0.1847	-0.1356	-0.1270	-0.1948	-0.4225	-0.5221	-0.7096	-0.5112	-0.0666	-0.6543	-0.6422	-0.4249	-0.6888	-0.2620
benzeno	0.8519	0.8517	0.6553	0.8356	1.0000	0.5714	0.7260	0.3975	0.3672	0.3384	0.4196	0.3122	-0.3269	-0.1888	-0.2423	-0.3475	-0.3445	-0.2931	-0.4270	-0.4704	-0.5373	-0.7062	-0.4713	-0.1743	-0.6262	-0.6431	-0.2511	-0.5973	-0.0692
c7c8 nao arom 1	0.7180	0.8043	0.8393	0.8710	0.5714	1.0000	0.7147	0.7980	0.5897	0.4898	0.3592	0.0826	0.2198	0.0230	-0.3742	-0.2122	-0.1122	-0.1359	-0.0975	-0.4326	-0.5496	-0.6025	-0.4961	-0.0165	-0.5757	-0.5638	-0.3815	-0.6203	-0.2863
tolueno	0.5762	0.6500	0.5096	0.7496	0.7260	0.7147	1.0000	0.8093	0.7372	0.7347	0.5087	0.2337	-0.0984	-0.2135	-0.5533	-0.5018	-0.3950	-0.3827	-0.3076	-0.7235	-0.7894	-0.7578	-0.7224	-0.1567	-0.7145	-0.7692	-0.4223	-0.7243	-0.3099
c7c8 nao arom 2	0.3910	0.5054	0.5730	0.6697	0.3975	0.7980	0.8093	1.0000	0.8519	0.6701	0.5152	0.2074	0.4246	0.1228	-0.4050	-0.2029	-0.1635	-0.2251	-0.0196	-0.4135	-0.5874	-0.6372	-0.6024	0.0140	-0.6645	-0.5651	-0.5688	-0.6804	-0.5003
etilbenzeno	0.2958	0.3851	0.4264	0.5427	0.3672	0.5897	0.7372	0.8519	1.0000	0.8246	0.7863	0.6134	0.3601	0.3732	-0.1597	0.0068	0.0343	-0.0065	-0.0315	-0.3977	-0.5646	-0.7448	-0.6857	-0.0984	-0.7694	-0.7491	-0.7369	-0.8503	-0.6029
p+m-xileno	0.3104	0.3686	0.2724	0.4590	0.3384	0.4898	0.7347	0.6701	0.8246	1.0000	0.7189	0.4156	0.0803	0.0659	-0.4116	-0.2410	0.0573	0.0576	-0.0570	-0.6798	-0.7459	-0.5849	-0.7450	-0.1016	-0.5748	-0.8106	-0.6132	-0.8115	-0.5360
o-xileno	0.2921	0.3137	0.3617	0.4196	0.4196	0.3592	0.5087	0.5152	0.7863	0.7189	1.0000	0.6949	0.2025	0.3459	-0.0286	0.0678	0.2337	0.2323	-0.0666	-0.2624	-0.5194	-0.6210	-0.5497	0.0059	-0.6510	-0.7437	-0.6923	-0.8505	-0.5651
estireno	0.2190	0.2103	0.1693	0.2495	0.3122	0.0826	0.2337	0.2074	0.6134	0.4156	0.6949	1.0000	0.1070	0.6622	0.5170	0.4506	0.2770	0.2842	-0.1110	0.0139	-0.0391	-0.5772	-0.3251	-0.3060	-0.5682	-0.5781	-0.6383	-0.6208	-0.4312
c9+ nao arom	-0.1422	-0.0675	0.2011	0.0358	-0.3269	0.2198	-0.0984	0.4246	0.3601	0.0803	0.2025	0.1070	1.0000	0.6393	0.2153	0.4758	0.4028	0.2881	0.5175	0.3905	0.2066	0.0408	-0.0009	0.4038	-0.0803	0.1493	-0.4622	-0.1687	-0.5175
c9a 1	-0.0536	-0.0311	0.1592	0.0170	-0.1888	0.0230	-0.2135	0.1228	0.3732	0.0659	0.3459	0.6622	0.6393	1.0000	0.7413	0.8690	0.6360	0.5807	0.3326	0.4887	0.3978	-0.1559	0.0104	0.0210	-0.2096	-0.0963	-0.5414	-0.3186	-0.4332
c9a 2	-0.1578	-0.2032	-0.1093	-0.2488	-0.2423	-0.3742	-0.5533	-0.4050	-0.1597	-0.4116	-0.0286	0.5170	0.2153	0.7413	1.0000	0.9030	0.5969	0.5836	0.2297	0.7693	0.7944	0.1688	0.4120	-0.0745	0.1393	0.2573	-0.2222	0.1268	-0.1345
c9a 3	-0.1727	-0.1892	0.0167	-0.1847	-0.3475	-0.2122	-0.5018	-0.2029	0.0068	-0.2410	0.0678	0.4506	0.4758	0.8690	0.9030	1.0000	0.7879	0.7494	0.4015	0.7493	0.7004	0.1918	0.3420	0.0845	0.1374	0.2255	-0.3840	-0.0060	-0.3295
c9a 4	-0.1095	-0.1299	0.0684	-0.1356	-0.3445	-0.1122	-0.3950	-0.1635	0.0343	0.0573	0.2337	0.2770	0.4028	0.6360	0.5969	0.7879	1.0000	0.9646	0.5721	0.5097	0.4130	0.2983	0.1820	0.2918	0.2181	0.0715	-0.4771	-0.1697	-0.4797
c9a 5	-0.0792	-0.1081	0.0580	-0.1270	-0.2931	-0.1359	-0.3827	-0.2251	-0.0065	0.0576	0.2323	0.2842	0.2881	0.5807	0.5836	0.7494	0.9646	1.0000	0.3920	0.4399	0.4092	0.3241	0.2520	0.1783	0.3026	0.0444	-0.4613	-0.1499	-0.4576
c9a 6	-0.2190	-0.1901	-0.0608	-0.1948	-0.4270	-0.0975	-0.3076	-0.0196	-0.0315	-0.0570	-0.0666	-0.1110	0.5175	0.3326	0.2297	0.4015	0.5721	0.3920	1.0000	0.4383	0.2788	0.2936	-0.0775	0.7334	0.0888	0.2659	-0.1941	-0.0013	-0.2755
c9a 7	-0.3561	-0.3942	-0.1629	-0.4225	-0.4704	-0.4326	-0.7235	-0.4135	-0.3977	-0.6798	-0.2624	0.0139	0.3905	0.4887	0.7693	0.7493	0.5097	0.4399	0.4383	1.0000	0.8852	0.5081	0.6803	0.3003	0.4165	0.6835	-0.0307	0.4378	-0.0813
c9a 8	-0.4043	-0.4464	-0.3409	-0.5221	-0.5373	-0.5496	-0.7894	-0.5874	-0.5646	-0.7459	-0.5194	-0.0391	0.2066	0.3978	0.7944	0.7004	0.4130	0.4092	0.2788	0.8852	1.0000	0.6340	0.7719	0.0737	0.6112	0.7538	0.1020	0.6319	0.0437
c9a 9	-0.6006	-0.6366	-0.5237	-0.7096	-0.7062	-0.6025	-0.7578	-0.6372	-0.7448	-0.5849	-0.6210	-0.5772	0.0408	-0.1559	0.1688	0.1918	0.2983	0.3241	0.2936	0.5081	0.6340	1.0000	0.7302	0.2642	0.9587	0.8438	0.3192	0.7980	0.0687
c9a 10	-0.3621	-0.4356	-0.2979	-0.5112	-0.4713	-0.4961	-0.7224	-0.6024	-0.6857	-0.7450	-0.5497	-0.3251	0.0009	0.0104	0.4120	0.3420	0.1820	0.2520	-0.0775	0.6803	0.7719	0.7302	1.0000	-0.0331	0.8052	0.7925	0.2784	0.7446	0.1625
c9a 11	-0.0540	-0.0598	0.0879	-0.0666	-0.1743	-0.0165	-0.1567	0.0140	-0.0984	-0.1016	0.0059	-0.3060	0.4038	0.0210	-0.0745	0.0845	0.2918	0.1783	0.7334	0.3003	0.0737	0.2642	-0.0331	1.0000	0.1213	0.3007	-0.1258	-0.0099	-0.2301
c9a 12	-0.5077	-0.5605	-0.4938	-0.6543	-0.6262	-0.5757	-0.7145	-0.6645	-0.7694	-0.5748	-0.6510	-0.5682	-0.0803	-0.2096	0.1393	0.1374	0.2181	0.3026	0.0888	0.4165	0.6112	0.9587	0.8052	0.1213	1.0000	0.7950	0.3519	0.8022	0.1334
c9a 13	-0.5485	-0.5840	-0.4428	-0.6422	-0.6431	-0.5638	-0.7692	-0.5651	-0.7491	-0.8106	-0.7437	-0.5781	0.1493	-0.0963	0.2573	0.2255	0.0715	0.0444	0.2659	0.6835	0.7538	0.8438	0.7925	0.3007	0.7950	1.0000	0.4174	0.8935	0.2180
c10+	-0.2927	-0.3351	-0.3945	-0.4249	-0.2511	-0.3815	-0.4223	-0.5688	-0.7369	-0.6132	-0.6923	-0.6383	-0.4622	-0.5414	-0.2222	-0.3840	-0.4771	-0.4613	-0.1941	-0.0307	0.1020	0.3192	0.2784	-0.1258	0.3519	0.4174	1.0000	0.6671	0.9328
indeno	-0.5657	-0.6114	-0.5878	-0.6888	-0.5973	-0.6203	-0.7243	-0.6804	-0.8503	-0.8115	-0.8505	-0.6208	-0.1687	-0.3186	0.1268	-0.0060	-0.1697	-0.1499	-0.0013	0.4378	0.6319	0.7980	0.7446	-0.0099	0.8022	0.8935	0.6671	1.0000	0.4663
naftaleno	-0.1149	-0.1654	-0.2614	-0.2620	-0.0692	-0.2863	-0.3099	-0.5003	-0.6029	-0.5360	-0.5651	-0.4312	-0.5175	-0.4332	-0.1345	-0.3295	-0.4797	-0.4576	-0.2755	-0.0813	0.0437	0.0687	0.1625	-0.2301	0.1334	0.2180	0.9328	0.4663	1.0000

As amostras nas posições 1, 30, 36, 43, 44, 57, 61 apresentaram ao menos um dado extremo, estando acima ou abaixo dos 3 desvios-padrão.

A Tabela 3 apresenta os valores do coeficiente de correlação de Pearson entre as 29 variáveis. É possível identificar alta correlação ($> 0,80$) entre algumas variáveis – sinalizadas em verde – bem como média correlação entre outras ($> 0,65$) – sinalizadas em amarelo.

A Tabela 4 contém valores da média, desvio padrão, valor mínimo e valor máximo de cada resultado y da destilação obtida para o controle de processo da *PYGAS* pesada.

Tabela 4: Estatística básica dos resultados de destilação das 68 amostras.

Parâmetro da Destilação	Unidade de expressão	máximo	mínimo	média	desvio padrão
PIE	°C	90,7	57,8	71,14	11,40
5%	°C	135,8	118,6	126,76	15,81
10%	°C	142,1	127,8	133,95	16,52
20%	°C	148,0	133,7	140,31	17,30
30%	°C	152,4	138,4	144,58	17,81
40%	°C	156,7	141,8	148,29	18,27
50%	°C	160,3	144,3	151,57	18,68
60%	°C	165,0	146,6	154,89	19,12
70%	°C	170,1	149,1	158,86	19,67
80%	°C	177,2	152,2	164,02	20,43
90%	°C	189,5	156,8	172,08	21,73
95%	°C	201,0	161,5	180,00	23,36
PFE	°C	214,6	177,4	195,59	25,26
RESÍDUO	% v/v	1,5	0,0	0,83	0,32

Ilustrativamente, a Figura 9 apresenta os gráficos de distribuição de cada resultado, com a linha média e linha de máximo e mínimo de carta controle, identificando os pontos extremos e acima ou abaixo dos 3 desvios-padrão.



Figura 9: Gráficos de distribuição de cada resultado y de cada parâmetro da destilação.

Somente a amostra na posição 43 apresentou resultado extremo, abaixo dos 3 desvios-padrão, sendo este na destilação 70%.

Como exemplo de resultados, abaixo se apresenta um cromatograma de uma amostra de *PYGAS* pesada na Figura 10, bem como o gráfico de destilação da mesma amostra, na Figura 11.

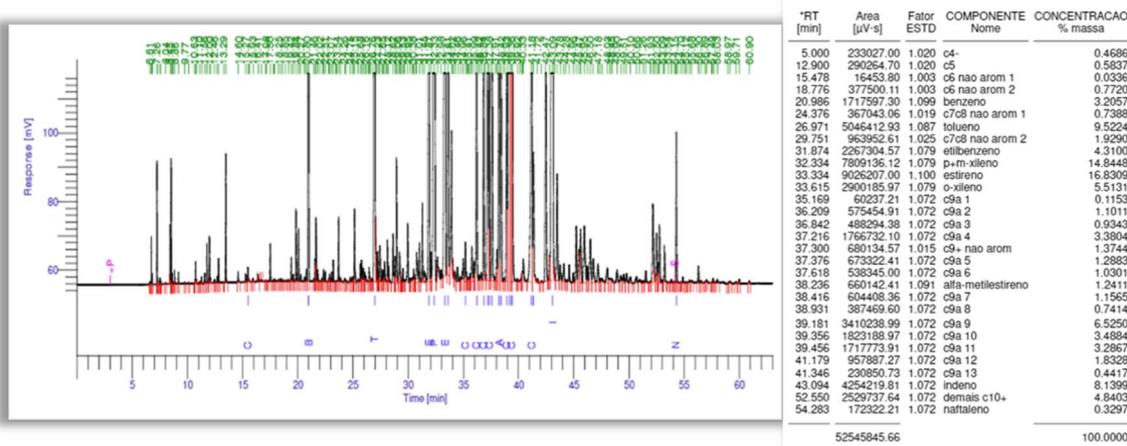


Figura 10: Cromatograma de uma amostra de PYGAS pesada, representado pelo gráfico e pela lista de componentes com as respectivas concentrações.

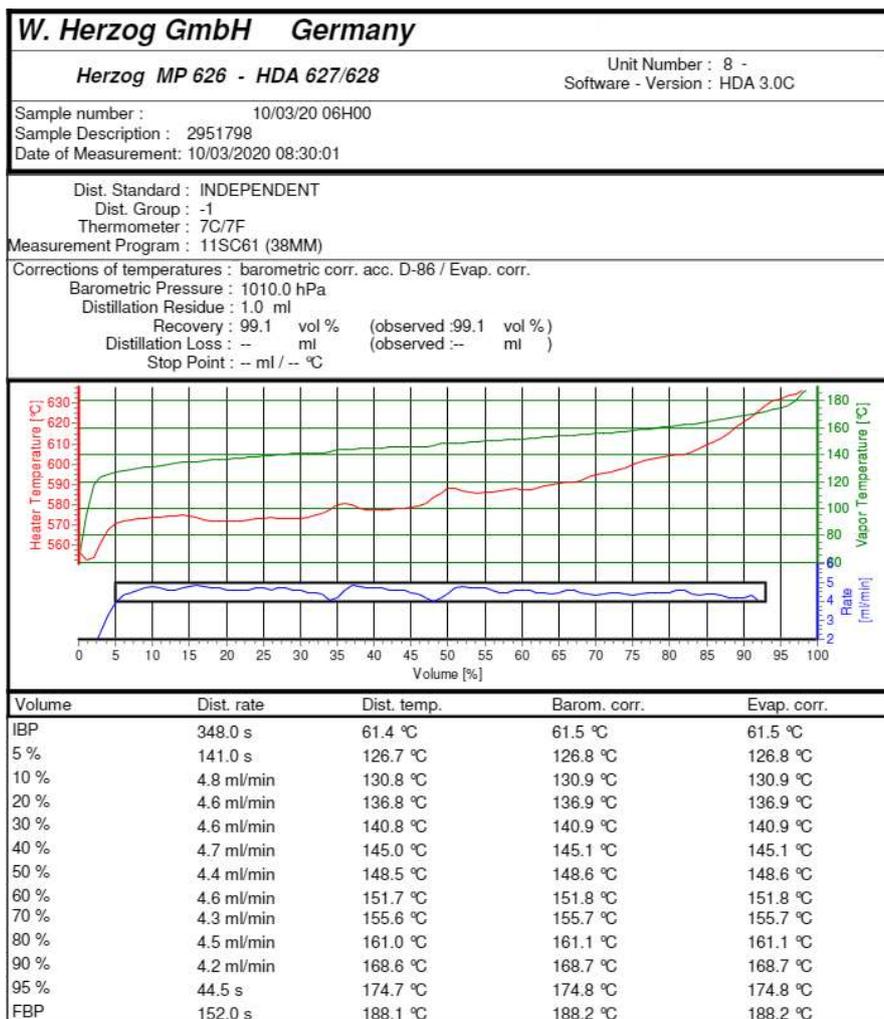


Figura 11: Gráfico de Destilação da mesma amostra de PYGAS pesada.

4.2 Pré processamento do conjunto de dados - autoescalamiento

O autoescalamiento faz-se necessário tanto para aplicação da PCA, quanto para os modelos de regressão MLR e PLS, pois, como pode ser identificado nos dados apresentados no subcapítulo anterior, há diferenças significativas nas concentrações de cada componente do resultado cromatográfico. A Figura 12 ilustra graficamente como cada variável x se apresenta dimensionalmente, sendo possível observar sua distribuição e desvios no quadro 1, enquanto no quadro 2 é possível ver como os dados ficam após o autoescalamiento.

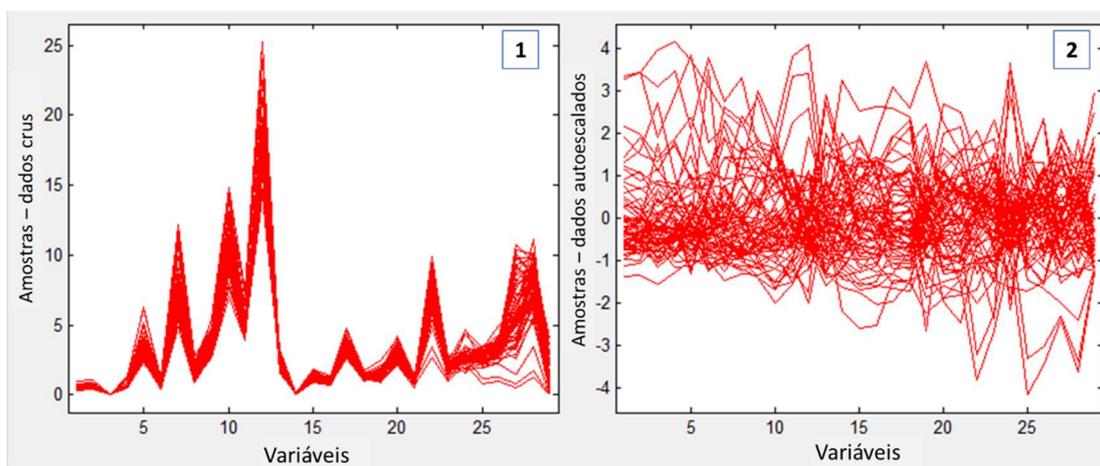


Figura 12: Distribuição dos dados da matriz X , graficamente. No Quadro 1, os valores originais de cada variável x . No quadro 2, os dados autoescalados de cada variável x .

O autoescalamiento foi utilizado para deixar todas as variáveis com a mesma importância (mesmo impacto) dentro do modelo. É possível observar que a variável 12 (Estireno) tem o maior valor médio (17,62%) podendo chegar à concentração de 25%, porém no autoescalamiento seu valor passa a ter o mesmo impacto que a variável 3 (C6 naro 1) que tem o menor valor médio (0,05%).

4.3 Separação das Amostras

Para a separação das amostras utilizou-se o algoritmo SPXY, ficando assim distribuídas as 68 amostras, como demonstrado nas Tabelas 5 e 6:

Tabela 5: Amostras do conjunto de predição e a respectiva posição no conjunto.

posição predição	nº conjunto inicial	posição predição	nº conjunto inicial
1	2	11	26
2	6	12	32
3	7	13	35
4	8	14	38
5	15	15	40
6	16	16	49
7	17	17	50
8	18	18	64
9	19	19	65
10	25	20	67

Tabela 6: Amostras do conjunto de calibração/validação e a respectiva posição no conjunto.

posição calibração	nº conjunto inicial	posição calibração	nº conjunto inicial	posição calibração	nº conjunto inicial
1	1	17	28	33	48
2	3	18	29	34	51
3	4	19	30	35	52
4	5	20	31	36	53
5	9	21	33	37	54
6	10	22	34	38	55
7	11	23	36	39	56
8	12	24	37	40	57
9	13	25	39	41	58
10	14	26	41	42	59
11	20	27	42	43	60
12	21	28	43	44	61
13	22	29	44	45	62
14	23	30	45	46	63
15	24	31	46	47	66
16	27	32	47	48	68

É possível identificar que todas as amostras que apresentaram algum dado extremo nos gráficos de distribuição, estão no conjunto de calibração/validação, sendo estas: 1, 30, 36, 43, 44, 57 e 61.

4.4 Análise de Componentes Principais

Na Figura 13 são exibidos os resultados da PCA conduzidos sobre os blocos de dados X (dados da cromatografia, variáveis independentes dos modelos de calibração ou preditores) e Y (dados da destilação, variáveis dependentes ou a serem preditas).

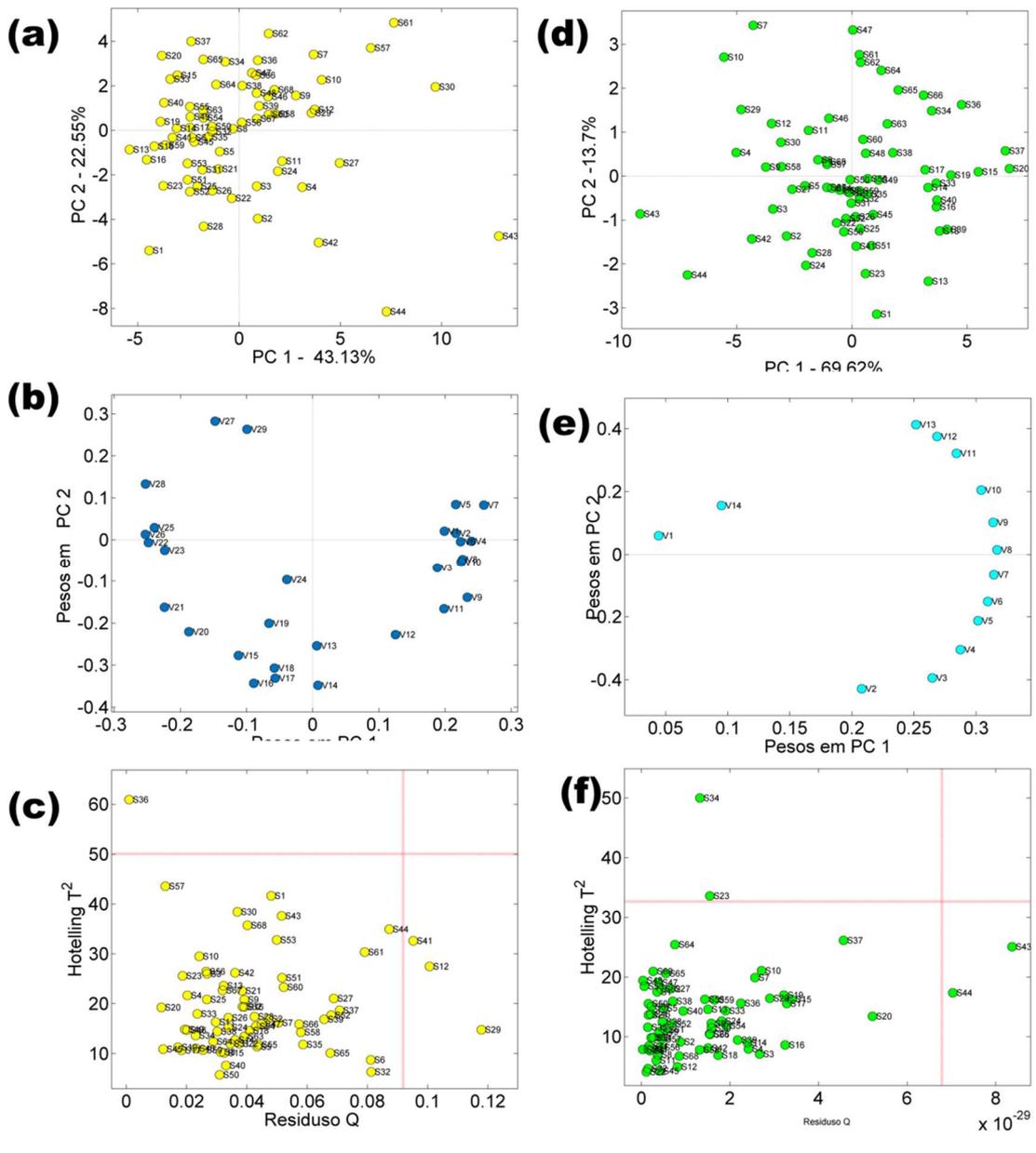


Figura 13: Resultados da análise de componentes principais sobre os blocos de dados X (a: escores, b: pesos e c: Resíduos Q versus T^2) e Y (d: escores, e: pesos e f: Resíduos Q versus T^2).

Nas Figuras 13a e 13d são exibidos a dispersão das amostras, para ambos os blocos de dados, no subespaço bidimensional definidos pelas duas primeiras componentes principais, responsáveis por 65% e 83% da variância explicada, respectivamente. Note que as amostras se apresentam com grupo homogêneo uni modal distribuído em torno de uma região central, caracterização de uma distribuição normal multivariada. Pela inspeção visual dos gráficos de escores não há evidência de amostras anômalas.

A inspeção dos gráficos de pesos (Figuras 13b e 13e) com exceção das variáveis V1 (ponto inicial) e V14 (resíduo) do bloco Y, todas as demais variáveis não exibem influência diferenciada sobre o conjunto de dados.

A confirmação final da presença/ausência de amostras anômalas foi realizada usando o gráfico de resíduos Q versus T^2 Hotelling, em que os limites estatísticos a 95% de confiança são representados pelas linhas pontilhadas vermelhas nas Figuras 13c e 13f. É possível notar que nenhuma amostra exibe ao mesmo tempo alto valor de resíduos Q e T^2 Hotelling, sendo, portanto, considerada uma amostra a ser retirada do conjunto de dados antes da modelagem.

4.5 MLR

MLR foi aplicado de modo a prever cada parâmetro de destilação. A Tabela 7 apresenta os resultados para as figuras de mérito da predição e a Figura 14 apresenta os gráficos de valores preditos *versus* referência de cada parâmetro.

Tabela 7: Apresentação dos resultados das figuras de mérito de cada parâmetro predito por MLR.

Parâmetro da Destilação	*RMSEP	SDV	*bias	r
PIE	7,51	7,402	2,09	0,4094
5	2,43	2,438	-0,51	0,7814
10	0,82	0,824	-0,15	0,9640
20	0,96	0,987	-0,01	0,9506
30	0,80	0,818	-0,09	0,9550
40	0,64	0,652	0,10	0,9726
50	0,73	0,753	0,01	0,9588

60	0,99	0,974	0,30	0,9449
80	0,92	0,610	-0,71	0,9849
90	0,78	0,764	-0,25	0,9810
95	1,45	1,479	-0,12	0,9835
PFE	2,29	2,351	-0,04	0,9438
RESIDUO	0,20	0,201	0,03	-0,1916

**Todos os valores estão expressos em °C, a exceção do resíduo que está expresso em % v/v.*

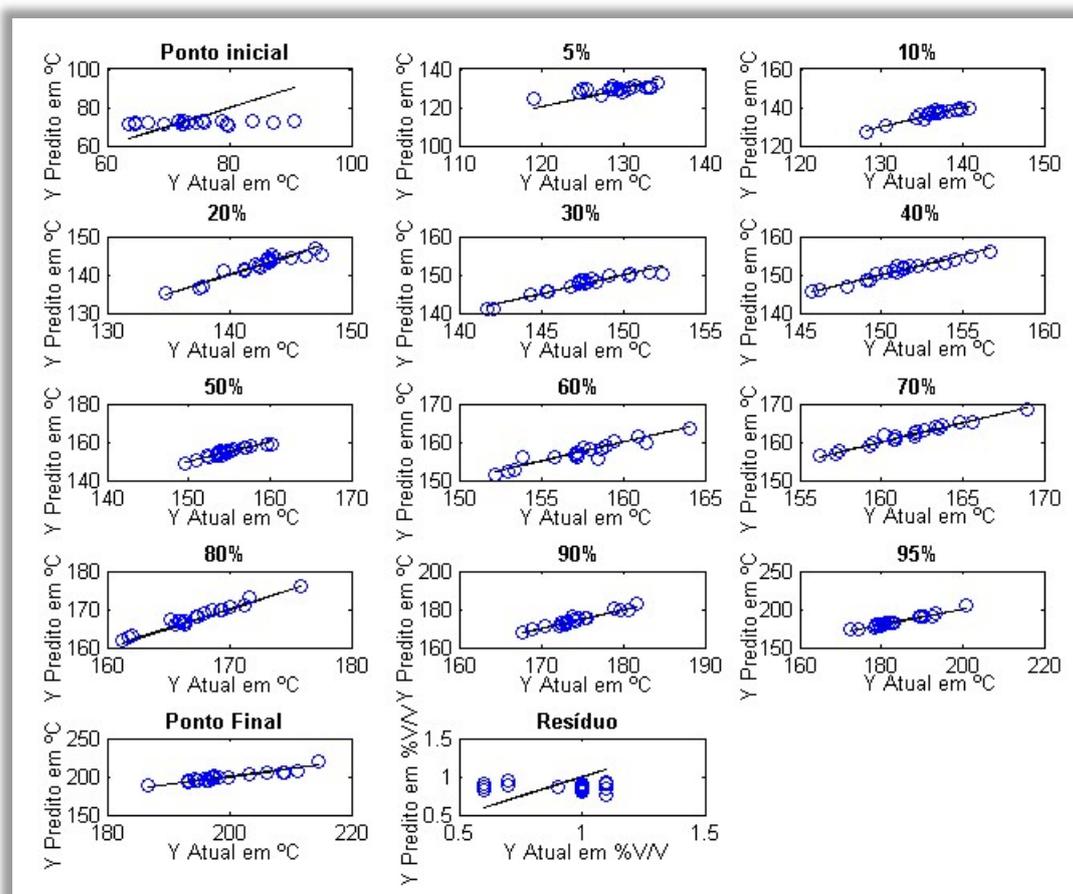


Figura 14: Gráficos de cada parâmetro dos valores atuais (referência) versus valores preditos pela equação obtida por MLR, para as 20 amostras do conjunto de predição. A linha preta representa a linha ideal, em que o valor predito seria idêntico ao valor da referência.

A Figura 14 ilustra os resultados do que é possível observar com os resultados das figuras de mérito da Tabela 7. Os pontos de destilação 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% e PFE apresentam um elevado valor de r (todos com $r > 0,9$), ou seja, o valor predito é semelhante à referência, com alguns valores maiores e outros menores. Também é possível identificar pelos gráficos destes pontos a boa capacidade do modelo de prever os resultados mais altos, médios e mais baixos de cada ponto. Os valores de RMSEP, SDV e *bias* também são baixos para estes pontos (lembrando que os valores de referência destes pontos variam de 127,8 °C, que é o menor valor para o ponto 10%, e 214,6 °C, que é o maior valor para o PFE). O destaque dentre estes pontos está para o *bias* do ponto 80%, que ficou em 0,7 (demais são menores que 0,3), que também pode ser evidenciado no gráfico, com a maioria dos pontos acima da linha preta, ou seja, a tendência de os valores preditos serem maiores que as referências.

Para destilação 5%, o valor de r ficou em 0,635. Os valores de RMSEP, SDV e *bias* não são tão elevados, porém no gráfico é possível verificar a dificuldade do modelo em prever corretamente valores abaixo de 127 °C e acima de 133 °C.

Para o PIE, o MLR foi incapaz de construir um modelo aceitável. Os valores de RMSEP e *bias* são altos e o r é $<0,5$. No gráfico é possível visualizar que o modelo matemático tende a dar quase sempre o valor de 70 °C, com pequenas variações em torno deste. Ou seja, o modelo foi incapaz de prever as situações que deveria dar 90 °C ou 64 °C, por exemplo.

Para o Resíduo de Destilação, ao comparar com o resultado médio que é de 0,8% v/v, identifica-se valores de RMSEP e *bias* altos, e o r é negativo e próximo à zero. O modelo mostrou-se incapaz de prever valores menores que 0,5% v/v. E para valores maiores, percebe-se a dificuldade do modelo, como exemplo, para um valor referência de 0,6% v/v, o modelo encontra valores que variam de 0,8 % v/v até 1,2 % v/v. Na busca de encontrar explicações para os modelos, a Figura 15 foi gerada para mostrar o impacto (peso) de cada variável definida para o modelo. A Tabela 8 contém os valores de b_0 das equações geradas, pois os b_0 's foram extrapolados para deixar mais evidente o impacto de cada variável.

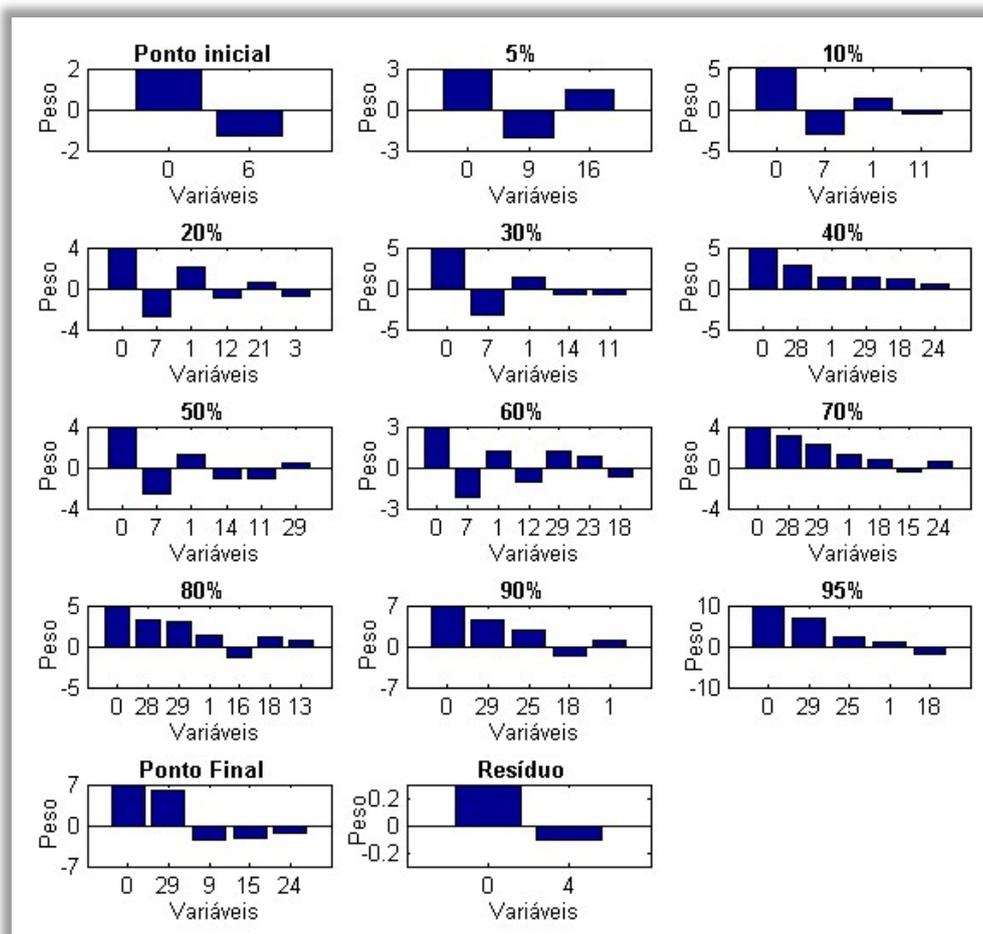


Figura 15: O peso de cada variável x definida para o modelo MLR de cada ponto da destilação. As variáveis definidas nos modelos são as seguintes: 0 é o b_0 da equação; 1 é C4-; 3 é c6 não arom 1; 4 é C6 não arom 2; 6 é c7c8 não arom 1; 7 é tolueno; 9 é etilbenzeno; 11 é o-xileno; 12 é estireno; 13 é c9+ não arom; 14 é c9a 1; 15 é c9a 2; 16 é c9a 3; 18 é c9a 5; 21 é c9a 8; 23 é c9a 10; 24 é c9a 11; 25 é c9a 12; 28 é indeno; 29 é naftaleno.

Tabela 8: Valores de b_0 de cada equação de destilação obtida por MLR.

Parâmetro da Destilação	Valor de b_0
PIE	71,56
5	128,8
10	136,0
20	142,4
30	146,7
40	150,4
50	153,8
60	157,1
70	161,2
80	166,6

90	174,7
95	182,7
PFE	198,5
<u>RESIDUO</u>	<u>0,8231</u>

Para o PIE, somente uma variável foi selecionada para o modelo, sendo essa o C7C8 naro 1. O b_0 é 71,6, e uma vez que há apenas uma variável, e essa tem impacto negativo, isso explica o fato para o valor sempre ficar em torno de 70°C. Além disso, físico-quimicamente os compostos de menor ponto de ebulição – C4-, C5, C6 naro 1 e 2 – têm maior impacto no valor do PIE e não foram selecionados pelo modelo, demonstrando ser um modelo matemático sem fundamento físico-químico. Importante lembrar que foi aplicada a rotina MLR “spa_gui”^{48, 49}, que seleciona as variáveis de impacto, descartando aquelas sem significância matemática.

Para destilação 5%, duas variáveis foram selecionadas, sendo uma positiva – C9a 3 – e outra negativa – etilbenzeno – e com maior impacto o etilbenzeno. O ponto de 5% na corrente de *PYGAS* pesada tem impacto da presença de compostos leves e benzeno e tolueno, que não estão na equação.

Os pontos de destilação 10%, 20%, 30%, 50% e 60% apresentam número maior de variáveis selecionadas e são semelhantes entre si, tanto as variáveis escolhidas, quanto o impacto que essas têm em cada modelo. Interessante a inversão do que se espera físico-quimicamente de que a variável 1 – C4- – tem impacto positivo na equação enquanto a variável 11 ou 12 – o-xileno ou estireno – tem impacto negativo, comparativamente. Apesar deste ponto de atenção, as variáveis definidas em cada modelo de fato têm importância físico-química, além de apenas matemática, podendo explicar os valores aceitáveis de figuras de mérito para estes pontos.

Os pontos de destilação 40%, 70%, 80%, 90% e 95% também apresentam número maior de variáveis selecionadas e são semelhantes entre si, tanto as variáveis escolhidas, quanto o impacto que essas têm em cada modelo. Nestes modelos, ou não há variável com impacto negativo, ou apenas uma com valor muito baixo. Apesar deste ponto de atenção, as variáveis definidas em cada modelo de fato têm importância físico-química, além de apenas matemática, podendo explicar os valores aceitáveis de figuras de mérito para estes pontos.

O PFE, apesar de ter variáveis um pouco diferentes dos demais pontos, tem em

sua equação as que impactam físico-quimicamente, em especial o naftaleno, o que explica as figuras de mérito com valores aceitáveis.

O Resíduo de Destilação está utilizando apenas uma variável e incorreta – C6 nario 1– que são compostos leves e não possuem impacto no Resíduo. Era esperado que para o Resíduo, a equação tivesse variáveis como indeno e/ou naftaleno. Essa ausência de variáveis significativas físico-quimicamente pode explicar os resultados das figuras de mérito insatisfatórios.

Avaliando todas as variáveis selecionadas em cada modelo, identifica-se que variáveis com alta correlação (correlação de Pearson >0,8) não foram selecionadas, o que é uma condição para funcionalidade do MLR.

4.6 PLS

Uma vez que o modelo MLR não se mostrou satisfatório para todos os pontos da destilação, aplicou-se então o modelo de regressão PLS.

Tabela 9: Resumo estatístico de calibração e validação cruzada da aplicação do PLS.

Parâmetro da Destilação	Calibração				Validação			
	VL	*RMSEC	REP%	r_{cal}	*RMSECV	REP%	*bias	r_{val}
PIE	1	6,5	9,21	0,326	7,6	10,80	0,21	-0,285
5%	1	3,0	2,32	0,591	3,1	2,42	0,02	0,515
10%	3	1,2	0,92	0,917	1,5	1,09	0,05	0,869
20%	5	0,9	0,63	0,965	1,2	0,83	0,00	0,929
30%	5	0,8	0,54	0,974	1,0	0,70	0,01	0,949
40%	4	0,8	0,52	0,974	1,0	0,67	0,01	0,954
50%	4	0,7	0,47	0,980	0,9	0,62	0,02	0,961
60%	4	0,7	0,46	0,983	0,9	0,60	0,02	0,968
70%	4	0,7	0,43	0,988	0,9	0,58	0,03	0,975
80%	4	0,8	0,49	0,988	1,1	0,66	0,05	0,975
90%	5	1,1	0,64	0,987	1,5	0,87	0,05	0,972
95%	6	1,5	0,81	0,988	1,9	1,07	0,09	0,975
PFE	4	2,1	1,07	0,976	2,7	1,38	0,12	0,954
RESÍDUO	1	0,3	42,14	0,304	0,3	43,57	0,00	0,125

*Todos os valores estão expressos em °C, a exceção do resíduo que está expresso em % v/v.

Para todos os modelos, as VLs foram escolhidas por meio da abordagem de Haaland e Thomas ⁴⁴. A Tabela 9 resume os resultados obtidos nas etapas de calibração e validação.

Paras os modelos de calibração e validação, o PLS indica, pelas figuras de mérito, bons resultados para destilação dos pontos 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% e PFE.

Assim como no MLR, o PLS não conseguiu modelos com bons resultados nas figuras de mérito para o PIE, 5% e resíduo.

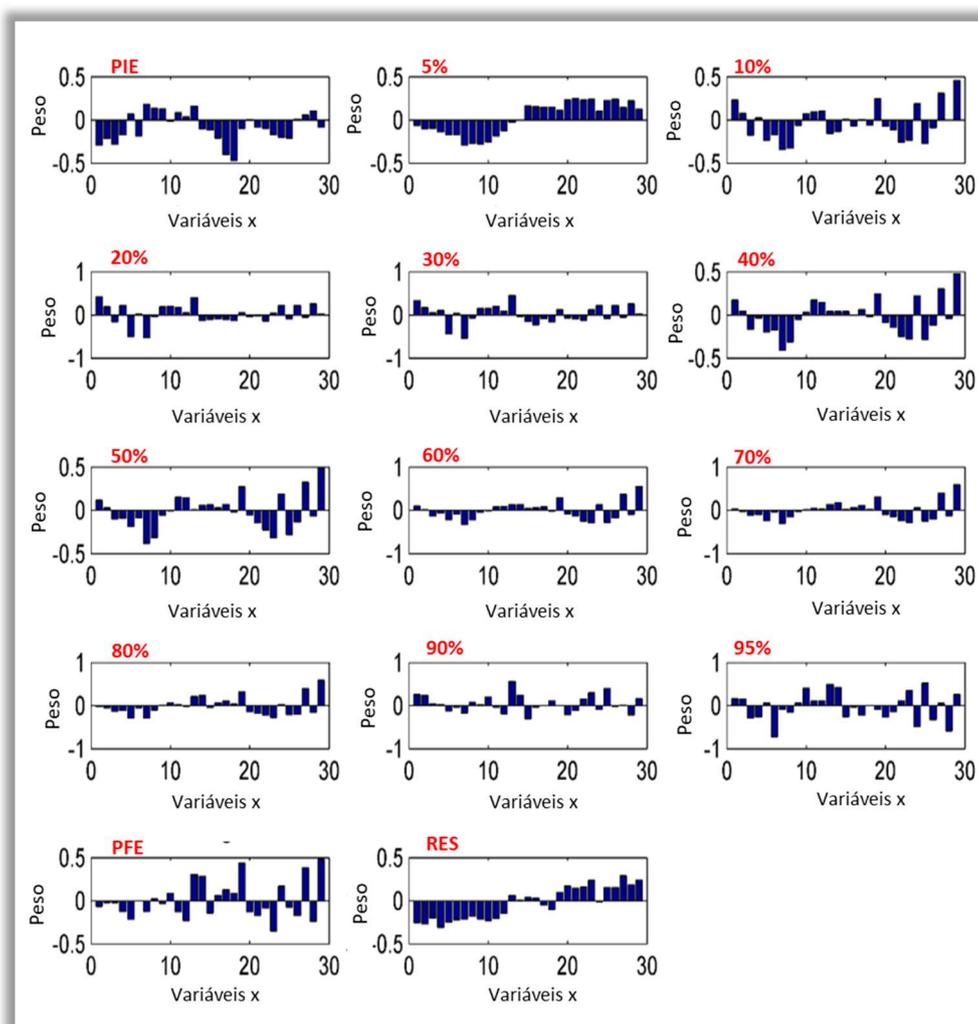


Figura 16: O peso de cada variável x para o modelo PLS de cada ponto da destilação.

O PLS utiliza todas as variáveis, então o peso que cada variável tem no modelo pode ser avaliado, a fim de entender se a resposta matemática pode ter uma

interpretação físico-química. A Figura 16 auxilia com informações para esta avaliação.

Conforme comentado na seção 2.2.5, espera-se uma correlação entre os resultados de cromatografia e os parâmetros de destilação. Quanto menor for o peso de uma variável da Figura 16, menor será seu impacto no modelo. Outro ponto a avaliar, é que as primeiras variáveis (como C4-, C5, por exemplo) são as de menor ponto de ebulição verdadeiro, enquanto as últimas variáveis têm o maior ponto de ebulição, conforme apresentado a partir do tipo de coluna cromatográfica utilizada, que separa HCs por ponto de ebulição. Na maioria dos pontos, essa relação é observada matematicamente para as variáveis de maior impacto em cada ponto da destilação modelado, ou seja, os compostos mais leves (primeiras variáveis) têm valor negativo (para baixar a temperatura) e os compostos mais pesados (últimas variáveis) têm valor positivo (para aumentar a temperatura).

Para o PIE, essa relação está coerente até a variável 13, porém nas variáveis 14 a 19 há uma inversão, com grande impacto no teor de compostos mais pesados na diminuição do PIE. Para o 5% e o resíduo, o peso das variáveis parece estar coerente, com as primeiras diminuindo o valor e as demais aumentando. O que chama a atenção é o peso semelhante para todas as variáveis, quando comparado aos demais modelos. Além disso, PIE, 5% e resíduo têm uma única VL e prevêm valores muito semelhantes (com r próximo a zero). A falta de correlação entre cromatografia e destilação gerou modelos PLS degenerados.

As figuras de mérito e gráficos para as amostras de predição estão na Tabela 10 e Figura 17. A predição foi realizada com o conjunto de 20 amostras, totalmente externo ao conjunto de calibração e validação.

Tabela 10: Resumo estatístico da predição na aplicação do PLS.

Parâmetros Destilação	*RMSEP	REP%	R ²	r	*bias	tbias
PIE	9,9	12,9	61,44	-0,1035	5,52	3,2237
5%	3,1	2,36	11,45	0,7397	0,35	0,5014
10%	1,5	1,09	61,35	0,8581	0,08	0,2188
20%	1,0	0,67	76,09	0,9436	0,23	0,9475
30%	0,8	0,53	98,23	0,9571	0,29	1,3545
40%	1,0	0,66	99,99	0,9343	0,39	1,4229
50%	0,7	0,44	99,99	0,9669	0,17	1,0004
60%	0,7	0,47	99,99	0,9700	0,28	1,3928
70%	0,8	0,49	99,99	0,9722	0,35	1,5138
80%	0,9	0,52	99,99	0,9816	0,50	1,7790
90%	0,9	0,49	99,99	0,9844	0,44	1,6673
95%	1,5	0,81	84,28	0,9708	0,05	0,1396
PFE	2,4	1,21	99,99	0,9263	0,02	0,0306
RES	0,2	24,78	32,68	0,0777	0,11	2,2181

*Todos os valores estão expressos em °C, a exceção do resíduo que está expresso em % v/v

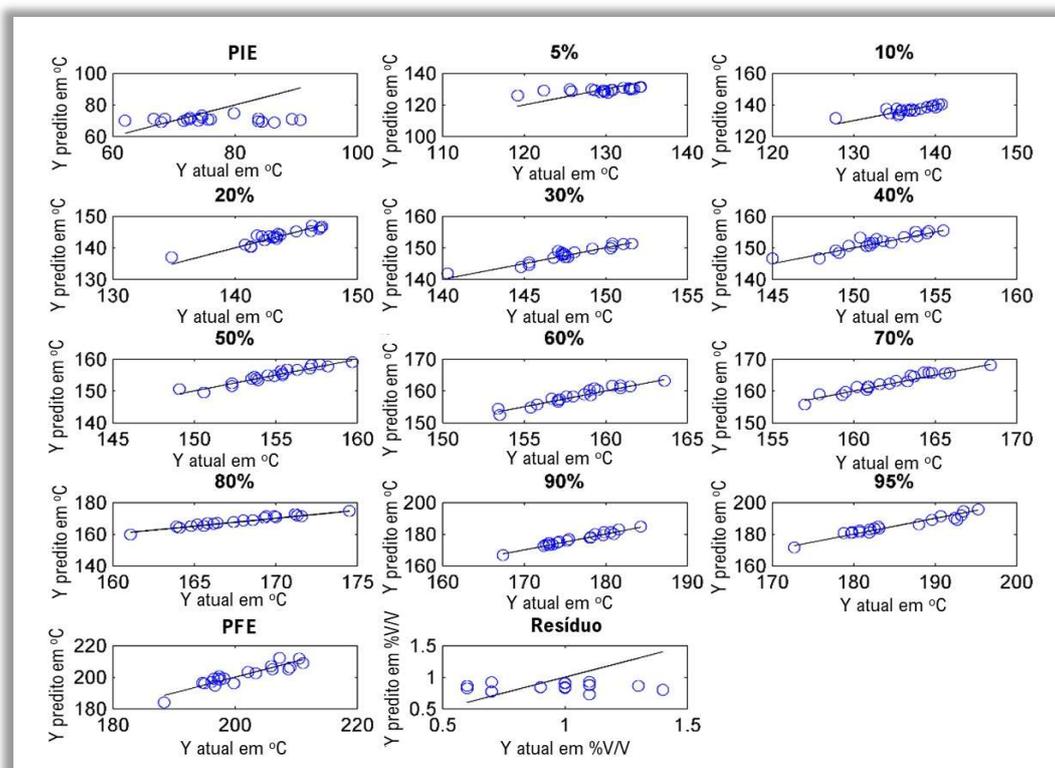


Figura 17: Gráfico dos resultados de destilação preditos versus referência para cada parâmetro de destilação para os modelos de predição obtidos por PLS.

Para os casos em que os modelos foram considerados válidos, foram obtidos valores de RMSEP e REP semelhantes aos encontrados na calibração e validação cruzada, indicando que os modelos têm a capacidade de prever amostras futuras com alta precisão. Estes mesmos modelos também tiveram valores de R^2 bem altos, próximos a 100 (ideal). O ponto de destilação 10% apresenta um resultado a ser observado para o valor de R^2 , uma vez que nas demais figuras de mérito apresentou resultados satisfatórios.

Além disso, a maioria dos modelos não apresentou erros sistemáticos para a destilação, pois o valor t_{bias} calculado foi inferior a $t_{crítico}$: 1,7291, segundo a norma ASTM E1655 ⁵¹. O valor de t_{bias} calculado ficou maior para PIE, resíduo e 80%, este último com comportamento semelhante na aplicação do MLR.

Assim como no MLR, o PLS foi incapaz de prever adequadamente os valores de PIE e Resíduo.

Para o Resíduo, tem-se uma justificativa a se destacar: ao avaliar como a análise de destilação é realizada, identifica-se a dificuldade deste ponto, pois o resíduo de destilação é uma medida manual, necessitando haver transferência do líquido não destilado do fundo do balão para uma proveta de 10 mL. O fundo do balão, além de conter este resíduo, contém pérolas de vidro. Para evitar que pérolas de vidro caiam na proveta, nem todo o resíduo é vertido. Porém o contrário também acontece: para garantir todo o volume de resíduo, pode haver transferência de pérola (s) de vidro para a proveta. Nas duas situações, a leitura contém erro intrínseco.

4.7 Comparação entre os modelos MLR e PLS

As figuras de mérito comparáveis de cada um dos modelos avaliados foram graficadas na Figura 18, a seguir, de modo a evidenciar a apresentação dos dados disponibilizados nas Tabelas 7 e 10, dos subcapítulos anteriores.

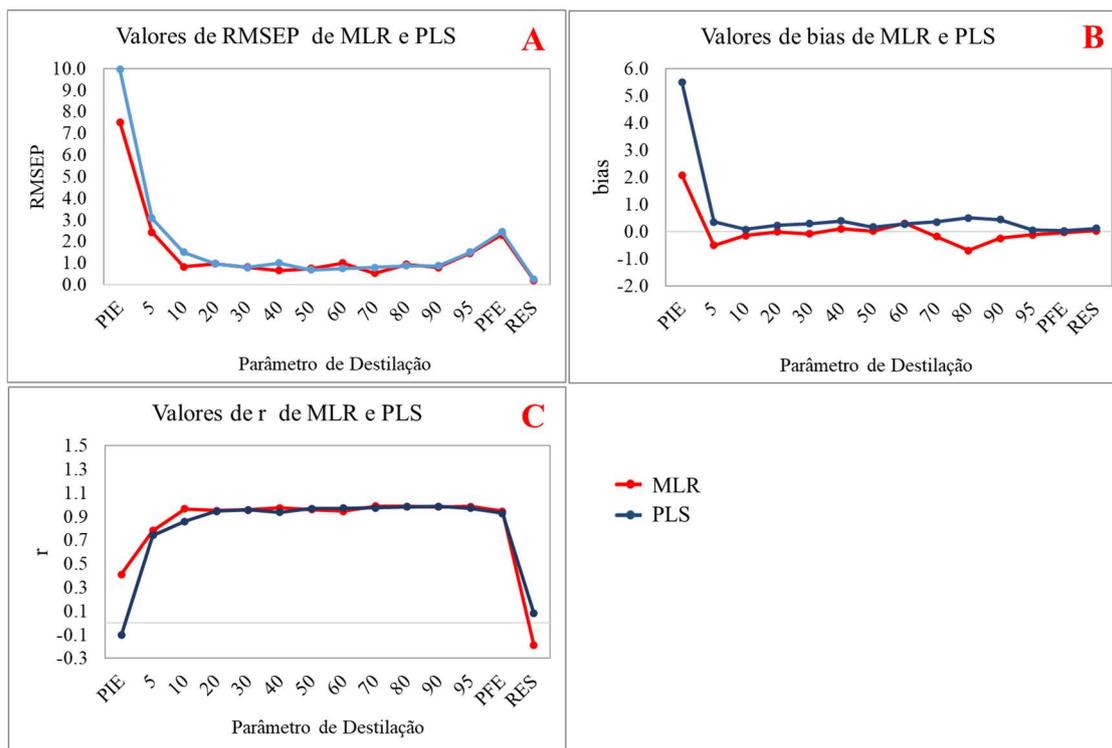


Figura 18: Figuras de mérito plotadas para MLR e PLS. Quadro A com os valores de RMSEP. Quadro B com os valores de *bias*. Quadro C com os valores de *r*.

Avaliando o que a Figura 18 apresenta, é possível observar que os valores de RMSEP obtidos por MLR ficam menores do que por PLS na maioria dos casos, ficando como exceção o ponto 60%, com o valor de RMSEP um pouco maior no MLR em relação ao PLS. Para os valores de *bias*, espera-se obter valores o mais próximo possível de zero, e nesta condição os resultados por MLR estão mais próximos deste critério, sendo exceção os pontos de destilação 5% e 80%. E para a correlação de Pearson, *r*, os valores mais próximos de 1 são os mais desejáveis, então, mais uma vez, o MLR apresentou resultados melhores ou iguais ao PLS, sendo exceção para o resíduo.

4.8 Comparação dos modelos com a Referência ASTM D86 e aplicação na indústria

As figuras de mérito apontam numericamente a validade de uma aplicação quimiométrica, porém na indústria tem-se a obrigatoriedade de aplicação de normas reconhecidas internacionalmente (ou nacionalmente, dependendo o caso), que podem ser mais ou menos restritivas que os valores apontados pelas figuras de mérito, sendo necessário avaliar a aplicação.

Como visto anteriormente no capítulo 3.2.1, os valores de repetibilidade e reprodutibilidade da norma ASTM D86 são fixos para os pontos PIE e PFE, porém calculados para os demais pontos, conforme a amostras. A Tabela 11 apresenta então os valores calculados de todos os pontos de destilação necessários para a *PYGAS* pesada.

Comparando os resultados de RMSEP do MLR e do PLS com os valores de repetibilidade e reprodutibilidade da referência, é possível evidenciar que a maioria dos resultados dos modelos estão dentro dos limites da repetibilidade, ficando apenas o PIE fora da faixa de aceitabilidade.

Tabela 11: Quadro demonstrativo entre valores de repetibilidade e reprodutibilidade do ASTM D86 de 2018 ¹ e RMSEP por MLR e PLS.

Parâmetro da Destilação	Repetibilidade °C	Reprodutibilidade °C	RMSEP MLR	RMSEP PLS
PIE	2,7	4,7	7,510	9,946
5	15,5	16,6	2,430	3,064
10	3,3	4,3	0,816	1,497
20	2,3	3,4	0,962	0,965
30	2,0	3,0	0,803	0,788
40	2,0	3,1	0,644	0,996
50	2,1	3,0	0,734	0,675
60	2,2	3,1	0,996	0,739
70	2,6	3,2	0,527	0,795
80	2,4	3,3	0,923	0,874
90	3,4	4,4	0,784	0,861
95	4,6	6,2	1,447	1,497
PFE	3,3	7,1	2,292	2,438
Resíduo			0,198	0,237

Para facilitar essa avaliação, foram elaboradas as Figuras 19 e 20, que contém gráficos com limites de repetibilidade e reprodutibilidade do ASTM D86 de 2018 ¹ e as 20 amostras de predição dos modelos de MLR e PLS.

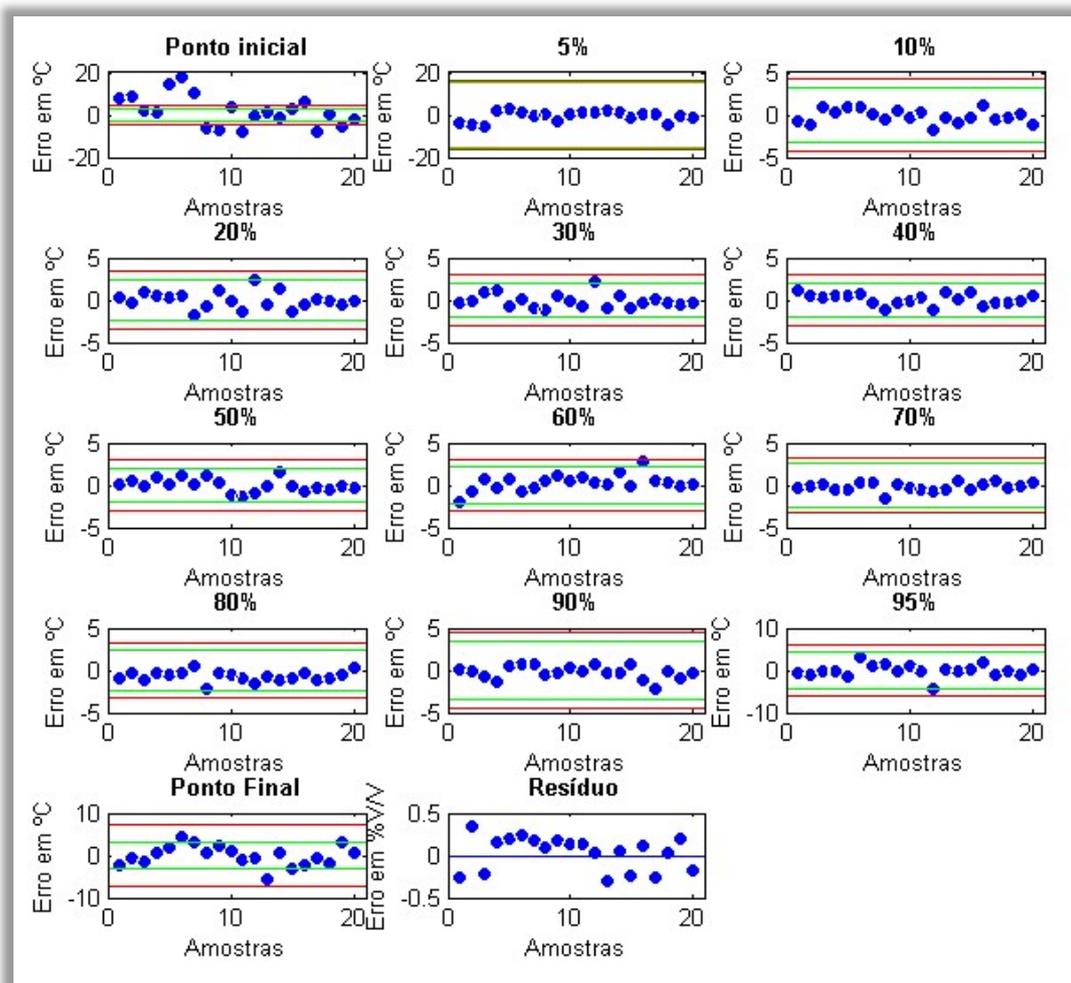


Figura 19: Resíduos do modelo MLR para o conjunto de predição, em comparação com a repetibilidade e reprodutibilidade do ASTM D86. A linha verde é a repetibilidade e a linha vermelha é a reprodutibilidade do ASTM D86 de 2018.

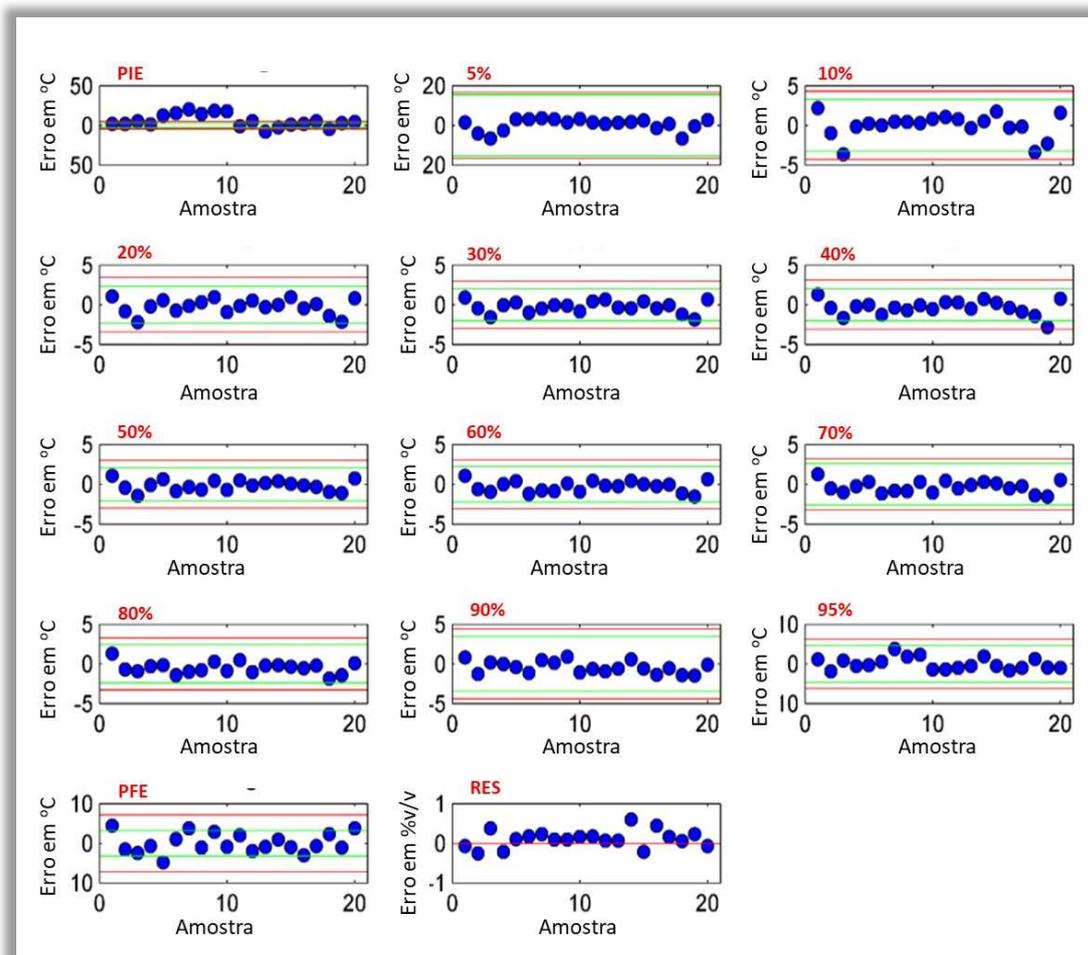


Figura 20: Resíduos do modelo PLS para o conjunto de predição, em comparação com a repetibilidade e reprodutibilidade do ASTM D86. A linha verde é a repetibilidade e a linha vermelha é a reprodutibilidade do ASTM D86 de 2018.

A destilação do ponto 5% apresentou resultados das figuras de mérito questionáveis, porém quando tem seus resultados comparados aos valores de repetibilidade e reprodutibilidade da referência, tem seu modelo validado. Os elevados valores de repetibilidade e reprodutibilidade para este ponto demonstram a complexidade dessa amostra para análise de destilação. A composição cromatográfica justifica o comportamento na destilação, pois há baixa concentração de compostos leves (hidrocarbonetos C4s e C5s) em uma mistura predominantemente pesada (HCs C8 e maiores), havendo um delta de temperatura elevado entre o ponto inicial e o 5%. A composição também explica a complexidade na determinação do PIE, que foi o único ponto que não passa nesta avaliação comparativa. O PIE tem forte influência das condições de temperatura nas quais a amostra e os materiais utilizados são mantidos,

uma vez que tem compostos voláteis como C4s e C5s. Como os dados utilizados neste trabalho são históricos de acompanhamento de processo, não foi feita qualquer seleção de dados e nem acompanhamento das análises para identificar possíveis desvios que pudessem justificar a dificuldade nas modelagens.

O resíduo de destilação não tem valores de repetibilidade e reprodutibilidade na referência, porém já foi discutido na aplicação do modelo PLS as dificuldades associadas às etapas manuais na determinação deste parâmetro, que se aplica também ao MLR.

Essa comparação dos erros dos valores preditos com a norma de referência, deixou evidente a possibilidade de aplicação dos modelos de calibração multivariada na indústria. O único ponto reprovado é o PIE, podendo inviabilizar todo este trabalho. Para o PIE aplicou-se, então um modelo utilizado pela engenharia, com uso de software Minitab, que não é objeto de estudo desta dissertação. A Figura 21 apresenta os exemplos de aplicação prática desta dissertação, na qual estão dispostos valores reais, obtidos conforme ASTM D86, e valores preditos, que são periodicamente comparados para revalidação ou atualização do modelo.

>>> Número da amostra				
>>> Ponto				
>>> Descrição	PYGAS PESADA	PYGAS PESADA	PYGAS PESADA	PYGAS PESADA
>>> Informações	-	-	-	-
>>> Data - Hora	10/08/2021 06:00	13/07/2021 06:00	15/06/2021 06:00	20/04/2021 06:00
>>> Tipo de amostra	Programada	Programada	Programada	Programada
>>> Status	Encerrado	Aprovado	Aprovado	Aprovado
DEST PIE	65	78.8	79.9	82.4
DEST 05%	128.5	126.3	123.5	126.5
DEST 10%	133.3	131.1	128	131.2
DEST 20%	138.6	137.7	133.7	138.1
DEST 30%	142.9	142.5	138.4	142.6
DEST 40%	146.4	147	143.8	147.9
DEST 50%	150	151	148.7	151.8
DEST 60%	153.5	154.9	153	155.9
DEST 70%	157.4	159.1	158.1	160.9
DEST 80%	161.8	164.6	164.5	167.3
DEST 90%	168.4	172.3	174.4	176.2
DEST 95%	174.5	179.4	182.2	185.1
DEST PFE	185.7	194.1	196.5	196.6
DEST RESIDUO	1.0	1.0	1.0	1.0
DEST PIE pred	64.2	69.2	79.1	79.2
DEST 05% pred	129.2	126.6	125.4	126.3
DEST 10% pred	134.1	132.1	127.9	132.1
DEST 20% pred	139.7	137.2	131.6	136.7
DEST 30% pred	143.3	142.1	138.1	142.2
DEST 40% pred	146.9	146.3	143.4	146.1
DEST 50% pred	149.5	150	149.5	151.6
DEST 60% pred	152.5	153.7	153.7	155.6
DEST 70% pred	156.5	158.1	157.7	159.7
DEST 80% pred	160.9	163.6	164.8	166.1
DEST 90% pred	167.4	172.1	174.7	174.4
DEST 95% pred	173.2	178.4	183.7	182.5
DEST PFE pred	186.4	193.3	197.9	197.8
DEST RESIDUO pred	1.1	1.0	0.8	1.0

Figura 21: Exemplos de resultados disponibilizados para Engenharia e Produção dos dados da amostra de *PYGAS* Pesada, com valores práticos realizados conforme ASTM D86 e valores preditos pelas equações.

5 CONCLUSÃO

A proposta deste trabalho foi obter os resultados da curva de destilação de amostras de *PYGAS* pesada a partir de resultados de cromatografia em fase gasosa, utilizando dados históricos de uma indústria petroquímica, aplicando calibração multivariada. Para tal, aplicou-se MLR e PLS, avaliando as figuras de mérito obtidas e comparando-as entre os dois métodos, bem como comparando os erros obtidos em cada ponto e método com os valores de repetibilidade e reprodutibilidade da norma de referência, o ASTM D86 de 2018.

Para o PIE, os valores de RMSEP ficaram elevados nos dois métodos, sendo 7,5 °C no MLR e 9,9 °C no PLS, e os valores preditos apresentaram erro em relação à referência acima do ASTM D86. Os dois métodos também apresentaram valores significativos de *bias*, sendo 2°C no MLR e 5°C no PLS. E a correlação de Pearson, r , valores muito baixos, de 0,4 pelo MLR e -0,1 pelo PLS. Os dois métodos se mostraram ineficientes para prever o valor de PIE em amostra de *PYGAS* pesada, tendendo a apresentar 70°C para todas as amostras.

Para destilação 5%, os valores de RMSEP ficaram superiores a 2°C pelos dois métodos, no MLR e PLS valores de r de 0,78 e 0,74, respectivamente. Pelo MLR o *bias* ficou em -0,5 °C e no PLS, 0,35°C. Apesar de valores das figuras de mérito questionáveis, ao comparar com o ASTM D86, evidencia-se que os erros da predição são muito inferiores aos valores da repetibilidade da referência, pois estes valores são elevados devido à complexidade da amostra, que apresenta poucos compostos leves.

Para destilação 10%, RMSEP do MLR apresentou resultados significativamente melhores que o PLS, sendo 0,8°C para o primeiro e 1,5 °C para o segundo. O r também apresentou resultado melhor para o MLR, sendo este de 0,96, enquanto no PLS foi de 0,85. Os dois métodos apresentaram valores não significativos de *bias*. No comparativo com a referência, o MLR apresentou todos os valores preditos com erros dentro da repetibilidade, enquanto o PLS apresentou 2 amostras acima da repetibilidade, mas abaixo da reprodutibilidade.

Para os pontos de destilação 20%, 30%, 40%, 50%, 60%, 70%, 80% e 90%, os valores obtidos para cada um dos métodos de calibração multivariada foi semelhante, com valores de RMSEP inferior a 1°C, valores de $r \geq 0,94$ e $bias \leq |0,4^\circ\text{C}|$, com exceção do ponto 80% de apresentou *bias* de -0,7 no MLR e 0,5 no PLS. Todos os

valores preditos ficaram com erros dentro da faixa de reprodutibilidade da referência, sendo que no MLR cada ponto de destilação apresentou no máximo um valor acima de repetibilidade, enquanto no PLS podem ter até dois valores acima da repetibilidade.

Para destilação 95%, os dois métodos apresentaram RMSEP de aproximadamente 1,5°C, $bias \leq |0,1^\circ\text{C}|$ e $r \geq 0,97$. Na comparação com ASTM D86, os dois métodos tiveram 19 pontos dentro da faixa de repetibilidade e apenas um ponto na faixa de reprodutibilidade.

Para PFE, todos os valores de erro ficaram dentro dos limites da reprodutibilidade do ASTM D86, com 4 valores fora dos limites de repetibilidade para o MLR e 5 para o PLS. Os valores de RMSEP ficaram semelhantes para os dois métodos, sendo 2,3°C pelo MLR e 2,4 °C para o PLS. Quanto ao *bias*, tanto MLR quanto PLS apresentaram valores inferiores a $|0,1^\circ\text{C}|$ e r de 0,94 e 0,92, respectivamente.

E finalmente para o Resíduo de Destilação, os dois métodos apresentaram valores de RMSEP de cerca de 0,2 % v/v, não sendo possível comparar os erros com a referência pois essa não tem valores de repetibilidade e reprodutibilidade para o resíduo. No MLR, o *bias* ficou em 0,03 % v/v e o r , em -0,19. No PLS o *bias* ficou em 0,11 % v/v e o r , em 0,08. O Resíduo de destilação apresentou os piores valores na figuras de mérito, sendo justificado pelo fato desta medida ser manual e apresentando erro por causa da utilização de pérolas de vidro.

Em todas as figuras de mérito, o MLR e o PLS apresentaram a mesma eficiência ou ineficiência par conseguir predizer adequadamente os valores da destilação. Em alguns pontos os valores de MLR se mostraram um pouco melhores que o PLS.

Apesar das dificuldades descritas, a estratégia aplicada mostrou-se eficiente e útil, pois reduz a probabilidade de exposição ao risco de incêndio e vapores tóxicos proporcionados pela análise de destilação, sendo substituída por uma análise que, além de mais segura, também utiliza menor quantidade de amostra. Este importante benefício reforça a necessidade de continuar avaliando a aplicação de outras ferramentas, a fim de melhorar os resultados de predição do PIE a serem aplicados na indústria.

REFERÊNCIAS BIBLIOGRÁFICAS

1. ASTM D86-20B. *Standard Test Method for Distillation of Petroleum Products and Liquid Fuels at Atmospheric Pressure*. ASTM International: West Conshohocken, PA, 2020.
2. van Leeuwen, J. A.; Jonker, R. J.; Gill, R. Octane number prediction based on gas chromatographic analysis with non-linear regression techniques. *Chem. Intell. Lab. Syst.* **1994**, 25, 325-340.
3. Flumignan, D. L.; Ferreira, F. O.; Tininis, A. G.; de Oliveira, J. E. Multivariate calibrations in gas chromatographic profiles for prediction of several physicochemical parameters of Brazilian commercial gasoline. *Chem. Intell. Lab. Syst.* **2008**, 92, 53–60.
4. de Godoy, L. A. F.; Pedroso, M. P.; Ferreira, E. C.; Augusto, F.; Poppi, R. J. Prediction of the physicochemical properties of gasoline by comprehensive two-dimensional gas chromatography and multivariate data processing. *J. Chromatogr. A*, **2011**; 1218, 1663-1667.
5. Nespeca, M. G.; Munhoz, J. F. V. L.; Flumignan, D. L.; de Oliveira, J. E.; Rapid and sensitive method for detecting adulterants in gasoline using ultra-fast gas chromatography and Partial Least Square Discriminant Analysis. *Fuel*, **2018**, 215, 204–211.
6. Mendes, G.; Aleme, H. G.; Barbeira, P. J. S. Determination of octane numbers in gasoline by distillation curves and partial. least squares regression. *Fuel*, **2012**, 97, 131–136.
7. Fakayode, S. O.; Mitchell, B. S.; Pollard, D. A. Determination of boiling point of petrochemicals by gas chromatography–mass spectrometry and multivariate regression analysis of structural activity relationship. *Talanta*, **2014**; 126, 151-156.
8. Rodrigues, E. V. A.; Silva, W.; Romão, S. R. C; Castro, E. V. R; Filgueiras, P. R. Determination of crude oil physicochemical properties by high-temperature gas chromatography associated with multivariate calibration. *Fuel*, **2018**; 220, 389–395.

9. Castaño, P.; Pawelec, B.; Fierro, J. L. G.; Arandes, J. M.; Bilbao, J.. Enhancement of pyrolysis gasoline hydrogenation over Pd-promoted Ni/SiO₂-Al₂O₃ catalysts. *Fuel*, **2007**, 86, 2262-2274.
10. Ren, T.; Patel, M.; Blok, K. Olefins from conventional and heavy feedstocks: Energy use in steam cracking and alternative processes. *Energy*, **2006**, 31,425-451.
11. Cavalcante, R.; de Faria, R.; Araújo, O. Q. F.; de Medeiros, J. Model for the First-Stage of PYGAS Upgrading: Experimental Procedure and Parameter Estimation. *Comput. Aided Chem. Eng.*, **2009**, 27, 627-632.
12. Cheng, Y. M.; Chang, J. R.; Wu, J. C. Kinetic study of pyrolysis gasoline hydrogenation over supported palladium catalyst. *Appl. Catal.*, **1986**, 24, 273-285.
13. Wang, Q.; Chen, J. Y.; Ming Pan, Chang He, Chang C. He, Bing J. Zhang, Qing L. Chen. A new sulfolane aromatic extractive distillation process and optimization for better energy utilization. *Chem. Eng. Process.*, **2018**, 128, 80-95.
14. James, A. T.; Martin, A. J. P. Gas-liquid partition chromatography. A technique for the analysis of volatile materials. *Analyst*, **1952**, 77, 915-932.
15. Blomberg, J.; Schoenmakers, P. J.; Brinkman, U. A. Th. Gas chromatographic methods for oil analysis. *J. Chromatogr. A*, **2002**, 972, 137-173.
16. McNair, H. M.; Bonelli, E. J. *Basic Gas Chromatography*, Varian Aerograph: Walnut Creek, 1969.
17. Entendendo o sistema de um Cromatógrafo Gasoso (CG). <https://www.dctech.com.br/entendendo-um-sistema-de-cromatografia-gasosa-cg/> (acesso em 04 Ago 2021).
18. Cienfuegos, F.; Vaitsman, D. *Análise Instrumental*, Interciência: Rio de Janeiro, 2000.
19. Hoh, E.; Mastovska, K. Large volume injection techniques in capillary gas chromatography. *J. Chromatogr. A*, **2008**, 1186, 2-15.
20. McWilliam, I.G; Dewar, R.A. In: Desty, D.H., Ed., *Gas Chromatography*, Butterworths: London, 1958, p. 142.

21. Yan, X. Sulfur and nitrogen chemiluminescence detection in gas chromatographic analysis. *J. Chromatogr. A*, **2002**, 976, 3–10.
22. Poole, C. F. Ionization-based detectors for gas chromatography. *J. Chromatogr. A*, **2015**, 1421, 137–153.
23. A espectrometria de massas e a química analítica. <https://revistaanalytica.com.br/a-espectrometria-de-massas-e-a-quimica-analitica>. Acesso em 04 Ago 2021.
24. Processamento de dados na espectrometria de massas. <https://revistaanalytica.com.br/processamento-de-dados-na-espectrometria-de-massas>. Acesso em 04 Ago 2021.
25. Jorgensen, A. D.; Picel, K. C.; Stamoudis, V. C. Prediction of gas chromatography flame ionization detector response factors from molecular structures. *Anal. Chem.*, **1990**; 62, 683-689.
26. Katritzky, A. R.; Ignatchenko, E. S.; Barcock, R. A.; Lobanov, V. S.; Karelson, M. Prediction of gas chromatographic retention times and response factors using a general qualitative structure-property relationships treatment. *Anal. Chem.*, **1994**; 66, 1799-1807.
27. Zhu, G.; Zhao, F.; Wang, D.; Xia, C. Extended effective carbon number concept in the quantitative analysis of multi-ethers using predicted response factors. *J. Chromatogr. A*, **2017**; 1513, 194-200.
28. Kállai, M.; Veres, Z.; Balla, J. Response of flame ionization detectors to different homologous series. *Chromatographia*, **2001**, 54, 511-517.
29. Scanlon, J.T.; Willis, D.E. Calculation of flame ionization detector relative response factors using the effective carbon number concept. *J. Chromatogr. Sci.* 1985; 23: 333-340
30. AC 8612 - Boiling Point Distribution. <https://www.paclp.com/lab-instruments/brand/ac/product/69/ac-8612-boiling-point-distribution>. Acesso em 28 Fev 2021.
31. Hydrocarbon Processing Industry Gas Chromatographs. <https://www.instrument-solutions.com/wp-content/uploads/Scion-Instruments-Hydrocarbon-Processing-Industry-Gas-Chromatographs-brochure.pdf> Acesso em 28 Fev 2021.

32. ASTM D850-21. *Standard Test Method for Distillation of Industrial Aromatic Hydrocarbons and Related Materials*. ASTM International: West Conshohocken, PA, 2016.
33. Santos, R. N. G.; Lima, E. R. A.; Paredes, M. L. L. ASTM D86 distillation curve: Experimental analysis and premises for literature modeling, *Fuel*, **2021**, 284, 1189581
34. Brereton, R. G.; *Chemometrics for Pattern Recognition*, 1 ed., Wiley: Chichester, 2009.
35. Brereton, R. G.; *Applied Chemometrics for Scientists*, 1 ed., Wiley: Chichester, 2007.
36. Schneider, M.; Vacaro, B. B.; Marcelo, M. C. A.; Ferrão, M. F.; Métodos de Análise Exploratória Aplicadas ao Ensino de Espectroscopia na Região do Infravermelho. *Rev. Virtual. Quím.*, **2018**, 10, 229-243.
37. Ferreira, M. M. C. *Quimiometria – Conceitos, Métodos e Aplicações*. Editora UNICAMP: Campinas, 2015.
38. Jiang, W.; Zhang, Z. M.; Yun, Y. O., Zhan, D. J.; Zheng, Y. B.; Liang, Y. Z.; Yang, Z. Y.; Yu, L.; Comparisons of five algorithms for chromatogram alignment; *Chromatographia*, **2013**, 76, 1067-1078.
39. Galvão, R. K. H.; Araújo, M. C. U.; Jose, G. E.; Pontes, M. J. C.; Silva, E.C.; Saldanha, T. C. B. A method for calibration and validation subset partitioning. *Talanta*, **2005**; 67, 736-740.
40. De Araújo Gomes, A.; Galvão, R. K. H.; De Araújo, M. C. U.; Vêras, G.; Da Silva, E. C. The successive projections algorithm for interval selection in PLS. *Microchem. J.*, **2003**, 110, 202–208.
41. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O. Quimiometria I: Calibração multivariada, um tutorial; *Quím. Nova*, **1999**, 22, 724-731.
42. Brereton, R. G.; Lloyd, G. R. Partial least squares discriminant analysis: taking the magic away. *J. Chemom.*, **2014**, 48, 213–225.
43. Andersson, M. A comparison of nine PLS1 algorithms. *J. Chemom.*, **2009**, 23, 518–529.

44. Haaland, D. M.; Thomas, E.V.; Partial Least-Squares Methods for Spectral Analyses: Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Anal. Chem.*, **1988**, 60, 1193–1202.
45. Instituto Nacional de Metrologia, Normalização e Qualidade Industrial (INMETRO); *Orientações sobre validação de métodos de ensaios químicos - DQO-CGRE-008*, 2003.
46. Valderrama, P.; Braga, J. W. B.; Poppi, R. J. Estado da arte de figuras de mérito em calibração multivariada. *Quim. Nova*, **2009**, 32, 1278-1287.
47. Ballabio, D. A MATLAB Toolbox for Principal Component Analysis and Unsupervised Exploration of Data Structure. *Chemom. Intell. Lab. Syst.*, **2015**, 149, 1–9.
48. Galvao, R. K. H.; Araujo, M. C. U.; Fargoso, W. D.; Silva, E. C.; Jose, G. E.; Soares, S. F. C.; Paiva, H. M.; A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm. *Chemom. Intell. Lab. Syst.*, **2008**, 92, 83-91.
49. Araujo, M. C. U.; Saldanha, T. C. B.; Galvao, R. K. H.; Yoneyama, T.; Chame, H. C.; Visani, V.; The successive projections algorithm for variable selection in spectroscopic multicomponent. *Chemom. Intell. Lab. Syst.*, **2001**, 57, 65-73.
50. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.*, **2001**, 58, 109–130.
51. ASTM E1655. *Standard Practices for Infrared Multivariate Quantitative Analysis*. ASTM International: West Conshohocken, PA, 2017.

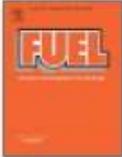
APÊNDICE A – PUBLICAÇÃO CIENTÍFICA



ELSEVIER

Fuel

Volume 303, 1 November 2021, 121298



Full Length Article

Exploring estimated hydrocarbon composition via gas chromatography and multivariate calibration to predict the pyrolysis gasoline distillation curve

Milla Beatrice Engelmann de Oliveira Garcia  , Bruna Cunha Dias, Adriano de Araújo Gomes  

Laboratório de Quimiometria e Instrumentação Analítica – LQIA, Instituto de Química da Universidade Federal do Rio Grande do Sul, Agronomia 9500, Porto Alegre, RS, Brazil

Received 25 March 2021, Revised 10 June 2021, Accepted 17 June 2021, Available online 25 June 2021.

DOI: <https://doi.org/10.1016/j.fuel.2021.121298>

