M 2021

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

ESTIMATIVA DE PROPRIEDADES DE MISTURAS DE GASOLINA E ETANOL COM BASE NA TECNOLOGIA NIR E EM CROMATOGRAFIA GASOSA

NUNO MIGUEL OLIVEIRA DIAS

DISSERTAÇÃO DE MESTRADO APRESENTADA
À FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO EM
ENGENHARIA QUÍMICA

Mestrado Integrado em Engenharia Química

Estimativa de Propriedades de Misturas de Gasolina e Etanol com base na Tecnologia NIR e em Cromatografia Gasosa

Dissertação de Mestrado

de

Nuno Miguel Oliveira Dias

Desenvolvida no âmbito da unidade curricular de Dissertação

realizado em

Petróleo de Portugal Petrogal, S.A. – Refinaria de Matosinhos



Orientador na FEUP: **Prof. Fernando Gomes Martins**

Coordenadores na Refinaria de Matosinhos: **Eng. Ricardo Albuquerque**

Dr. Jorge Ribeiro





Março de 2021

DOCUMENTO CONFIDENCIAL. UTILIZAR APENAS PARA PROPÓSITOS DE AVALIAÇÃO.

Agradecimentos

Começo por agradecer ao Professor Fernando Martins pelo tempo que disponibilizou para me orientar e ensinar, pelas ideias e sugestões e pelo que me permitiu aprender ao longo deste trabalho.

Agradeço ao Engenheiro Ricardo Albuquerque pelo acompanhamento diário, pelo interesse que sempre demonstrou e pelo conhecimento que partilhou comigo, não só a nível académico e profissional como a nível pessoal, por exemplo pelo empenho e dedicação que coloca no que faz e a constante vontade de aprender que demonstra. Agradeço, também, ao Dr. Jorge Ribeiro por me ter recebido no Laboratório de que é responsável, pelo interesse demonstrado e por se ter mostrado sempre disponível para me ajudar no que precisasse. Agradecer ainda a todas pessoas que trabalham no Laboratório da Refinaria de Matosinhos pela simpatia desde o primeiro dia, pela ajuda, pelo bom ambiente e pela boa disposição.

Obrigado a todos os meus amigos, os do secundário, os da faculdade e os de futebol, e em especial, ao Agostinho, ao André Santos, ao Bruno, ao André Sousa e ao Zé por estarem sempre presentes, me permitirem abstrair-me das preocupações e por me conseguirem fazer rir e ficar bem-disposto qualquer que seja a situação.

Obrigado à minha família por estarem sempre presentes. Obrigado à minha irmã, Daniela, e à Nadia por estarem sempre disponíveis para me ouvir, me ajudar e, essencialmente, por me aturarem sempre, com tudo de bom e de mau que isso implica. Aos meus pais pela preocupação, pelo interesse em tudo o que faço e, principalmente, pela educação e pelo exemplo que são para mim, obrigado!

Esta tese é vossa.

O Prof. Fernando Gomes Martins, orientador desta dissertação, é membro integrado do LEPABE – Laboratório de Engenharia de Processos, Ambiente Biotecnologia e Energia, financiado por: Unidade de Investigação UID/EQU/00511/2020 - Laboratório de Engenharia de Processos, Ambiente, Biotecnologia e Energia – LEPABE - financiado por fundos nacionais através da FCT/MCTES (PIDDAC).



Resumo

A otimização do processo de produção de um produto é um fator de elevada importância no sucesso de uma empresa, sendo que essa otimização pode ser conseguida, por exemplo, diminuindo a duração do processo de controlo de qualidade. Aliado a isto, na Refinaria de Matosinhos, surgiu a necessidade de conseguir uma análise rápida de gasolinas após a adição de etanol às mesmas, para voltar a determinar várias propriedades cujos valores podem ser alterados devido a essa adição. Para a caraterização de uma gasolina de modo convencional são necessárias determinar mais de 20 propriedades, sendo essa determinação bastante demorada. Assim, o objetivo deste trabalho é a construção de modelos matemáticos para a estimativa de propriedades físico-químicas de gasolinas, combinando modelos quimiométricos com a espetroscopia de infravermelho próximo e com a cromatografia gasosa, por se tratarem de técnicas de análise rápidas, aplicáveis à análise de sistemas químicos multicomponente e que, quando combinadas com os modelos matemáticos, permitem a determinação simultânea de várias propriedades das gasolinas.

Este trabalho iniciou-se com a análise de amostras de gasolinas etanolizadas para a criação de uma base de dados, tendo sido realizadas análises pelos métodos convencionais e pela espetroscopia de infravermelho próximo, para a obtenção dos valores das propriedades pretendidas (pressão de vapor, RON, MON, E70, E100, E150, FBP, teor em etanol, aromáticos, benzeno, MTBE, olefinas e oxigénio), dos dados de cromatografia gasosa e dos espetros. De seguida, foram aplicadas várias metodologias quimiométricas, incluindo a divisão de dados em conjuntos de calibração e teste, a aplicação de pré-tratamentos e posterior treino e validação, com o conjunto de calibração, de modelos de regressão para cada tipo de dados. Por último, realizou-se a previsão das propriedades das amostras do conjunto de teste usando os modelos desenvolvidos, determinaram-se os indicadores de desempenho dos modelos e realizou-se a comparação entre os valores previstos e os valores obtidos pelos métodos.

De acordo com os resultados obtidos, comprova-se que a combinação de quimiometria com a espetroscopia de infravermelho próximo e com a cromatografia gasosa são duas alternativas muito válidas que podem ser implementadas na substituição dos métodos convencionais. A utilização destas alternativas permite a redução dos custos e do tempo de análise associados ao processo de controlo de qualidade, sem prejudicar a exatidão do mesmo, otimizando os processos de controlo de qualidade e de produção da gasolina e possibilitando a análise após a adição do etanol, tal como pretendido.

Palavras Chave: Gasolina com etanol, propriedades físico-químicas, espetroscopia de infravermelho próximo, cromatografia gasosa, quimiometria.

Abstract

Optimizing the production process of a product is a highly important factor in the success of a company, and this can be achieved, for example, by shortening the duration of the quality control process. Allied to this, at the Matosinhos Refinery, the need arose to achieve a quick analysis of gasolines after the addition of ethanol to them, in order to re-determine several properties whose values may be altered due to this addition. To characterize a gasoline in a conventional way, it is necessary to determine more than 20 properties, and this determination takes a long time. Therefore, the objective of this work is the development of mathematical models for the estimation of physical-chemical properties of gasoline, combining chemometric models with near-infrared spectroscopy and gas chromatography, since these are fast analysis techniques, applicable to the analysis of multi-component chemical systems and, when combined with mathematical models, allow the simultaneous determination of several properties of gasoline.

This work began with the analysis of samples of ethanolised gasolines for the creation of a database. Analysis was carried out using conventional methods and near-infrared spectroscopy to obtain the values of the desired properties (vapor pressure, RON, MON, E70, E100, E150, FBP, ethanol content, aromatics, benzene, MTBE, olefins, and oxygen), the gas chromatography data and the spectra. Various chemometric methodologies were then applied, including the division of data into calibration and test sets, the application of pre-treatments and the subsequent train and validation of regression models through the calibration set for each type of data. Finally, the sample properties of the test set were predicted using the developed models, the performance indicators of the models were determined, and a comparison was made between the predicted values and the values obtained by conventional methods.

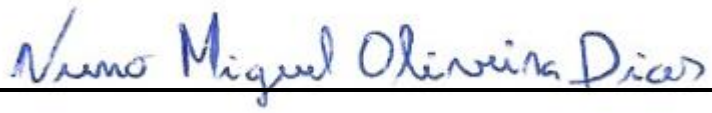
According to the results obtained, the combination of chemometrics with near-infrared spectroscopy and gas chromatography are two valid alternatives that can be implemented to replace conventional methods. The use of these alternatives allows the reduction of costs and analysis time associated with the quality control process, without compromising its accuracy, optimizing the quality control and production processes of petrol and enabling analysis after the addition of ethanol, as intended.

Keywords: Gasoline with ethanol, physicochemical properties, near-infrared spectroscopy, gas chromatography, chemometrics.

Declaração

Declara, sob compromisso de honra, que este trabalho é original e que todas as contribuições não originais foram devidamente referenciadas com identificação da fonte.

Porto, 8 de fevereiro de 2021



(Nuno Miguel Oliveira Dias)

Índice

1 Introdução 1

1.1 Enquadramento e Apresentação do Projeto 1

1.2 Apresentação da Empresa 3

1.3 Contributos do Autor para o Trabalho 3

1.4 Organização da Tese 3

2 Contexto e Estado da Arte 4

2.1 Espetroscopia de Infravermelho Próximo (NIR) 4

2.2 Cromatografia Gasosa 7

2.3 Quimiometria 8

2.3.1 Modelos de Regressão 9

2.3.2 Quimiometria e Espetroscopia NIR 10

2.3.3 Quimiometria e Cromatografia Gasosa 11

3 Materiais e Métodos 13

3.1 Técnicas de Caraterização Físico-Química 13

3.1.1 Pressão de Vapor – EN 13016-1 13

3.1.2 Curva de Destilação – ISO 3405 14

3.1.3 RON e MON – ISO 5164 e ISO 5163 15

3.1.4 Cromatografia – ISO 22854 17

3.2 Métodos Espetroscópicos - NIR 19

3.3 Quimiometria 19

3.3.1 Divisão de Dados 20

3.3.2 Pré-tratamento de Dados 21

3.3.3 Modelos de Regressão 23

3.3.4 Indicadores de Desempenho dos Modelos 27

3.3.5 Validação 28

4 Resultados e Discussão 30

4.1 Dados 30

4.2 Divisão e Pré-tratamento dos Dados 32

4.3 Construção dos Modelos 33

4.4 Previsão de Propriedades com base em Espetroscopia NIR 34

4.4.1 Pressão de Vapor 34

4.4.2 RON e MON 36

4.4.3 Caraterísticas da Destilação 38

4.4.4 Composição da Gasolina 39

4.5 Previsão de Propriedades com base em Cromatografia Gasosa 40

4.5.1 Pressão de Vapor 40

4.5.2 RON e MON 42

4.5.3 Caraterísticas da Destilação 43

5 Conclusões 45

6 Avaliação do Trabalho Realizado 47

6.1 Objetivos Realizados 47

6.2 Trabalho Futuro 47

6.3 Apreciação Final 47

7 Referências 48

Apêndice A - Tabelas e Gráficos de Resultados 51

A.1 RON e MON com base em Espetroscopia NIR 51

A.2 Caraterísticas da Destilação com base em Espetroscopia NIR 54

A.3 Composição da Gasolina com base em Espetroscopia NIR 60

A.4 Tabela de Resultados da Cromatografia Gasosa 68

A.5 RON e MON com base em Cromatografia Gasosa 71

A.6 Caraterísticas da Destilação com base em Cromatografia Gasosa 74

Lista de Figuras

Figura 2.1 – Exemplo de um espetro de NIR de uma gasolina. 6

Figura 3.1 - Representação esquemática do aparelho de medição da pressão de vapor. 14

Figura 3.2 – Representação esquemática de um aparelho de destilação atmosférica. 15

Figura 3.3 - Representação de um carburador utilizado na determinação do RON e MON (International Organization for Standardization, 2014). 17

Figura 3.4 – Representação esquemática de um cromatógrafo gasoso (International Organization for Standardization, 2016). 18

Figura 3.5 – Esquema resumo das várias etapas do processo de criação de modelos. 20

Figura 3.6 - Representação gráfica da divisão de dados realizada pela SVR (Leal et al., 2020). 26

Figura 4.1 - Exemplos de espetros NIR de gasolinas. 34

Figura 4.2 - Comparação entre os valores da previsão e os valores do método para a pressão de vapor (NIR). 36

Figura 4.3 - Comparação entre os valores da previsão e os valores do método para o RON (NIR). 37

Figura 4.4 - Comparação entre os valores da previsão e os valores do método para o E100 (NIR). 38

Figura 4.5 - Comparação entre os valores da previsão e os valores do método para o teor em etanol. 40

Figura 4.6 - Comparação entre os valores da previsão e os valores do método para a pressão de vapor (GC). 42

Figura 4.7 - Comparação entre os valores da previsão e os valores do método para o RON (GC). 43

Figura 4.8 - Comparação entre os valores da previsão e os valores do método para o E100 (GC). 44

Figura A.1 - Comparação entre os valores da previsão e os valores do método para o MON (NIR). 53

Figura A.2 - Comparação entre os valores da previsão e os valores do método para o E70 (NIR). 58

Figura A.3 - Comparação entre os valores da previsão e os valores do método para o E150 (NIR). 58

Figura A.4 - Comparação entre os valores da previsão e os valores do método para o FBP (NIR). 59

Figura A.5 - Comparação entre os valores da previsão e os valores do método para o teor em aromáticos. 66

Figura A.6 - Comparação entre os valores da previsão e os valores do método para o teor em benzeno. 66

Figura A.7 - Comparação entre os valores da previsão e os valores do método para o teor em MTBE.. 67

Figura A.8 - Comparação entre os valores da previsão e os valores do método para o teor em olefinas. 67

Figura A.9 - Comparação entre os valores da previsão e os valores do método para o teor em oxigénio. 68

Figura A.10 - Comparação entre os valores da previsão e os valores do método para o MON (GC). 73

Figura A.11 - Comparação entre os valores da previsão e os valores do método para o E70 (GC)..... 78

Figura A.12 - Comparação entre os valores da previsão e os valores do método para o E150 (GC). 78

Figura A.13 - Comparação entre os valores da previsão e os valores do método para o FBP (GC). 79

Lista de Tabelas

Tabela 1.1 – Especificações das propriedades das gasolinas 2

Tabela 2.1 - Sinais de absorção no espetro NIR de alguns dos principais compostos presentes em produtos petrolíferos 6

Tabela 2.2 – Exemplos de metodologias quimiométricos 9

Tabela 3.1 – Condições de operação dos ensaios de determinação de RON e MON 16

Tabela 3.2 – Parâmetros estatísticos de avaliação dos modelos. 28

Tabela 4.1 - Tabela resumo das principais caraterísticas da base de dados para NIR 31

Tabela 4.2 - Tabela resumo das principais caraterísticas da base de dados para CG 32

Tabela 4.3 – Tabela de resultados da previsão da pressão de vapor (NIR) 35

Tabela 4.4 – Limites de repetibilidade de reprodutibilidade para o RON e o MON 36

Tabela 4.5 - Tabela dos melhores resultados da previsão de RON e MON (NIR) 37

Tabela 4.6 - Tabela dos melhores resultados da previsão das características de destilação (NIR) 38

Tabela 4.7 - Limites de repetibilidade e reprodutibilidade das propriedades de cromatografia gasosa 39

Tabela 4.8 - Tabela dos melhores resultados da previsão de propriedades de cromatografia gasosa .. 39

Tabela 4.9 – Tabela de resultados da previsão da pressão de vapor (GC) 41

Tabela 4.10 - Tabela dos melhores resultados da previsão de RON e MON (GC) 42

Tabela 4.11 - Tabela dos melhores resultados da previsão das caraterísticas da destilação (GC) 43

Tabela A.1 - Tabela de resultados da previsão do RON (NIR) 51

Tabela A.2 - Tabela de resultados da previsão do MON (NIR) 52

Tabela A.3 - Tabela de resultados da previsão do E70 (NIR) 54

Tabela A.4 - Tabela de resultados da previsão do E100 (NIR) 55

Tabela A.5 - Tabela de resultados da previsão do E150 (NIR) 56

Tabela A.6 - Tabela de resultados da previsão do FBP (NIR) 57

Tabela A.7 - Tabela de resultados da previsão do teor em aromáticos 60

Tabela A.8 - Tabela de resultados da previsão do teor em benzeno 61

Tabela A.9 - Tabela de resultados da previsão do teor em etanol 62

Tabela A.10 - Tabela de resultados da previsão do teor em MTBE 63

Tabela A.11 - Tabela de resultados da previsão do teor em olefinas 64

Tabela A.12 - Tabela de resultados da previsão do teor em oxigénio 65

Tabela A.13 - Tabela de resultados de cromatografia gasosa (Parte 1) *69*

Tabela A.14 - Tabela de resultados de cromatografia gasosa (Parte 2) *70*

Tabela A.15 - Tabela de resultados da previsão do RON (GC)..... *71*

Tabela A.16 - Tabela de resultados da previsão do MON (GC) *72*

Tabela A.17 - Tabela de resultados da previsão do E70 (GC) *74*

Tabela A.18 - Tabela de resultados da previsão do E100 (GC) *75*

Tabela A.19 - Tabela de resultados da previsão do E150 (GC) *76*

Tabela A.20 - Tabela de resultados da previsão do FBP (GC) *77*

Notação e Glossário

A_{PG}	Área de Sinal Corrigido do Grupo Parcial	
b	Coeficiente de Regressão	
C	Matriz de Pesos de y	
C_A	Concentração do Composto Absorvente	$mol.L^{-1}$
c	Velocidade da Luz no Vazio	$m.s^{-1}$
E	Matriz de Erros Residuais de X	
E_P	Energia dos Fotões	J
e	Vetor de Erro	
F	Vetor de Erros Residuais de y	
$F_{RR,\ PG}$	Fator de Resposta Relativo do Composto/Grupo	
G	Números de Grupos de Validação Cruzada	
G^0	Energia Livre de Gibbs	
h	Constante de Planck	J.s
I	Intensidade da Luz Transmitida	cd
I_0	Intensidade da Luz Incidente	cd
K_D	Constante de Distribuição	
k'	Fator de Capacidade	
l	Distância Percorrida pela Luz no Meio Absorvente	cm
M_i	Massa Atómica do Elemento i	$g.mol^{-1}$
n_i	Número de Átomos do Elemento	
P	Matriz de Cargas de X	
R	Constante Molar dos Gases Ideais	$J.mol^{-1}\text{-}K^{-1}$
T	Matriz de Pontuações de X	
t_0	Tempo Morto	
t_r	Tempo de Retenção	
T_λ	Transmitância	
U	Vetor de Pontuações de y	
X	Matriz de Variáveis de Entrada	
X	Variável de Entrada/Vetor de Variáveis de Entrada	
y	Variável Resposta/ Vetor de Variáveis Resposta	
Z	Componente Principal	

Letras gregas

α	Fator de Separação	
ε_A	Coeficiente de Absorção Molar	$L.mol^{-1}.cm^{-1}$
λ	Comprimento de Onda	nm
v	Frequência da Luz	Hz
ρ_{PG}	Massa Volúmica do Grupo Parcial	$kg.m^{-3}$
w	Concentração Mássica do Elemento/Grupo de Hidrocarbonetos	

Índices

i	índice ou contador

Lista de Siglas

AOTF	Filtros Sintonizáveis Acústico-Óticos
ASVP	Pressão de Vapor do Ar Saturado
CFR	*Cooperative Fuel Research*
CV	Validação Cruzada
DVPE	Pressão de Vapor Seco Equivalente
E70	Percentagem Evaporada a 70 ºC
E100	Percentagem Evaporada a 100 ºC
E150	Percentagem Evaporada a 150 ºC
FBP	Ponto Final de Ebulição

GC	Cromatografia Gasosa
ISO	*International Organization for Standardization*
MLR	Regressão Linear Múltipla
MON	*Motor Octane Number*
MTBE	Éter metil-terc-butílico
MSC	Correção de Dispersão Multiplicativa
MSE	Erro Quadrático Médio
NIR	Infravermelho-próximo
PC	Componente Principal
PG	Grupo Parcial
PCA	Análise por Componentes Principais
PLS	Mínimos Quadrados Parciais
R^2	Coeficiente de Determinação
RMSE	Raiz do Erro Quadrático Médio
RON	*Research Octane Number*
RVP	Pressão de Vapor de Reid
SG	Savitzky-Golay
SNV	Variação Padrão
SS	Normalização Padrão
SVM	Máquinas de Vetores de Suporte
SVR	Máquinas de Vetores de Suporte aplicadas à Regressão

1 Introdução

1.1 Enquadramento e Apresentação do Projeto

Este trabalho insere-se no processo de otimização da produção de gasolinas com incorporação de etanol, mais especificamente, no desenvolvimento de metodologias analíticas que permitam caracterizar, de forma rápida e rigorosa, várias propriedades das gasolinas, importantes para controlar o processo de adição. A incorporação, em linha, de etanol nas gasolinas exige um controlo rigoroso de modo a garantir que as especificações do produto são cumpridas garantindo, ao mesmo tempo, o mínimo *give away*. Assim, o desenvolvimento de técnicas analíticas que permitam obter resultados rápidos revela-se de importância determinante.

O controlo de qualidade de combustíveis, como as gasolinas, é regulado por uma norma europeia, *EN 228:2008*, e os ensaios para caraterização das mesmas efetuam-se de acordo com os métodos analíticos referidos na respetiva legislação e que seguem normativos internacionais (Organização Internacional de Normalização (ISO)). A caraterização das gasolinas envolve a determinação de mais de 20 propriedades físico-químicas diferentes, sendo que muitas dessas determinações são demoradas e necessitam de quantidades de amostra relevantes. O facto deste controlo ser feito não só ao produto final, mas também durante a produção exige tempo e recursos significativos pelo que se justifica encontrar metodologias alternativas fiáveis que permitam otimizar este processo.

A gasolina é uma mistura complexa de hidrocarbonetos, obtida a partir do crude e que, normalmente, contém entre 150 a 200 compostos orgânicos. Os hidrocarbonetos presentes na gasolina têm de 4 a 12 átomos de carbono nas suas estruturas, com pontos de ebulição entre 0 e 215 ºC, aproximadamente, e podem ser divididos por 4 categorias: parafinas, olefinas, nafténicos e aromáticos. Além destes compostos, as gasolinas podem conter outros componentes como etanol, éteres (por exemplo, MTBE), entre outros (Speight, 2015).

As propriedades das gasolinas são avaliadas de forma a garantir não só um bom desempenho do combustível, mas também a segurança no seu manuseamento e armazenamento. Desse modo, existem valores limite, superiores e/ou inferiores, definidos e que têm de ser cumpridos para que as gasolinas possam ser comercializadas. Os valores limite de especificação de várias propriedades da gasolina podem ser consultados na Tabela 1.1.

Tabela 1.1 – Especificações das propriedades das gasolinas

	Unidades	Valor Mínimo	Valor Máximo
Pressão de Vapor			
Inverno	kPa	60	90
Verão	kPa	-	60
Abril/Outubro	kPa	45	90
RON		95	-
MON		85	-
Destilação			
Evaporado a 70 ºC (E70)			
Inverno	% V/V	22,0	50,0
Verão	% V/V	20,0	48,0
Abril/Outubro	% V/V	20,0	50,0
Evaporado a 100 ºC (E100)	% V/V	46,0	-
Evaporado a 150 ºC (E150)	% V/V	75,0	-
Ponto final de ebulição (FBP)	ºC	-	210,0
Aromáticos	% V/V	-	35,0
Benzeno	% V/V	-	1,0
Etanol	% V/V	-	10,0
Olefinas	% V/V	-	18,0
Oxigénio	% m/m	-	2,7
Éteres, com nº de átomos de carbono > 5	% V/V	-	22,0

A identificação das propriedades físico-químicas das gasolinas pode ser alcançada a partir da identificação dos compostos presentes na mistura, uma vez que existe relação entre as propriedades e a sua composição química. Tendo isto em conta, as técnicas escolhidas para servir de base à previsão das propriedades foram a espetroscopia NIR e a cromatografia gasosa, por se tratar de duas técnicas que necessitam de uma pequena quantidade de amostra, são aplicáveis à análise de sistemas químicos multicomponente, requerem pouco tempo de ocupação de analista e, no caso do NIR, envolve um tempo de análise reduzido.

O objetivo deste trabalho é a construção de modelos matemáticos para a previsão, de forma rápida, de propriedades físico-químicas de gasolinas. Regressão por componentes principais

(PCR), regressão por mínimos quadrados parciais (PLS) e máquinas de vetores de suporte aplicadas à regressão (SVR) foram os métodos estatísticos utilizados no desenvolvimento desses modelos que se baseiam em dados espetroscópicos e cromatográficos. Estes modelos foram utilizados para a previsão das seguintes propriedades da gasolina: pressão de vapor, número de octano de pesquisa (RON) e de motor (MON), características da destilação (E70, E100, E150 e FBP), teor em etanol, aromáticos, benzeno, olefinas, oxigénio, e éter metil-terc-butílico.

1.2 Apresentação da Empresa

A Refinaria de Matosinhos iniciou a sua atividade em 1969, é de natureza *hydroskimming* e está inserida na categoria das refinarias de especialidades. Trata-se de uma refinaria com elevada flexibilidade produtiva, resultando numa vasta gama de produtos comerciais refinados e em importantes matérias primas para a indústria química e petroquímica. Atualmente, emprega de forma direta, cerca de 400 trabalhadores, sendo também uma fonte geradora de emprego de forma indireta. Tem capacidade de processamento de 110 mil barris de petróleo por dia e engloba no seu complexo industrial quatro fábricas de produtos (Combustíveis, Aromáticos e Solventes, Lubrificantes e Óleos Base) e uma fábrica de utilidades, interligadas entre si. Conta com interligação ao terminal petroleiro de Leixões e ao aeroporto Francisco Sá Carneiro.

1.3 Contributos do Autor para o Trabalho

Nas empresas existe uma constante necessidade de evolução e de otimização dos seus processos, sendo a redução do tempo necessário à sua realização um dos pontos chave do sucesso e garantia de competitividade. Com este trabalho, pretendeu-se criar ferramentas que permitam prever as várias propriedades de gasolinas etanolizadas com recurso à combinação de métodos experimentais, como o NIR e a cromatografia gasosa, com técnicas estatísticas, de forma a criar modelos que nos permitam prever as propriedades desejadas, sem a necessidade de realizar um elevado número de ensaios, possibilitando uma análise rápida da gasolina após a adição de etanol e reduzindo o tempo necessário à análise de gasolinas na sua globalidade.

1.4 Organização da Tese

Este trabalho encontra-se dividido em 6 capítulos. No primeiro capítulo descreve-se o problema em estudo e os principais objetivos deste trabalho. No segundo capítulo abordam-se os principais fundamentos teóricos, assim como o estado atual das técnicas usadas neste trabalho. Os procedimentos experimentais das várias técnicas usadas neste trabalho apresentam-se no terceiro capítulo. O quarto capítulo apresenta e discute os resultados e as conclusões retiradas dessa análise são enunciadas no quinto capítulo. Por último, aborda-se a avaliação em relação ao cumprimento dos objetivos previamente definidos e as perspetivas futuras do trabalho.

2 Contexto e Estado da Arte

De forma a permanecer competitivas, as empresas devem estar constantemente à procura de formas de otimizar os seus processos e de aumentar os lucros. No caso de uma refinaria, essa procura por novas soluções tem como objetivos aumentar a eficiência processual, maximizar a produtividade e reduzir os custos de produção. Nesse sentido, um fator importante é a otimização do tempo de produção, que pode ser conseguida, por exemplo, diminuindo a duração do processo de controlo de qualidade. A otimização deste processo revela elevada importância porque de modo a caraterizar um produto, por exemplo uma gasolina, são necessárias medir mais de 20 propriedades físico-químicas, sendo que grande parte dessas determinações é demorada, dispendiosa e necessita de grandes quantidades de amostra. Além disso, na Refinaria de Matosinhos, surgiu a necessidade de conseguir uma análise rápida de gasolinas após a adição de etanol às mesmas, para voltar a determinar várias propriedades cujos valores podem ser alterados devido a essa adição, tornando mais urgente a implementação de metodologias que permitam determinar as várias propriedades em simultâneo e com reduzido tempo de análise.

2.1 Espetroscopia de Infravermelho Próximo (NIR)

O princípio da espetroscopia NIR baseia-se na absorção de radiação eletromagnética por parte dos eletrões da matéria. Assim, um espetro de NIR é o resultado das interações existentes entre ondas eletromagnéticas e a matéria em análise, como, por exemplo, a interação existente entre a radiação e as vibrações das ligações químicas intramoleculares. Este tipo de espetroscopia interage, principalmente, com compostos orgânicos e água, e vem sendo implementado com sucesso como um método rápido de Controlo de Qualidade em diferentes indústrias (Gauglitz e Moore, 2014). No âmbito deste tipo de aplicações, as vantagens desta análise face aos métodos convencionais, residem no facto de ser uma análise que despende menos tempo, é necessária uma quantidade menor de amostra e, quando devidamente processados, os dados fornecidos pelos espetros de NIR permitem caraterizar diversas propriedades com apenas uma medição. Apresenta ainda outras vantagens, como o facto de ser necessária pouca ou nenhuma preparação da amostra. Por outro lado, a principal desvantagem reside no facto da seletividade natural do NIR ser baixa devido à sobreposição de várias bandas (Eldin, 2011). Do ponto de vista da aplicabilidade industrial esta técnica apresenta vantagens como o facto de poder ser implementada em locais diferentes ao longo do processo e de, como já foi referido, não ser necessária uma preparação individual da amostra. Além destes dois fatores, existem outros critérios práticos que influenciam favoravelmente a escolha do espetrofotómetro NIR, como a rapidez de resposta, o custo e a portabilidade (Gauglitz e Moore, 2014).

Teoricamente, existem duas formas de descrever a radiação eletromagnética: como uma onda eletromagnética ou como uma corrente de fotões. No caso da onda eletromagnética, a interação entre a onda e as moléculas descreve-se com base na teoria eletromagnética composta pelas equações de Maxwell. A intensidade da luz transmitida (I) pode ser descrita de acordo com a Lei de Beer-Lambert, como mostra a Equação 2.1, onde I_0 representa a intensidade da luz incidente, ε_A é o coeficiente de absorção molar, C_A é a concentração do composto absorvente e l é a distância percorrida pela luz no meio absorvente.

$$I = I_0 \times e^{-\varepsilon_A C_A l} \tag{2.1}$$

A Equação 2.1 também pode ser escrita na forma logarítmica, tal como está expresso na Equação 2.2, em que T_λ é a transmitância ao comprimento de onda λ.

$$\log\left(\frac{I}{I_0}\right) = \log\left(\frac{1}{T_\lambda}\right) = -\varepsilon_A C_A l \tag{2.2}$$

No segundo caso, recorre-se ao uso das leis de ótica quântica para descrever o comportamento dos fotões e a sua interação com as moléculas. A energia dos fotões calcula-se a partir da Equação 2.3, na qual h representa a constante de Planck ($h = 6{,}626 \times 10^{-34}$ J.s) e v é a frequência da luz.

$$E_p = h \times v \tag{2.3}$$

A velocidade da luz no vazio (c) e o comprimento de onda (λ) estão relacionados entre si tal como é expresso pela Equação 2.4 (Gauglitz e Moore, 2014).

$$v = \frac{c}{\lambda} \tag{2.4}$$

A espetroscopia NIR vem sendo muito usada na análise de produtos de refinação de petróleos e petroquímica, porque se trata de uma técnica rápida, não destrutiva e que abrange uma gama alargada de números de onda (4 000 cm^{-1} a 12 500 cm^{-1}). Os sinais de absorção de alguns dos principais compostos presentes em produtos petrolíferos estão expressos na Tabela 2.1. As absorções no NIR dependem de vibrações harmónicas, que ocorrem quando um modo vibracional é excitado e a mudança no número quântico é diferente de ±1, e de bandas vibracionais, que se verificam quando um fotão excita dois modos vibracionais ou mais. Esta técnica de análise é extremamente sensível a mudanças de temperatura, que influenciam as bandas de absorção dos grupos funcionais. Tal como já foi referido, o espetro resultante de uma análise NIR é bastante complexo e é necessário o uso de técnicas quimiométricas para extrair informação relevante do espetro (Gauglitz e Moore, 2014). Na Figura 2.1 encontra-se representado um exemplo de um espetro típico de uma análise NIR.

Tabela 2.1 - Sinais de absorção no espetro NIR de alguns dos principais compostos presentes em produtos petrolíferos

Ligação	C-H				O-H
Grupo	Hidrocarbonetos Alifáticos	Hidrocarbonetos Aromáticos	Hidrocarbonetos Olefínicos		Álcoois e Fenóis
Gama de Números de Onda (cm^{-1})	9100-8000, 2º harmónico de extensão CH	≈8700, 2º harmónico de extensão CH	Ligações Duplas *cis*	≈8400, 2º harmónico de extensão CH	≈7150, 1º harmónico de extensão CH
	7700-6900, banda combinatória	7100-6900, banda combinatória		≈5900, 1º harmónico de extensão CH	≈10000, 2º harmónico de extensão CH
	6300-5500, 1º harmónico de extensão CH	≈5900, 1º harmónico de extensão CH		≈4500; ≈4600, banda combinatória	≈5000, banda combinatória
	5000-4100, banda combinatória	≈4000; ≈4600, banda combinatória	Ligações Duplas Terminais	≈6100, 1º harmónico de extensão CH	
				≈4700, banda combinatória	

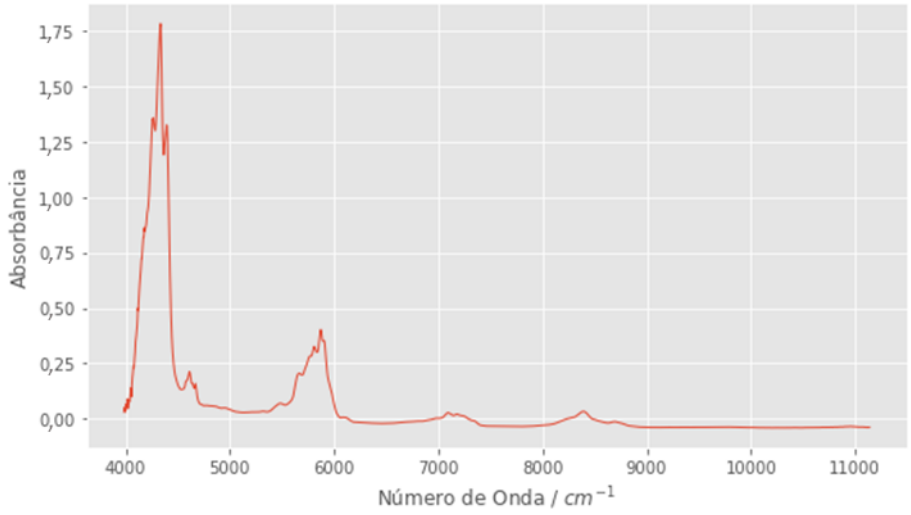


Figura 2.1 – Exemplo de um espetro de NIR de uma gasolina.

A configuração habitual de um espetrofotómetro NIR inclui uma fonte de luz, um monocromador, detetores de reflexão difusa e um detetor de transmitância. A fonte de luz é, normalmente, uma lâmpada de tungsténio ou de tungsténio-halogéneo. No caso do monocromador, existem várias opções, umas com partes móveis, como os espetrofotómetros FT-NIR, e outras sem partes móveis, como é o caso dos filtros sintonizáveis acústico-óticos

(AOTF). As caraterísticas mais importantes de um espetrofotómetro NIR são: a exatidão e a precisão do comprimento de onda, a exatidão, a precisão e a linearidade fotométrica, a resolução e a sensibilidade ao ruído (Burns e Ciurczak, 2008).

2.2 Cromatografia Gasosa

A cromatografia gasosa (GC) é uma técnica de separação que se baseia nas diferenças no ponto de ebulição ou na pressão de vapor de ebulição, e na polaridade dos compostos para realizar a separação (Stauffer et al., 2019). É uma técnica com capacidade de tratar misturas com elevada complexidade, sendo por isso um método analítico usual em várias áreas, tais como a indústria química e petroquímica. Tem como vantagens a precisão e exatidão dos ensaios aliadas à elevada capacidade de separação. Além disso, o elevado número de fases estacionárias, os detetores com diferentes sensibilidades específicas e as técnicas de alta temperatura e alta resolução que possibilitam a análise de macro- e microcomponentes são, também, vantagens da cromatografia gasosa. Por outro lado, a deterioração da coluna e a contaminação da mesma podem causar alguns problemas à análise como instabilidade na linha de base e pequenas modificações no tempo de retenção, constituindo-se como as principais desvantagens desta técnica (Forgács e Cserháti, 2003).

Existem alguns parâmetros que permitem caracterizar a separação como as constantes de distribuição (K_D), o fator de capacidade (k'), o fator de separação (α) e o número de resolução (R). As constantes de distribuição (K_D) quantificam a razão entre a concentração do composto na fase estacionária e na fase móvel, como descrito na Equação 2.5.

$$K_D = \frac{Concentração\ do\ Composto\ na\ Fase\ Estacionária}{Concentração\ do\ Composto\ na\ Fase\ Móvel} \tag{2.5}$$

A distribuição pode ser descrita, de forma generalizada, pela Equação 2.6, em que G^0 é a energia livre de Gibbs, R é a constante dos gases ideais e T é a temperatura da coluna.

$$\ln(K_D) = \frac{-G^0}{R\,T} \tag{2.6}$$

O fator de capacidade (k') é o tempo que um composto permanece na fase estacionária em relação à fase móvel, e é calculado a partir do tempo de retenção (t_r) e do tempo morto (t_0). O primeiro, tempo de retenção (t_r), é o intervalo de tempo no qual um composto se encontra na fase estacionária, enquanto que o segundo, tempo morto (t_0), é o tempo necessário para que um composto não retido viaje ao longo da coluna. A Equação 2.7 expressa a forma de cálculo do fator de capacidade.

$$k' = \frac{t_r - t_0}{t_0} \tag{2.7}$$

O fator de separação (α) é a razão entre a razão de partição do último composto a eluir (k_2) e a razão de partição dos compostos que eluem anteriormente (k_1), e expressa-se pela Equação 2.8.

$$\alpha = \frac{k_2}{k_1} \tag{2.8}$$

Por último, o número de resolução (R) descreve a eficácia de separação de dois picos vizinhos. Este parâmetro pode calcular-se de duas formas diferentes, expressas nas Equações 2.9 e 2.10, em que t_{R1} e t_{R2} são os tempos de retenção dos picos 1 e 2, respetivamente, w_{h1} e w_{h2} são as larguras dos picos 1 e 2 a meio da altura dos mesmos, respetivamente, e w_{b1} e w_{b2} são as larguras dos picos 1 e 2 a na base dos mesmos, respetivamente (Forgács e Cserháti, 2003).

$$R = \frac{1{,}18 \cdot (t_{R2} - t_{R1})}{(w_{h1} + w_{h2})} \tag{2.9}$$

$$R = \frac{2 \cdot (t_{R2} - t_{R1})}{(w_{b1} + w_{b2})} \tag{2.10}$$

Os resultados da cromatografia podem apresentar-se sob a forma de % (m/m) ou % (V/V) de um grupo ou composto. Para se obter esses resultados é necessário calcular, previamente, o fator de resposta relativo ($F_{RR,PG}$) de cada composto/grupo. A Equação 2.11 expressa a fórmula de cálculo desse fator, em que M_C representa a massa atómica do carbono, n_C o número de átomos de carbonos do composto, M_H a massa atómica do hidrogénio e n_H o número de átomos de hidrogénio do composto.

$$F_{RR,PG} = \frac{[(M_C \cdot n_C) + (M_H \cdot n_H)] \cdot 0{,}7487}{M_C \cdot n_C} \tag{2.11}$$

Após o cálculo deste fator, é possível calcular a %(m/m) do grupo, de acordo com a Equação 2.12, em que w_{PG} é a %(m/m) grupo e A_{PG} é a área de sinal corrigido do grupo parcial (PG).

$$w_{PG} = \frac{100 \cdot A_{PG} \cdot F_{RR,PG}}{\sum_i A_{PG,i} \cdot F_{RR,PG,i}} \tag{2.12}$$

A %(V/V) do grupo calcula-se a partir da %(m/m) do mesmo, através da Equação 2.13, em que ϕ_{PG} é a %(V/V) do grupo e ρ_{PG} é a massa volúmica do grupo, em kg/m^3.

$$\phi_{PG} = \frac{100 \cdot \frac{w_{PG}}{\rho_{PG}}}{\sum_i \left(\frac{w_{PG,i}}{\rho_{PG,i}}\right)} \tag{2.13}$$

2.3 Quimiometria

A quimiometria apresenta-se como uma boa solução para o problema em estudo nesta dissertação, uma vez que se trata de uma área da química que usa métodos estatísticos e

matemáticos na otimização de processos e para extrair informação relevante de dados experimentais, através da aplicação de lógicas formais para avaliar e interpretar os dados (Brereton, 2003). Métodos de análise como, por exemplo, espetroscopia e a cromatografia medem diversas variáveis para cada amostra, fornecendo, assim, bases de dados de elevadas dimensões suscetíveis de serem tratadas e interpretadas. Apesar de serem usadas grandes quantidades de dados, existem métodos de análise estatística multivariável que permitem estudar, simultaneamente, todas essas variáveis, tendo em consideração a informação sobre a relação entre elas. Essa informação pode ser usada para, por exemplo, construir um modelo com equações de regressão que explicam como é que as variáveis resposta se relacionam com as variáveis experimentais. Os procedimentos quimiométricos de processamento, avaliação e interpretação são definidos como metodologias quimiométricas. Na Tabela 2.2 encontram-se alguns exemplos desse tipo de metodologias (Otto, 2017).

Tabela 2.2 – Exemplos de metodologias quimiométricos

Metodologias Quimiométricas	
Classificação	Otimização
Criação de Modelos	Processamento de Imagem
Estatística Descritiva e Inferencial	Processamento de Sinal
Informação e Teoria de Sistemas	Projeto Experimental
Métodos de Inteligência Artificial	Reconhecimento de Padrões

Neste trabalho, as metodologias quimiométricas utilizadas para a análise dos dados são Processamento de Sinal e Criação de Modelos. O Processamento de Sinal consiste na diminuição da influência do ruído e eliminação de problemas como o *offset*, através de pré-tratamentos matemáticos, e, consequente, melhoria na busca de informação útil. A Criação de Modelos é um processo de calibração de modelos de regressão que associam as variáveis resposta às propriedades físico-químicas quantificáveis da amostra.

2.3.1 **Modelos de Regressão**

Os modelos de regressão permitem correlacionar dois grupos de variáveis: variáveis de entrada e variáveis de resposta. Com o uso deste tipo de modelos pretende obter-se uma relação entre estes dois tipos de variáveis que permita prever o valor da variável resposta de uma amostra a partir dos seus valores das variáveis de entrada.

A regressão linear múltipla (MLR) é um dos modelos mais diretos na categoria dos modelos lineares e consiste na procura de equações de regressão que relacionem as variáveis de entrada com as variáveis resposta. A aplicação deste modelo apresenta algumas restrições, como por exemplo: é necessário que o número de observações de calibração seja superior ao número de

variáveis de entrada, quando isto não acontece, é necessário selecionar uma parte das variáveis de entrada para ser usada na calibração, correndo o risco de ser omitida informação relevante, durante a fase de seleção das variáveis a usar (Miller e Miller, 2010).

O modelo de regressão por componentes principais (PCR) é a junção da análise por componentes principais (PCA) ao modelo de MLR. É também ele um modelo linear, e que permite solucionar de melhor forma a restrição da MLR relacionada com o número de observações de calibração e o número de variáveis de entrada. A PCA irá reduzir o número de variáveis de entrada de forma a que apenas se utilizem as necessárias para explicar a maior parte da variabilidade, sendo, assim, mais fácil ter um número de observações superior ao número de variáveis de entrada, sem que seja perdida informação importante (Miller e Miller, 2010).

Ainda no grupo dos modelos lineares, o modelo de regressão por mínimos quadrados parciais (PLS) apresenta-se como um dos mais utilizados na análise de dados espetroscópicos e consiste na procura pelo espaço que melhor explica a variabilidade no eixo das ordenadas (y), descrevendo também a variabilidade no eixo das abcissas (x). Este modelo permite analisar a importância relativa de cada uma das variáveis independentes no cálculo da previsão da variável resposta e, normalmente, apresenta melhores resultados quando existe um maior número de variáveis independentes do que o número de observações da variável de saída (Miller e Miller, 2010).

Por último, o modelo das máquinas de vetor de suporte aplicadas à regressão (SVR), embora menos conhecido, é mais uma alternativa que apresenta potencial para a obtenção de bons resultados na criação de modelos de regressão para dados espetrais. É um método simples, poderoso, capaz de processar relações não-lineares e vetores de entrada com elevadas dimensões (Hastie et al., 2008).

2.3.2 Quimiometria e Espetroscopia NIR

Os modelos quimiométricos têm sido cada vez mais aplicados a dados espetroscópicos e em várias áreas da ciência, como por exemplo, na caracterização de combustíveis e na área farmacêutica. Na indústria petroquímica, mais especificamente no desenvolvimento e caraterização de combustíveis, a espetroscopia NIR destaca-se por permitir a realização de uma análise multicomponente, rápida e não destrutiva.

Nesta área, a combinação de NIR com modelos de regressão multivariável teve o seu primeiro registo no ano de 1988, na determinação do MON de uma gasolina a partir de espetros NIR e de MLR (Muller, 1988). Desde essa altura, existem diversas publicações sobre o uso de NIR na medição de propriedades de gasolina. Em 2001, foi realizado um estudo para estimar o RON de uma gasolina, usando NIR e dois métodos quimiométricos diferentes, MLR e regressão de Ridge (Chung et al., 2001). Mais tarde, em 2007, realizou-se um estudo que comparou o desempenho

de 6 modelos de calibração lineares e não-lineares, aplicados a dados espetroscópicos de NIR, na previsão de 6 propriedades de uma gasolina (massa volúmica, ponto inicial de ebulição, ponto final de ebulição e temperatura final de ebulição para 10, 50 e 90 % de volume recuperado) (Balabin et al., 2007). Em 2017, foi desenvolvido um trabalho com objetivo de aferir se um aparelho NIR portátil podia ser usado com a mesma fiabilidade de um aparelho NIR de bancada na previsão de propriedades de gasolinas e de misturas de gasolina com biodiesel, sendo que as últimas contêm etanol. As propriedades em estudo foram as características da destilação (temperaturas de destilação com 10, 50 e 90 % de volume recuperado, e o volume recuperado no ponto final de ebulição), a massa volúmica e o teor em biodiesel das misturas. Os modelos de regressão foram obtidos através de regressão por mínimos quadrados parciais (PLS) e obtiveram bons desempenhos, exceto para a massa volúmica e o conteúdo em biodiesel (da Silva et al., 2017).

Mais recentemente, em 2019, foi desenvolvido um simulador, baseado na combinação de espetros NIR e modelos de regressão obtidos por PLS, que prevê, com fiabilidade, o RON de uma potencial gasolina final a partir da mistura dos espetros NIR das várias correntes de hidrocarbonetos disponíveis para preparar a gasolina. O simulador melhora a eficiência do processo de preparação da gasolina porque permite estimar o valor de RON duma potencial gasolina, antes da sua preparação (da Silva et al., 2019).

Uma abordagem inovadora para prever o RON de uma gasolina usando NIR, que consiste na combinação de um algoritmo para reduzir a dimensão dos dados espetroscópicos e um modelo de redes neuronais artificias (ANN), obteve resultados mais exatos do que os obtidos com os outros modelos e, além disso, com menor tempo de processamento (Wang et al., 2020).

No âmbito da determinação de propriedades de gasolina, foi desenvolvido um estudo com o objetivo de comparar o rigor dos modelos usando dados de RMN e de NIR. Para isso, os modelos foram desenvolvidos usando a técnica SVR, para os dois tipos de dados, e, após isso, foram realizadas estimativas para um conjunto de dados de teste. Os resultados mostraram que os modelos obtidos a partir de dados de NIR obtém melhores desempenhos, apesar das diferenças não serem muito significativas. Além disso, o nível de precisão obtido com os modelos de SVR é adequado e pode permitir a substituição dos métodos convencionais de análise para, pelo menos, 11 parâmetros diferentes de qualidade da gasolina (Leal et al., 2020).

2.3.3 Quimiometria e Cromatografia Gasosa

Devido a vantagens como a precisão e exatidão dos ensaios aliadas à elevada capacidade de separação, a cromatografia gasosa (GC) é, frequentemente, utilizada na análise de petróleo, que pode conter mais de cem mil componentes, e seus derivados. Esta aplicabilidade e as reconhecidas vantagens do método fazem com que este seja aplicado em diferentes áreas da

indústria petroquímica, como a investigação, o desenvolvimento de produtos e o controlo da qualidade (Bartle e Myers, 2002).

Em 2006, realizou-se um estudo com o objetivo de prever e classificar a qualidade de uma gasolina, considerando o RON, a partir de dados de cromatografia gasosa (GC) e espetroscopia FTIR combinados com técnicas quimiométricas como redes neuronais artificiais (ANN), redes neuronais artificiais híbridas, máquinas de vetores de suporte (SVM) e regressão linear múltipla (MLR). No que diz respeito à previsão da qualidade, os resultados obtidos apresentaram uma excelente correlação entre a composição química e os valores do RON. Sendo que, também na classificação das gasolinas, os resultados obtidos apresentaram elevados desempenhos (Brudzewski et al., 2006).

Em 2007, desenvolveu-se uma investigação que pretendia estudar a presença de solventes em gasolinas através da combinação de dados de cromatografia gasosa e quimiometria. Neste caso, as técnicas usadas na classificação das gasolinas foram a análise por componentes principais (PCA) e a análise linear discriminante (LDA). Os modelos obtidos conseguem detetar com exatidão a presença de solventes nas gasolinas, tendo, no entanto, algumas dificuldades na deteção em amostras com concentrações baixas de solventes (Skrobot et al., 2007). Também em 2007, uma combinação de cromatografia gasosa (GC) e do método de modelação suave independente de analogia de classe (*Soft Independent Modelling for Class Analogy*, SIMCA) foi usada na classificação de gasolinas. Os modelos obtidos conseguiram classificar com sucesso, se a gasolina estava de acordo com a regulamentação, em 96 das 100 amostras de treino e em 47 das 50 amostras de teste (Flumignan et al., 2007).

Em 2011, foi desenvolvido um estudo com o intuito de prever propriedades físico-químicas de gasolinas (características da destilação e massa volúmica) usando cromatografia gasosa a duas dimensões, combinada com modelos de regressão. Os modelos utilizados foram a regressão por mínimos quadrados parciais (PLS) e uma variante da PLS envolvendo intervalos de sinergia (siPLS). Os resultados demonstraram que os modelos obtidos por esta estratégia podem ser usados para a previsão das propriedades em causa (de Godoy et al., 2011).

Mais recentemente, em 2018, foi investigada a possibilidade de usar cromatografia gasosa rápida juntamente com regressão por mínimos quadrados parciais (PLS) para prever várias propriedades (RON, conteúdo total de aromáticos, conteúdo em benzeno, tolueno, xileno, metanol e em outros oxigenados) de gasolinas em poucos minutos. Os resultados demonstram que os modelos desenvolvidos podem ser uma alternativa fiável aos métodos convencionais na previsão da qualidade de uma gasolina (Lee et al., 2018).

3 Materiais e Métodos

Neste capítulo descrevem-se os procedimentos necessários para a análise das amostras de gasolina através de técnicas físico-químicas e de espetroscopia de infravermelho próximo (NIR). Relativamente a cada uma das técnicas realizadas, destaca-se a sua importância na avaliação da gasolina e o seu contributo para o controlo e monitorização dos lotes. Além disso, são também descritas as técnicas utilizadas na criação e otimização dos modelos usados na estimativa de propriedades da gasolina.

3.1 Técnicas de Caraterização Físico-Química

Os ensaios realizados, descritos neste capítulo, são os que se encontram especificados na Norma Europeia (EN) 228:2012, e que são também sugeridos na diretiva 98/70/EC, relacionada com a qualidade do petróleo e dos combustíveis.

Importa referir que, com o intuito de obter resultados mais exatos, os ensaios para a determinação da pressão de vapor, das caraterísticas de destilação e do RON e MON, foram realizados em duplicado, de forma consecutiva. O resultado para essa propriedade é registado como a média das duas medições caso não seja ultrapassado o limite de repetibilidade, que consiste no valor máximo da diferença entre os resultados de dois testes obtidos através da realização correta do método, para o mesmo material de teste e sujeita às mesmas condições de operação (por ex.: operador, local e equipamento).

3.1.1 **Pressão de Vapor – EN 13016-1**

O ensaio referente à pressão de vapor mede a pressão exercida pela amostra, em vácuo, que consiste na pressão parcial de produtos, componentes ou alimentações de petróleo, sem a presença de água não-dissolvida, e na pressão parcial de ar dissolvido na amostra. Desta forma, a pressão de vapor é um critério de classificação da segurança no manuseamento, armazenamento e transporte de uma gasolina, sendo os valores limite para esta propriedade, estabelecidos de modo a prevenir o fenómeno de cavitação no decorrer das operações de transferência.

Ainda no que diz respeito aos valores limite, importa referir que a pressão de vapor deve estar contida num intervalo de valores de referência, uma vez que produtos cuja pressão de vapor esteja fora dos limites de especificação, acarretam problemas no desempenho. No caso da gasolina, um valor de pressão de vapor demasiado elevado, pode causar uma evaporação muito rápida da gasolina, fazendo com que o caudal que está a entrar no sistema diminua, podendo surgir uma situação de bloqueio de vapor que impede o funcionamento do sistema (por ex.: um motor de um carro). No que diz respeito a um valor de pressão demasiado baixo, irá fazer com

que o produto evapore de forma mais lenta do que o pretendido, no caso de um carro, iria levar a dificuldades no arranque e a uma aceleração deficiente.

O método utilizado para a determinação da pressão de vapor em gasolinas é o EN 13016-1 que determina, de forma direta, a pressão de vapor do ar saturado (ASVP), que resulta da pressão parcial da gasolina e da pressão parcial do ar dissolvido. No entanto, o valor de pressão de vapor da gasolina que se pretende é o valor equivalente à pressão de vapor seco de Reid (DVPE), em quilo pascal (kPa), que pode ser determinado de acordo com a Equação 3.1, :

$$DVPE = 0{,}965 \times ASVP - 3{,}78 \quad (3.1)$$

O método consiste na injeção de uma amostra de ar frio saturado (entre 0 e 1 ºC) numa câmara controlada termo estaticamente e com um volume interno 5 vezes superior ao volume de amostra injetado. A medição irá ocorrer a uma temperatura de 37,8 ºC, sendo que, assim que a amostra atingir essa temperatura, é medida a pressão total na câmara usando um sensor de pressão e um indicador, sendo essa medição correspondente ao valor de pressão de vapor do ar saturado (ASVP). É importante referir que esta medição deve ser a primeira a ser realizada, pois o valor desta propriedade é muito suscetível a alterações devido a perdas por evaporação (European Committee for Standardization, 2018). O equipamento usado na medição da pressão de vapor foi *Eravap*, da *Eralytics*, e pode ser representado esquematicamente pela Figura 3.1.

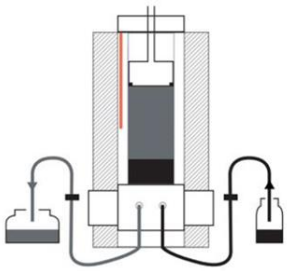


Figura 3.1 - Representação esquemática do aparelho de medição da pressão de vapor.

3.1.2 Curva de Destilação - ISO 3405

As caraterísticas de destilação de hidrocarbonetos fornecem informações importantes em relação à segurança no uso e ao desempenho dos compostos, em especial quando se está a lidar com combustíveis ou solventes, sendo, por isso, importante determiná-las. Essas informações acerca da composição e do comportamento (caraterísticas de armazenamento e desempenho) do composto, neste caso um combustível, são retiradas do intervalo de temperaturas durante o qual este se encontra em ebulição.

Os produtos petrolíferos, na sua maioria, têm especificações que impõem valores limite para as várias caraterísticas de destilação, que permitem controlar o desempenho final do produto e, também, evitar situações indesejáveis como a formação de vapores que podem originar

misturas explosivas em conjunto com o ar ou que podem escapar para a atmosfera sob a forma de emissões.

O método ISO 3405 adequa-se à determinação das caraterísticas de destilação de destilados médios e leves que sejam derivados do petróleo e que apresentem ponto inicial de ebulição superior a 0 ºC e ponto final de ebulição inferior a 400 ºC, categoria na qual se incluem as gasolinas. Dentro desta categoria existe ainda uma divisão, feita com base no valor de pressão de vapor de Reid (RVP) da gasolina a ser testada e que irá influenciar as condições operatórias da destilação, nomeadamente a temperatura do banho do condensador. Assim, para gasolinas com RVP igual ou superior a 65,5 kPa a temperatura do banho do condensador deve estar entre 0 e 1 ºC e para gasolinas com RVP inferior a 65,5 kPa deve encontrar-se entre 0 e 5 ºC, mantendo-se as restantes condições.

O método consiste na destilação de uma porção da amostra de gasolina (100 mL), que deve ser mantida a uma temperatura inferior a 10 ºC, antes da realização do ensaio. Durante o ensaio, são realizadas observações e registos sistemáticos das leituras do termómetro e do volume de condensado recuperado, que permitem obter a informação pretendida acerca da destilação. No final da destilação, é medido o volume de resíduo no balão e são registadas as perdas que ocorreram durante a mesma (International Organization for Standardization, 2019). O equipamento usado na determinação da curva de destilação foi *Optidist*, da *Herzog*, e encontra-se representado esquematicamente na Figura 3.2.

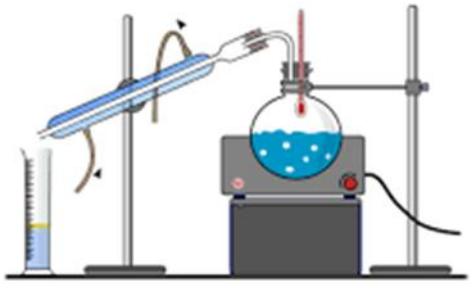


Figura 3.2 – Representação esquemática de um aparelho de destilação atmosférica.

Para desenvolver os modelos de calibração usando o NIR, as caraterísticas que se pretende são a percentagem de volume evaporado a 70, 100 e 150 ºC (E70, E100 e E150, respetivamente), e o ponto final de ebulição (FBP), que representa a temperatura máxima lida no termómetro durante toda a destilação e que, normalmente, só ocorre após a evaporação total do líquido.

3.1.3 **RON e MON – ISO 5164 e ISO 5163**

Os números de octano, RON e MON, de uma gasolina são dois parâmetros importantes na avaliação da sua qualidade, uma vez que medem as suas características antidetonantes. Um combustível com números de octano baixos apresenta maior tendência para a autodetonação

por compressão, gerando pulsos de pressão no interior do cilindro do motor que conduzem ao aumento do consumo de combustível e, no pior cenário, podem mesmo danificar o motor.

Os métodos de análise mais utilizados para a determinação dos valores de RON e MON de um combustível são a ISO 5164 e a ISO 5163, respetivamente (International Organization for Standardization, 2014; International Organization for Standardization, 2014). Os dois métodos são aplicáveis para valores compreendidos entre 0 e 120, no entanto a gama de valores de trabalho é de 40 a 120. Os valores destas duas propriedades fornecem informação acerca das características de detonação dos combustíveis em motores automóveis sob diferentes condições de operação. No caso do RON, as condições de operação são mais suaves, enquanto que para o MON, as condições de operação são mais severas, originando que, para a mesma amostra, o valor do RON seja superior ao do MON.

O procedimento deste método consiste em colocar a amostra a operar num *Cooperative Fuel Research Motor* (CRF), com um rácio combustível-ar que maximize a sua detonação, e compará-la com diferentes misturas de combustíveis de referência, de forma a determinar qual dos combustíveis de referência produz a mesma intensidade de detonação com a mesma razão de compressão do motor. Os combustíveis de referência, normalmente, usados são misturas volumétricas de iso-octano e n-heptano, que têm, por convenção, um RON e MON de 100 e de 0, respetivamente. É a composição desta mistura que permite determinar o número de octanas da mistura e, consequentemente, o da amostra. As condições de operação deste ensaio mais relevantes estão descritas na Tabela 3.1.

Tabela 3.1 – Condições de operação dos ensaios de determinação de RON e MON

	RON	MON
Rotações do Motor (rpm)	600 ± 6	900 ± 9
Ângulo de Motor de Detonação (º btdc [a])	13	Variável de acordo com a taxa de compressão do motor
Temperatura de admissão de ar (º F)	Variável de acordo com a pressão atmosférica	100
Temperatura de mistura ar-combustível (º F)	Não tem	300

[a] btdc – *before top dead center*.

Encontra-se representado esquematicamente na Figura 3.3 um carburador, um dos elementos presentes no aparelho utilizado para a determinação do RON e MON.

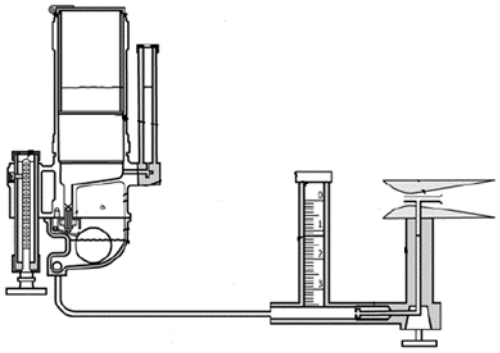


Figura 3.3 - Representação de um carburador utilizado na determinação do RON e MON (International Organization for Standardization, 2014).

3.1.4 **Cromatografia – ISO 22854**

Alguns dos compostos presentes na formulação de uma gasolina são prejudiciais para a saúde humana e têm impactes ambientais, por exemplo na poluição do ar, nomeadamente, os compostos aromáticos, em especial o benzeno, e os compostos olefínicos. Além disso, estes mesmos compostos são propensos a reagir com oxigénio atmosférico absorvido e entre si durante o armazenamento, causando alterações nas caraterísticas físico-químicas da mistura e, consequentemente, diminuindo a qualidade da gasolina. Principalmente, devido aos impactes ambientais e aos efeitos na saúde humana, foram estabelecidos limites máximos para a concentração destes compostos na gasolina. Assim, a análise cromatográfica de uma amostra de gasolina tem grande importância no controlo de qualidade da mesma, uma vez que, este ensaio permite determinar a composição exata e pormenorizada da mistura, tornando possível avaliar se a gasolina se encontra de acordo com a especificação relativa à concentração de vários compostos.

A ISO 22854 é o método mais usual para a determinação da concentração de hidrocarbonetos e compostos oxigenados numa gasolina. Este método consiste na realização de uma cromatografia gasosa que determina a concentração dos diferentes grupos de hidrocarbonetos que a constituem. Esta separação dos compostos ocorre após a injeção e consequente vaporização da amostra, e a deteção é realizada através de um detetor de ionização por chama. As amostras de gasolina são suscetíveis a perdas por evaporação e, por isso, a amostra deve ser mantida a uma temperatura entre os 0 e os 5 ºC, antes da realização do ensaio. A concentração mássica, w, dos grupos de hidrocarbonetos está relacionada com a área dos picos correspondentes a esse grupo, sendo calculada multiplicando a área do pico por um fator de resposta relativo e normalizando os valores para 100 %, tal como é descrito na Equação 3.2, onde A_{PG} e $F_{RR,PG}$ são, respetivamente, a área corrigida e o fator de resposta para cada grupo parcial (PG). PG representa o grupo de compostos que pertencem a um grupo de hidrocarbonetos e que têm um número de carbonos iguais entre si.

$$w_{PG} = \frac{100 \cdot A_{PG} \cdot F_{RR,PG}}{\sum_i (A_{PG,i} \cdot F_{RR,PG,i})} \tag{3.2}$$

A conversão de concentração mássica para concentração volúmica é realizada a partir da densidade de cada grupo parcial.

No caso específico do conteúdo de oxigénio, w_O, presente em todos os compostos oxigenados que foram identificados, i, é utilizada a Equação 3.3, em que n_O é o número de átomos de oxigénio na molécula, M_O é a massa atómica do oxigénio, M_i é a massa molecular do composto oxigenado e w_i é a concentração mássica do composta na mistura.

$$w_O = \sum_i \left(\frac{n_O \cdot M_O}{M_i} w_i \right) \tag{3.3}$$

O equipamento usado na realização dos ensaios de cromatografia gasosa foi o Reformulyzer M3 da Analytical Controls. A Figura 3.4 ilustra a configuração típica de um equipamento de cromatografia gasosa, que consiste num injetor, um detetor de ionização por chama, colunas com diferentes funções, armadilhas e catalisadores de hidrogenação. Tipicamente, o processo de separação começa com a absorção de álcoois e de aromáticos com ponto de ebulição elevado numa armadilha (Zona 1). Após essa etapa, os restantes aromáticos são separados dos outros compostos numa coluna polar (Zona 2). Uma outra armadilha (Zona 3) é usada para recolher os éteres. O passo seguinte é separar as olefinas dos compostos saturados e é realizado numa armadilha de olefinas (Zona 4). Depois, é usada uma coluna de peneiros moleculares (Zona 5) para separar os hidrocarbonetos saturados em parafinas e nafténicos, de acordo com o seu número de carbonos. Os éteres são eluídos da armadilha (Zona 3), separados e detetados de acordo com o seu ponto de ebulição. As olefinas, também são eluídas, hidrogenadas na coluna de platina (Zona 6), separadas e detetadas usando uma peneira molecular (Zona 7). Por fim, os álcoois e os aromáticos são eluídos da coluna polar (Zona 2) e da armadilha (Zona 1), respetivamente, depois, são separados numa coluna não-polar (Zona 8) e detetados de acordo com o seu ponto de ebulição (International Organization for Standardization, 2016).

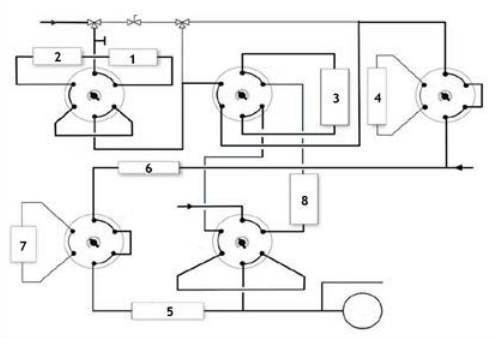


Figura 3.4 – Representação esquemática de um cromatógrafo gasoso (International Organization for Standardization, 2016).

3.2 Métodos Espetroscópicos - NIR

O princípio base da espetroscopia NIR são vibrações de diferentes tipos de vibrações que se verificam entre os átomos presentes nas moléculas. Na prática, o ensaio consiste numa injeção de amostra numa coluna associada a um espetrofotómetro, sendo os resultados obtidos sob a forma de um espetro. Um espetrofotómetro, regra geral, é constituído pelos seguintes componentes: uma sonda de fibra ótica NIR, um interferômetro, um detetor InGaAs e uma lâmpada de halogéneo como fonte de feixe de luz. Além destes componentes, usa-se um divisor de feixe para separar o feixe de luz que entra em dois feixes distintos, que, mais tarde, serão recombinados, focados numa fibra ótica e enviados para a amostra, já como feixe de luz único. Em relação aos dois feixes, importa referir que um deles tem um caminho ótico constante, enquanto que o caminho ótico do outro é variado com recurso a meios mecânicos. Um interferograma é produzido na recombinação de feixes e é obtido um espetro de frequência após a transformação de Fourier (Burns e Ciurczak, 2008). Neste caso, o espetrofotómetro utilizado foi o *NR800 Fourier Transformed Near-Infrared Analyzer*, da *Yokogawa Eletric Corporation* (*Yokogawa EA*). Enquanto que o controlo e a monitorização do espetrofotómetro é realizada com um computador usando uma aplicação informática adequada, tendo sido usado o *SPECTLAND 2.09*, também ela da *Yokogawa EA*, que permite ver as condições do espetrofotómetro, definir os parâmetros de recolha e dar início à medição.

No que diz respeito ao procedimento, é importante manter a amostra numa temperatura entre 0 e 5 ºC, antes da análise, de modo a que não existam perdas por evaporação, não sendo necessária mais nenhuma etapa prévia de preparação da amostra. As condições de recolha do NIR, que tal como já foi referido são escolhidas na aplicação informática, são as seguintes: alcance espetral de 11000 a 4000 cm^{-1}, resolução de 4 cm^{-1} (resolução máxima selecionável) e 128 leituras. Antes da injeção da amostra, recolhe-se um espetro da célula em branco para medir a quantidade de luz que não é resultado da absorção da amostra. Assim, o espetro da amostra resultante da análise é automaticamente corrigido pelo espetrofotómetro, tendo por base o espetro da célula em branco, o que leva a que a informação fornecida pelo espetro resultante dessa correção seja exclusivamente da amostra.

3.3 Quimiometria

O principal objetivo deste trabalho é a estimativa de propriedades de gasolinas com etanol com base na tecnologia NIR e em cromatografia gasosa. Uma vez que não existe nenhuma equação matemática simples que permita proceder a essa estimativa diretamente, é necessário o desenvolvimento de modelos de calibração multivariável. Esta secção descreve os métodos utilizados na divisão de dados, as técnicas de pré-tratamento de dados, os métodos estatísticos

multivariáveis utilizados no desenvolvimento de modelos, as técnicas utilizadas na otimização dos mesmos, e os indicadores de desempenho dos modelos.

As várias etapas associadas à criação de modelos utilizadas neste trabalho estão representadas esquematicamente na Figura 3.5.

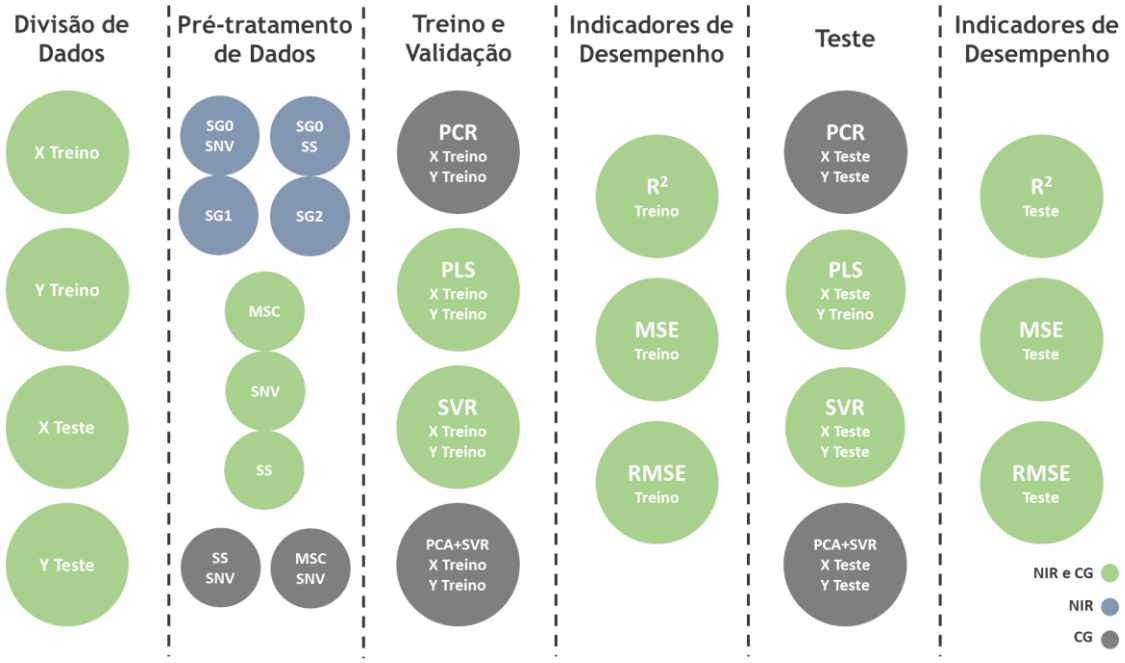


Figura 3.5 – Esquema resumo das várias etapas do processo de criação de modelos.

As etapas descritas anteriormente foram realizadas recorrendo à aplicação informática *Python* com especial importância para a ferramenta *Scikit-Learn* (Pedregosa et al., 2011).

3.3.1 Divisão de Dados

O desempenho dos modelos não depende apenas da etapa de calibração ou do tipo de método estatístico usado na sua criação. Existem mais fases deste processo que têm grande influência no desempenho final dos modelos, como é o caso da divisão de dados.

A importância da divisão de dados reside no facto de ser necessária a existência de vários conjuntos de dados, que permitam não só calibrar o modelo, mas também avaliar o desempenho deste na presença de novas amostras. O objetivo é que o modelo obtido apresente bons desempenhos para amostras que não tenham sido usadas na sua calibração e, para isso, não devem ser usadas todas as amostras presentes na base de dados na etapa de calibração. De outra forma, existe um risco elevado de que o modelo seja demasiado dependente desses valores e obtenha elevado desempenho para o conjunto de dados de calibração, mas não responda de forma adequada perante novas amostras.

Quando a base de dados disponível tem um número relativamente elevado de amostras, a opção mais adequada para a divisão de dados é a criação de 3 conjuntos, que não têm valores em comum e são tratados de forma independente: o conjunto de treino, conjunto de maior

dimensão e que é utilizado na criação do modelo; o conjunto de validação, usado para definir a complexidade ótima do modelo e evitar que o modelo fique demasiado dependente dos valores das amostras do conjunto de treino (sobreajustamento); e o conjunto de teste, utilizado para avaliar a resposta do modelo na presença de novas amostras. Existem abordagens alternativas que formam dois conjuntos, conjunto de calibração e conjunto de teste, sendo aplicada validação cruzada no conjunto de calibração, funcionando assim como conjunto de treino e validação, ou ainda outra hipótese em que se aplica validação cruzada ao conjunto total de dados, formando apenas conjuntos de treino e teste através da validação cruzada(Browne, 2000).

A divisão dos dados pelos 3 conjuntos pode ser feita de diversas formas, através de um algoritmo simples, de forma aleatória ou com um procedimento mais elaborado. A abordagem mais usada e a forma mais simples de fazer a divisão dos dados é a divisão aleatória dos dados pelos 3 conjuntos, tendo apenas em conta a quantidade de valores que se pretende em cada conjunto de dados. Esta abordagem não contempla nenhum critério de seleção, pelo que a representatividade do conjunto inicial de dados em cada um dos subconjuntos não está assegurada.

O procedimento adotado neste trabalho baseia-se nos valores da variável de saída, ou seja, nos valores da propriedade em estudo obtidos pelo método clássico e, tendo a conta que a base de dados não é muito extensa, foi escolhida a abordagem que forma dois conjuntos de valores, conjunto de calibração e conjunto de teste, e usa a validação cruzada no primeiro desses dois conjuntos.

3.3.2 Pré-tratamento de Dados

Após a recolha adequada dos dados, o pré-tratamento de dados é a etapa mais importante antes da calibração dos modelos. A variância presente nas variáveis de entrada que não se relaciona com as variáveis de resposta, pode reduzir o poder de previsão do modelo em causa. Assim, para melhorar a taxa de sucesso dos modelos de calibração, é importante reduzir a variância sistemática que não transmite qualquer informação acerca da variável de saída nem está relacionada com a mesma, e esse é o principal objetivo da etapa de pré-tratamento de dados.

A variância sistemática que se pretende remover pode estar relacionada com a linha de base, com efeitos de dispersão multiplicativos e com regiões de conteúdo com pouca informação relevante. No caso da espetroscopia NIR, por vezes, existem situações em que picos não são visíveis devido à sua baixa intensidade ou porque estão sobrepostos por picos de maior intensidade e mais amplos, o que se pode dever, por exemplo, a fenómenos de dispersão. Além disso, os parâmetros operacionais do aparelho podem influenciar diretamente o espetro de NIR (Burns e Ciurczak, 2008).

O primeiro tratamento a ser aplicado aos dados do espetro NIR é a correção da linha de base. A linha de base de um espetro pode ser definida como a média do ruído presente no espetro e que, idealmente, deveria ser uma linha reta com intensidade zero. A correção da linha de base é uma etapa importante no pré-tratamento de dados porque as variações experimentais de espectros individuais dentro de grandes conjuntos de dados podem impedir a construção de modelos quantitativos precisos. Neste caso, a correção da linha de base dos espetros foi feita automaticamente através do programa *Spectland* 2.09 (*Yokogawa Eletric Corp*.).

De seguida, serão descritas, de forma sucinta, as técnicas de pré-tratamento de dados utilizadas após a correção da linha de base nos espetros NIR e algumas das suas vantagens e desvantagens. Algumas destas técnicas foram também aplicadas aos dados da cromatografia gasosa, que não estão sujeitos a nenhum tratamento prévio.

A técnica Savitzky-Golay (SG) usa o alisamento de forma a reduzir o ruído presente no espetro e tem 3 parâmetros ajustáveis: o grau da derivada, o grau do polinómio e o tamanho da janela. Os parâmetros são escolhidos pelo utilizador e devem ser otimizados para cada conjunto de dados, além disso, deve ser respeitada a seguinte restrição na escolha desses parâmetros: a ordem da derivada não pode ser superior à ordem do polinómio. O procedimento desta técnica é o seguinte: primeiro, ajustar um polinómio à janela de valores usando os mínimos quadrados lineares, depois, é calculada a derivada do polinómio e esse valor pode ser usado como estimativa da derivada do ponto central da janela e este procedimento repete-se para todos os pontos do espetro de forma sequencial (Rinnan et al., 2009).

A técnica de variação padrão (SNV) obtém o espetro corrigido a partir da Equação 3.4, em que a_0 é o valor médio do espetro em causa e a_1 é o desvio padrão do mesmo. O pré-tratamento usando SNV é feito amostra a amostra, o que se torna numa vantagem no caso de existirem inconsistências nos dados. Por outro lado, a forma como os parâmetros usados na correção são estimados faz com que se torne sensível ao ruído (Rinnan et al., 2009).

$$x_{corr} = \frac{x_{org} - a_0}{a_1} \tag{3.4}$$

Correção de dispersão multiplicativa (MSC) é uma técnica que não só reduz a variabilidade física devido à dispersão, mas também ajusta as diferentes linhas de base presentes nas amostras. É uma das técnicas mais usadas para os espetros NIR e baseia-se em 2 passos: determinação dos coeficientes de correção que descrevem a dispersão e correção dos espetros recorrendo a esses coeficientes. Estes passos estão descritos na Equação 3.5 e na Equação 3.6, respetivamente, em que x_{org} é o espetro original da amostra, x_{ref} é o espetro de referência usado, e é a parte não-modelada de x_{org}, x_{corr} é o espetro corrigido e b_0 e $b_{ref,1}$ são os parâmetros escalares, diferentes para cada amostra. Na maioria dos casos, o espetro de referência é o espetro médio do conjunto de calibração. Depois da aplicação de MSC, cada espetro da base de dados tem a

mesma amplitude. MSC é uma técnica útil e estável se for aplicada a dados homogéneos e representativos (Rinnan et al., 2009).

$$x_{org} = b_0 + b_{ref,1} \cdot x_{ref} + e \tag{3.5}$$

$$x_{corr} = \frac{x_{org} - b_0}{b_{ref,1}} = x_{ref} + \frac{e}{b_{ref,1}} \tag{3.6}$$

Ainda dentro do mesmo tipo de técnicas de pré-tratamento foi usada a normalização padrão (SS) que utiliza a Equação 3.7 para obter o espetro corrigido. A Equação 3.7 assemelha-se à Equação 3.6, no entanto, nesta técnica os parâmetros não são obtidos amostra a amostra e, assim, c_0 representa o valor médio de todos os espetros e c_1 é o desvio padrão da totalidade dos espetros utilizados (Pedregosa et al., 2011). O facto de a obtenção dos parâmetros ser feita de forma global e não amostra a amostra, como nas técnicas MSC e SNV, resulta num espetro corrigido diferente do obtido por essas técnicas.

$$x_{corr} = \frac{x_{org} - c_0}{c_1} \tag{3.7}$$

3.3.3 **Modelos de Regressão**

Com o objetivo de prever várias caraterísticas da gasolina a partir dos dados de NIR e de cromatografia gasosa, foram construídos modelos de regressão por componentes principais (PCR), por mínimos quadrados parciais (PLS), usando máquinas de vetores de suporte (SVR) e a combinação de análise por componentes principais (PCA) e regressão usando máquinas de vetores de suporte (SVR).

No caso dos modelos de regressão por componentes principais, o primeiro passo é a redução do número de variáveis de entrada, aplicando a análise por componentes principais (PCA) aos dados originais e realizando, de seguida, uma regressão linear múltipla ou uma regressão recorrendo às máquinas de vetores de suporte, em que a variável resposta é a propriedade que se pretende estimar e as variáveis de entrada são os componentes principais (PC). A análise por componentes principais (PCA) é uma técnica utilizada na redução da quantidade de dados correlacionados. O objetivo é encontrar componentes principais Z_1, Z_2, ..., Z_n que são combinações lineares das variáveis originais, tal como se expressa nas Equações 3.8 e 3.9.

$$Z_1 = a_{11}X_1 + a_{12}X_2 + \cdots a_{1n}X_n \tag{3.8}$$

(...)

$$Z_n = a_{n1}X_1 + a_{n2}X_2 + \cdots a_{nn}X_n \tag{3.9}$$

Assim, obtém-se um número de PC's igual ao número de variáveis de entrada dos dados originais, no entanto, os coeficientes que explicam os PC's a partir dos dados originais são calculados de forma a que não exista correlação entre eles, ao contrário do que acontece com

as variáveis dos dados originais. Além disso, a determinação dos PC's é realizada de maneira que o primeiro componente principal (Z_1) explique a maior parte da variância presente nos dados originais, seguido do segundo componente principal (Z_2) e assim sucessivamente. Este facto permite que, apesar do número de componentes principais total ser igual ao número de variáveis originais, o número de PC's usado seja inferior ao número de variáveis originais, sem perda significativa da informação original. Os PC's são, normalmente, obtidos a partir da matriz de covariância, podendo também ser obtidos a partir da matriz de correlação nos casos em que os dados estão normalizados. Esta segunda hipótese, em que os dados são normalizados previamente, tem especial interesse para os casos em que as variáveis têm diferentes escalas ou de forma a nivelar o contributo de cada variável no caso de uma das variáveis ter uma variância muito superior às restantes (Miller e Miller, 2010).

A regressão linear múltipla (MLR) procura determinar equações de regressão, similares à Equação 3.10, que permitam relacionar as variáveis de entrada com a variável resposta. Este método tem como principal restrição o facto de ser necessário um número de observações superior ao número de variáveis de entrada (Miller e Miller, 2010).

$$y_1 = b_{01} + b_{11}Z_1 + b_{12}Z_2 + \cdots b_{1n}Z_n \tag{3.10}$$

A regressão por componentes principais (PCR), que consiste na aplicação de PCA aos dados iniciais, seguida de regressão linear múltipla aos conjunto de componentes principais, apresenta-se como uma boa alternativa ao uso da regressão linear múltipla de forma isolada, pois permite contornar a principal restrição associada à MLR, preservando grande parte da variabilidade presente nos dados originais e, consequentemente, melhorando os resultados da regressão.

A regressão por mínimos quadrados parciais (PLS) é um modelo de regressão com capacidade para processar matrizes de dados com grandes dimensões e com variáveis colineares, e para gerar modelos em situações em que o número de variáveis é superior ao número de observações, sendo, por isso, bastante usada no âmbito da espetroscopia.

Tal como no PCR, o PLS utiliza combinações lineares das variáveis de entrada diferentes das variáveis de entrada originais. No entanto, em PCR os PC's são escolhidos de forma a manter o máximo da variância do conjunto de variáveis de entrada, enquanto na PLS, além de se manter o máximo da variância do conjunto original, é atribuído um maior peso às variáveis que apresentarem maior correlação com as variáveis resposta, pois serão importantes para melhorar a capacidade de previsão do modelo. Assim, as combinações lineares das variáveis de entrada são calculadas e escolhidas de acordo com três critérios: elevada correlação com a variável resposta, elevada variância em relação ao conjunto original e ausência de correlação entre si. Estes três critérios permitem que se obtenha um novo conjunto de variáveis, designadas por variáveis latentes, que explica grande parte da variância do conjunto original, tem elevada

correlação com o conjunto das variáveis resposta e tem uma dimensão bastante menor do que o conjunto original de variáveis de entrada (Miller e Miller, 2010). Tem ainda outras vantagens como ser capaz de lidar com a falta de dados e ruído de medição presente nos mesmos, e o facto de considerar erros tanto no espetro como nas variáveis resposta, assumindo que o erro é igualmente distribuído. A sua principal desvantagem é o facto de assumir que a relação existente entre as variáveis de entrada é linear (Brereton, 2003). Para obter melhor resultados é necessário otimizar o número de variáveis latentes a usar, uma vez que o uso de um número de variáveis latentes desadequado irá fornecer um modelo com reduzida capacidade preditiva. Por outro lado, um número demasiado elevado pode levar a que o modelo só faça boas previsões das amostras que foram usadas na sua calibração. Esta otimização é realizada na etapa de validação, descrita na Secção 3.3.5.

Neste caso, existe apenas uma variável resposta e, portanto, a regressão é feita a partir de uma matriz de variáveis de entrada (X) com dimensão $n \times m$, em que n é o número de amostras e m o número de variáveis de entrada, e um vetor de variáveis resposta (y) de tamanho n. O objetivo deste tipo de regressão é obter uma relação linear entre X e y através de um vetor de coeficientes de regressão (b) e de um vetor de erro (f), tal como descrito na Equação 3.11.

$$y = X\,b + f \tag{3.11}$$

De acordo com o algoritmo, tanto a matriz X como o vetor y são descritos como o produto de um conjunto de fatores ortogonais (pontuações) e um conjunto de cargas, tendo em conta informação proveniente tanto de X como de y, expresso nas Equações 3.12, 3.13, 3.14, 3.15 e 3.16, em que T é a matriz de pontuações de X, W é a matriz de pesos de X, P é a matriz de cargas de X, E é a matriz de erros residuais de X, c é o vetor dos pesos de y, f é o vetor de erros residuais de y e b é o vetor dos coeficientes de regressão. O principal objetivo deste procedimento é maximizar a covariância entre as pontuações de X e y (Wold et al., 2001; Brereton, 2003).

$$T = X\,W \tag{3.12}$$

$$X = T\,P + E \tag{3.13}$$

$$y = T\,c + f \tag{3.14}$$

$$b = W\,c \tag{3.15}$$

$$y = X\,b + f \tag{3.16}$$

O modelo de PLS (*PLSRegression*) utilizado no *Python* provém da ferramenta *Scikit-Learn* (Pedregosa et al., 2011).

As máquinas de vetores de suporte (SVM) foram inicialmente utilizadas como uma ferramenta de classificação binária, consistindo na distribuição não-linear dos vetores de entrada num espaço caraterístico de grande dimensão. Este método de classificação pode ser utilizado no estudo de problemas de regressão e tem sido cada vez mais utilizado com esse objetivo. Uma das principais vantagens da aplicação deste modelo na regressão é que a forma como o algoritmo está construído garante a convergência para soluções ótimas, porque se baseia em programação quadrática com restrições lineares. Apesar disso, esta técnica também permite processar dados não-lineares com valores de erro mais reduzidos (Hastie et al., 2008).

O objetivo da SVR é, a partir de um conjunto de dados na forma $\{(X_1, y_1), \dots, (X_n, y_n)\}$, encontrar uma função $f(X)$ que origine valores que quando comparados com os valores objetivo (y_i) tenha um erro inferior a ε. A Equação 3.17 ilustra a formulação generalizada da equação $f(X)$ que se pretende obter a partir da SVR, em que $\langle \cdot, \cdot \rangle$ representa o produto escalar em X (Leal et al., 2020).

$$f(X) = \langle w, X \rangle + b \; com \; w \in X, b \in \mathbb{R} \tag{3.17}$$

Quanto mais parecida com uma reta for a resposta da função de aproximação $f(X)$, menor será a sensibilidade das previsões a perturbações nos valores de X. Essa semelhança da resposta de $f(X)$ a uma reta pode ser otimizada minimizando o valor da norma, $\|w\|^2$. Pretende-se encontrar uma reta ou um plano, dependendo da dimensão dos dados, que consiga manter todas as previsões dos pontos de treino dentro da sua margem ε, com pequenos ganhos de w. A Equação 3.18 ilustra a representação deste objetivo sob a forma de um problema de otimização convexo (Leal et al., 2020).

$$\begin{gathered} Minimizar \; \frac{1}{2}\|w\|^2 \\ Sujeito\ a \begin{cases} y_i - \langle w, X \rangle - b_0 \leq \varepsilon \\ \langle w, X \rangle + b_0 - y_i \leq \varepsilon \end{cases} \end{gathered} \tag{3.18}$$

A Figura 3.6 ilustra graficamente a divisão de dados realizada pela SVR.

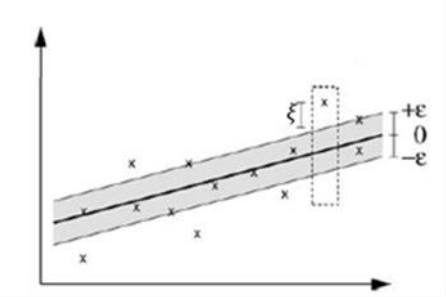


Figura 3.6 - Representação gráfica da divisão de dados realizada pela SVR (Leal et al., 2020).

Em diversas situações, a função obtida não aproxima os valores de previsão com a precisão desejada e, por isso, de forma a melhorar a capacidade preditiva do modelo são adicionados

três variáveis às equações do problema de otimização, C, ξ e ξ_i^*. A otimização da desempenho do modelo é conseguida pela otimização dos valores utilizados nos seus parâmetros. As variáveis ξ e ξ_i^* são chamadas variáveis folga e quantificam o desvio em relação à reta/plano dos pontos que se encontrem fora do valor de precisão desejado ε. Por seu lado, o parâmetro C funciona como uma penalização em relação aos valores que se encontram fora do valor de precisão pretendido, quantificando a tolerância existente em relação aos mesmo. Assim, quanto maior o valor de C maior será a tolerância existente (Sharp, 2020). Neste trabalho, os valores de C e ε serão otimizados com o objetivo de melhorar o desempenho do modelo. O problema de otimização passa a ser representado de acordo com a Equação 3.19 (Leal et al., 2020).

$$Minimizar\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_{\ i} + \xi_i^*)$$
$$Sujeito\ a \begin{cases} y_i - \langle w, X \rangle - b_0 \leq \varepsilon + \xi_{\ i} \\ \langle w, X \rangle + b_0 - y_i \leq \varepsilon + \xi_i^* \\ \xi_{\ i}, \xi_i^* \geq 0 \end{cases} \tag{3.19}$$

No caso do modelo SVR da ferramenta *Python* utilizada, também possível otimizar a performance através da escolha do *kernel*. O *kernel* especifica o tipo de função usada na construção do modelo e podem ser usadas as seguintes opções: $\{'rbf', 'linear', poly', 'sigmoid', 'precomputed'\}$.

O modelo de SVR (*svm.SVR*) utilizado no *Python* provém da ferramenta *Scikit-Learn* (Pedregosa et al., 2011).

3.3.4 Indicadores de Desempenho dos Modelos

De forma a avaliar o desempenho dos modelos obtidos foram calculados e analisados vários parâmetros estatísticos. Os parâmetros calculados avaliam os modelos de duas formas: avaliação da razão entre a variância residual e a variância dos dados observados, e avaliação da diferença entre os valores previstos e os valores determinados por métodos convencionais como, no caso deste trabalho, a determinação da pressão de vapor, do RON, do MON e as caraterísticas de destilação. Apesar de avaliarem diferentes métricas, os parâmetros usados neste trabalho têm todos como base o erro de previsão, $y_i - \hat{y}_i$. Os critérios em causa não são utilizados apenas para avaliar o modelo final, mas são também usados para a criação e otimização dos modelos (Varmuza e Filzmoser, 2009). A Tabela 3.2 apresenta as fórmulas de cálculo e as definições dos parâmetros usados neste trabalho.

Tabela 3.2 – Parâmetros estatísticos de avaliação dos modelos.

Parâmetro	Fórmula de Cálculo	Definição
Coeficiente de Determinação (R^2)	$R^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	Medida da qualidade da previsão fornecida pelo modelo em relação à utilização da média dos valores de resposta. Valores de R^2 elevados indicam melhores previsões e erros menores.
Erro Quadrático Médio (MSE)	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	Média aritmética dos erros quadráticos. É usada preferencialmente durante a criação e otimização do modelo. Valores de MSE menores, indicam erros menores e melhores previsões.
Raiz do Erro Quadrático Médio (RMSE)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	Descreve a diferença entre os valores previstos e observados. Valores de RMSE menores, indicam erros menores e melhores modelos de previsão.

3.3.5 **Validação**

O desempenho do modelo está relacionado com a complexidade do mesmo, sendo que, normalmente, existe um nível ótimo de complexidade, a partir do qual aumentar a complexidade não irá originar melhorias no desempenho, por exemplo, no caso do PLS, existe um número ótimo de variáveis latentes para cada caso que se está a tratar e pode não ser o número de variáveis latentes mais elevado que se testou. Além disso, a capacidade de um modelo prever valores de amostras externas ao conjunto de calibração não é diretamente proporcional à complexidade do modelo, isto é, existe um valor limite de complexidade do modelo a partir do qual a capacidade de previsão do modelo irá melhorar para as amostras do conjunto de calibração, mas irá piorar para amostras externas ao conjunto de calibração. Este fenómeno tem o nome de sobreajustamento e o que acontece é que o modelo se ajusta de tal forma aos valores presentes no conjunto de calibração que tem dificuldade em prever de forma exata valores diferentes dos de calibração. É, por isso, necessário encontrar esse ótimo de complexidade do modelo que otimiza os resultados de previsão de valores externos à calibração. Na etapa de validação, avalia-se o modelo de forma a obter o grau ótimo de complexidade, sendo por isso uma etapa com grande influência no desempenho final do modelo (Balabin et al., 2010).

Tal como já foi referido, neste trabalho, os dados foram divididos em dois conjuntos: calibração e teste. Assim, a validação foi feita com recurso à aplicação de validação cruzada no conjunto de calibração.

A validação cruzada procura otimizar a complexidade do modelo, encontrar a complexidade ótima dentro das várias hipóteses testadas e formar um modelo a partir dessa complexidade ótima. Este procedimento consiste na divisão de um conjunto de valores em dois conjuntos, em que o maior é usado para calibração e o outro funciona como conjunto de teste, repetindo este processo um determinado número de vezes. Para aplicar este procedimento é necessário definir o tamanho pretendido para cada conjunto e o número de repetições que pretendemos que se realize, isto é, se se definir um número de repetições de 10, serão formados 10 conjuntos de treino e 10 conjuntos de teste diferentes, sendo que os valores só fazem parte do conjunto de teste por uma vez e nunca o mesmo valor se encontra no conjunto de treino e de teste numa mesma repetição (Browne, 2000). Desta forma, dependendo do tipo de modelo em causa, realiza-se validação cruzada para cada um dos conjuntos de parâmetros que se pretende testar e, a partir dos valores do erro quadrático médio (MSE), determina-se o conjunto de parâmetros ótimo, que deve ser usado na construção do modelo.

4 Resultados e Discussão

Neste capítulo apresentam-se os dados recolhidos, descrevem-se as várias técnicas aplicadas aos dados ao longo do processo do desenvolvimento de modelos e discutem-se as previsões de cada propriedade obtidas pelo modelo com melhor desempenho. Este capítulo encontra-se dividido em duas grandes secções, sendo a primeira relacionada com as previsões dos modelos a partir de espetros NIR e, a segunda, com as previsões dos modelos a partir de dados de cromatografia gasosa.

A avaliação do desempenho dos modelos realiza-se através de vários fatores, sendo um deles a comparação entre os valores dos limites de repetibilidade e de reprodutibilidade dos métodos convencionais com a diferença entre o valor previsto e o valor obtido pelo método. A repetibilidade e a reprodutibilidade são duas medidas de precisão dos métodos, sendo que a repetibilidade corresponde à diferença entre resultados de dois testes independentes, obtidos através da realização correta do método, para o mesmo material de teste e sujeita às mesmas condições de operação (por ex.: operador, local e equipamento), enquanto que para a reprodutibilidade os dois testes são sujeitos a condições de operação diferentes. Cada método tem limites de repetibilidade e de reprodutibilidade caraterísticos, definidos pelas respetivas normas, sendo que os resultados de duas análises da mesma amostra realizadas pelos métodos convencionais só podem ser aceites mediante o cumprimento desses limites.

4.1 Dados

A base de dados da espetroscopia NIR é composta por um conjunto de 100 amostras de gasolinas, analisadas pelos vários métodos convencionais e pela espetroscopia NIR. Uma vez que na Refinaria de Matosinhos, a adição de etanol às gasolinas produzidas só irá começar a ocorrer durante o ano de 2021, as amostras analisadas são provenientes de estudos realizados anteriormente na refinaria com o intuito de aferir a influência da adição de etanol às gasolinas e de estudos formulados no âmbito deste trabalho. Assim, cerca de 80% das amostras foram formuladas para o desenvolvimento deste trabalho e as restantes 20% das amostras são provenientes desse estudo realizado previamente. No caso dos estudos realizados no decorrer deste trabalho, as mesmas foram formuladas de forma a que os valores das propriedades determinadas se encontrassem dentro de intervalos de valores obtidos habitualmente nas gasolinas produzidas na refinaria e de modo a distribuírem-se o mais uniformemente possível ao longo desses intervalos de valores. As principais caraterísticas dos valores de cada propriedade encontram-se resumidas na Tabela 4.1. Por último, importa referir que existem propriedades nas quais o total de valores é inferior a 100 porque não existiam registos dos valores dessas propriedades para algumas das amostras do estudo anterior.

Tabela 4.1 - Tabela resumo das principais caraterísticas da base de dados para NIR

Propriedade	Unidades	Núm. de Amostras	Mín.	Máx.	Méd.
RON		100	87,0	99,8	94,1
MON		100	80,1	88,0	84,5
Pressão de Vapor	kPa	100	43,7	79,1	61,6
Evaporado a 70 ºC	% V/V	100	17,1	45,8	31,0
Evaporado a 100 ºC	% V/V	100	40,9	61,8	52,0
Evaporado a 150 ºC	% V/V	100	74,4	93,0	87,7
Ponto Final de Ebulição	ºC	80	177,1	200,0	189,4
Aromáticos	% V/V	100	24,4	40,8	32,1
Benzeno	% V/V	100	0,6	2,1	1,1
Etanol	% V/V	100	0,0	9,2	3,4
MTBE	% V/V	99	0,1	19,3	4,1
Olefinas	% V/V	100	1,1	13,6	8,7
Oxigénio	% m/m	100	0,0	4,8	2,0

No caso da cromatografia gasosa, a base de dados é composta por um conjunto de 93 amostras de gasolinas, todas elas presentes na base de dados do NIR. Assim, neste caso, cerca de 85% das amostras foram desenvolvidas no âmbito deste trabalho e os restantes 15% são provenientes do estudo realizado anteriormente. Os valores das propriedades também se encontram devidamente distribuídos no intervalo de valores de interesse de cada propriedade. A Tabela 4.2 resume as principais caraterísticas dos valores de cada propriedade presente nessa base de dados. Por último, importa referir que para algumas das propriedades, o total de valores é inferior a 93 porque não existiam registos dos valores dessas propriedades para algumas das amostras do estudo anterior.

Tabela 4.2 - Tabela resumo das principais caraterísticas da base de dados para CG

Propriedade	Unidades	Núm. de Amostras	Mín.	Máx.	Méd.
RON		93	87,0	98,9	94,0
MON		93	80,1	87,8	84,8
Pressão de Vapor	kPa	93	43,7	79,1	61,2
Evaporado a 70ºC	% V/V	93	17,1	44,8	30,9
Evaporado a 100ºC	% V/V	93	40,9	61,8	51,8
Evaporado a 150ºC	% V/V	93	74,4	93,0	87,7
Ponto Final de Ebulição	ºC	74	177,1	200,0	189,2

4.2 Divisão e Pré-tratamento dos Dados

A divisão dos dados realizou-se com base nos valores da variável de saída, ou seja, com base nos valores da propriedade que se pretendia estimar, e, tendo em conta que a base de dados não é muito extensa, foi escolhida a abordagem que forma dois conjuntos de valores: conjunto de calibração e conjunto de teste. De forma a garantir que a variabilidade dos dois conjuntos se assemelhava à do conjunto inicial de valores, a divisão dos dados pelos dois conjuntos foi realizada da seguinte forma: i) ordenação dos valores do menor para o maior; ii) em cada 5 valores, o terceiro valor é colocado no conjunto de teste e os restantes 4 são colocados no conjunto de calibração. Tanto no caso da base de dados referente ao uso dos espetros NIR como na base de dados de cromatografia gasosa, esta divisão resulta num conjunto de calibração que contém cerca de 80% da base de dados inicial e um conjunto de teste que contém os restantes 20%. No caso das propriedades para as quais existem um menor número de registos, a divisão foi feita da mesma forma, mas as percentagens de dados em cada conjunto podem variar ligeiramente em relação aos conjuntos das propriedades que totalizam o número máximo de amostras, mantendo valores muito próximos das proporções obtidas nesses conjuntos. Importa referir que, como a divisão foi feita de igual forma independentemente do número total de amostras, existem propriedades para as quais o último intervalo de valores não está completo, originando situações em que o valor mais elevado para essa propriedade esteja no conjunto de teste.

No que diz respeito aos pré-tratamentos, nos dados provenientes do método espetroscópico foram aplicados todos os pré-tratamentos que foram explicados anteriormente. Além da correção de linha de base realizada pela aplicação informática associada ao espetrofotómetro, foram aplicados os seguintes pré-tratamentos: correção de dispersão multiplicativa (MSC), variação padrão (SNV), normalização padrão (SS), Savitzky-Golay (SG) e combinações de SG com

SNV e com SS. No caso do SG, existe a necessidade de definir alguns parâmetros para a aplicação do método, tendo sido definido uma janela de 11 pontos, um polinómio de terceiro grau e 3 categorias de ordem de derivação diferentes - sem derivar (SG0), primeira derivada (SG1) e segunda derivada (SG2). Nos casos em que existe combinação de SG com outros pré-tratamentos, SNV e SS, foi utilizado o SG com ordem de derivada igual a 0.

A matriz dos dados obtida a partir da cromatografia gasosa não é composta por espetros e apresenta menores dimensões, pelo que não foram aplicados os mesmos pré-tratamentos que no caso anterior. Neste caso, os pré-tratamentos aplicados foram os seguintes: MSC, SNV e SS, para os modelos em que não se realiza PCA, e SNV, combinação de MSC e de SNV e combinação de SS com SNV para os casos em que se realiza PCA.

4.3 Construção dos Modelos

Após a divisão dos dados e aplicação dos devidos pré-tratamentos, a etapa seguinte é a da construção dos modelos recorrendo aos métodos de regressão já apresentados. Tal como foi mencionado na secção anterior, foram criados 2 conjuntos de dados, conjunto de calibração e conjunto de teste, sendo que a opção por esta abordagem se deveu ao número reduzido de amostras existente. Para otimizar o(s) parâmetro(s) dos métodos de regressão, foi aplicada validação cruzada no conjunto de calibração, funcionando assim esse conjunto para treino e validação. Para a aplicação da validação cruzada recorreu-se a uma ferramenta existente no *Python*, o *GridSearch*, que permite realizar a validação cruzada testando todos os parâmetros ou combinações de parâmetros que se pretende, neste caso, o número de variáveis latentes, para o PLS, e $kernel,\ C$ e ε, para o SVR. O *GridSearch* realiza diversas iterações de validação cruzada para todas as combinações de parâmetros possíveis tendo em conta os parâmetros que se pretende testar e, após isso, escolhe a combinação de parâmetros que cria o modelo com menor MSE do conjunto de calibração, uma vez que o MSE foi a métrica escolhida para fazer essa avaliação. De seguida, é utilizada a melhor combinação de parâmetros para construir o modelo e estimar os valores da propriedade para o conjunto de calibração e de teste. Por fim, calculam-se os indicadores de desempenho do modelo (R^2, MSE e RMSE) para os pontos do conjunto de calibração e do conjunto de teste, e analisa-se graficamente o desempenho do modelo de regressão, por representação dos valores da propriedade obtidos experimentalmente e dos valores da propriedade estimados pelo modelo de regressão com melhor desempenho.

No caso em que a previsão das propriedades é feita com base nos espetros de NIR, os métodos de regressão utilizados foram PLS e SVR. Enquanto que no outro caso, foram utilizados os seguintes métodos: PCR, PLS, SVR e a combinação de PCA com SVR.

No caso do PLS, o parâmetro que se pretende otimizar é o número de variáveis latentes. Este parâmetro foi testado para o intervalo de valores inteiros de 1 até 31, uma vez que ao longo

do processo de desenvolvimento dos modelos se pode verificar que um aumento no limite superior desse intervalo não traria melhorias significativas no desempenho dos modelos e iria aumentar o tempo de processamento dos dados. Para o SVR, os parâmetros otimizados foram: $kernel = \{'linear', 'rbf'\}$, $C = \{0{,}001;\ 0{,}01;\ 0{,}1;\ 1;\ 10;\ 100;\ 1000\}$ e $\varepsilon = \{0{,}1;\ 0{,}2;\ 0{,}3;\ 0{,}4;\ 0{,}5;\ 0{,}6;\ 0{,}7;\ 0{,}8; 0{,}9\}$. Na análise por componentes principais (PCA), o número de PC's que se considerou foi 7, tanto na análise feita pelo método PCR como na análise feita por combinação de PCA com SVR, uma vez que se trata do último componente com valor próprio superior a 1 e porque os 7 PC's apresentavam uma variância acumulada superior a 90% da variância do conjunto original de valores.

4.4 Previsão de Propriedades com base em Espetroscopia NIR

Nesta secção, as previsões são obtidas a partir da conjugação dos modelos e do espetro NIR da amostra. Na Figura 4.1 encontram-se representados vários exemplos de espetros típicos de uma análise NIR de gasolina, em que cada linha representa um espetro de uma amostra diferente.

Comparando a Figura 4.1 com a Figura 2.1, é possível verificar que o espetro não é utilizado na sua totalidade, apenas são utilizados os dados correspondentes aos valores de número de onda entre 4460 e 9000 cm^{-1}, por se tratar da zona do espetro que contém informação relevante.

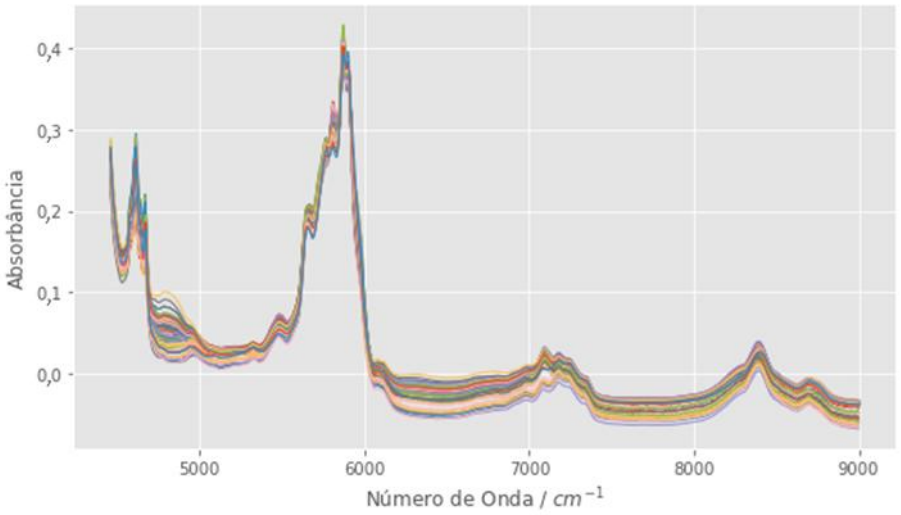


Figura 4.1 - Exemplos de espetros NIR de gasolinas.

4.4.1 Pressão de Vapor

A Tabela 4.3 apresenta os resultados para cada uma das tentativas de previsão da pressão de vapor. Na Tabela 4.3, e nas seguintes tabelas de resultados, a coluna *Parâmetros* apresenta os parâmetros otimizados utilizados na criação dos modelos com os métodos PLS e SVR, no caso do PLS, o parâmetro é o número de variáveis latentes, enquanto que no SVR é apresentado um conjunto de parâmetros sob a seguinte forma: $C;\ \varepsilon; kernel$. A verde encontra-se destacada a

metodologia que obteve melhor desempenho para o conjunto de teste que, neste caso, foi a combinação de SG com grau de derivada igual a 0, seguido de SS e com SVR como modelo de regressão, com a seguinte combinação de parâmetros: $C = 10$, $\varepsilon = 0{,}1$ e $kernel = 'linear'$. Os valores limite de repetibilidade e de reprodutibilidade para a pressão de vapor, de acordo com a EN 13016-1, são ±0,90 e ±1,58 kPa, respetivamente.

Tabela 4.3 – Tabela de resultados da previsão da pressão de vapor (NIR)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	23	1,000	0,02	0,15	0,870	9,01	3,00
SNV e PLS	24	1,000	0,02	0,13	0,872	8,84	2,97
SS e PLS	24	0,999	0,04	0,19	0,903	6,70	2,59
SG0+SNV e PLS	24	1,000	0,03	0,16	0,879	8,36	2,89
SG0+SS e PLS	24	0,999	0,04	0,21	0,904	6,66	2,58
SG1 e PLS	23	1,000	0,00	0,03	0,872	8,84	2,97
SG2 e PLS	4	0,798	14,33	3,79	0,651	24,12	4,91
MSC e SVR	1000; 0,1; linear	0,657	24,39	4,94	0,602	27,48	5,24
SNV e SVR	1000; 0,2; linear	0,997	0,25	0,50	0,884	8,04	2,83
SS e SVR	10; 0,1; linear	1,000	0,01	0,10	0,907	6,39	2,53
SG0+SNV e SVR	1000; 0,3; linear	0,997	0,22	0,47	0,905	6,53	2,56
SG0+SS e SVR	**10; 0.1; linear**	**1,000**	**0,01**	**0,10**	**0,910**	**6,20**	**2,49**
SG1 e SVR	1000; 0,1; linear	0,858	10,11	3,18	0,744	17,67	4,20
SG2 e SVR	100; 0,1; rbf	0,732	19,03	4,36	0,615	26,63	5,16

Na Figura 4.2 encontra-se representada a comparação entre os valores da previsão, obtida pelo modelo com melhor desempenho, e os valores do método, e os valores dos limites de repetibilidade e de reprodutibilidade do mesmo. É possível perceber que em relação aos pontos do conjunto de calibração, os valores da previsão se aproximam muito dos valores do método e que, nos pontos de teste, a maior parte se encontra dentro do limite de reprodutibilidade e alguns deles encontram-se até dentro do limite de repetibilidade.

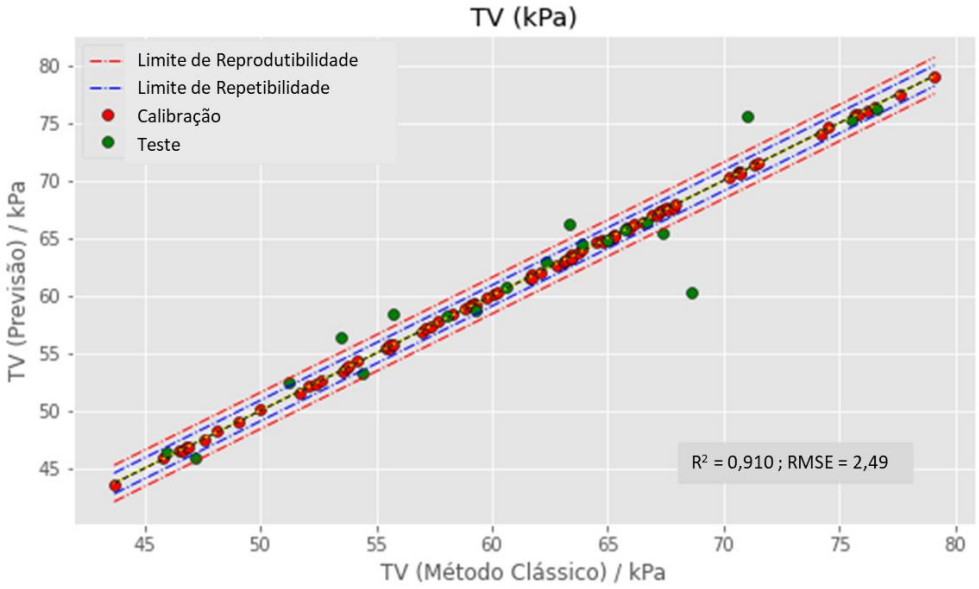


Figura 4.2 - Comparação entre os valores da previsão e os valores do método para a pressão de vapor (NIR).

4.4.2 **RON e MON**

Os valores limite de repetibilidade e de reprodutibilidade dos métodos de teste do RON e MON, referidos na ISO 5164 e 5163, respetivamente, podem ser consultados na Tabela 4.4.

Tabela 4.4 - Limites de repetibilidade de reprodutibilidade para o RON e o MON

Característica	Número de Octano	Repetibilidade	Reprodutibilidade
MON	80 a 90	±0,2	±0,9
RON	90 a 100	±0,2	±0,7

A Tabela 4.5 apresenta os resultados das metodologias com os melhores desempenhos na previsão do RON e MON, para as amostras do conjunto de teste. No caso do RON, a metodologia que obteve melhor desempenho foi a combinação de SNV com SVR, com a seguinte combinação de parâmetros: $C = 1000$, $\varepsilon = 0{,}2$ e $kernel = 'linear'$. Para o MON, a combinação de SG, com grau de derivada igual a 1, e SVR, com os parâmetros $C = 1000$, $\varepsilon = 0{,}1$ e $kernel = 'rbf'$, foi a que obteve melhor desempenho. Os resultados de cada uma das metodologias usadas na previsão do RON e MON podem ser encontrados no Apêndice A.1.

Tabela 4.5 - Tabela dos melhores resultados da previsão de RON e MON (NIR)

Propriedade	Metodologia	Parâmetros	Calibração			Teste		
			R^2	MSE	RMSE	R^2	MSE	RMSE
RON	SNV e SVR	1000; 0.2; linear	0,995	0,03	0,17	0,923	0,39	0,62
MON	SG1 e SVR	1000; 0.1; rbf	0,992	0,01	0,12	0,947	0,10	0,31

Na Figura 4.3 representa-se a comparação entre os valores da previsão, do modelo com melhor desempenho, e os valores do método para o RON, e os valores dos limites de repetibilidade e de reprodutibilidade do mesmo. Em relação aos pontos do conjunto de calibração, é possível perceber que os valores da previsão se aproximam muito dos valores do método, estando todos eles dentro das linhas do limite de reprodutibilidade e a grande maioria dentro das linhas limite de repetibilidade. Nos pontos de teste, quase a totalidade dos pontos encontra-se dentro do limite de reprodutibilidade.

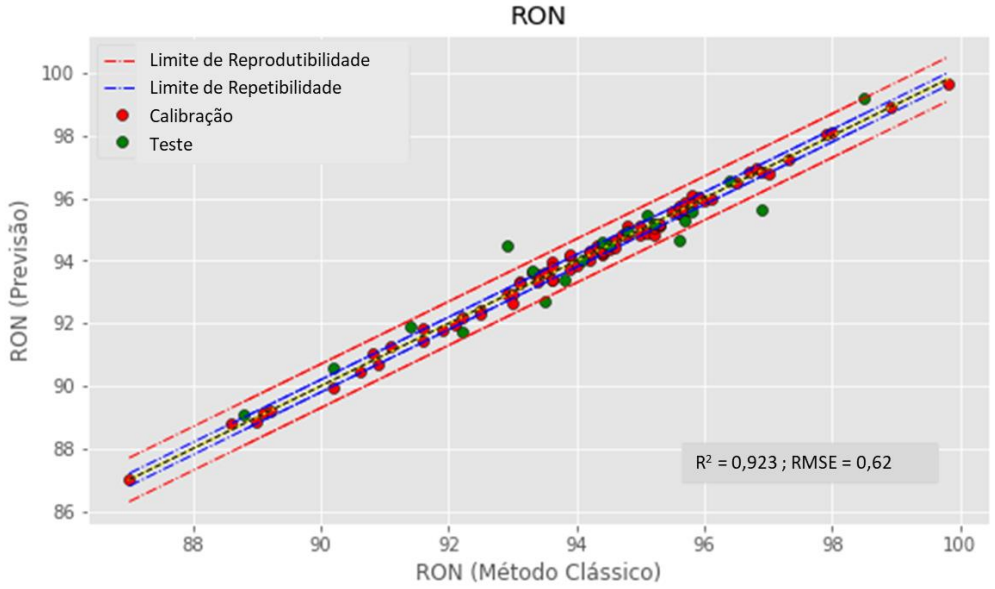


Figura 4.3 - Comparação entre os valores da previsão e os valores do método para o RON (NIR).

A comparação entre os valores da previsão com melhor desempenho e os valores do método para o MON encontra-se no Apêndice A.1.

4.4.3 Caraterísticas da Destilação

A Tabela 4.6 apresenta os melhores resultados, com base nos indicadores de desempenho dos conjunto de teste, para cada uma das caraterísticas da destilação avaliadas, E70, E100, E150, FBP. Os resultados de todas as metodologias usadas na previsão do E70, E100, E150 e FBP podem ser encontrados no Apêndice A.2.

Tabela 4.6 - Tabela dos melhores resultados da previsão das características de destilação (NIR)

Propriedade	Metodologia	Parâmetros	Calibração			Teste		
			R^2	MSE	RMSE	R^2	MSE	RMSE
E70	MSC e PLS	15	0,986	0,50	0,71	0,975	0,91	0,95
E100	SNV e SVR	100; 0,1; linear	0,997	0,06	0,25	0,980	0,40	0,63
E150	SS e SVR	1; 0,1; linear	0,998	0,02	0,15	0,990	0,13	0,35
FBP	SNV e SVR	1000; 0,1; linear	0,977	0,47	0,68	0,873	2,81	1,67

Na Figura 4.4 encontra-se representada a comparação entre os valores da melhor previsão e os valores do método para o E100. É possível verificar que, tanto para o conjunto de calibração como para o de teste, os valores previstos aproximam-se bastante dos valores do método.

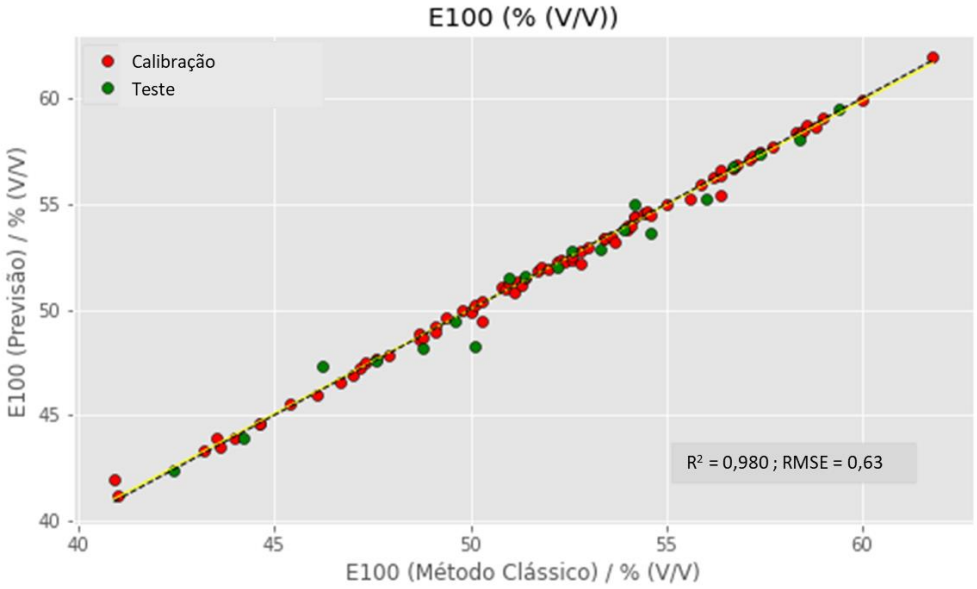


Figura 4.4 - Comparação entre os valores da previsão e os valores do método para o E100 (NIR).

As comparações entre os valores da melhor previsão e os valores do método para o E70, E150 e FBP encontram-se no Apêndice A.2.

4.4.4 Composição da Gasolina

Os limites de repetibilidade e reprodutibilidade, referidos na ISO 22854, para as propriedades provenientes da cromatografia gasosa encontram-se representados na Tabela 4.7, em que W é a média dos dois resultados que estão a ser comparados e w_B é o teor em benzeno.

Tabela 4.7 - Limites de repetibilidade e reprodutibilidade das propriedades de cromatografia gasosa

Componente ou Grupo		Repetibilidade	Reprodutibilidade
Aromáticos, %V/V		$\pm 0{,}0095 \cdot W + 0{,}1952$	$\pm 0{,}0450 \cdot W + 0{,}1384$
Olefinas, %V/V		$\pm 0{,}0185 \cdot W + 0{,}1415$	$\pm 0{,}1176 \cdot W + 0{,}5118$
Benzeno, %V/V	$w_B \geq 0{,}8$	$\pm 0{,}0147 \cdot W + 0{,}0031$	$\pm 0{,}0777 \cdot W - 0{,}0250$
	$w_B < 0{,}8$	$\pm 0{,}02$	$\pm 0{,}04$
MTBE, %V/V		$\pm 0{,}0193 \cdot W + 0{,}0024$	$\pm 0{,}0251 \cdot W + 0{,}3515$
Etanol, %V/V		$\pm 0{,}0193 \cdot W + 0{,}0024$	$\pm 0{,}0251 \cdot W + 0{,}3515$
Oxigénio, %m/m		$\pm 0{,}04$	$\pm 0{,}31$

A Tabela 4.8 apresenta os melhores resultados, com base no desempenho dos modelos para os pontos do conjunto de teste, para o teor em aromáticos, teor em benzeno, teor em etanol, teor em MTBE, teor em olefinas e teor em oxigénio. Os resultados de todas metodologias usadas na previsão destas propriedades podem ser encontrados no Apêndice A.3.

Tabela 4.8 - Tabela dos melhores resultados da previsão de propriedades de cromatografia gasosa

Propriedade	Metodologia	Parâmetros	Calibração			Teste		
			R^2	MSE	RMSE	R^2	MSE	RMSE
Aromáticos	SNV e SVR	10; 0,3; linear	0,990	0,12	0,35	0,983	0,22	0,47
Benzeno	SNV e PLS	20	0,999	0,00	0,01	0,982	0,00	0,04
Etanol	SG1 e PLS	5	0,979	0,10	0,31	0,965	0,16	0,40
MTBE	SS e PLS	8	0,972	0,18	0,43	0,976	0,19	0,43
Olefinas	MSC e SVR	1000; 0,1; linear	0,998	0,01	0,12	0,994	0,05	0,23
Oxigénio	SG0+SS e SVR	0,01; 0,1; linear	0,993	0,01	0,07	0,993	0,01	0,07

A comparação entre os valores da melhor previsão e os valores do método para o teor em etanol, encontra-se representada na Figura 4.5, juntamente com os valores dos limites de repetibilidade e reprodutibilidade do mesmo. É possível verificar que, tanto para o conjunto de calibração como para o de teste, os valores previstos aproximam-se bastante dos valores do método e que quase a totalidade dos pontos se encontra dentro das linhas dos valores limite de reprodutibilidade. As comparações entre os valores da previsão com melhor desempenho e os valores do método para as restantes propriedades encontram-se no Apêndice A.3.

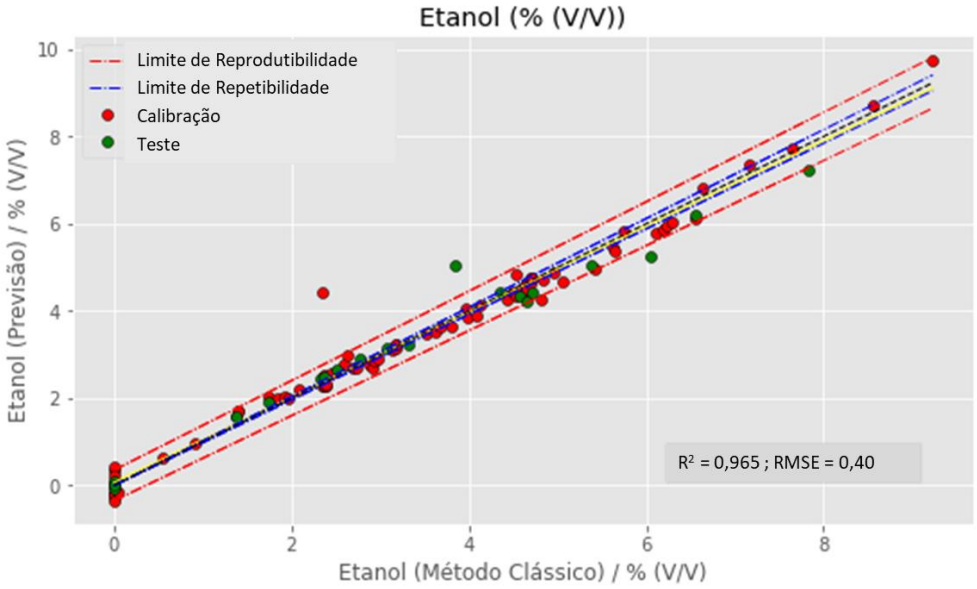


Figura 4.5 - Comparação entre os valores da previsão e os valores do método para o teor em etanol.

4.5 Previsão de Propriedades com base em Cromatografia Gasosa

Nesta secção, as previsões são obtidas a partir da conjugação dos modelos e dos resultados de cromatografia gasosa das amostras. Esses resultados são obtidos sob a forma de uma tabela com os valores do teor de diversos componentes ou grupos de componentes presentes nessas amostras. Neste caso, a tabela com os resultados proveniente da análise cromatográfica não é usada na sua totalidade porque alguns desses componentes ou grupos não se encontram presentes nas gasolinas, sendo por isso retiradas essas colunas, nas quais só constam valores iguais a zero. No Apêndice A.4 pode ser vista, como exemplo, uma parte da tabela usada na construção de modelos.

4.5.1 Pressão de Vapor

A Tabela 4.9 apresenta os resultados de cada uma das metodologias usadas na previsão da pressão de vapor com base na cromatografia gasosa. A verde encontra-se destacada a metodologia que obteve melhor desempenho nos pontos do conjunto de teste que, neste caso, foi a combinação de MSC e SVR, com a seguinte combinação de parâmetros: $C = 1000$, $\varepsilon = 0{,}2$ e $kernel = 'rbf'$.

Tabela 4.9 – Tabela de resultados da previsão da pressão de vapor (GC)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC+SNV e PCR	-	0,819	12,56	3,54	0,633	28,21	5,31
SNV e PCR	-	0,819	12,56	3,54	0,633	28,21	5,31
SS+SNV e PCR	-	0,537	32,12	5,67	0,473	40,46	6,36
MSC e PLS	8	0,975	1,72	1,31	0,935	4,96	2,23
SNV e PLS	10	0,979	1,43	1,20	0,932	5,20	2,28
SS e PLS	7	0,973	1,85	1,36	0,925	5,75	2,40
MSC e SVR	**1000; 0,2; rbf**	**0,994**	**0,43**	**0,65**	**0,965**	**2,69**	**1,64**
SNV e SVR	1000; 0,1; rbf	0,994	0,40	0,64	0,962	2,94	1,71
SS e SVR	1; 0,7; linear	0,976	1,64	1,28	0,933	5,16	2,27
MSC+SNV e PCA+SVR	100; 0,1; rbf	0,959	2,85	1,69	0,513	37,35	6,11
SNV e PCA+SVR	100; 0,1; rbf	0,959	2,85	1,69	0,513	37,35	6,11
SS+SNV e PCA+SVR	1000; 0,4; rbf	0,997	0,18	0,42	0,121	67,52	8,22

Na Figura 4.6 encontra-se representada a comparação entre os valores da previsão obtidos pelo modelo com melhor desempenho e os valores do método, e os valores dos limites de repetibilidade e de reprodutibilidade do mesmo. É possível perceber que, em relação aos pontos do conjunto de calibração, os valores da previsão se aproximam muito dos valores do método e que, nos pontos de teste, grande parte se encontra dentro dos limites de reprodutibilidade e alguns deles encontram-se dentro dos limites de repetibilidade.

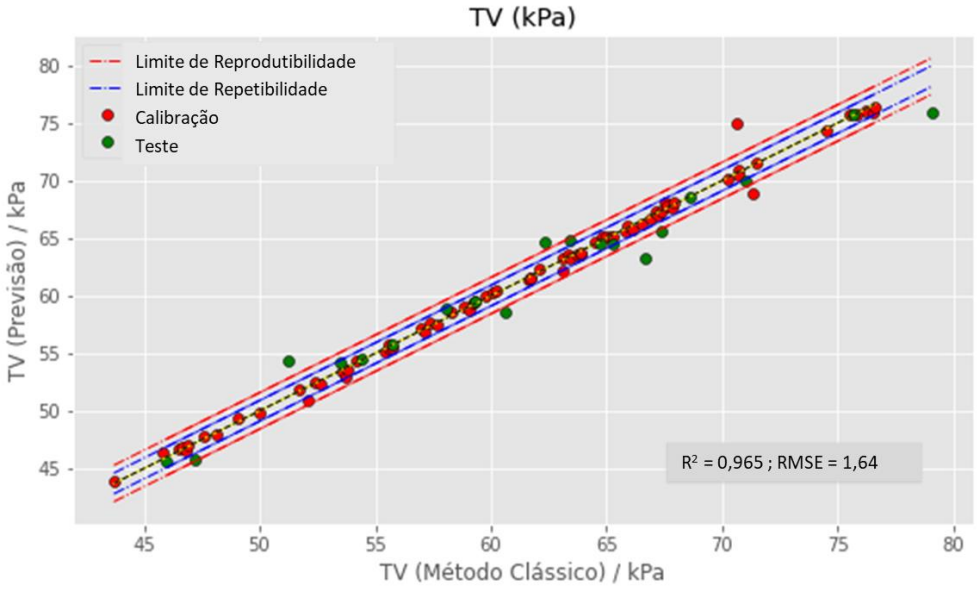


Figura 4.6 - Comparação entre os valores da previsão e os valores do método para a pressão de vapor (GC).

4.5.2 **RON e MON**

A Tabela 4.10 apresenta os resultados das metodologias com os melhores desempenhos, para os pontos do conjunto de teste, na previsão do RON e MON a partir da cromatografia gasosa. Para o RON e o MON, a metodologia que obteve melhor desempenho, em ambos os casos, foi a combinação de SS com SVR. Os resultados de cada uma das metodologias usadas na previsão do RON e MON podem ser encontrados no Apêndice A.5.

Tabela 4.10 - Tabela dos melhores resultados da previsão de RON e MON (GC)

Propriedade	Metodologia	Parâmetros	Calibração			Teste		
			R^2	MSE	RMSE	R^2	MSE	RMSE
RON	SS e SVR	10; 0.1; linear	0,982	0,09	0,30	0,973	0,14	0,38
MON	SS e SVR	1; 0.2; linear	0,935	0,11	0,33	0,893	0,21	0,46

Na Figura 4.7 encontra-se representada a comparação entre os valores da melhor previsão e os valores do método para o RON e os valores dos limites de repetibilidade e de reprodutibilidade do mesmo. Em relação aos pontos do conjunto de calibração, é possível perceber que os valores da previsão se aproximam muito dos valores do método, estando quase todos dentro das linhas do limite de reprodutibilidade e a grande parte dentro das linhas limite de repetibilidade. Nos pontos de teste, quase a totalidade dos pontos encontra-se dentro do limite de reprodutibilidade.

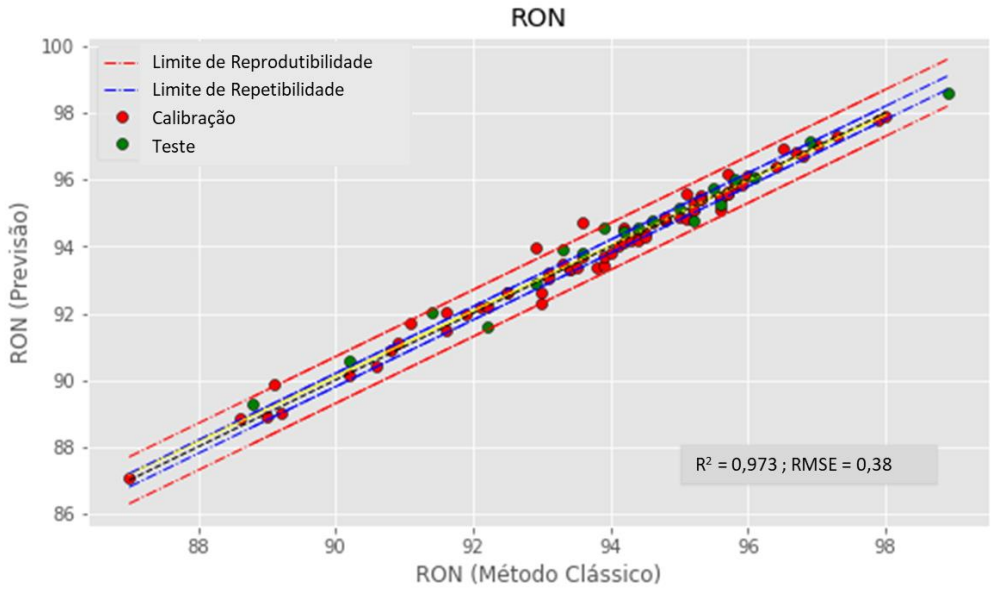


Figura 4.7 - Comparação entre os valores da previsão e os valores do método para o RON (GC).

A comparação entre os valores da melhor previsão e os valores do método, com base em cromatografia gasosa, para o MON, encontra-se representado no Apêndice A.5.

4.5.3 Caraterísticas da Destilação

A Tabela 4.11 apresenta os melhores resultados, segundo os indicadores de desempenho do conjunto de teste, para a previsão do E70, E100, E150, FBP com base na cromatografia gasosa. Os resultados de todas as metodologias usadas na previsão das caraterísticas da destilação podem ser encontrados no Apêndice A.6.

Tabela 4.11 - Tabela dos melhores resultados da previsão das caraterísticas da destilação (GC)

Propriedade	Metodologia	Parâmetros	Calibração			Teste		
			R^2	MSE	RMSE	R^2	MSE	RMSE
E70	MSC+SNV e PCA+SVR	1000; 0,7; linear	0,968	1,00	1,00	0,940	2,24	1,50
E100	SNV e SVR	100; 0,5; linear	0,991	0,2	0,44	0,986	0,34	0,58
E150	SS e SVR	10; 0,1; linear	0,998	0,03	0,17	0,849	2,13	1,46
FBP	MSC e PLS	6	0,912	1,85	1,36	0,907	2,20	1,48

A comparação entre os valores da melhor previsão e os valores do método para o E100 encontra-se representada na Figura 4.8. É possível verificar que todos os valores previstos se aproximam bastante dos valores do método.

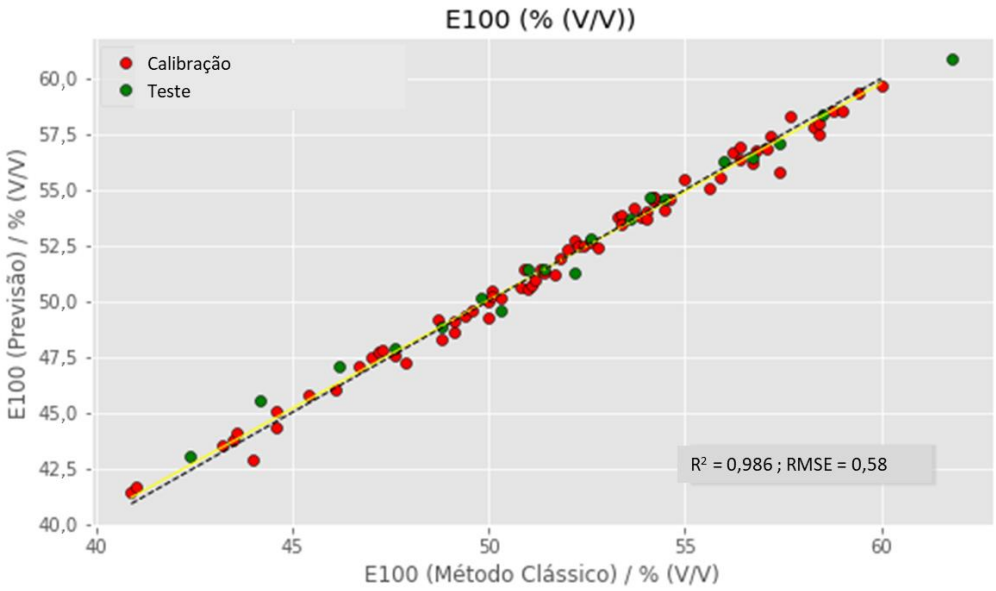


Figura 4.8 - Comparação entre os valores da previsão e os valores do método para o E100 (GC).

A comparação entre os valores da previsão com melhor desempenho e os valores do método para o E70, E150 e FBP encontram-se representados no Apêndice A.6.

5 Conclusões

Durante esta dissertação, desenvolveram-se modelos de calibração multivariável para a previsão de propriedades de gasolinas etanolizadas, com base em dados de espetroscopia NIR e de cromatografia gasosa. O desenvolvimento desses modelos tem como intuito fornecer novas alternativas de determinação das caraterísticas das gasolinas de forma mais rápida e igualmente eficaz, quando comparadas com os métodos convencionais atualmente utilizados, possibilitando a obtenção de diversas propriedades da gasolina de forma simultânea e com uma redução no tempo necessário à análise, e otimizando o processo de controlo de qualidade do produto.

Verifica-se que as metodologias utilizadas se afiguram como boas alternativas na previsão das propriedades analisadas. No caso das propriedades em que se desenvolveram modelos a partir do NIR e da cromatografia gasosa, nenhuma das duas técnicas se destaca, uma vez que das 7 propriedades em análise, 4 obtêm melhores previsões a partir da cromatografia gasosa (pressão de vapor, RON, E100 e FBP) e 3 obtêm melhores previsões a partir dos espetros de NIR (MON, E70 e E150). No caso das caraterísticas da cromatografia gasosa previstas a partir do NIR, todos os modelos apresentam bons desempenhos, com especial destaque para a previsão do teor em etanol, uma vez que a determinação do teor em etanol de forma rápida era um dos principais objetivos do trabalho.

Um aspeto com grande influência no desempenho dos modelos é a base de dados usada para o desenvolvimento dos mesmos. A dimensão da base de dados e a adequação dos dados nela presentes ao trabalho que se pretende desenvolver são fatores cruciais para o bom desempenho dos modelos, sendo necessária uma base de dados com elevado número de amostras e com variabilidade nos valores das propriedades que se pretende estimar. Tendo em conta os bons desempenhos obtidos pelos modelos desenvolvidos, apesar do número reduzido de amostras, pode concluir-se que a continuação do trabalho através da atualização contínua da base de dados com novas amostras poderá permitir aumentar o intervalo de valores avaliados para cada propriedade, tornando os modelos mais abrangentes e conferindo-lhes capacidade de analisar gasolinas com composições mais diversificadas.

Previamente à aplicação dos modelos de regressão multivariável, foram aplicadas técnicas de pré-tratamento de dados, tanto aos dados dos espetros NIR como aos dados de cromatografia gasosa, de forma a retirar o máximo de informação possível dos mesmos. Foram usadas várias técnicas e combinações das mesmas dependendo do tipo de dados e do método de regressão a aplicar após o pré-tratamento. De todas as técnicas aplicadas, a correção de dispersão

multiplicativa (MSC), a variação padrão (SNV) e a normalização padrão (SS) foram as que mais vezes constaram na metodologia com melhor capacidade de previsão.

No que diz respeito aos modelos de regressão multivariável, foram utilizadas as seguintes alternativas: regressão por componentes principais (PCR), regressão por mínimos quadrados parciais (PLS), máquinas de vetores de suporte aplicadas à regressão (SVR) e a combinação da análise por componentes principais (PCA) com as máquinas de vetores de suporte aplicadas à regressão (PCA+SVR). No caso da previsão com base nos espetros NIR, em que só foram aplicados os métodos PLS e SVR, o método de regressão com melhores resultados em 9 das 13 propriedades determinadas foi o SVR. No caso da previsão com base na cromatografia gasosa, em que foram usadas as 4 opções, o SVR obteve os melhores resultados para 5 das 7 propriedades determinadas. De forma geral, as máquinas de vetores de suporte aplicadas à regressão foram a opção que obteve melhores desempenhos na calibração dos modelos.

Conclui-se que as duas técnicas, espetroscopia NIR e cromatografia gasosa, podem ser utilizadas como base para a previsão de diversas propriedades da gasolina com bom desempenho. Afigurando-se assim como duas alternativas com muito potencial face aos métodos de análise convencionais para a determinação de propriedades como a pressão de vapor, RON, MON, características da destilação e, no caso da espetroscopia NIR, para a determinação da composição das gasolinas. Ainda no que diz respeito ao NIR, esta técnica apresenta grandes vantagens porque, além de permitir prever todas as propriedades em causa, possibilita essa previsão utilizando uma reduzida quantidade de amostra e em cerca de 15 minutos, enquanto que recorrendo à determinação através dos métodos convencionais, além de ser necessária uma maior quantidade de amostra, a combinação das várias análises necessárias resulta num processo de várias horas para a determinação das propriedades. Assim, a otimização do processo de produção pode ser conseguida com a substituição dos métodos convencionais por estas duas técnicas, reduzindo o tempo necessário para a determinação das propriedades, os custos associados à quantidade de amostra e à manutenção de equipamentos, sem prejudicar o controlo de qualidade das gasolinas.

6 Avaliação do Trabalho Realizado

6.1 Objetivos Realizados

O objetivo desta dissertação era a previsão de propriedades de gasolinas etanolizadas com base na espetroscopia NIR e em cromatografia gasosa, recorrendo ao desenvolvimento de modelos de calibração multivariável. A importância deste trabalho reside no facto de a aplicação destas técnicas de análise de gasolinas permitir a determinação simultânea de várias propriedades utilizando uma menor quantidade recursos do que os despendidos nos métodos convencionais habitualmente usados, com maior destaque, para a redução no tempo necessário à análise da amostra.

Os modelos desenvolvidos, tanto com base no NIR como com base na cromatografia gasosa, apresentam boa capacidade preditiva das propriedades avaliadas neste trabalho. Conclui-se que foram cumpridos os objetivos propostos para este trabalho, permitindo, assim, a determinação de diversas propriedades de gasolinas etanolizadas com a realização de um único ensaio.

6.2 Trabalho Futuro

O trabalho desenvolvido nesta dissertação apresenta bons desempenhos para previsão das propriedades consideradas neste estudo. Futuramente, uma atualização contínua da base de dados, com a adição de novas amostras que aumentem a variabilidade da mesma, poderá ser benéfica, pois a composição das gasolinas pode variar bastante e porque, analisando os resultados, se prevê que, com esse aumento da variabilidade da base de dados, os modelos se possam tornar mais abrangentes e generalistas, tendo capacidade para prever características de gasolinas com variadas composições e aumentando o intervalo de valores avaliados para cada propriedade, mantendo o bom desempenho.

6.3 Apreciação Final

Os resultados obtidos apresentam uma boa capacidade na previsão de propriedades e são bastante promissores, uma vez que com um aumento do número de amostras e da variabilidade na base de dados se prevê que os modelos se possam tornar mais abrangentes e generalistas. Este trabalho demonstra que este tipo de análises poderá ter um papel relevante na redução do tempo de análise sem influenciar a qualidade da mesma, permitindo a otimização do processo de produção.

7 Referências

Balabin, R. M., Safieva, R. Z., & Lomakina, E. I. (2007). Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. *Chemometrics and Intelligent Laboratory Systems*, *88*, 183–188. https://doi.org/10.1016/j.chemolab.2007.04.006

Balabin, R. M., Safieva, R. Z., & Lomakina, E. I. (2010). Gasoline Classification using Near-infrared (NIR) Spectroscopy Data: Comparison of Multivariate Techniques. *Analytica Chimica Acta*, *671*, 27–35. https://doi.org/10.1016/j.aca.2010.05.013

Bartle, K. D., & Myers, P. (2002). History of Chromatography. *Trends in Analytical Chemistry*, *21*, 547–557. https://doi.org/10.1038/1661000b0

Brereton, R. G. (2003). *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. John Wiley & Sons, Ltd.

Browne, M. W. (2000). Cross-Validation Methods. *Journal of Mathematical Psychology*, *44*, 108–132.

Brudzewski, K., Kesik, A., Kołodziejczyk, K., Zborowska, U., & Ulaczyk, J. (2006). Gasoline Quality Prediction using Gas Chromatography and FTIR Spectroscopy: An Artificial Intelligence Approach. *Fuel*, *85*, 553–558. https://doi.org/10.1016/j.fuel.2005.07.019

Burns, D. A., & Ciurczak, E. W. (2008). *Handbook of Near-Infrared Analysis* (D. A. Burns & E. W. Ciurczak (eds.); 3ª Edição). CRC Press. https://doi.org/10.1201/9781420007374-33

Chung, H., Lee, H., & Jun, C. H. (2001). Determination of Research Octane Number using NIR spectral data and ridge regression. *Bulletin of the Korean Chemical Society*, *22*(1), 37–42.

da Silva, N. C., Caribé de Góes Massa, A. R., Domingos, D., Amigo, J. M., das Virgens Rebouças, M., Pasquini, C., & Pimentel, M. F. (2019). NIR-based Octane Rating Simulator for use in Gasoline Compounding Processes. *Fuel*, *243*, 381–389. https://doi.org/10.1016/j.fuel.2019.01.145

da Silva, N. C., Cavalcanti, C. J., Honorato, F. A., Amigo, J. M., & Pimentel, M. F. (2017). Standardization from a Benchtop to a Handheld NIR Spectrometer using Mathematically Mixed NIR Spectra to Determine Fuel Quality Parameters. *Analytica Chimica Acta*, *954*, 32–42. https://doi.org/10.1016/j.aca.2016.12.018

de Godoy, L. A. F., Pedroso, M. P., Ferreira, E. C., Augusto, F., & Poppi, R. J. (2011). Prediction of the Physicochemical Properties of Gasoline by Comprehensive Two-dimensional Gas Chromatography and Multivariate Data Processing. *Journal of Chromatography A*, *1218*, 1663–1667. https://doi.org/10.1016/j.chroma.2011.01.056

Eldin, A. B. (2011). Near Infra Red Spectroscopy. Em I. Akyar (Ed.), *Wide Spectra of Quality Control* (1ª Edição, pp. 237–248). InTech.

European Committee for Standardization. (2018). *Liquid petroleum products - Vapour pressure - Part 1: Determination of air saturated vapour pressure (ASVP) and calculated dry vapour pressure equivalente (DVPE) (EN No. 13016-1:2018)* (pp. 1–16).

Flumignan, D. L., Tininis, A. G., Ferreira, F. de O., & de Oliveira, J. E. (2007). Screening Brazilian C Gasoline Quality: Application of the SIMCA Chemometric Method to Gas Chromatographic Data. *Analytica Chimica Acta*, *595*, 128–135. https://doi.org/10.1016/j.aca.2007.02.049

Forgács, E., & Cserháti, T. (2003). Gas Chromatography. Em M. Lees (Ed.), *Food Authenticity and Traceability* (1ª Edição, pp. 197–217). Woodhead Publishing. https://doi.org/10.1533/9781855737181.1.197

Gauglitz, G., & Moore, D. S. (2014). *Handbook of Spectroscopy* (G. Gauglitz & D. S. Moore (eds.); 2ª Edição). Wiley-VCH.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). Support Vector Machines and Flexible Discriminants. Em *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2ª Edição, Vol. 2, pp. 1–764). Springer. https://doi.org/10.1111/j.1751-5823.2009.00095_18.x

International Organization for Standardization. (2014a). *Petroleum products — Determination of knock characteristics of motor and aviation fuels — Motor method (ISO Standard No. 5163:2014)* (pp. 1–26).

International Organization for Standardization. (2014b). *Petroleum products — Determination of knock characteristics of motor fuels — Research method (ISO Standard No. 5164:2014)* (pp. 1–24).

International Organization for Standardization. (2016). *Liquid petroleum products — Determination of hydrocarbon types and oxygenates in automotive-motor gasoline and in ethanol (E85) automotive fuel — Multidimensional gas chromatography method (ISO Standard No. 22854:2016)* (pp. 1–19).

International Organization for Standardization. (2019). *Petroleum and related products from natural or synthetic sources - Determination of distillation characteristics at atmospheric pressure (ISO Standard No. 3405:2019)* (pp. 1–40).

Leal, A. L., Silva, A. M. S., Ribeiro, J. C., & Martins, F. G. (2020). Using Spectroscopy and Support Vector Regression to Predict Gasoline Characteristics: A Comparison of 1H NMR and NIR. *Energy and Fuels*, A-I. https://doi.org/10.1021/acs.energyfuels.0c01889

Lee, D. M., Lee, D. H., & Hwang, I. H. (2018). Gasoline Quality Assessment Using Fast Gas Chromatography and Partial Least-Squares Regression for the Detection of Adulterated Gasoline. *Energy and Fuels*, *32*, 10556–10562. https://doi.org/10.1021/acs.energyfuels.8b02368

Miller, J. N., & Miller, J. C. (2010). Multivariate Analysis. Em *Statistics and Chemometrics for Analytical Chemistry* (6ª Edição, Vol. 6, pp. 221–248). Pearson. https://doi.org/10.7861/clinmedicine.14-6-677

Muller, D. (1988). Hydrocarbon Fuel Process Analysis by Acusto Optic Tunable Filter Near-infrared Spectroscopy. *American Chemical Society*, *43*, 287–319.

Otto, M. (2017). What is Chemometrics? Em *Chemometrics: Statistics and Computer Application in Analytical Chemistry* (3ª Edição, pp. 1–13). Wiley-VCH.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Duborg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the Most Common Pre-processing Techniques for Near-infrared Spectra. *Trends in Analytical Chemistry*, *28*(10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Sharp, T. (2020). *An Introduction to Support Vector Regression (SVR)*. Towards Data Science. https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2

Skrobot, V. L., Castro, E. V. R., Pereira, R. C. C., Pasa, V. M. D., & Fortes, I. C. P. (2007). Use of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) in Gas Chromatographic (GC) Data in the Investigation of Gasoline Adulteration. *Energy and Fuels*, *21*, 3394–3400. https://doi.org/10.1021/ef0701337

Speight, J. G. (2015). Gasoline. Em *Handbook of Petroleum Product Analysis* (2ª Edição, pp. 104–126). John Wiley & Sons, Inc.

Stauffer, E., A. Dolan, J., & Newman, R. (2019). Gas Chromatography and Gas Chromatography-Mass Spectrometry. Em P. Worsfold, A. Townshend, C. Poole, & M. Miró (Eds.), *Encyclopedia of Analytical Science* (3ª Edição, pp. 235–293). Elsevier. https://doi.org/10.1016/B978-0-12-409547-2.14103-4

Varmuza, K., & Filzmoser, P. (2009). Calibration. Em *Introduction to Multivariate Statistical Analysis in Chemometrics* (1ª Edição). CRC Press.

Wang, S., Liu, S., Zhang, J., Che, X., Wang, Z., & Kong, D. (2020). Feasibility Study on Prediction of Gasoline Octane Number using NIR Spectroscopy Combined with Manifold Learning and Neural Network. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *228*, 1–8. https://doi.org/10.1016/j.saa.2019.117836

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A Basic Tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*, 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1

Apêndice A - Tabelas e Gráficos de Resultados

A.1 RON e MON com base em Espetroscopia NIR

A Tabela A.1 apresenta os resultados para cada uma das metodologias usadas na previsão do RON com base na espetroscopia NIR. A verde encontra-se destacada a metodologia que obteve melhor desempenho para as amostras do conjunto de teste.

Tabela A.1 - Tabela de resultados da previsão do RON (NIR)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	13	0,982	0,09	0,31	0,917	0,41	0,64
SNV e PLS	13	0,982	0,10	0,31	0,917	0,42	0,65
SS e PLS	15	0,984	0,09	0,29	0,911	0,44	0,67
SG0+SNV e PLS	13	0,981	0,10	0,32	0,915	0,43	0,65
SG0+SS e PLS	15	0,983	0,09	0,30	0,910	0,45	0,67
SG1 e PLS	10	0,995	0,03	0,17	0,886	0,57	0,75
SG2 e PLS	5	0,948	0,28	0,53	0,654	1,73	1,31
MSC e SVR	1000; 0,5; linear	0,958	0,22	0,47	0,894	0,53	0,73
SNV e SVR	**1000; 0,2; linear**	**0,995**	**0,03**	**0,17**	**0,923**	**0,39**	**0,62**
SS e SVR	0,1; 0,3; linear	0,976	0,13	0,36	0,916	0,42	0,65
SG0+SNV e SVR	10; 0,2; linear	0,971	0,16	0,39	0,898	0,51	0,71
SG0+SS e SVR	0,1; 0.4; linear	0,973	0,15	0,38	0,912	0,44	0,66
SG1 e SVR	1000; 0,3; rbf	0,989	0,06	0,24	0,900	0,50	0,71
SG2 e SVR	1000; 0,1; rbf	0,998	0,01	0,09	0,833	0,83	0,91

A Tabela A.2 apresenta os resultados para cada uma das tentativas de previsão do MON com base na espetroscopia NIR. A verde encontra-se destacada a metodologia que obteve melhor desempenho com base nos indicadores referentes ao conjunto de teste.

Tabela A.2 - Tabela de resultados da previsão do MON (NIR)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	10	0,912	0,16	0,40	0,855	0,26	0,51
SNV e PLS	10	0,913	0,16	0,40	0,864	0,25	0,50
SS e PLS	14	0,943	0,11	0,33	0,860	0,25	0,50
SG0+SNV e PLS	10	0,911	0,17	0,41	0,865	0,24	0,49
SG0+SS e PLS	7	0,853	0,27	0,52	0,877	0,22	0,47
SG1 e PLS	5	0,896	0,19	0,44	0,921	0,14	0,38
SG2 e PLS	10	0,985	0,03	0,17	0,760	0,43	0,66
MSC e SVR	1000; 0,1; linear	0,913	0,16	0,40	0,898	0,18	0,43
SNV e SVR	100; 0,2; linear	0,970	0,06	0,24	0,818	0,33	0,57
SS e SVR	0,01; 0,2; linear	0,882	0,22	0,47	0,892	0,20	0,44
SG0+SNV e SVR	100;0,2; linear	0,965	0,06	0,25	0,833	0,30	0,55
SG0+SS e SVR	0,01; 0.2; linear	0,882	0,22	0,47	0,892	0,20	0,44
SG1 e SVR	**1000; 0,1; rbf**	**0,992**	**0,01**	**0,12**	**0,947**	**0,10**	**0,31**
SG2 e SVR	1000; 0.2; rbf	0,981	0,03	0,19	0,780	0,40	0,63

Na Figura A.1 encontra-se representada a comparação entre os valores da melhor previsão e os valores do método para o MON, e os valores dos limites de repetibilidade e de reprodutibilidade do mesmo.

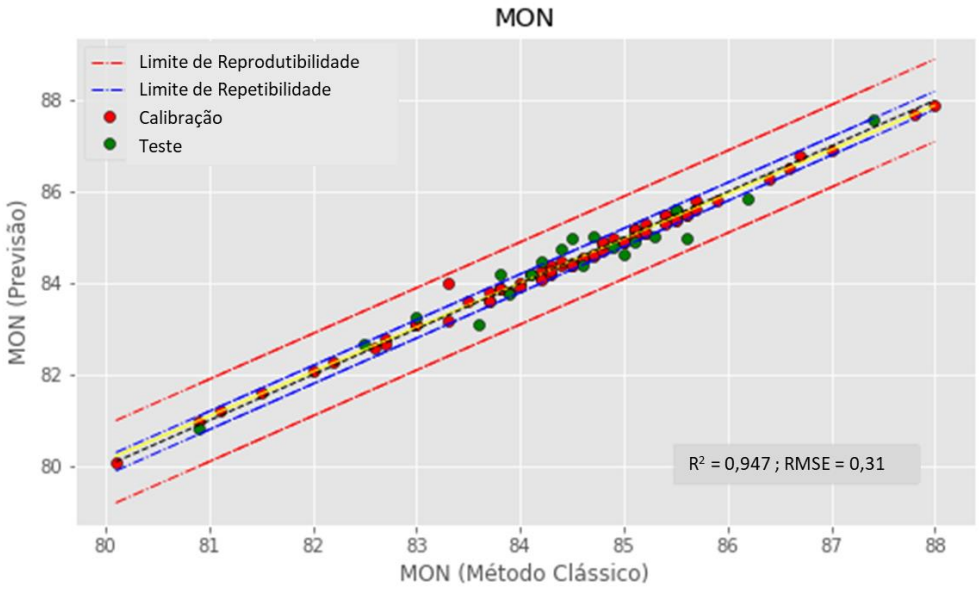


Figura A.1 - Comparação entre os valores da previsão e os valores do método para o MON (NIR).

A.2 Caraterísticas da Destilação com base em Espetroscopia NIR

A Tabela A.3 apresenta os resultados para cada uma das metodologias usadas na previsão do E70 com base na espetroscopia NIR. A verde encontra-se destacada a metodologia que obteve melhor desempenho segundo os indicadores do conjunto de teste.

Tabela A.3 - Tabela de resultados da previsão do E70 (NIR)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	**15**	**0,986**	**0,50**	**0,71**	**0,975**	**0,91**	**0,95**
SNV e PLS	15	0,985	0,52	0,72	0,973	1,00	1,00
SS e PLS	16	0,982	0,65	0,81	0,972	1,01	1,00
SG0+SNV e PLS	15	0,984	0,57	0,76	0,971	1,06	1,03
SG0+SS e PLS	16	0,981	0,68	0,83	0,967	1,19	1,09
SG1 e PLS	7	0,976	0,85	0,92	0,924	2,77	1,66
SG2 e PLS	10	0,986	0,51	0,71	0,700	10,95	3,31
MSC e SVR	1000; 0,6; linear	0,930	2,46	1,57	0,958	1,54	1,24
SNV e SVR	100; 0,1; linear	0,981	0,69	0,83	0,973	1,00	1,00
SS e SVR	1; 0,1; linear	0,981	0,68	0,83	0,969	1,13	1,06
SG0+SNV e SVR	100; 0,1; linear	0,979	0,76	0,87	0,966	1,23	1,11
SG0+SS e SVR	1; 0.1; linear	0,980	0,70	0,84	0,966	1,24	1,12
SG1 e SVR	1000; 0,1; rbf	0,975	0,89	0,94	0,944	2,06	1,44
SG2 e SVR	1000; 0.1; rbf	1,000	0,01	0,10	0,755	8,96	2,99

A Tabela A.4 apresenta os resultados para cada uma das metodologias usadas na previsão do E100 com base na espetroscopia NIR. A verde encontra-se destacada a metodologia que obteve melhor desempenho para as amostras do conjunto de teste.

Tabela A.4 - Tabela de resultados da previsão do E100 (NIR)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	20	0,999	0,02	0,15	0,968	0,66	0,81
SNV e PLS	19	0,998	0,03	0,18	0,976	0,49	0,70
SS e PLS	23	0,999	0,02	0,13	0,969	0,62	0,79
SG0+SNV e PLS	19	0,998	0,04	0,19	0,977	0,47	0,69
SG0+SS e PLS	23	0,999	0,02	0,15	0,969	0,63	0,79
SG1 e PLS	8	0,990	0,20	0,45	0,947	1,07	1,04
SG2 e PLS	14	0,999	0,02	0,14	0,897	2,09	1,45
MSC e SVR	1000; 0,8; linear	0,965	0,74	0,86	0,959	0,83	0,91
SNV e SVR	**100; 0,1; linear**	**0,997**	**0,06**	**0,25**	**0,980**	**0,40**	**0,63**
SS e SVR	1; 0,1; linear	0,997	0,06	0,24	0,978	0,44	0,66
SG0+SNV e SVR	100; 0,1; linear	0,997	0,07	0,26	0,978	0,46	0,68
SG0+SS e SVR	1; 0.1; linear	0,997	0,07	0,26	0,979	0,43	0,65
SG1 e SVR	1000; 0,1; rbf	0,994	0,13	0,36	0,974	0,53	0,73
SG2 e SVR	1000; 0.1; rbf	1,000	0,01	0,10	0,937	1,29	1,14

A Tabela A.5 apresenta os resultados para cada uma das metodologias usadas na previsão do E150 com base na espetroscopia NIR. A verde encontra-se destacada a metodologia que obteve melhor desempenho com base nos indicadores referentes ao conjunto de teste.

Tabela A.5 - Tabela de resultados da previsão do E150 (NIR)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	19	0,998	0,02	0,15	0,984	0,21	0,46
SNV e PLS	20	0,999	0,02	0,13	0,987	0,17	0,41
SS e PLS	21	0,999	0,02	0,13	0,989	0,15	0,38
SG0+SNV e PLS	20	0,998	0,02	0,14	0,986	0,18	0,42
SG0+SS e PLS	21	0,998	0,02	0,15	0,988	0,15	0,39
SG1 e PLS	27	1,000	0,00	0,00	0,914	1,11	1,05
SG2 e PLS	26	1,000	0,00	0,00	0,805	2,51	1,58
MSC e SVR	1000; 0,1; linear	0,988	0,15	0,39	0,990	0,13	0,37
SNV e SVR	100; 0,1; linear	0,998	0,02	0,16	0,990	0,13	0,36
SS e SVR	**1; 0,1; linear**	**0,998**	**0,02**	**0,15**	**0,990**	**0,13**	**0,35**
SG0+SNV e SVR	1000; 0,2; linear	0,996	0,05	0,23	0,985	0,19	0,44
SG0+SS e SVR	1; 0.1; linear	0,998	0,03	0,17	0,990	0,13	0,36
SG1 e SVR	1000; 0,1; rbf	0,999	0,02	0,13	0,986	0,17	0,42
SG2 e SVR	1000; 0.1; rbf	0,999	0,01	0,09	0,916	1,08	1,04

A Tabela A.6 apresenta os resultados para cada uma das metodologias usadas na previsão do FBP com base na espetroscopia NIR. A verde encontra-se destacada a metodologia que obteve melhor desempenho para as amostras do conjunto de teste.

Tabela A.6 - Tabela de resultados da previsão do FBP (NIR)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	20	0,999	0,02	0,13	0,858	3,14	1,77
SNV e PLS	28	1,000	0,00	0,01	0,843	3,47	1,86
SS e PLS	10	0,920	1,64	1,28	0,742	5,69	2,39
SG0+SNV e PLS	28	1,000	0,00	0,01	0,856	3,18	1,78
SG0+SS e PLS	10	0,918	1,69	1,30	0,741	5,71	2,39
SG1 e PLS	6	0,990	0,21	0,46	0,696	6,70	2,59
SG2 e PLS	11	1,000	0,00	0,03	0,811	4,16	2,04
MSC e SVR	1000; 0,6; linear	0,834	3,39	1,84	0,736	5,81	2,41
SNV e SVR	**1000; 0,1; linear**	**0,977**	**0,47**	**0,68**	**0,873**	**2,81**	**1,67**
SS e SVR	10; 0,8; linear	0,975	0,52	0,72	0,831	3,73	1,93
SG0+SNV e SVR	1000; 0,7; linear	0,981	0,39	0,63	0,830	3,74	1,93
SG0+SS e SVR	10; 0.8; linear	0,975	0,51	0,72	0,836	3,62	1,90
SG1 e SVR	1000; 0,1; rbf	0,956	0,91	0,95	0,794	4,54	2,13
SG2 e SVR	1000; 0.1; rbf	1,000	0,01	0,10	0,851	3,28	1,81

Na Figura A.2 encontra-se representada a comparação entre os valores da melhor previsão e os valores do método para o E70.

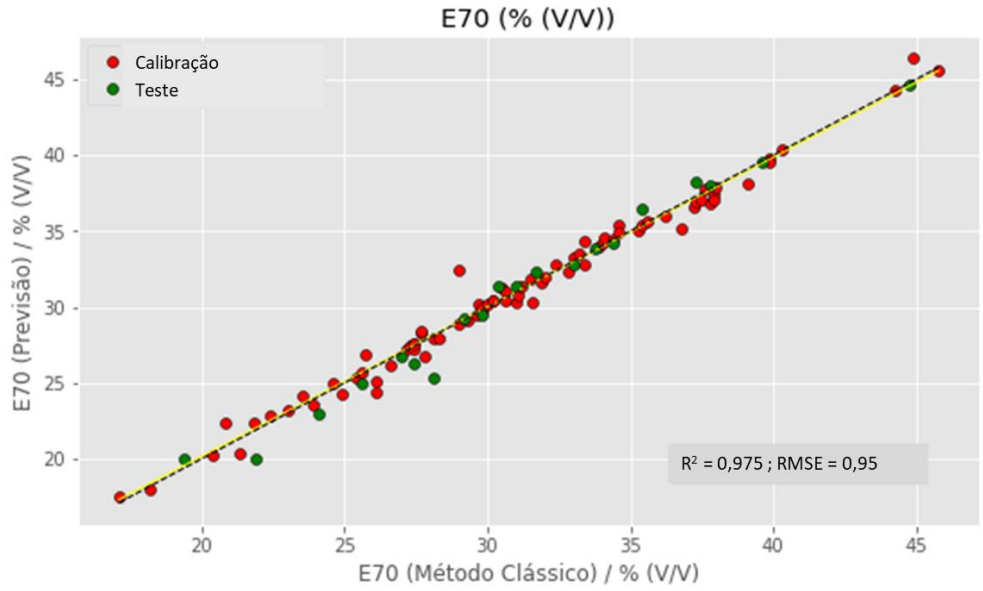


Figura A.2 - Comparação entre os valores da previsão e os valores do método para o E70 (NIR).

Na Figura A.3 encontra-se representada a comparação entre os valores da previsão com melhor desempenho e os valores do método para o E150.

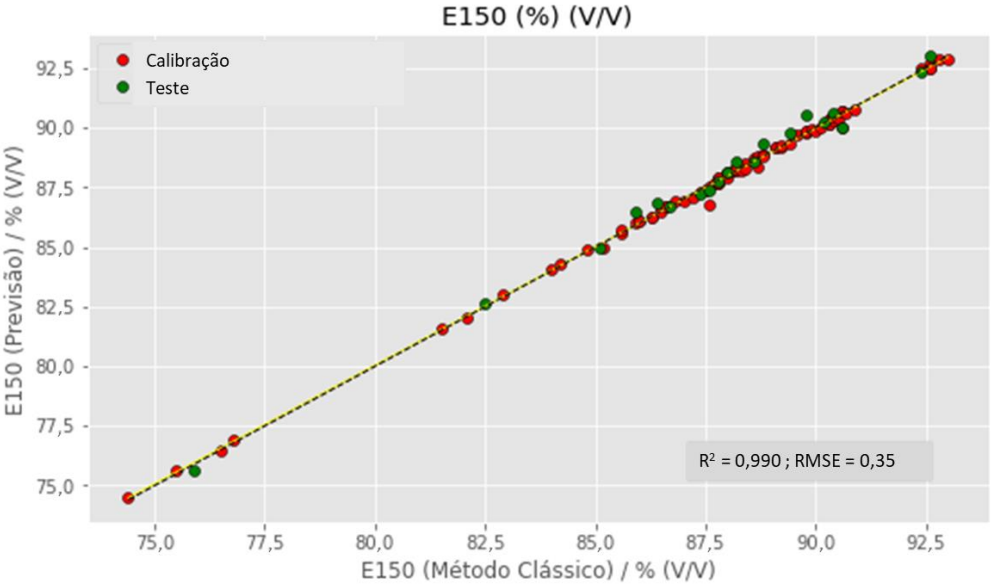


Figura A.3 - Comparação entre os valores da previsão e os valores do método para o E150 (NIR).

Os valores limite de repetibilidade e de reprodutibilidade para o ponto final de ebulição, de acordo com a ISO 3405, são ±3,3 e ±7,1 kPa, respetivamente.

Na Figura A.4 encontra-se representada a comparação entre os valores da melhor previsão e os valores do método para o FBP, e os valores dos limites de repetibilidade e de reprodutibilidade do mesmo.

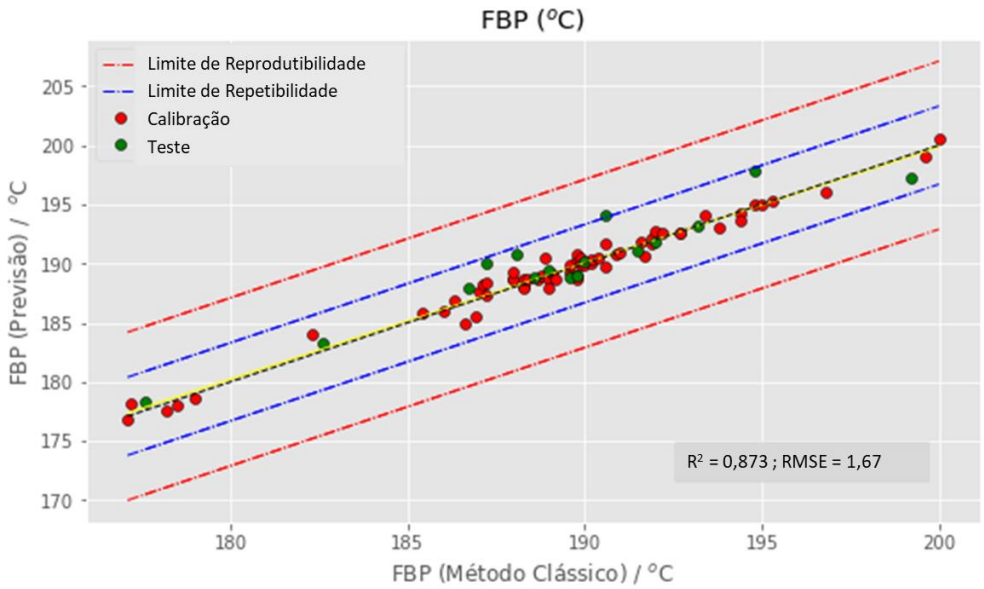


Figura A.4 - Comparação entre os valores da previsão e os valores do método para o FBP (NIR).

A.3 Composição da Gasolina com base em Espetroscopia NIR

A Tabela A.7 apresenta os resultados para cada uma das metodologias usadas na previsão do teor em aromáticos com base na espetroscopia NIR. A verde encontra-se destacada a metodologia que obteve melhor desempenho para as amostras do conjunto de teste.

Tabela A.7 - Tabela de resultados da previsão do teor em aromáticos

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	9	0,987	0,16	0,40	0,979	0,27	0,52
SNV e PLS	10	0,988	0,15	0,39	0,980	0,26	0,51
SS e PLS	10	0,987	0,16	0,40	0,967	0,43	0,66
SG0+SNV e PLS	10	0,987	0,15	0,39	0,980	0,26	0,51
SG0+SS e PLS	10	0,987	0,16	0,40	0,967	0,43	0,66
SG1 e PLS	19	1,000	0,00	0,02	0,942	0,76	0,87
SG2 e PLS	17	1,000	0,00	0,03	0,892	1,40	1,18
MSC e SVR	1000; 0,2; linear	0,990	0,13	0,36	0,983	0,22	0,47
SNV e SVR	**10; 0,3; linear**	**0,990**	**0,12**	**0,35**	**0,983**	**0,22**	**0,47**
SS e SVR	0,01; 0,1; linear	0,986	0,17	0,41	0,975	0,32	0,57
SG0+SNV e SVR	10; 0,2; linear	0,990	0,13	0,35	0,982	0,23	0,48
SG0+SS e SVR	0,01; 0.1; linear	0,986	0,17	0,41	0,975	0,32	0,57
SG1 e SVR	1000; 0,4; rbf	0,992	0,10	0,32	0,970	0,38	0,62
SG2 e SVR	1000; 0.1; rbf	0,999	0,01	0,09	0,953	0,61	0,78

A Tabela A.8 apresenta os resultados para cada uma das metodologias usadas na previsão do teor em benzeno com base na espetroscopia NIR. A verde encontra-se destacada a metodologia que obteve melhor desempenho com base nos indicadores do conjunto de teste.

Tabela A.8 - Tabela de resultados da previsão do teor em benzeno

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	19	0,999	0,00	0,01	0,980	0,00	0,04
SNV e PLS	**20**	**0,999**	**0,00**	**0,01**	**0,982**	**0,00**	**0,04**
SS e PLS	19	0,999	0,00	0,01	0,977	0,00	0,04
SG0+SNV e PLS	20	0,999	0,00	0,01	0,981	0,00	0,04
SG0+SS e PLS	19	0,999	0,00	0,01	0,976	0,00	0,05
SG1 e PLS	21	1,000	0,00	0,00	0,971	0,00	0,05
SG2 e PLS	22	1,000	0,00	0,00	0,973	0,00	0,05
MSC e SVR	1000; 0,1; linear	0,958	0,00	0,06	0,944	0,00	0,07
SNV e SVR	10; 0,1; linear	0,959	0,00	0,06	0,942	0,01	0,07
SS e SVR	0,1; 0,1; linear	0,952	0,00	0,07	0,920	0,01	0,08
SG0+SNV e SVR	10; 0,1; linear	0,959	0,00	0,06	0,941	0,01	0,07
SG0+SS e SVR	0,1; 0.1; linear	0,952	0,00	0,07	0,919	0,01	0,08
SG1 e SVR	100; 0,1; rbf	0,952	0,00	0,07	0,942	0,00	0,07
SG2 e SVR	10; 0.1; rbf	0,946	0,01	0,07	0,936	0,01	0,07

A Tabela A.9 apresenta os resultados para cada uma das metodologias usadas na previsão do teor em etanol com base na espetroscopia NIR. A verde encontra-se destacada a metodologia que obteve melhor desempenho para as amostras do conjunto de teste.

Tabela A.9 - Tabela de resultados da previsão do teor em etanol

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	6	0,963	0,18	0,42	0,959	0,19	0,43
SNV e PLS	6	0,963	0,18	0,42	0,958	0,19	0,44
SS e PLS	7	0,960	0,19	0,43	0,963	0,17	0,41
SG0+SNV e PLS	6	0,963	0,18	0,42	0,958	0,19	0,44
SG0+SS e PLS	7	0,960	0,19	0,43	0,963	0,17	0,41
SG1 e PLS	**5**	**0,979**	**0,10**	**0,31**	**0,965**	**0,16**	**0,40**
SG2 e PLS	18	1,000	0,00	0,03	0,802	0,90	0,95
MSC e SVR	100; 0,1; rbf	0,959	0,19	0,44	0,957	0,20	0,44
SNV e SVR	100; 0,1; rbf	0,960	0,19	0,44	0,957	0,20	0,44
SS e SVR	0,001; 0,1; linear	0,956	0,21	0,46	0,957	0,20	0,44
SG0+SNV e SVR	100; 0,1; rbf	0,960	0,19	0,44	0,957	0,20	0,44
SG0+SS e SVR	0,001; 0.1; linear	0,956	0,21	0,46	0,957	0,20	0,44
SG1 e SVR	100; 0,1; rbf	0,964	0,17	0,41	0,952	0,22	0,47
SG2 e SVR	1000; 0.1; rbf	0,998	0,01	0,10	0,786	0,98	0,99

A Tabela A.10 apresenta os resultados para cada uma das metodologias usadas na previsão do teor em MTBE com base na espetroscopia NIR. A verde encontra-se destacada a metodologia que obteve melhor desempenho com base nos indicadores do conjunto de teste.

Tabela A.10 - Tabela de resultados da previsão do teor em MTBE

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	10	0,979	0,14	0,37	0,965	0,27	0,52
SNV e PLS	10	0,980	0,13	0,37	0,964	0,28	0,53
SS e PLS	**8**	**0,972**	**0,18**	**0,43**	**0,976**	**0,19**	**0,43**
SG0+SNV e PLS	10	0,979	0,14	0,37	0,964	0,28	0,53
SG0+SS e PLS	8	0,972	0,18	0,43	0,976	0,19	0,43
SG1 e PLS	16	1,000	0,00	0,03	0,975	0,19	0,43
SG2 e PLS	21	1,000	0,00	0,02	0,903	0,75	0,86
MSC e SVR	1000; 0,1; linear	0,975	0,17	0,41	0,973	0,21	0,46
SNV e SVR	10; 0,1; linear	0,975	0,16	0,41	0,972	0,21	0,46
SS e SVR	0,01; 0,1; linear	0,972	0,19	0,43	0,970	0,23	0,48
SG0+SNV e SVR	10; 0,1; linear	0,975	0,17	0,41	0,972	0,21	0,46
SG0+SS e SVR	0,01; 0.1; linear	0,972	0,19	0,43	0,970	0,23	0,48
SG1 e SVR	1000; 0,1; rbf	0,995	0,03	0,18	0,973	0,21	0,46
SG2 e SVR	1000; 0.1; rbf	0,999	0,01	0,09	0,941	0,45	0,67

A Tabela A.11 apresenta os resultados para cada uma das metodologias usadas na previsão do teor em olefinas com base na espetroscopia NIR. A verde encontra-se destacada a metodologia que obteve melhor desempenho para as amostras do conjunto de teste.

Tabela A.11 - Tabela de resultados da previsão do teor em olefinas

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	15	0,999	0,01	0,10	0,992	0,07	0,26
SNV e PLS	16	0,999	0,01	0,09	0,993	0,06	0,25
SS e PLS	15	0,998	0,01	0,12	0,991	0,08	0,28
SG0+SNV e PLS	16	0,999	0,01	0,09	0,993	0,06	0,25
SG0+SS e PLS	15	0,998	0,02	0,12	0,991	0,08	0,28
SG1 e PLS	15	1,000	0,00	0,04	0,987	0,12	0,34
SG2 e PLS	30	1,000	0,00	0,00	0,936	0,57	0,76
MSC e SVR	**1000; 0,1; linear**	**0,998**	**0,01**	**0,12**	**0,994**	**0,05**	**0,23**
SNV e SVR	100; 0,1; linear	0,999	0,01	0,09	0,993	0,07	0,26
SS e SVR	1; 0,1; linear	0,999	0,01	0,08	0,989	0,10	0,31
SG0+SNV e SVR	100; 0,1; linear	0,999	0,01	0,09	0,992	0,07	0,26
SG0+SS e SVR	10; 0.1; linear	0,999	0,01	0,08	0,988	0,11	0,33
SG1 e SVR	1000; 0,1; rbf	0,999	0,01	0,09	0,994	0,05	0,23
SG2 e SVR	1000; 0.1; rbf	0,999	0,01	0,09	0,970	0,27	0,52

A Tabela A.12 apresenta os resultados para cada uma das metodologias usadas na previsão do teor em oxigénio com base na espetroscopia NIR. A verde encontra-se destacada a metodologia que obteve melhor desempenho com base nos indicadores do conjunto de teste.

Tabela A.12 - Tabela de resultados da previsão do teor em oxigénio

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC e PLS	9	0,994	0,00	0,07	0,991	0,01	0,08
SNV e PLS	13	0,998	0,00	0,04	0,992	0,01	0,08
SS e PLS	7	0,992	0,01	0,08	0,992	0,01	0,07
SG0+SNV e PLS	13	0,998	0,00	0,04	0,992	0,01	0,08
SG0+SS e PLS	7	0,992	0,01	0,08	0,992	0,01	0,07
SG1 e PLS	8	0,997	0,00	0,05	0,984	0,01	0,10
SG2 e PLS	18	1,000	0,00	0,01	0,748	0,17	0,42
MSC e SVR	100; 0,1; linear	0,993	0,01	0,07	0,993	0,01	0,07
SNV e SVR	1000; 0,1; rbf	0,996	0,00	0,06	0,992	0,01	0,08
SS e SVR	0,01; 0,1; linear	0,993	0,01	0,07	0,993	0,01	0,07
SG0+SNV e SVR	1000; 0,1; rbf	0,996	0,00	0,06	0,991	0,01	0,08
SG0+SS e SVR	**0,01; 0.1; linear**	**0,993**	**0,01**	**0,07**	**0,993**	**0,01**	**0,07**
SG1 e SVR	1000; 0,1; rbf	0,993	0,01	0,08	0,985	0,01	0,10
SG2 e SVR	100; 0.1; rbf	0,990	0,01	0,09	0,915	0,06	0,24

Na Figura A.5 encontra-se representada a comparação entre os valores da melhor previsão e os valores do método para o teor em aromáticos, e os valores dos limites de repetibilidade e de reprodutibilidade do mesmo.

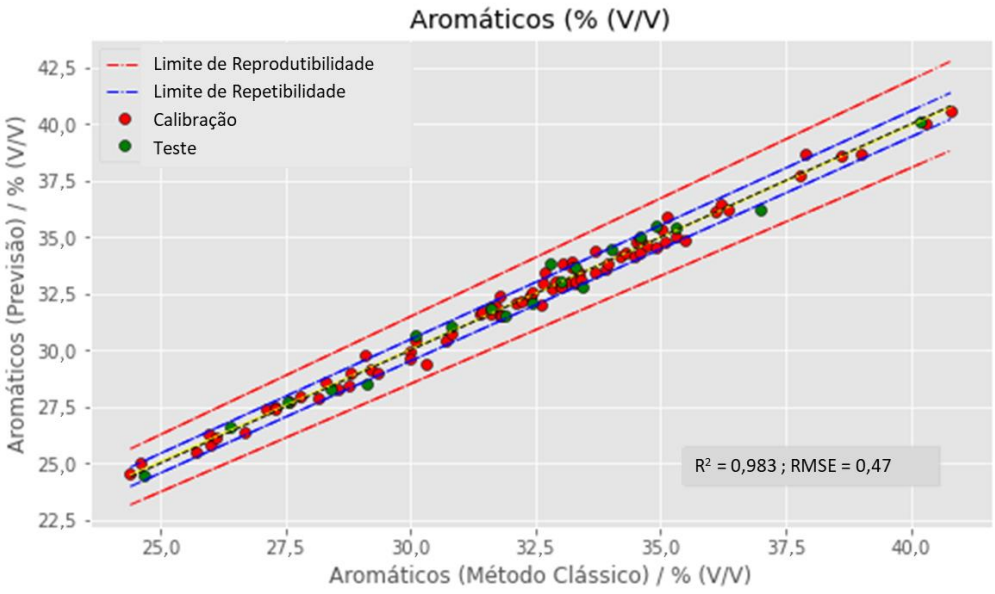


Figura A.5 - Comparação entre os valores da previsão e os valores do método para o teor em aromáticos.

Na Figura A.6 encontra-se representada a comparação entre os valores da previsão com melhor desempenho e os valores do método para o teor em benzeno, e os valores dos limites de repetibilidade e de reprodutibilidade do mesmo.

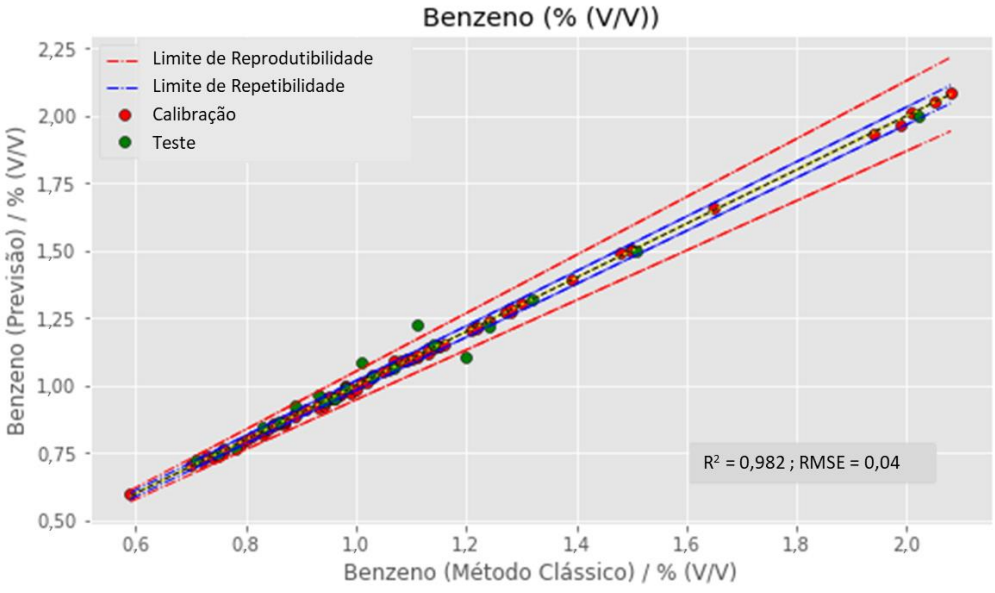


Figura A.6 - Comparação entre os valores da previsão e os valores do método para o teor em benzeno.

Na Figura A.7 encontra-se representada a comparação entre os valores da melhor previsão e os valores do método para o teor em MTBE, e os valores dos limites de repetibilidade e de reprodutibilidade do mesmo.

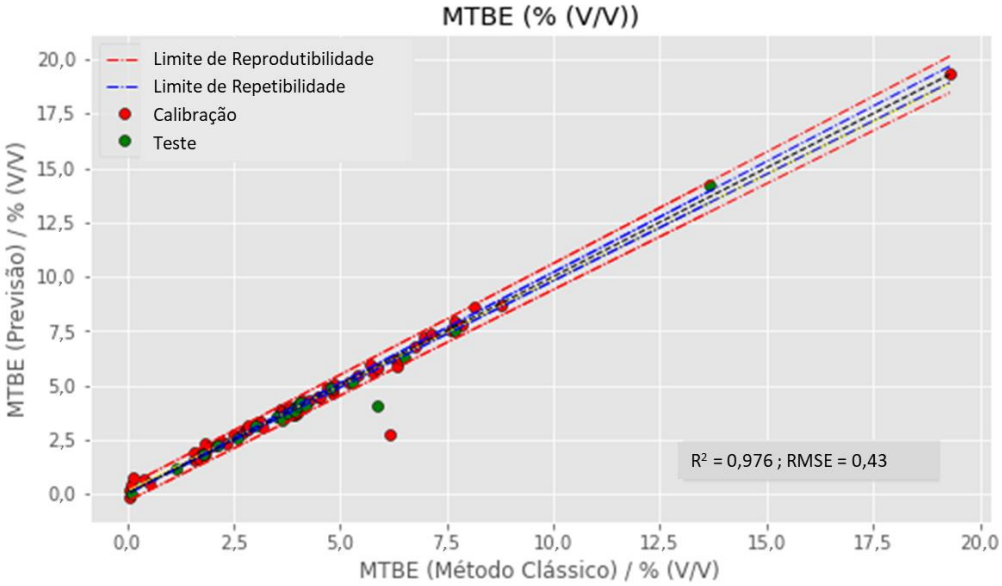


Figura A.7 - Comparação entre os valores da previsão e os valores do método para o teor em MTBE.

Na Figura A.8 encontra-se representada a comparação entre os valores da previsão com melhor desempenho e os valores do método para o teor em olefinas, e os valores dos limites de repetibilidade e de reprodutibilidade do mesmo.

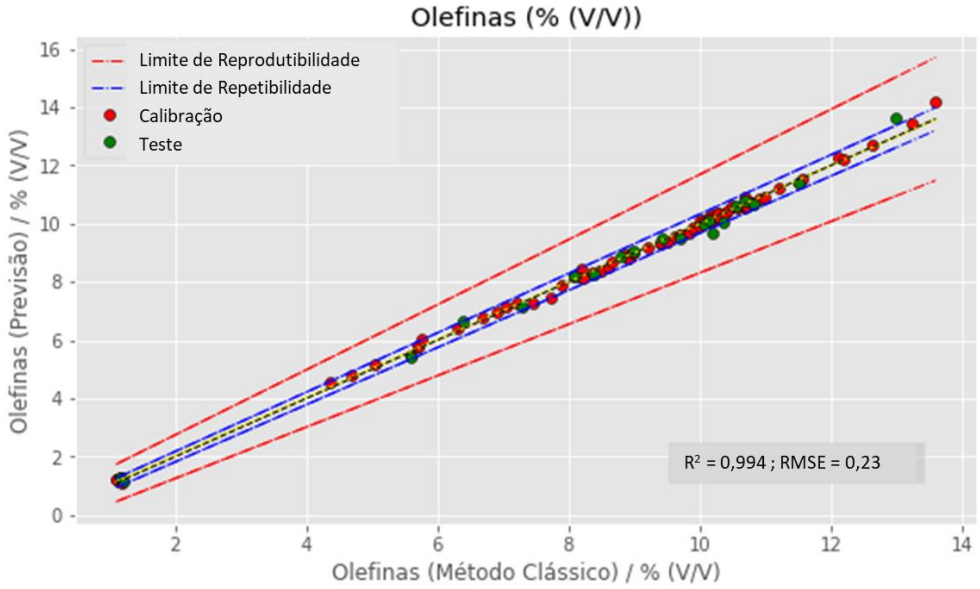


Figura A.8 - Comparação entre os valores da previsão e os valores do método para o teor em olefinas.

Na Figura A.9 encontra-se representada a comparação entre os valores da melhor previsão e os valores do método para o teor em oxigénio, e os valores dos limites de repetibilidade e de reprodutibilidade do mesmo.

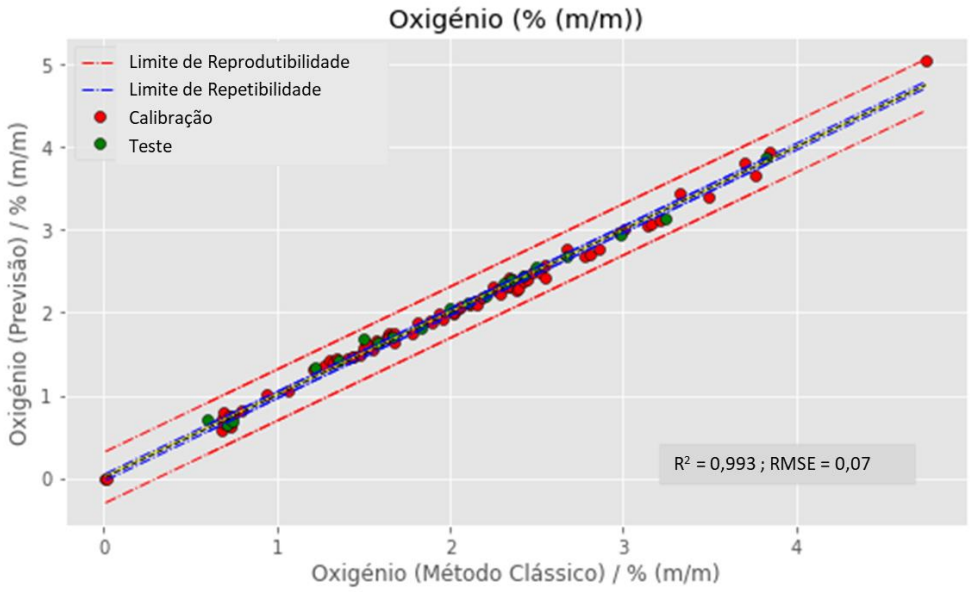


Figura A.9 - Comparação entre os valores da previsão e os valores do método para o teor em oxigénio.

A.4 Tabela de Resultados da Cromatografia Gasosa

Na Tabela A.13 encontra-se representado um exemplo de uma tabela com os dados cromatográficos utilizados para a previsão de propriedades a partir de cromatografia gasosa.

Tabela A.13 - Tabela de resultados de cromatografia gasosa (Parte 1)

Nº Amostra	Tolueno	Aromáticos C9	Aromáticos C8	Aromáticos C11+	Aromáticos C10	Benzeno	Ciclo Olefinas C9	Ciclo Olefinas C8	Ciclo Olefinas C7	Ciclo Olefinas C6	Ciclo Olefinas C5	Naftenos C9	Naftenos C8	Naftenos C7	Naftenos C6	Naftenos C5	Naftenos C10
42915	5.72	6.15	9.28	0.34	1.97	1.22	0.11	0.29	0.48	0.36	0.1	0.46	0.91	1.36	1.86	0.83	0.16
42824	6.34	6.45	8.36	0.33	1.9	1.03	0.14	0.35	0.54	0.46	0.12	0.51	1.02	1.64	1.71	0.6	0.19
42823	7.69	6.84	9.13	0.37	2.02	1.24	0.13	0.33	0.54	0.42	0.1	0.51	1.01	1.58	1.66	0.57	0.18
42911	7.17	6.66	8.74	0.35	1.94	1.1	0.12	0.31	0.51	0.41	0.09	0.49	0.97	1.57	1.67	0.59	0.17
42914	6.91	7.6	9.44	0.42	2.3	0.89	0.12	0.32	0.52	0.4	0.08	0.51	1.0	1.71	1.56	0.5	0.18
40023	7.74	6.7	9.35	0.35	1.53	0.71	0.07	0.19	0.3	0.22	0.06	0.4	0.83	1.7	1.5	0.56	0.12
41684	6.72	9.41	11.2	0.36	2.82	1.07	0.25	0.57	0.63	0.66	0.18	0.82	1.44	2.18	2.61	0.52	0.37
41687	9.01	8.04	9.72	0.35	2.31	1.39	0.17	0.43	0.63	0.5	0.13	0.67	1.18	1.8	1.65	0.5	0.31
1661	7.03	6.86	8.7	0.37	2.05	1.01	0.11	0.29	0.47	0.35	0.1	0.45	0.91	1.42	1.52	0.56	0.16
41683	5.89	8.29	9.91	0.72	3.08	0.89	0.13	0.35	0.58	0.44	0.11	0.58	1.14	1.52	1.75	0.63	0.2
37172	7.91	7.06	10.34	0.25	1.48	0.76	0.09	0.23	0.39	0.28	0.08	0.59	1.06	2.08	1.72	0.55	0.22
1663	7.68	13.38	9.55	0.65	3.07	1.16	0.09	0.22	0.37	0.29	0.08	0.39	0.79	1.14	1.39	0.57	0.13
42913	6.6	9.37	11.71	0.52	2.88	1.3	0.17	0.42	0.63	0.51	0.14	0.68	1.21	1.66	1.52	0.45	0.3
41688	8.82	7.5	10.27	0.42	2.23	1.5	0.11	0.28	0.45	0.34	0.1	0.47	0.93	1.45	1.57	0.59	0.16
42912	6.05	9.14	9.58	0.56	3.03	0.86	0.19	0.47	0.65	0.56	0.16	0.74	1.29	1.96	1.83	0.43	0.34
671	9.9	9.24	11.6	0.54	2.8	1.24	0.15	0.38	0.59	0.45	0.11	0.66	1.18	1.64	1.37	0.33	0.29
7409	18.5	5.81	7.61	0.22	1.93	1.07	0.12	0.33	0.56	0.41	0.13	0.61	1.12	1.78	1.88	0.51	0.23
41686	7.51	7.75	10.77	0.38	2.4	1.28	0.16	0.4	0.58	0.48	0.13	0.64	1.14	1.67	1.61	0.53	0.28
41685	9.9	7.9	11.02	0.34	2.21	1.51	0.18	0.43	0.61	0.51	0.13	0.66	1.16	1.72	1.57	0.47	0.29
7413	14.25	7.29	9.98	0.27	2.21	1.13	0.14	0.35	0.58	0.43	0.12	0.67	1.26	2.12	1.84	0.4	0.26
37166	9.22	7.05	11.96	0.24	1.07	0.78	0.0	0.03	0.04	0.02	0.0	0.26	0.58	1.73	1.55	0.65	0.0
35487	8.88	9.21	12.42	0.39	2.5	1.14	0.14	0.37	0.62	0.45	0.13	0.78	1.35	2.27	1.7	0.31	0.33
37165	11.64	8.25	8.88	0.38	2.35	0.93	0.13	0.33	0.55	0.39	0.12	0.57	1.08	1.62	1.5	0.41	0.22
7407	8.9	6.74	11.52	0.15	1.0	0.85	0.0	0.0	0.03	0.02	0.0	0.19	0.49	1.51	1.87	0.72	0.0
36356	13.73	7.58	9.74	0.38	2.28	1.03	0.14	0.36	0.6	0.44	0.13	0.69	1.22	1.85	1.51	0.34	0.32
41242	6.29	9.29	10.63	0.77	3.96	0.87	0.2	0.48	0.64	0.58	0.18	0.82	1.49	2.08	1.68	0.4	0.37
41241	5.56	7.86	9.32	0.6	3.01	0.87	0.13	0.33	0.52	0.38	0.12	0.59	1.08	1.38	1.6	0.6	0.25
38582	11.58	8.07	9.88	0.39	2.28	0.98	0.13	0.34	0.57	0.41	0.12	0.66	1.16	1.82	1.52	0.35	0.3
42916	8.56	7.12	9.62	0.39	2.15	1.27	0.12	0.3	0.5	0.37	0.11	0.51	0.97	1.51	1.51	0.51	0.18
7414	8.94	7.43	11.73	0.22	1.56	0.98	0.06	0.15	0.25	0.18	0.06	0.42	0.84	1.87	1.81	0.55	0.13

Tabela A.14 - Tabela de resultados de cromatografia gasosa (Parte 2)

Olefinas C9	Olefinas C8	Olefinas C7	Olefinas C6	Olefinas C5	Olefinas C4	Olefinas C10	MTBE	Etanol	Parafinas C9	Parafinas C8	Parafinas C7	Parafinas C6	Parafinas C5	Parafinas C4	Parafinas C10	Saturados C11+
0.23	0.51	1.04	1.99	3.13	0.2	0.15	1.65	2.39	1.13	8.05	8.13	17.74	19.24	1.71	0.49	0.6
0.28	0.65	1.36	2.47	3.7	0.21	0.15	1.58	5.43	1.18	7.85	11.67	17.18	12.13	1.28	0.55	0.61
0.26	0.6	1.24	2.28	3.48	0.21	0.14	2.12	2.97	1.19	8.77	10.65	16.21	12.54	1.79	0.55	0.7
0.25	0.58	1.22	2.21	3.37	0.2	0.13	2.21	4.81	1.15	8.18	10.81	16.45	12.68	1.73	0.52	0.63
0.25	0.58	1.2	2.1	2.79	0.14	0.16	3.1	3.33	1.27	10.29	11.16	15.3	10.49	2.08	0.57	0.72
0.14	0.35	0.69	1.31	2.02	0.14	0.05	3.77	2.33	1.1	11.72	9.97	15.31	13.91	3.82	0.4	0.65
0.45	0.89	1.48	3.2	4.74	0.28	0.24	1.91	2.74	1.69	5.52	6.8	11.45	13.81	0.99	0.9	1.03
0.32	0.65	1.25	2.53	4.45	0.26	0.17	3.16	2.34	1.39	7.67	6.77	12.29	13.8	2.56	0.76	0.87
0.23	0.52	1.05	1.92	2.96	0.2	0.15	2.84	5.39	1.16	10.55	8.74	14.35	14.05	2.31	0.51	0.69
0.28	0.61	1.12	2.28	4.09	0.36	0.14	0.15	6.05	1.4	8.12	6.81	12.83	16.44	1.33	0.63	1.14
0.17	0.39	0.74	1.46	2.58	0.21	0.05	3.52	2.39	1.29	10.33	7.13	12.88	15.93	4.64	0.55	0.67
0.19	0.44	0.86	1.56	2.46	0.2	0.12	1.83	3.18	1.04	7.18	7.38	13.33	14.93	3.51	0.42	0.43
0.33	0.67	1.24	2.54	4.46	0.3	0.18	2.34	6.17	1.53	8.89	6.12	10.58	11.57	1.32	0.75	0.95
0.22	0.48	0.93	1.83	3.13	0.23	0.12	2.57	4.64	1.13	7.79	7.0	13.38	15.8	2.47	0.49	0.6
0.36	0.72	1.34	2.83	4.81	0.32	0.19	3.85	5.87	1.48	7.25	6.49	10.92	11.96	1.88	0.82	1.03
0.31	0.65	1.17	2.24	3.48	0.25	0.16	3.12	2.64	1.63	13.11	6.84	9.42	8.39	2.3	0.75	1.08
0.23	0.49	0.99	2.23	4.4	0.27	0.12	0.08	0.0	1.26	6.1	5.74	12.25	15.97	5.46	0.7	0.91
0.31	0.64	1.19	2.43	4.2	0.27	0.16	2.33	3.8	1.48	9.5	6.66	12.27	13.43	2.08	0.7	0.85
0.35	0.69	1.26	2.55	4.32	0.25	0.18	1.56	2.69	1.55	10.42	6.71	11.46	11.88	1.53	0.74	1.21
0.25	0.53	1.04	2.31	4.58	0.26	0.13	0.11	0.0	1.6	10.87	6.57	11.11	13.84	1.46	0.79	1.26
0.0	0.13	0.26	0.32	0.36	0.06	0.0	4.27	0.0	0.92	12.69	7.88	14.46	18.11	4.83	0.17	0.34
0.27	0.58	1.12	2.4	4.46	0.26	0.18	4.06	0.0	1.59	7.89	6.81	9.67	10.95	5.07	0.83	0.88
0.24	0.53	1.03	2.14	3.88	0.24	0.13	6.98	2.37	1.23	7.11	6.57	10.86	11.93	3.91	0.66	0.81
0.0	0.11	0.24	0.32	0.35	0.07	0.0	3.97	0.0	0.89	11.21	7.53	14.62	19.72	6.18	0.17	0.59
0.25	0.56	1.1	2.36	4.49	0.27	0.14	4.08	0.56	1.41	7.29	6.84	10.54	11.4	4.64	0.79	0.93
0.37	0.76	1.35	2.97	4.81	0.38	0.19	2.22	2.92	1.7	5.1	6.25	9.72	13.06	5.41	0.89	1.09
0.25	0.54	0.97	2.04	3.72	0.34	0.13	3.97	4.56	1.47	10.81	5.64	12.11	16.19	1.49	0.65	0.91
0.24	0.53	1.03	2.19	4.12	0.31	0.16	5.19	1.74	1.42	8.3	6.71	10.49	11.37	3.95	0.75	0.91
0.24	0.51	0.99	1.99	3.35	0.24	0.13	2.78	7.65	1.2	9.25	6.92	12.41	13.27	2.04	0.54	0.77
0.11	0.29	0.57	1.11	2.08	0.15	0.04	3.92	0.0	1.31	13.16	7.48	12.77	16.5	2.03	0.44	0.86

A.5 RON e MON com base em Cromatografia Gasosa

A Tabela A.15 apresenta os resultados para cada uma das metodologias usadas na previsão do RON com base na cromatografia gasosa. A verde encontra-se destacada a metodologia que obteve melhor desempenho para as amostras do conjunto de teste.

Tabela A.15 - Tabela de resultados da previsão do RON (GC)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC+SNV e PCR	-	0,913	0,43	0,66	0,915	0,45	0,67
SNV e PCR	-	0,913	0,43	0,66	0,915	0,45	0,67
SS+SNV e PCR	-	0,747	1,26	1,12	0,753	1,31	1,14
MSC e PLS	6	0,949	0,25	0,50	0,931	0,37	0,61
SNV e PLS	16	0,973	0,13	0,37	0,967	0,18	0,42
SS e PLS	7	0,960	0,20	0,45	0,944	0,30	0,55
MSC e SVR	10; 0,1; linear	0,959	0,21	0,45	0,957	0,23	0,48
SNV e SVR	1000; 0,1; linear	0,971	0,14	0,38	0,969	0,17	0,41
SS e SVR	**10; 0,1; linear**	**0,982**	**0,09**	**0,30**	**0,973**	**0,14**	**0,38**
MSC+SNV e PCA+SVR	1000; 0,1; linear	0,912	0,44	0,66	0,923	0,41	0,64
SNV e PCA+SVR	1000; 0,1; linear	0,912	0,44	0,66	0,923	0,41	0,64
SS+SNV e PCA+SVR	10; 0,9; linear	0,742	1,28	1,13	0,762	1,26	1,12

A Tabela A.16 apresenta os resultados para cada uma das metodologias usadas na previsão do MON com base na cromatografia gasosa. A verde encontra-se destacada a metodologia que obteve melhor desempenho com base nos indicadores do conjunto de teste.

Tabela A.16 - Tabela de resultados da previsão do MON (GC)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC+SNV e PCR	-	0,770	0,39	0,62	0,808	0,38	0,62
SNV e PCR	-	0,770	0,39	0,62	0,808	0,38	0,62
SS+SNV e PCR	-	0,747	0,43	0,66	0,731	0,54	0,73
MSC e PLS	6	0,893	0,18	0,43	0,883	0,23	0,48
SNV e PLS	6	0,893	0,18	0,43	0,884	0,23	0,48
SS e PLS	4	0,894	0,18	0,42	0,889	0,22	0,47
MSC e SVR	100; 0,3; rbf	0,955	0,08	0,28	0,833	0,33	0,58
SNV e SVR	100; 0,3; rbf	0,957	0,07	0,27	0,838	0,32	0,57
SS e SVR	**1; 0,2; linear**	**0,935**	**0,11**	**0,33**	**0,893**	**0,21**	**0,46**
MSC+SNV e PCA+SVR	1000; 0,7; linear	0,758	0,41	0,64	0,813	0,37	0,61
SNV e PCA+SVR	1000; 0,7; linear	0,758	0,41	0,64	0,813	0,37	0,61
SS+SNV e PCA+SVR	1000; 0,3; linear	0,742	0,44	0,66	0,738	0,52	0,72

Na Figura A.10 encontra-se representada a comparação entre os valores da melhor previsão e os valores do método para o MON, e os valores dos limites de repetibilidade e de reprodutibilidade do mesmo.

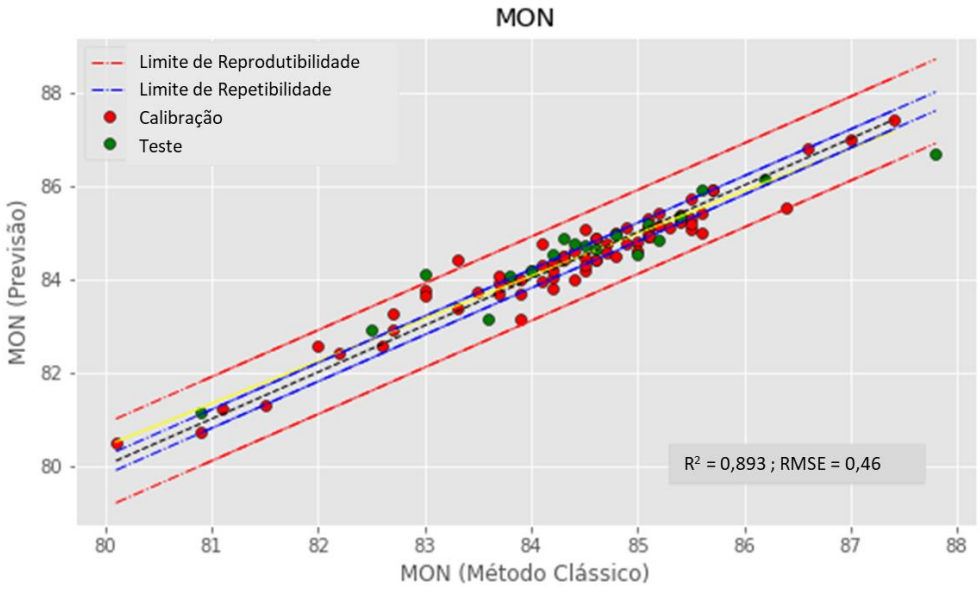


Figura A.10 - Comparação entre os valores da previsão e os valores do método para o MON (GC).

A.6 Caraterísticas da Destilação com base em Cromatografia Gasosa

A Tabela A.17 apresenta os resultados para cada uma das metodologias usadas na previsão do MON com base na cromatografia gasosa. A verde encontra-se destacada a metodologia que obteve melhor desempenho para as amostras do conjunto de teste.

Tabela A.17 - Tabela de resultados da previsão do E70 (GC)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC+SNV e PCR	-	0,970	0,96	0,98	0,937	2,35	1,53
SNV e PCR	-	0,970	0,96	0,98	0,937	2,35	1,53
SS+SNV e PCR	-	0,695	9,63	3,10	0,345	24,27	4,93
MSC e PLS	11	0,978	0,69	0,83	0,910	3,33	1,83
SNV e PLS	11	0,981	0,60	0,78	0,912	3,26	1,81
SS e PLS	9	0,975	0,78	0,89	0,930	2,60	1,61
MSC e SVR	100; 0,3; rbf	0,990	0,31	0,56	0,932	2,52	1,59
SNV e SVR	1000; 0,4; rbf	0,995	0,15	0,39	0,918	3,04	1,74
SS e SVR	1; 0,8; linear	0,973	0,84	0,92	0,924	2,83	1,68
MSC+SNV e PCA+SVR	**1000; 0,7; linear**	**0,968**	**1,00**	**1,00**	**0,940**	**2,24**	**1,50**
SNV e PCA+SVR	1000; 0,7; linear	0,968	1,00	1,00	0,940	2,24	1,50
SS+SNV e PCA+SVR	100; 0,6; rbf	0,990	0,33	0,58	0,509	18,18	4,26

A Tabela A.18 apresenta os resultados para cada uma das metodologias usadas na previsão do MON com base na cromatografia gasosa. A verde encontra-se destacada a metodologia que obteve melhor desempenho com base nos indicadores do conjunto de teste.

Tabela A.18 - Tabela de resultados da previsão do E100 (GC)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC+SNV e PCR	-	0,976	0,50	0,71	0,974	0,61	0,78
SNV e PCR	-	0,976	0,50	0,71	0,974	0,61	0,78
SS+SNV e PCR	-	0,870	2,75	1,66	0,825	4,09	2,02
MSC e PLS	13	0,992	0,16	0,41	0,984	0,37	0,61
SNV e PLS	13	0,992	0,17	0,41	0,986	0,33	0,57
SS e PLS	20	0,994	0,12	0,34	0,981	0,45	0,67
MSC e SVR	10; 0,5; linear	0,990	0,21	0,46	0,987	0,31	0,56
SNV e SVR	**100; 0,5; linear**	**0,991**	**0,20**	**0,44**	**0,986**	**0,34**	**0,58**
SS e SVR	10; 0,2; linear	0,993	0,14	0,37	0,987	0,31	0,56
MSC+SNV e PCA+SVR	10; 0,6; linear	0,975	0,53	0,73	0,974	0,61	0,78
SNV e PCA+SVR	10; 0,6; linear	0,975	0,53	0,73	0,974	0,61	0,78
SS+SNV e PCA+SVR	1; 0,9; linear	0,866	2,83	1,68	0,828	4,03	2,01

A Tabela A.19 apresenta os resultados para cada uma das metodologias usadas na previsão do MON com base na cromatografia gasosa. A verde encontra-se destacada a metodologia que obteve melhor desempenho para as amostras do conjunto de teste.

Tabela A.19 - Tabela de resultados da previsão do E150 (GC)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC+SNV e PCR	-	0,980	0,28	0,53	0,828	2,42	1,56
SNV e PCR	-	0,980	0,28	0,53	0,828	2,42	1,56
SS+SNV e PCR	-	0,871	1,80	1,34	0,666	4,71	2,17
MSC e PLS	12	0,994	0,08	0,28	0,841	2,24	1,50
SNV e PLS	12	0,994	0,08	0,28	0,838	2,28	1,51
SS e PLS	14	0,997	0,03	0,19	0,847	2,16	1,47
MSC e SVR	10; 0,4; linear	0,993	0,09	0,31	0,836	2,31	1,52
SNV e SVR	100; 0,4; linear	0,993	0,10	0,32	0,837	2,30	1,52
SS e SVR	**10; 0,1; linear**	**0,998**	**0,03**	**0,17**	**0,849**	**2,13**	**1,46**
MSC+SNV e PCA+SVR	1000; 0,4; linear	0,978	0,30	0,55	0,831	2,38	1,54
SNV e PCA+SVR	1000; 0,4; linear	0,978	0,30	0,55	0,831	2,38	1,54
SS+SNV e PCA+SVR	1; 0,1; linear	0,815	2,57	1,60	0,647	4,98	2,23

A Tabela A.20 apresenta os resultados para cada uma das metodologias usadas na previsão do MON com base na cromatografia gasosa. A verde encontra-se destacada a metodologia que obteve melhor desempenho com base nos indicadores do conjunto de teste.

Tabela A.20 - Tabela de resultados da previsão do FBP (GC)

Metodologia	Parâmetros	Calibração			Teste		
		R^2	MSE	RMSE	R^2	MSE	RMSE
MSC+SNV e PCR	-	0,776	4,71	2,17	0,567	10,20	3,19
SNV e PCR	-	0,776	4,71	2,17	0,567	10,20	3,19
SS+SNV e PCR	-	0,761	5,02	2,24	0,809	4,49	2,12
MSC e PLS	**6**	**0,912**	**1,85**	**1,36**	**0,907**	**2,20**	**1,48**
SNV e PLS	6	0,930	1,48	1,22	0,871	3,03	1,74
SS e PLS	18	0,979	0,44	0,67	0,900	2,35	1,53
MSC e SVR	1000; 0,1; rbf	0,987	0,27	0,52	0,799	4,74	2,18
SNV e SVR	1000; 0,1; rbf	0,988	0,25	0,50	0,829	4,03	2,01
SS e SVR	1; 0,8; linear	0,941	1,23	1,11	0,895	2,48	1,58
MSC+SNV e PCA+SVR	1000; 0,1; rbf	0,998	0,04	0,19	0,576	10,00	3,16
SNV e PCA+SVR	1000; 0,1; rbf	0,998	0,04	0,19	0,576	10,00	3,16
SS+SNV e PCA+SVR	0,1; 0,8; linear	0,661	7,11	2,67	0,713	6,76	2,60

Na Figura A.11 encontra-se representada a comparação entre os valores da melhor previsão e os valores do método para o E70.

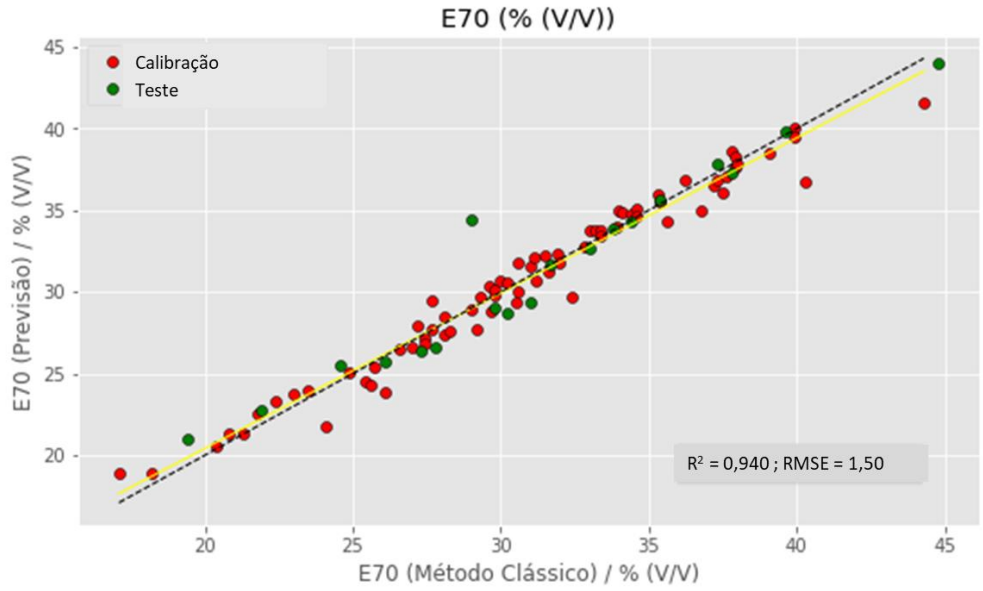


Figura A.11 - Comparação entre os valores da previsão e os valores do método para o E70 (GC).

Na Figura A.12 encontra-se representada a comparação entre os valores da previsão com melhor desempenho e os valores do método para o E150.

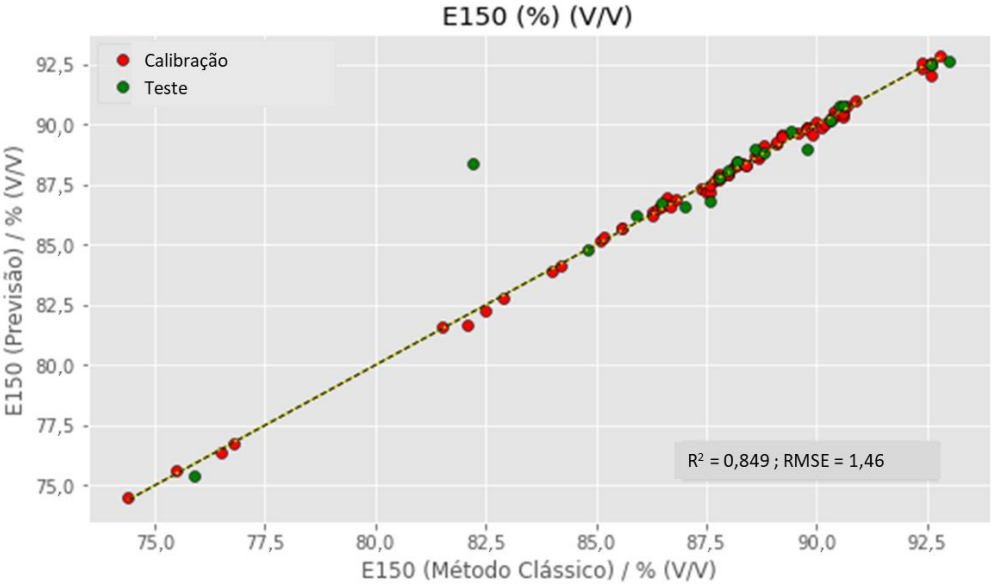


Figura A.12 - Comparação entre os valores da previsão e os valores do método para o E150 (GC).

Na Figura A.13 encontra-se representada a comparação entre os valores da melhor previsão e os valores do método para o FBP, e os valores dos limites de repetibilidade e reprodutibilidade do mesmo.

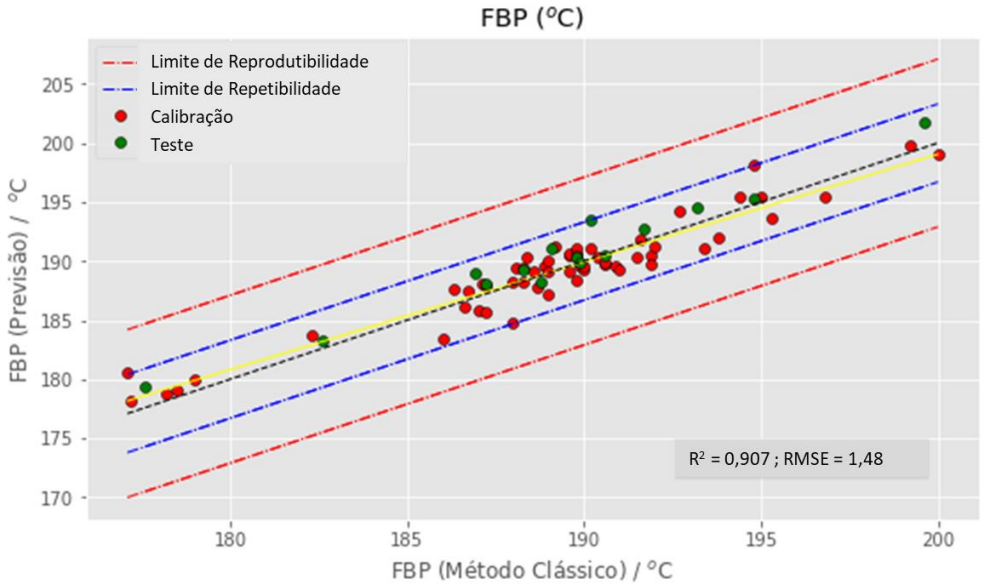


Figura A.13 - Comparação entre os valores da previsão e os valores do método para o FBP (GC).