



UNIVERSIDADE FEDERAL DO MARANHÃO – UFMA

Fundação Instituída nos termos da Lei nº 5.152, de 21/10/1966 – São Luís - Maranhão

Centro de Ciências Exatas e Tecnologia – CCET
Programa de Pós-Graduação em Engenharia Elétrica - PPGE



Análise Multivariada da Qualidade de Combustíveis Comerciais no Estado do Maranhão

Morgana Cristhya Silva dos Santos

São Luís
2025

Morgana Cristhya Silva dos Santos

Análise Multivariada da Qualidade de Combustíveis Comerciais no Estado do Maranhão

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica do Centro de Ciências Exatas e Tecnologias, da Universidade Federal do Maranhão, como requisito para a obtenção do grau de Mestre em Engenharia Elétrica.

Orientador: Allan Kardec Duailibe Barros Filho

São Luís
2025

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Cristhya Silva dos Santos, Morgana.

Análise Multivariada da Qualidade de Combustíveis
Comerciais no Estado do Maranhão / Morgana Cristhya Silva
dos Santos. - 2025.

51 f.

Orientador(a): Allan Kardec Duailibe Barros Filho.

Dissertação (Mestrado) - Programa de Pós-graduação em
Engenharia Elétrica/ccet, Universidade Federal do
Maranhão, São Luís, 2025.

1. Combustíveis. 2. Qualidade. 3. Análise
Multivariada. I. Kardec Duailibe Barros Filho, Allan. II.
Título.

CCET- Centro de Ciências Exatas e Tecnologias
Universidade Federal do Maranhão

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica intitulada *Análise Multivariada da Qualidade de Combustíveis Comerciais no Estado do Maranhão* de autoria de Morgana Cristhya Silva dos Santos.

Prof. Dr. Allan Kardec Duailibe Barros Filho
Orientador
Programa de Pós Graduação em Engenharia de Elétrica-UFMA

Prof. Dr. Ewaldo Eder Carvalho Santana
Membro da banca
Programa de Pós Graduação em Engenharia de Elétrica-UFMA

Prof. Dr. Jonathan Araújo Queiroz
Membro da banca
Universidade Estadual do Maranhão – UEMA

São Luís
2025

Agradecimentos

Primeiramente, toda a minha gratidão a Deus.

Gratidão aos familiares que me apoiaram e encorajaram ao longo da minha vida.

Agradeço ao meu orientador, professor Allan Kardec, pela sua orientação e apoio ao longo de toda minha jornada acadêmica no mestrado.

Expresso minha gratidão também, aos colegas e amigos.

Meus agradecimentos também a UFMA e ao Programa de Formação de Recursos Humanos PRH-54.1 da ANP, pelo apoio dado durante este processo.

“A persistência é o caminho do êxito.”
(Charles Chaplin)

Resumo

A qualidade dos combustíveis é um fator crucial para o desempenho dos motores e a mitigação dos impactos ambientais. No Maranhão, onde a distribuição de combustíveis possui grande importância estratégica, variações na qualidade podem gerar prejuízos econômicos e ambientais significativos. Este estudo utilizou técnicas de Análise Multivariada para avaliar parâmetros físico-químicos de combustíveis comercializados no Maranhão, empregando Análise de Componentes Principais (PCA) e Análise de Componentes Independentes (ICA). Os resultados indicaram que a PCA foi mais eficiente na separação dos grupos de diesel, evidenciando correlações significativas entre as variáveis analisadas, enquanto a ICA apresentou menor capacidade de distinção. Além disso, a identificação de *outliers* sugeriu variações na formulação, possíveis contaminações ou falhas experimentais. No caso das gasolinas, a PCA revelou padrões bem definidos entre os tipos comum e aditivado, indicando que algumas variáveis físico-químicas são determinantes para essa diferenciação. Já a análise do etanol hidratado por ICA mostrou um comportamento homogêneo, com algumas amostras discrepantes, sugerindo adulteração ou inconsistências na qualidade. Com base nesses resultados, este trabalho propõe estudos futuros para aprofundar a identificação das variáveis mais relevantes, aprimorar as técnicas de análise e fortalecer os processos de controle e fiscalização da qualidade dos combustíveis não só no estado do Maranhão, como em outras regiões do país. Dessa forma, a aplicação da Análise Multivariada se mostra uma ferramenta eficaz para garantir maior segurança, eficiência e conformidade com os padrões regulatórios.

Palavras-chave: Combustíveis; Qualidade; Análise Multivariada.

Abstract

The quality of fuels is a crucial factor for engine performance and the mitigation of environmental impacts. In Maranhão, where the distribution of fuels holds great strategic importance, variations in quality can cause significant economic and environmental damage. This study used Multivariate Analysis techniques to evaluate the physicochemical parameters of fuels marketed in Maranhão, employing Principal Component Analysis (PCA) and Independent Component Analysis (ICA). The results indicated that PCA was more efficient in separating the diesel groups, showing significant correlations between the proven variations, while ICA showed a lower capacity for distinction. Moreover, the identification of outliers suggested variations in formulation, possible contaminations, or experimental failures. In the case of gasoline, PCA revealed well-defined patterns between common and additive types, indicating that some physicochemical variables are determinants for this differentiation. An analysis of hydrated ethanol using ICA showed a homogeneous behavior, with some discrepant samples, indicating adulteration or inconsistencies in quality. Based on these findings, this work proposes future studies for in-depth results in identifying the most relevant variables, improving analysis techniques, and strengthening the processes of control and inspection of fuel quality not only in the state of Maranhão but also in other regions of the country. In this way, the application of Multivariate Analysis proves to be an effective tool to ensure greater safety, efficiency, and compliance with regulatory standards.

Keywords: Fuels; Quality; Multivariate Analysis.

Lista de Figuras

Figura 1: Unidades da federação que possuem cobertura do PMQC.....	14
Figura 2: Resultados das análises feitas pelo PMQC, sobre características físico-químicas que mais frequentemente apresentam resultados fora das especificações, referentes às amostras de óleo diesel coletadas no Estado do Maranhão, no período de 2017 a 2023.....	16
Figura 3: Resultados das análises feitas pelo PMQC, sobre características físico-químicas que mais frequentemente apresentam resultados fora das especificações, referentes às amostras de gasolina coletadas no Estado do Maranhão, no período de 2017 a 2023.....	17
Figura 4: Resultados das análises feitas pelo PMQC, sobre características físico-químicas que mais frequentemente apresentam resultados fora das especificações, referentes as amostras de etanol coletadas no Estado do Maranhão, no período de 2017 a 2023..	17
Figura 5: Esquema ilustrativo das características analisadas pelas técnicas de PCA e ICA, nesta pesquisa.	27
Figura 6: Gráfico bidimensional com os dados das amostras de óleo diesel comum S10 e S500, após a aplicação de PCA.	31
Figura 7: Gráfico de pesos dos quatro primeiros componentes principais em relação às variáveis originais do conjunto de dados de óleo diesel analisado por PCA.....	32
Figura 8: Gráfico bidimensional com os dados das amostras de óleo diesel comum S10 e S500, após a aplicação de ICA, veículos a GNV.	33
Figura 9: Gráfico bidimensional com os dados das amostras de gasolina tipo C comum e aditivada, após a aplicação de PCA	36
Figura 10: Gráfico de pesos dos quatro primeiros componentes principais em relação às variáveis originais do conjunto de dados de gasolina analisado por PCA.....	37
Figura 11: Gráfico bidimensional com os dados das amostras de gasolina tipo C comum e aditivada, após a aplicação de ICA	38
Figura 12: Gráfico bidimensional com os dados das amostras de etanol hidratado, após a aplicação de PCA.....	40
Figura 13: Gráfico de pesos dos quatro primeiros componentes principais em relação às variáveis originais do conjunto de dados de etanol hidratado, analisado por PCA.....	41
Figura 14: Gráfico bidimensional com os dados das amostras de etanol hidratado, após a aplicação de ICA	43

Lista de Tabelas

Tabela 1: Correspondentes numéricos atribuídos às cores das amostras para aplicação de PCA no Matlab.....28

Tabela 2: Valores de variância explicada e variância acumulada obtidos com a aplicação de PCA no conjunto de dados de óleo diesel comum S10 e S500.....29

Tabela 3: Valores de variância explicada e variância acumulada obtidos com a aplicação de PCA no conjunto de dados com amostras de gasolina tipo C comum e aditivada.....35

Tabela 4: Valores de variância explicada e variância acumulada obtidos com a aplicação de PCA no conjunto de dados com amostras de etanol hidratado.....39

Sumário

1. Introdução.....	10
1.1. Contextualização do problema.....	12
1.2. Objetivos do trabalho.....	12
1.2.1 Objetivo Geral	12
1.2.2 Objetivos específicos	12
2. Fundamentação Teórica	13
2.1 Combustíveis.....	13
2.2 Qualidade de Combustíveis: Conceitos e Regulamentações.....	13
2.1.2. Parâmetros Físico-Químicos de Qualidade	15
2.2 Comercialização de Combustíveis	18
2.3 Trabalhos Relacionados	21
3. Análise Estatística Multivariada.....	22
3.1 Análise de Componentes Principais (PCA)	23
3.2 Análise de Componentes Independentes (ICA)	24
4. Métodos	26
5. Resultados e Discussão	29
5.1 Modelagem por PCA para o Óleo Diesel.....	29
5.2 Modelagem por ICA para o Óleo Diesel.....	33
5.3 Modelagem por PCA para a Gasolina.....	34
5.4 Modelagem por ICA para a Gasolina.....	38
5.5 Modelagem por PCA para o Etanol	39
5.6 Modelagem por ICA para o Etanol Hidratado	42
6. Conclusão.....	44
6.1 Trabalhos futuros	45
Referências	46

1. Introdução

A qualidade dos combustíveis é fundamental para o desempenho dos motores e a redução de emissões de poluentes, sendo um tema de interesse em engenharia, economia e sustentabilidade. O estado do Maranhão destaca-se como um polo de distribuição de combustíveis no Brasil, devido à proximidade com refinarias e acesso a mercados regionais e internacionais. Contudo, a variação na qualidade dos combustíveis afeta o desempenho dos motores e o meio ambiente (BRASIL, 2023).

No Maranhão, a crescente demanda por combustíveis fósseis, aliada aos desafios de fiscalização, exige um estudo aprofundado sobre a qualidade dos combustíveis comercializados. A origem e a logística de distribuição, além das condições de armazenamento, influenciam as propriedades dos produtos, o que torna essencial a aplicação de métodos estatísticos avançados para garantir a conformidade com as normas e a qualidade para o mercado (IBP, 2024).

No Brasil, a ANP instituiu o Programa de Monitoramento de Qualidade de Combustíveis (PMQC) para identificar inconformidades e garantir a qualidade dos produtos vendidos. Dados de 2020 mostram que o Maranhão contava com 1.477 postos e monitorava 170 municípios, com 1.192 amostras analisadas (AGUIAR, et al., 2008). No Maranhão, o Laboratório de Análise e Pesquisa em Química Analítica do Petróleo da Universidade Federal do Maranhão (LAPQAP/UFMA) realiza essas análises, criando um banco de dados de alta dimensionalidade.

A pesquisa sobre a qualidade de combustíveis tem evoluído em sintonia com o desenvolvimento de técnicas de análise multivariada. Na última década, a aplicação dessas técnicas tornou-se crucial para compreender a complexidade dos dados de qualidade de combustíveis, especialmente em contextos onde a variabilidade nos parâmetros físico-químicos é alta e influencia diretamente o desempenho dos combustíveis.

Para avaliar a qualidade dos combustíveis, diversos parâmetros são analisados, como densidade, viscosidade, teor de enxofre, ponto de fulgor e octanagem, além do potencial de poluição e eficiência energética. A análise integrada desses parâmetros exige métodos estatísticos robustos, como a Análise Multivariada, que possibilita a identificação de padrões, correlações e anomalias nos dados (OLIVEIRA, 2021; RIBEIRO e SCHIRMER, 2017). Técnicas como a Análise de Componentes Principais (PCA), Análise de Componentes Independentes (ICA), análise de agrupamento e análise de regressão multivariada são amplamente empregadas.

A Análise de Componentes Principais (PCA) é uma das ferramentas estatísticas mais

utilizadas para este fim, pois permite reduzir a dimensionalidade dos dados sem perda significativa de informações (ALVES, 2019). A PCA é eficaz em identificar padrões e correlações em grandes conjuntos de dados, facilitando o monitoramento de conformidade e destacando amostras que fogem ao padrão (BLANCO, 2018). Estudos recentes apontam que essas técnicas são eficazes para detectar contaminação e adulteração de combustíveis, o que é essencial para a preservação do desempenho dos motores e a proteção ambiental (MOURA, 2019). Assim, na aplicação de PCA no Maranhão propicia uma análise exploratória dos dados de qualidade de combustíveis, permitindo a redução da dimensionalidade e a identificação de padrões e correlações (ALVES, 2019). Esta técnica é útil para analisar dados multivariados e possibilita a identificação de agrupamentos de amostras e de correlações entre indicadores de qualidade.

Por outro lado, o uso da técnica de Análise de Componentes Independentes (ICA), é outra abordagem que pode ser aplicada à análise de combustíveis, especialmente para prever parâmetros de qualidade a partir da transformação de uma possível mistura de sinais em um número de componentes independentes (ICs), sem reduzir as dimensões da mistura. De modo geral, ao aplicar ICA, tem-se como objetivo a separação de um conjunto de sinais obtidos a partir de sinais misturados, mesmos sem conhecer alguma informação sobre os sinais originais (CHOI et al., 2005).

Desse modo, a aplicação dessas técnicas de análise multivariada permitem criar modelos preditivos que ajudam na tomada de decisão sobre a qualidade do produto em diferentes etapas da cadeia de abastecimento. Além disso, a combinação de tecnologias de aprendizado de máquina com métodos de análise multivariada tem ampliado a capacidade de tratamento e análise de grandes volumes de dados, trazendo maior precisão e agilidade para a fiscalização e controle de qualidade.

A análise multivariada da qualidade de combustíveis no Maranhão visa entender os fatores que impactam a qualidade dos combustíveis, identificando correlações entre variáveis e propondo melhorias no controle de qualidade. Isso permite ações mais eficazes de fiscalização, especialmente em um mercado heterogêneo e sujeito a variações na procedência e qualidade dos combustíveis (RIBEIRO e SCHIRMER, 2017). Além disso, o uso de técnicas multivariadas auxilia na identificação de lotes de combustíveis com desvios de qualidade.

Assim, este trabalho propõe o uso do PCA e ICA como ferramentas para análise exploratória dos dados de qualidade de combustíveis do Maranhão, visando identificar padrões e correlações e, assim, contribuir para a melhoria da fiscalização e controle de qualidade desses produtos no estado.

1.1. Contextualização do problema

A qualidade dos combustíveis comercializados no Estado do Maranhão é um fator crucial para o desempenho de motores e a mitigação dos impactos ambientais. No entanto, há variações significativas nos parâmetros físico-químicos desses combustíveis, o que pode comprometer tanto a segurança dos consumidores quanto a conformidade com as normas vigentes. Além disso, fatores como a logística de distribuição, o armazenamento inadequado e o transporte em condições adversas podem agravar a degradação da qualidade do produto final.

Neste cenário, surge a necessidade de compreender quais são as principais variáveis que influenciam a qualidade dos combustíveis e de que forma elas se inter-relacionam. Para isso, a utilização de uma metodologia robusta, como a análise multivariada, se torna essencial. Através dessa abordagem, é possível investigar simultaneamente diferentes parâmetros e identificar padrões ocultos ou correlações não evidentes em análises convencionais.

1.2. Objetivos do trabalho

1.2.1 Objetivo Geral

Aplicar a Análise Multivariada para análise exploratória de dados de qualidade de combustíveis afim de obter a redução da dimensionalidade dos dados, e identificar possíveis padrões e correlação entre os indicadores de qualidade de combustíveis do estado do Maranhão.

1.2.2 Objetivos específicos

- Analisar os parâmetros físico-químicos mais relevantes na qualidade dos combustíveis e identificar padrões ou grupos de combustíveis que apresentem maiores variações na qualidade, utilizando técnicas de Análise de Componentes Principais e Análise de Componentes Independentes.
- Comparar os resultados obtidos com a técnica de Análise de Componentes Principais aos obtidos com a técnica de Análise de Componentes Independentes.
- Oferecer subsídios técnicos para a melhoria das práticas de fiscalização e controle da qualidade no estado do Maranhão.

2.Fundamentação Teórica

2.1 Combustíveis

Inicialmente, para abordar a questão da análise multivariada da qualidade de combustíveis comerciais, seja no Estado do Maranhão, ou em qualquer outra parte do território nacional, é imprescindível fazer uma definição do elemento “combustível”, bem como determinar suas fontes geradoras e suas diversas formas existentes utilizadas no país.

Os combustíveis são substâncias que, ao sofrerem combustão, liberam calor, o qual pode ser utilizado para diversas aplicações. Esse calor pode ser aproveitado para acionar turbinas em usinas termelétricas, transformando-se em energia cinética, ou ainda para fornecer potência a motores de veículos (EPE, 2018). De maneira geral, os combustíveis desempenham um papel essencial como fontes energéticas, sendo amplamente utilizados em processos industriais para aquecimento, na geração de eletricidade e, principalmente, na alimentação de motores a combustão interna (MIRANDA, 2013).

Conforme destaca Miranda (2013), os combustíveis se sobressaem entre os diversos materiais energéticos disponíveis, dada sua relevância para o funcionamento da sociedade moderna. Do ponto de vista químico, um combustível pode ser descrito como qualquer substância capaz de reagir com o oxigênio do ar, resultando na produção de calor, chama e gases. Esse processo, denominado combustão, caracteriza-se pela liberação de energia térmica de maneira controlada e aproveitável (DIONYSIO; MEIRELLES, 2016).

As fontes de combustíveis classificam-se em renováveis e não renováveis. Os renováveis, como biogás, biodiesel, lenha, carvão vegetal e etanol, derivam de fontes vegetais e são reabastecidos naturalmente. Já os não renováveis, como carvão mineral, gasolina e urânio, originam-se de materiais fósseis ou nucleares e possuem taxas de reposição extremamente lentas, podendo levar milhões de anos para se formarem.

Independentemente do valor energético ou do método de produção e uso, os combustíveis são essenciais para a vida moderna, portanto, aplicar mais técnicas para elevar sua qualidade antes de serem comercializados será altamente útil para aqueles que os utilizam.

2.2 Qualidade de Combustíveis: Conceitos e Regulamentações

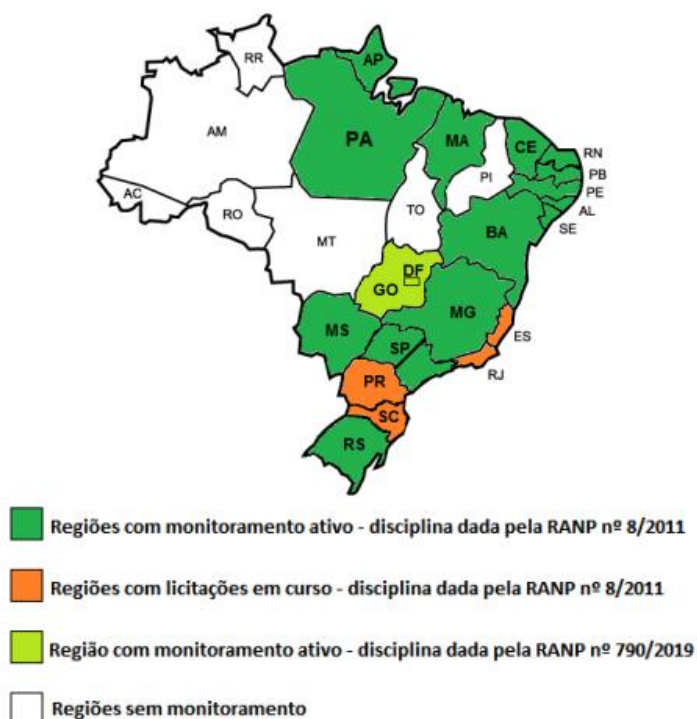
No Brasil, a Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) é responsável por regulamentar e fiscalizar a qualidade dos combustíveis. A ANP estabelece

especificações físico-químicas que garantem a qualidade mínima necessária para o desempenho esperado dos combustíveis, conforme detalhado em suas diretrizes (ANP, 2020a).

No que diz respeito ao monitoramento da qualidade dos combustíveis, é importante mencionar o Programa de Monitoramento de Qualidade de Combustíveis (PMQC), lançado em janeiro de 1998, com o objetivo de diminuir os índices de não conformidade dos produtos regulados (ANP, 2020b). Este programa oferece uma visão geral da qualidade dos combustíveis, envolvendo a coleta, transporte e análises físico-químicas dos combustíveis líquidos utilizados em veículos. Os resultados das análises são consolidados em relatórios mensais disponibilizados no portal da ANP, permitindo um planejamento mais eficiente das ações fiscais, tanto pela ANP quanto por outros órgãos que colaboram com a Agência, como Ministérios Públicos, Procons e Secretarias de Fazenda (PORTAL DA ANP, 2024).

Atualmente, o PMQC é gerido pela Resolução N° 8, de 09 de fevereiro de 2011, que determina que a coleta e o transporte das amostras de combustíveis são de responsabilidade dos laboratórios contratados pela ANP. Estes laboratórios, que realizam o serviço de coleta e análise das amostras de combustíveis, são determinados por processo licitatório (SBQ, 2022). Os locais de coleta são selecionados por sorteio, e as datas de recolhimento das amostras são mantidas em sigilo. Atualmente, encontram-se cobertos pelo PMQC as seguintes unidades da federação:

Figura 1 – Unidades da federação que possuem cobertura do PMQC.



Fonte: SBQ, 2022.

É relevante destacar que, em cada visita a um revendedor de combustíveis, os representantes do laboratório contratado coletam pelo menos uma amostra de cada tipo de combustível disponível no local. Por fim, é importante ressaltar que o PMQC é um programa focado na análise da qualidade dos produtos, assegurando que estejam em conformidade com as normas regulamentadoras, no entanto, não possui caráter fiscalizatório e, portanto, caso identifique não conformidades, não tem a capacidade de autuar o revendedor.

2.1.2. Parâmetros Físico-Químicos de Qualidade

Os parâmetros físico-químicos de qualidade dos combustíveis são de suma importância para garantir que eles estejam munidos de todos os requisitos básicos de desempenho e segurança. Esses parâmetros podem variar de acordo com o tipo de combustível, mas os principais parâmetros estabelecidos pelas regulamentadoras dessas substâncias, incluem: densidade, destilações, massa específica, ponto de fulgor, número de octano, composição de hidrocarbonetos, teor de enxofre, fluidez, viscosidade, teor alcoólico, teor de biodiesel, cor, etc.

Imperioso destacar que, esses são somente alguns exemplos comuns de parâmetros de qualidade para combustíveis, que devem ser aplicados de forma conjunta ou separadamente a cada tipo de combustível. Além disso, devido à sua grande relevância para a sociedade em geral, os padrões ou parâmetros dos combustíveis comercializados no Brasil, são regulamentados e constantemente monitorados pelos órgãos de controle nacional, como a Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), tudo com o enfoque primordial de garantir que os combustíveis distribuídos nacionalmente, atendam de forma satisfatória, os padrões ambientais e de desempenho exigidos.

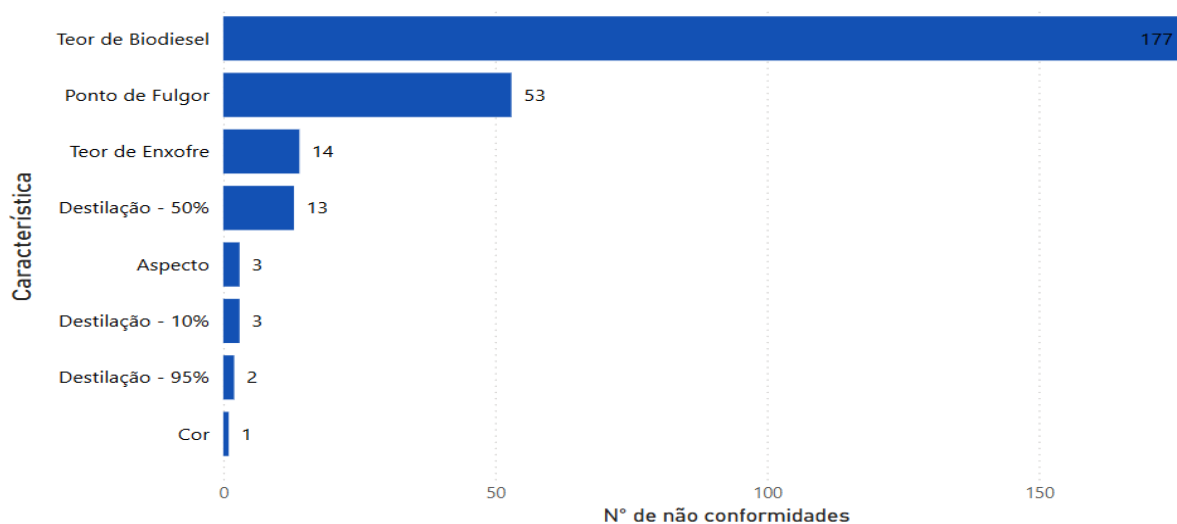
Além disso, é importante ressaltar, que a ANP realiza a publicação dos boletins mensais do PMQC. Esses boletins trazem mensalmente os resultados do Programa, que desde outubro de 2020, também são divulgadas pelo portal do Painel Dinâmico do PMQC (ANP, 2021).

Nas figuras apresentadas a seguir (Figuras 2, 3 e 4), são apresentadas as principais características físico-químicas que mais contribuem para resultados fora das especificações normativas no estado do Maranhão, conforme publicação disponível no Painel Dinâmico do PMQC (2025). Os valores apresentados nessas figuras, referem-se ao período de janeiro de 2017 a dezembro de 2023, o mesmo período considerado para coleta dos dados analisados no

presente trabalho.

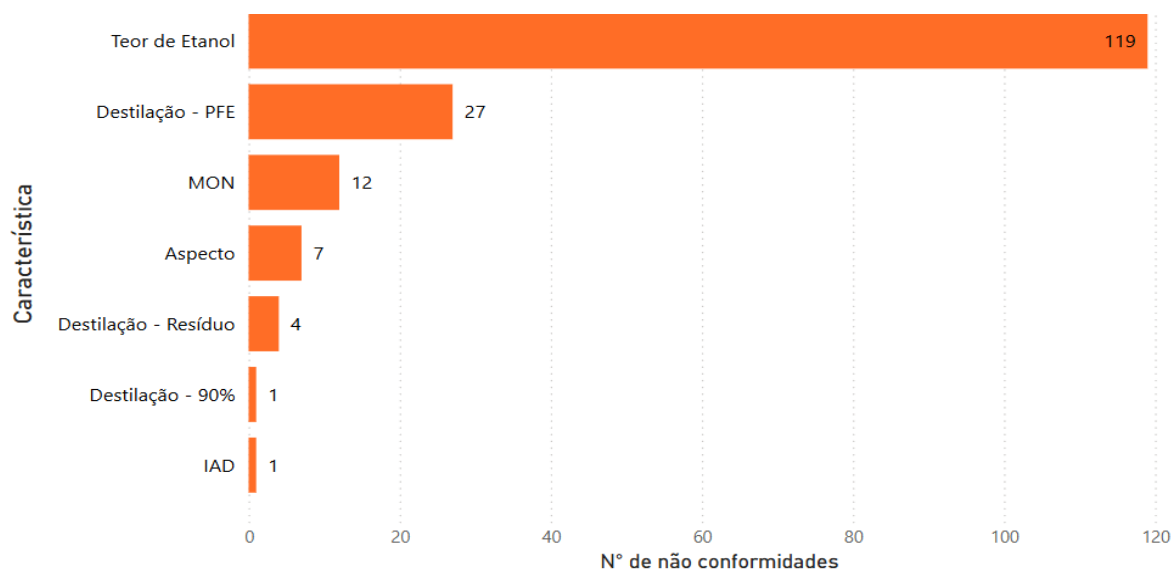
De acordo com os dados disponibilizados no Painel Dinâmico do PMQC (2025), no Maranhão, as características físico-químicas que mais frequentemente apresentam resultados fora da especificação definida pelas normas regulamentadoras, são: teor de biodiesel e ponto de fulgor, para o óleo diesel; teor de etanol e destilação – PFE, para a gasolina; teor alcoólico e massa específica a 20° C, para o etanol.

Figura 2 – Resultados das análises feitas pelo PMQC, sobre características físico-químicas que mais frequentemente apresentam resultados fora das especificações, referentes às amostras de óleo diesel coletadas no Estado do Maranhão, no período de 2017 a 2023.



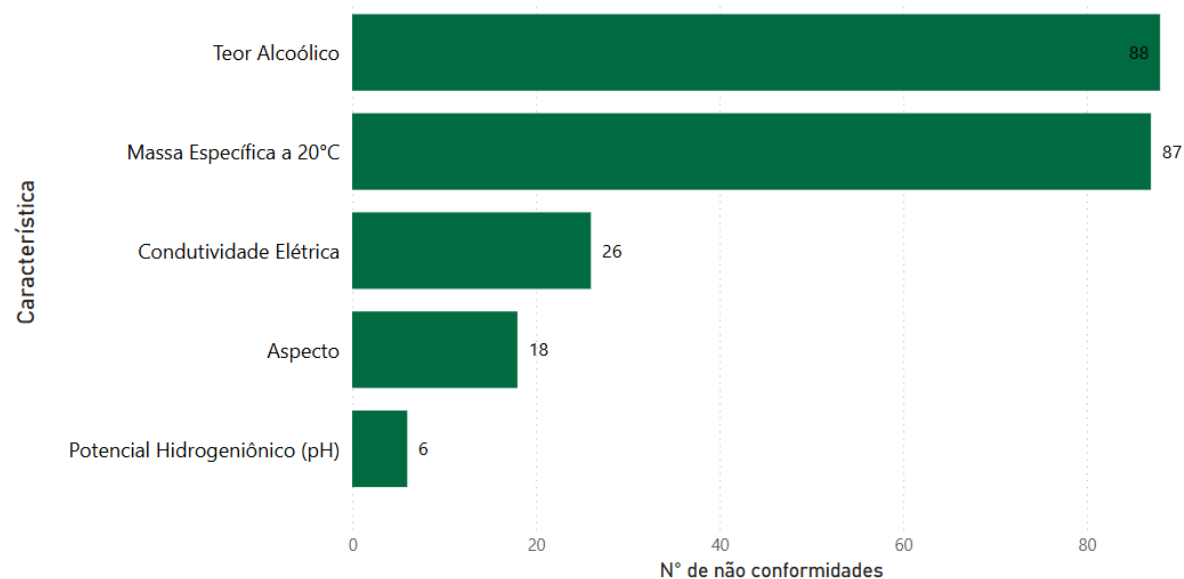
Fonte: PMQC, 2025.

Figura 3 – Resultados das análise feitas pelo PMQC, sobre características físico-químicas que mais frequentemente apresentam resultados fora da especificações, referentes às amostras de gasolina coletadas no Estado do Maranhão, no período de 2017 a 2023.



Fonte: PMQC, 2025.

Figura 4 – Resultados das análise feitas pelo PMQC, sobre características físico-químicas que mais frequentemente apresentam resultados fora da especificações, referentes às amostras de etanol coletadas no Estado do Maranhão, no período de 2017 a 2023.



Fonte: PMQC, 2025.

2.1.3. Normas Técnicas e Regulamentação no Brasil

A Agência Nacional de Petróleo (ANP) foi estabelecida em 1997, conforme a Lei nº 9.478, que trata da política energética do Brasil e regula as atividades relacionadas ao monopólio do petróleo, além de criar o Conselho Nacional de Política Energética e a própria ANP, entre outras diretrizes.

A ANP, vinculada ao Ministério de Minas e Energia (MME), tem a responsabilidade de elaborar e implementar resoluções, normas e portarias referentes a cada tipo de combustível autorizado para comercialização nos postos. Antes de serem contratados pela ANP, os laboratórios que analisam a qualidade dos combustíveis passam por um rigoroso processo de licitação. Após a seleção, esses laboratórios realizam análises detalhadas para garantir a segurança dos consumidores que utilizam esses combustíveis.

Além disso, a ANP é comprometida com a saúde pública e a proteção ambiental. Seus regulamentos e atividades de fiscalização asseguram que sejam atendidas as novas exigências ambientais e que os motores funcionem adequadamente. Dessa forma, as características de qualidade dos produtos regulados são revisadas e, quando necessário, constantemente reformuladas (PORTAL DA ANP, 2024).

Contudo, a regulamentação dos combustíveis derivados do petróleo no Brasil não se limita à Lei nº 9.478/97. A Constituição Federal Brasileira, em seu artigo 177, determina que o petróleo, o gás natural e seus derivados são monopólio da União, que tem a responsabilidade de fiscalizar e regulamentar as atividades de exploração, produção e comercialização desses produtos, garantindo sua qualidade. Além disso, existe um extenso arcabouço normativo que regula diversas práticas relacionadas à extração, qualificação e comercialização de combustíveis no Brasil (CONSTITUIÇÃO FEDERAL, 1988).

2.2 Comercialização de Combustíveis

Em fevereiro de 2024 a produção do diesel no Brasil cresceu cerca de 8,5% em comparação ao ano anterior, e segue superando as expectativas de crescimento no ano corrente (IPB, 2024). O Maranhão reflete esse consumo diesel nacional, principalmente devido as intensa atividades agrícolas e transporte de cargas.

De acordo com os dados apresentado na Síntese Mensal de Comercialização de Combustíveis, publicada pela ANP em dezembro de 2024, as vendas de óleo diesel B

diminuíram 6,1% em dezembro de 2024 em comparação com o mesmo mês em 2023. Mas, em termos do total acumulado do ano, houve um aumento de 2,7%. Já as vendas de gasolina C apresentaram uma queda de 2,1% em dezembro de 2024, quando comparadas ao mesmo período de 2023 (ANP, 2024). De modo semelhante, o volume acumulado de vendas ao longo de 2024 foram 4,0% menores do que em 2023. Por outro lado, as vendas do etanol hidratado, registraram um aumento de 3,9% em dezembro de 2024, quando comparadas ao mesmo período de 2023. No acumulado do ano de 2024 em relação a 2023, a alta foi de 33,4% (ANP, 2024).

2.2.1 Óleo Diesel

O óleo diesel é um combustível líquido derivado de petróleo, utilizado em motores ciclo Diesel, que possuem como características combustão interna e ignição por compressão (PETROBRAS, 2023). Quanto à sua composição química, o diesel possui cadeias compostas de 8 a 16 carbono e apresenta menores concentrações de nitrogênio, enxofre e oxigênio (RIBEIRO e SCHIRMER, 2017).

No entanto, essas características podem ser transformadas com a adição de outras substâncias. Pois, o óleo diesel é formulado através da mistura de diversas correntes como gasóleos, nafta pesada, diesel leve e diesel pesado, provenientes das diversas etapas de processamento do petróleo bruto (PETROBRAS, 2023).

Os óleos diesel podem ser classificados com base na adição de biodiesel. O óleo diesel A, é um combustível obtido a partir do refino do petróleo e do processamento de gás natural, sem a presença de biodiesel em sua composição. Já o óleo Diesel B, resulta da mistura do biodiesel ao Óleo Diesel A, seguindo a proporção determinada pela legislação vigente (PETROBRAS, 2023).

Considerando a aplicação rodoviária, existem dois tipos principais de óleo diesel: S10 e S500. O óleo diesel S10, apresenta um teor máximo de enxofre de 10 mg/kg e é destinado a veículos que atendem às exigências das fases P7 e posteriores do Programa de Controle da Poluição do Ar por Veículos Automotores (PROCONVE). Por outro lado, o óleo Diesel S500 possui um teor de enxofre de até 500 mg/kg (PETROBRAS, 2023).

2.2.2 Gasolina

A gasolina é um dos combustíveis mais utilizados no Brasil, sendo composta por uma mistura inflamável e complexa de centenas de hidrocarbonetos. Esses compostos são obtidos a partir do refino de frações do petróleo e possuem moléculas com cadeias que variam de 5 a 12 átomos de carbono (ANP, 2020c). No entanto, é importante ressaltar que a composição da gasolina é variável e depende principalmente da origem do petróleo e dos processos utilizados em seu refino (RIBEIRO e SCHIRMER, 2017).

No Brasil, a Agência Nacional do Petróleo (ANP) estabelece três categorias de gasolina: tipo A, tipo B e tipo C. Sendo que, a gasolina tipo A é produzida em refinarias, e a gasolina tipo B é destinada exclusivamente ao uso das forças armadas (DA ROCHA et al., 2017). Portanto, a gasolina do tipo C é a única disponível nos postos revendedores (ANP, 2020c). Por isso, somente a gasolina do tipo C, foi considerada como objeto de estudo do presente trabalho.

Segundo a ANP (2020c), a gasolina C é definida como aquela que possui etanol anidro combustível na quantidade estipulada pela legislação atual. O etanol anidro é incorporado a toda a gasolina vendida no Brasil, na proporção de 27% em volume. De acordo com a Lei nº 9.478, de 6 de agosto de 1997, essa percentagem deve ser determinada pelo Poder Executivo, variando entre 18% e 27,5% (MME, 2022). A adição de etanol anidro à gasolina, melhora as características da gasolina, proporcionando uma melhor octanagem, resistindo a uma compressão maior, e ainda reduz seu custo (DO NASCIMENTO et al., 2021).

Além dos tipos de gasolina, este combustível pode ser classificado como gasolina comum ou aditivada. A diferença entre esta última da primeira, é que ela possui um maior número de octano, permitindo maiores taxas de compressão no motor (RIBEIRO e SCHIRMER, 2017).

2.2.3 Etanol

O etanol é um biocombustível, gerado principalmente a partir de fontes de matéria-prima que são ricas em sacarose ou amido, como a cana-de-açúcar e o milho, respectivamente (SANTOS, 2012). Composto apenas por uma molécula de C_2H_5OH , o etanol é considerado uma substância pura.

No mercado brasileiro, o etanol é comercializado como etanol anidro (AEAC) e

etanol combustível hidratado (AEHC) (RIBEIRO e SCHIRMER, 2017). O que diferencia esses dois tipos, é o conteúdo de água no combustível. Sendo que, o AEAC tem baixos níveis de água, cerca de 0,5%, e o AEHC possui cerca de 5% de água em volume. Além disso, o AEHC é obtido nas fábricas de etanol a partir das colunas de destilação, enquanto o AEAC é obtido da remoção do excesso de água (RIBEIRO, 2017).

O etanol anidro, conforme discutido anteriormente neste estudo, é adicionado a toda a gasolina comercializada no Brasil, devendo respeitar a proporção de 27% em volume, conforme a legislação vigente (BORGES, 2024). Adicionalmente, está em tramitação um projeto de lei (PL 4.516/2023), denominado “Combustível do Futuro”, que propõe um aumento desse percentual obrigatório de etanol anidro para 30% (CAMPOS JÚNIOR, 2024). Essa é uma ação que auxilia na redução da dependência de petróleo e seus derivados.

2.3 Trabalhos Relacionados

A análise multivariada da qualidade dos combustíveis comerciais vem sendo objeto de diversos estudos acadêmicos, visando avaliar e assegurar a conformidade dos combustíveis com os padrões estabelecidos para cada combustível, de acordo com suas características individuais e coletivas.

No estudo de Aguiar (2008) três métodos estatísticos foram aplicados para investigar o perfil físico-químico da gasolina no estado de Rondônia, em uma parceria entre o LABCOM/UNIR e a ANP. A análise incluiu parâmetros como curva de destilação, densidade, teor de álcool, presença de hidrocarbonetos e índices de octanagem, considerando os padrões de qualidade estabelecidos pela ANP (Portaria n. 309 de 27/12/2001). Embora o PCA e a Análise de Cluster não tenham sido eficazes para distinguir amostras conforme ou não conforme, eles foram úteis para identificar a distribuição de frações leves, médias e pesadas. A análise de correlação permitiu uma compreensão detalhada das propriedades físico-químicas, mostrando forte correlação, especialmente entre os pares IAD e RON, IAD e MON, T90% e PFE, T90% e Res, T10% e T50%, T50% e T90% ($p < 0,05$).

Já o trabalho de Côrrea (2009), aborda pesquisas nas áreas de Engenharia de Eletricidade e Química Analítica, focando no aprimoramento das análises químicas do Programa de Monitoramento da Qualidade de Combustíveis (PMQC) da ANP, realizado no Maranhão pelo LAPQAP. O estudo propõe melhorias no Sistema Inteligente de Monitoramento e Controle da Qualidade de Combustíveis (SIMCQC), sugerindo a adição de novos agentes a

uma sociedade multiagente para ampliar as técnicas de análise. Apresenta também a criação de uma ontologia de combustíveis, que usa a metodologia 101 para representar o conhecimento na área e facilitar a comunicação entre os agentes. A pesquisa utiliza Inteligência Artificial, o motor de inferência JESS, ontologias e o middleware JADE, além da metodologia PASSI para modelar o sistema.

Outro exemplo de estudo desse tema, é o relevante projeto denominado "Sistema Integrado de Monitoramento e Controle da Qualidade de Combustíveis no Estado do Maranhão", que foi desenvolvido em 2004 pelo pesquisador Delano Brandes Marques como parte de sua dissertação de mestrado em Ciência da Computação na Universidade Federal do Maranhão (UFMA). Este projeto, tema de dissertação de mestrado do curso de pós graduação em engenharia de eletricidade, propôs a criação de um Sistema Integrado que faz uso de técnicas de geoprocessamento e inteligência artificial para monitorar e potencializar a qualidade dos combustíveis no estado. O estudo científico destacou a importância do exame de dados estatísticos objetivando amparar a adoção de determinadas decisões pertinentes à indústria petrolífera no Estado do Maranhão.

Portanto, estudo que tem como principal foco a análise multivariada empregada para avaliar a qualidade dos combustíveis comerciais, são de grande valia, pois proporcionam ferramentas eficientes para monitorar e controlar a qualidade e a conformidade dos combustíveis com os padrões preestabelecidos.

3. Análise Estatística Multivariada

A análise estatística multivariada é uma ferramenta importante na manipulação e interpretação de dados que envolvem múltiplas variáveis. De acordo com Gomes e Calvalcante (2017), o uso dessa ferramenta possibilita o agrupamento de variáveis similares, investiga a dependência entre variáveis agrupadas, relaciona variáveis observadas objetivando prever uma ou mais variável, além de formular testes de hipóteses (FERREIRA, 1996).

De acordo com Batarelli Júnior (2016), os métodos que realizam a análise de mais de duas variáveis para cada observação da amostra são considerados parte da análise multivariada. De modo geral, as variáveis selecionadas para cada observação podem ser quantitativas (discretas ou contínuas) ou qualitativas (ordinais ou normais) (BETARELLI JUNIOR, 2016). Caso múltiplas variáveis sejam mensuradas simultaneamente em cada elemento da amostra, a aplicação da estatística multivariada pode ser considerada adequada

para sua análise. Como essas variáveis geralmente apresentam correlação entre si, um grande número de variáveis torna a interpretação mais complexa, dificultando o uso de métodos univariados convencionais. (MINGOTI, 2005).

Segundo Mingoti (2005), a análise estatística multivariada se divide em dois grandes grupos: um primeiro, consistindo em técnicas exploratórias de sintetização (ou simplificação) da estrutura de variabilidade dos dados, e um segundo, consistindo em técnicas de inferência estatística. Em linhas gerais, pode-se considerar que a Análise de Componentes Principais e a Análise de Componentes Independentes fazem parte do primeiro grupo, onde se encontram as técnicas de estatística multivariada responsáveis por simplificar a análise da variabilidade de um grande volume de dados (MINGOTI, 2005).

3.1 Análise de Componentes Principais (PCA)

A Análise de Componentes Principais (PCA) é um método estatístico essencial na redução de dimensionalidade, tendo um papel fundamental na análise exploratória de dados (Sanguansat, 2012). A PCA aplicada a análise de dados com uma grande quantidade de variáveis pode ser descrita como apresentado por Correa (2021), como um método que avalia as inter-relações com o “objetivo de reconhecer padrões na distribuição das amostras, avaliar a relação entre amostras e as variáveis e ainda detectar a presença de amostras que apresentam um comportamento distinto (*outliers*)”.

O PCA é uma técnica de redução de dimensionalidade transformando um conjunto de dados com grande número de variáveis em um conjunto menor. A redução dos dados aplicando PCA é obtida através da combinação linear das correlações das variáveis originais. Desta forma obtém-se uma representação menor através das componentes principais (do inglês, *Principal Component-PC's*) resultantes. O modelo matemático por trás do PCA é baseado na *Singular Value Decomposition (SVD)*, e pode ser descrito da seguinte forma, sendo que a matriz X é decomposta de acordo com a Equação 1 (CORREA, 2021; PARENTE, 2022).

$$X_{n \times m} = U_{n \times n} \Sigma_{n \times m} V_{m \times m}^T \quad \text{Eq. (1)}$$

Sendo que, n é o número de amostras e m é o número de atributos iniciais que compõem o conjunto de dados de interesse, X , que pode ser expandido pela Equação 2 (PARENTE, 2022):

$$X_{n \times m} = \begin{bmatrix} | & | & | \\ u_1 & \cdots & u_n \\ | & | & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & \sigma_p & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} - & v_1 & - \\ - & \vdots & - \\ - & v_m & - \end{bmatrix}^T \quad \text{Eq. (2)}$$

O pressuposto do PCA é que a primeira componente deve ter o máximo de variância explicada

e a segunda a variância não explicada na primeira componente (CORREA, 2021). Enquanto $U_{n \times n}$ e $V_{m \times m}$, são ortogonais e Σ é uma matriz diagonal composta pelos valores singulares, de modo que $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p$ (PARENTE, 2022). Tem-se também, que P é matriz de pesos em que os elementos de cada coluna correspondem aos coeficientes das combinações lineares das variáveis originais, como mostra a Equação 3 (CORREA, 2021).

$$P = X \cdot V^T = X_{n \times m} \cdot \begin{bmatrix} | & | & | & | \\ v_1 & v_2 & \dots & v_p \\ | & | & | & | \end{bmatrix} = \begin{bmatrix} PC_{1,1} & PC_{1,2} & \dots & PC_{1,q} \\ PC_{2,1} & PC_{2,2} & \dots & PC_{2,q} \\ \vdots & \vdots & \ddots & \vdots \\ PC_{n,1} & PC_{n,2} & \dots & PC_{n,q} \end{bmatrix} \quad \text{Eq. (3)}$$

Os resultados da aplicação do PCA podem ser visualizados através de gráficos de *scores*, onde podem ser vistas as principais relações entre as variáveis. Com os diferentes agrupamentos de amostras são identificados as semelhanças e diferenças, tendências e *outliers*.

3.2 Análise de Componentes Independentes (ICA)

Segundo Roberts e Everson (2001), a Análise de Componentes Independentes (ICA) é uma técnica estatística avançada utilizada para desvendar estruturas complexas em conjuntos de dados multivariados. Em termos simples, a Análise de Componentes Independentes (ICA) pode ser definida como uma técnica estatística que separa um sinal complexo (ou observado) em seus componentes independentes de origem. Assim sendo, o ICA se apresenta como uma abordagem viável para a Separação Segra de Fontes (BSS, do inglês, *Blind Source Separation*), uma tarefa essencial em muitos contextos de processamento de informações.

Quanto à relação do ICA com o PCA, é possível afirmar que o ICA busca expressar a mistura de sinais que não são diretamente independentes em uma forma não linear no número de componentes independentes (IC, do inglês, *Independent Components*), que são menores do que as dimensões originais da mistura. No entanto, se a redução de informação for necessária, um pré-processamento da mistura é feito pela aplicação do PCA. Afirma-se ainda, que o ICA está interessado em informações não gaussianas, enquanto o PCA está procurando informações gaussianas. Em outras palavras, o ICA pode revelar informações mais significativas em dados não gaussianos do que o PCA (GE e SONG, 2007).

De acordo com Hyvarinen e Oja (2000), a análise de componentes independentes (ICA) é uma técnica estatística e computacional para revelar fatores ocultos que fundamentam conjuntos de variáveis aleatórias, medições ou sinais. Para introduzir o algoritmo de ICA assume-se que l variáveis medidas, $x(k) = [x_1(k), x_2(k), \dots, x_l(k)]$, na amostra k podem ser

expressas como combinações lineares de r componentes independentes desconhecidas $[s_1, s_2, \dots, s_r]^T$ (onde $r \leq l$); a relação entre eles é dada pela Equação 4 (GE e SONG, 2007):

$$X = A.S + E \quad \text{Eq. (4)}$$

Considerando que n é o número de medições, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbf{R}^{l \times n}$ é a matriz de dados, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r] \in \mathbf{R}^{l \times r}$ é a matriz de mistura, $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n] \in \mathbf{R}^{r \times n}$ é a matriz de componentes independentes e $\mathbf{E} \in \mathbf{R}^{l \times n}$ é a matriz residual. O problema básico do ICA é estimar a componente original $\mathbf{S}$ e a matriz de mistura $\mathbf{A}$ de $\mathbf{X}$ (GE e SONG, 2007).

Assim, o objetivo do ICA é calcular uma matriz de separação $\mathbf{W}$ para que os componentes da matriz de dados reconstruída $\mathbf{S}$ se tornem, o máximo possível, independentes uns dos outros. Dado como:

$$\hat{\mathbf{S}} = \mathbf{W}.\mathbf{X} \quad \text{Eq. (5)}$$

Antes da aplicação do algoritmo de ICA, a matriz de dados $\mathbf{X}$ deve ser branqueada, para que sejam eliminadas as correlações cruzadas entre variáveis aleatórias. Para isto, foi utilizado um método conhecido para branqueamento, que é a decomposição de autovalor. Considerando $\mathbf{x}(\mathbf{k})$ com sua covariância $\mathbf{R}_x = \mathbf{E}\{\mathbf{x}(\mathbf{k}).\mathbf{x}(\mathbf{k})^T\}$, a decomposição de autovalor de $\mathbf{R}_x$ é dada pela Equação 6 (GE e SONG, 2007):

$$\mathbf{R}_x = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \quad \text{Eq. (6)}$$

A transformação do branqueamento é expressa como visto na Equação 7:

$$\mathbf{z}(\mathbf{k}) = \mathbf{Q}.\mathbf{x}(\mathbf{k}) \quad \text{Eq. (7)}$$

Onde $\mathbf{Q} = \mathbf{\Lambda}^{-1/2}.\mathbf{U}^T$. De maneira simples, pode-se verificar que $\mathbf{R}_z = \mathbf{E}\{\mathbf{z}(\mathbf{k}).\mathbf{z}(\mathbf{k})^T\}$, é a matriz identidade sob essa transformação. Depois da transformação de branqueamento, temos que (GE e SONG, 2007):

$$\mathbf{z}(\mathbf{k}) = \mathbf{Q}.\mathbf{x}(\mathbf{k}) = \mathbf{Q}.\mathbf{A}.\mathbf{s}(\mathbf{k}) = \mathbf{B}.\mathbf{s}(\mathbf{k}) \quad \text{Eq. (8)}$$

Onde $\mathbf{B}$ é uma matriz ortogonal, o que resulta na seguinte verificação:

$$\mathbf{E}\{\mathbf{z}(\mathbf{k}).\mathbf{z}(\mathbf{k})^T\} = \mathbf{B}.\mathbf{E}\{\mathbf{s}(\mathbf{k}).\mathbf{s}(\mathbf{k})^T\}.\mathbf{B}^T = \mathbf{B}.\mathbf{B}^T = \mathbf{I} \quad \text{Eq. (9)}$$

Assim, pode ser feita a seguinte estimativa:

$$\hat{\mathbf{s}}(\mathbf{k}) = \mathbf{B}^T.\mathbf{z}(\mathbf{k}) = \mathbf{B}^T.\mathbf{Q}.\mathbf{x}(\mathbf{k}) \quad \text{Eq. (10)}$$

Por fim, das equações 2 e 7, obtém-se a relação entre $\mathbf{W}$ e $\mathbf{B}$:

$$\mathbf{W} = \mathbf{B}^T.\mathbf{Q} \quad \text{Eq. (11)}$$

4. Métodos

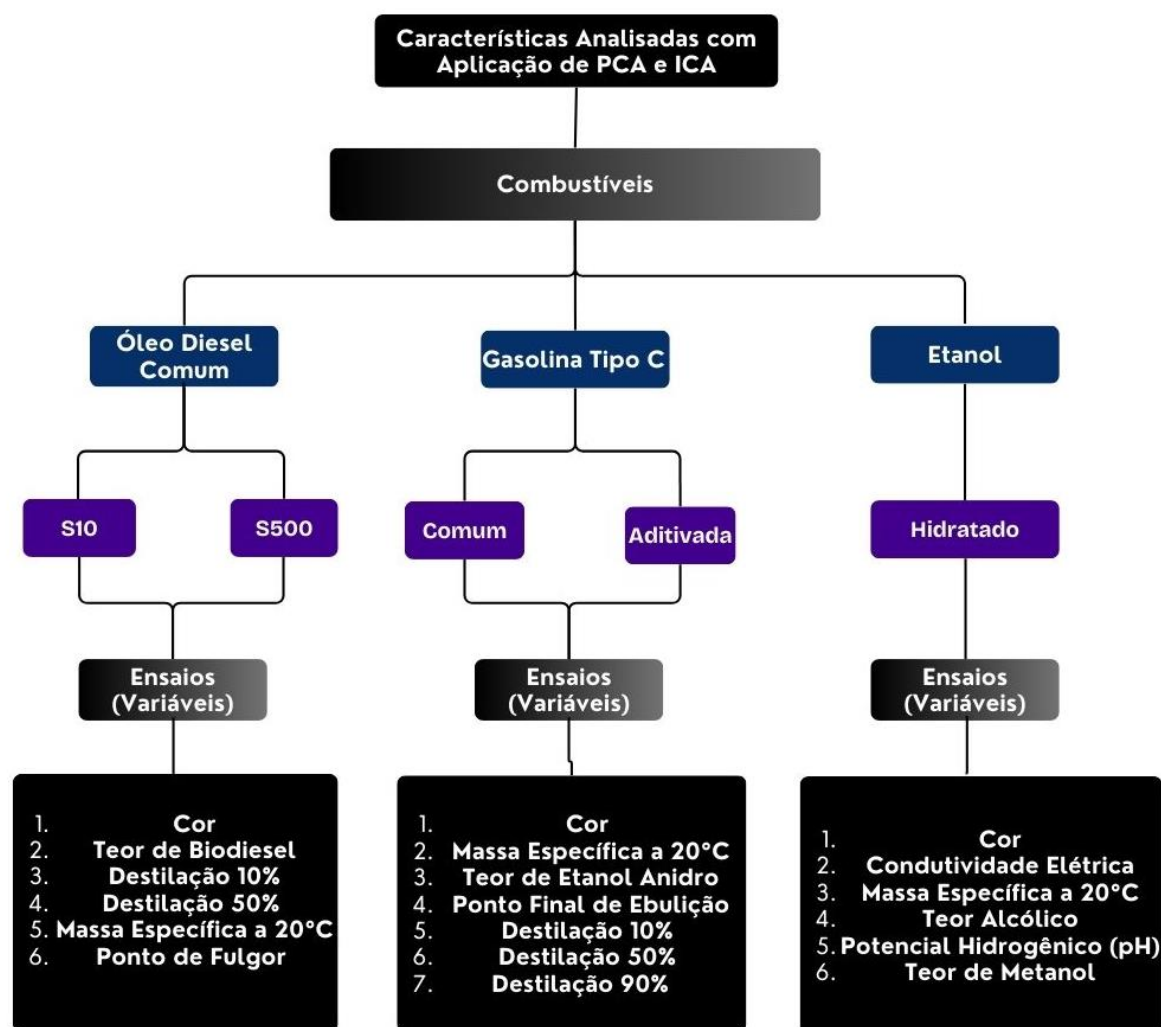
Para o desenvolvimento desta pesquisa, inicialmente foi realizado o levantamento bibliográfico sobre o tema abordado. Após este levantamento ser realizado, foram adquiridos os dados sobre a qualidade dos combustíveis no Estado do Maranhão, disponíveis no banco de dados aberto da ANP (ANP, 2023b). Estes dados, são obtidos através do Programa de Monitoramento de Qualidade de Combustíveis (PMQC) da ANP.

O PMQC fornece dados sobre análise da qualidade dos combustíveis óleo diesel, etanol hidratado e gasolina. Para esta pesquisa, foram selecionados dados referentes ao monitoramento da qualidade do combustível óleo diesel comum S10 e S500, gasolina do tipo C comum e aditivada e etanol hidratado, cujas as amostras foram coletados em diversos postos de combustíveis, de diferentes municípios do Maranhão entre os anos 2017 e 2023.

Após a seleção dos dados, os mesmos foram tratados de acordo com os objetivos deste trabalho. Assim, foram selecionadas amostras de óleo diesel comum S10 e S500, gasolina comum e aditivada e etanol hidratado, e os ensaios correspondentes a estas amostras. Estes ensaios compõem o grupo de variáveis utilizadas para a aplicação das técnicas de PCA e ICA.

A Figura 5, mostra um esquema que contém os combustíveis analisados neste estudo e os seus respectivos ensaios. Foram desconsiderados do estudo as amostras que apresentaram dados nulos para os ensaios. Para a aplicação de PCA e ICA, foram considerados como variáveis para as amostras de óleo diesel comuns selecionadas, os ensaios de (1) cor, (2) teor de biodiesel, (3) destilação - 10%, (4) destilação - 50%, (5) massa específica a 20°C e (6) ponto de fulgor. Enquanto que, para as amostras de gasolina, que possuem uma maior quantidade de dados disponíveis para análise, foram consideradas sete variáveis: (1) cor, (2) massa específica a 20°C, (3) teor de etanol anidro, (4) ponto final de ebulição, (5) destilação - 10%, (6) destilação - 50% e (7) destilação - 90%. Já para as amostras de etanol hidratado, foram consideradas as variáveis (1) cor, (2) condutividade elétrica, (3) massa específica a 20° C, (4) teor alcólico, (5) potencial hidrogênico (pH) e (6) teor de metanol.

Figura 5 – Esquema ilustrativo das características analisadas pelas técnicas de PCA e ICA, nesta pesquisa.



Fonte: Produzido pela autora.

Como o resultado para o ensaio cor é encontrado no conjunto de dados como uma palavra e não como um número, para a aplicação de PCA, estes resultados foram substituídos por valores numéricos, conforme mostra a Tabela 1.

Tabela 1 – Correspondentes numéricos atribuídos às cores das amostras para aplicação de PCA no Matlab.

Cor da Amostra	Número Correspondente
Amarela	1
Amarelado	2
Incolor	3
Laranja	4
Rosa	5
Verde	6
Vermelho	7

Fonte: Produzido pela autora.

Após a seleção das características as serem analisadas com aplicação de PCA, foram montadas matrizes de dados, sendo que, esta matriz contém em suas linhas amostras dos combustíveis analisados e nas colunas as variáveis correspondentes a estas amostras, formando uma matriz de dimensões: 4751x7, para o óleo diesel; 5532x8, para a gasolina; e 394x7, para o etanol. Com a matriz formada, foi implementado um algoritmo no software *Matlab R2023b* (versão acadêmica) para a aplicação de PCA e ICA no conjunto de dados estudado. Os resultados obtidos serão apresentados através de gráficos e tabelas.

5. Resultados e Discussão

A aplicação das técnicas de PCA e ICA nos conjuntos de dados referente a qualidade dos combustíveis no Estado do Maranhão fornecidos pelos boletins do PMQC, disponíveis no site da ANP, possibilitou a aquisição de resultados em formas de gráficos e tabelas referente à aplicação dessa técnica nos dados em estudo. Assim, foi possível realizar a análise dos resultados de forma precisa, contribuindo desta forma para um melhor exame dos resultados obtidos.

5.1 Modelagem por PCA para o Óleo Diesel

A Tabela 2 mostra a porcentagem dos valores de variância explicada e variância acumulada, e também, a quantidade de componentes principais (PC) obtidas com aplicação do método PCA no conjunto de dados de óleo diesel comum S10 e S500. Essa tabela possui três colunas, sendo que: a primeira coluna apresenta o número do PC extraído; a segunda coluna mostra a porcentagem de variância que é explicada por cada PC; enquanto que a terceira coluna, indica variância acumulada, que é composta pela soma da variância explicada pelos PC's, mostrando quantos PC's são necessários para representar a maior parte das informações contidas no conjunto de dados analisados.

Tabela 2 - Valores de variância explicada e variância acumulada obtidos com a aplicação de PCA no conjunto de dados de óleo diesel comum S10 e S500.

PC	Variância Explicada (%)	Variância Acumulada (%)
1	37,42	37,42
2	21,97	59,39
3	16,83	76,22
4	9,47	85,69
5	7,65	93,34
6	6,66	100,00

Fonte: Produzido pela autora.

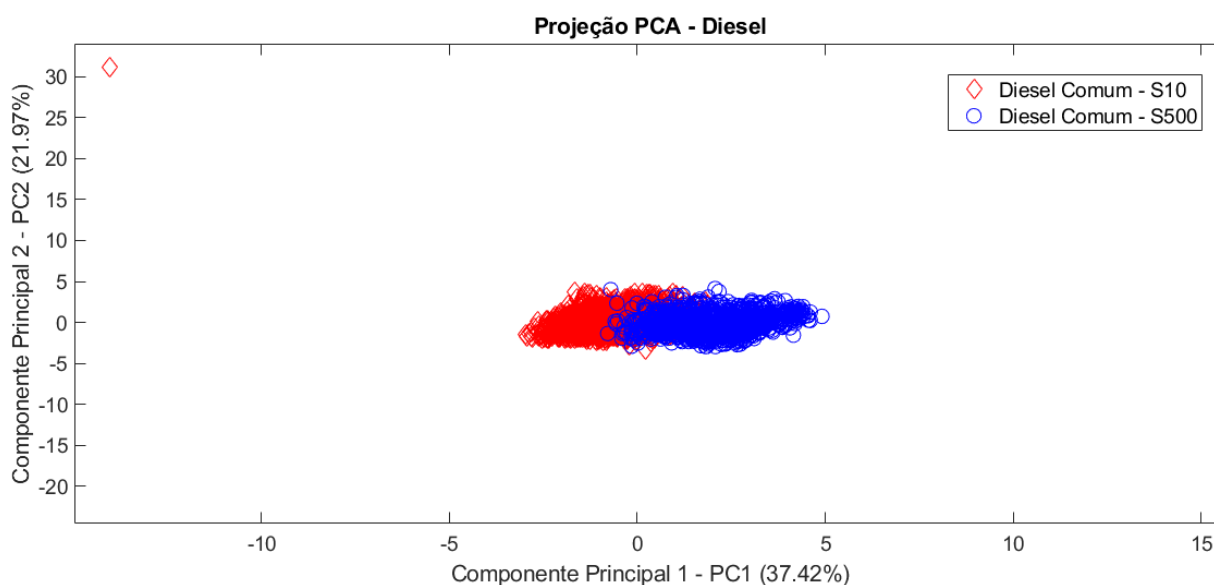
De acordo com os valores apresentados na Tabela 2, observa-se que foram encontrados 6 componentes principais. Ao aplicar PCA, espera-se que a quantidade de componentes principais seja dada pelo número de variáveis subtraído de 1 unidade (SANTOS et al., 2024). Já que, um dos principais objetivos da PCA é redução da dimensionalidade dos dados. Portanto, ao aplicar PCA a um conjunto de dados, pretende-se que o número máximo de componentes principais encontrado, seja inferior ao número de variáveis consideradas na análise (FOLLI et al., 2023).

Assim, como foram consideradas 6 variáveis para aplicação de PCA, que é o mesmo número de componentes principais obtido, as componentes que possuem menor valor de variância explicada (PC5 e PC6) foram desconsideradas deste estudo, por não contribuírem, de forma significativa, para explicação do conjunto de dados analisado. Pois, a eliminação dos PC's que possuem menor variância explicada, permite reduzir o número de variáveis a serem analisadas (DUARTE, 1998).

Analisando a Tabela 2, observa-se que os componentes principais que mais contribuem para a representação das informações contidas no conjunto de dados analisados, são as: PC1 (com 37,42%), PC2 (com 21,97%), PC3 (com 16,83%) e PC4 (com 9,47). Juntas, estas componentes principais representam 85,69% das informações contidas nos dados analisados. Assim, verifica-se que dependendo da quantidade de informações que se queira representar, pode-se considerar um número menor de variáveis, e conseqüentemente, reduzir a dimensionalidade do conjunto de dados analisados (SANTOS et al., 2024). Desse modo, programas que são responsáveis pela análise da qualidade dos combustíveis (tais como, o PMQC), podem utilizar a técnica de PCA para reduzir o volume de suas bases de dados e analisar uma menor quantidade de variáveis.

Agora será analisada a Figura 6, que mostra um gráfico de dispersão resultante da aplicação de PCA a um conjunto de dados contendo amostras de óleo diesel S10 e S500, coletadas em diversos municípios do estado do Maranhão.

Figura 6 – Gráfico bidimensional com os dados das amostras de óleo diesel comum S10 e S500, após a aplicação de PCA.



Fonte: Produzido pela autora.

Analisando a Figura 6, observa-se no gráfico de dispersão, que o eixo *x* representa o primeiro componente principal (PC1), explicando 37,42% da variabilidade dos dados de óleo diesel comum. Enquanto, o eixo *y* representa o segundo componente principal (PC2), que explica 21,97% da variabilidade desses dados. No gráfico, os pontos vermelhos em formato de losango representam o diesel comum S10 e os pontos azuis em formato de círculo representam o diesel comum S500.

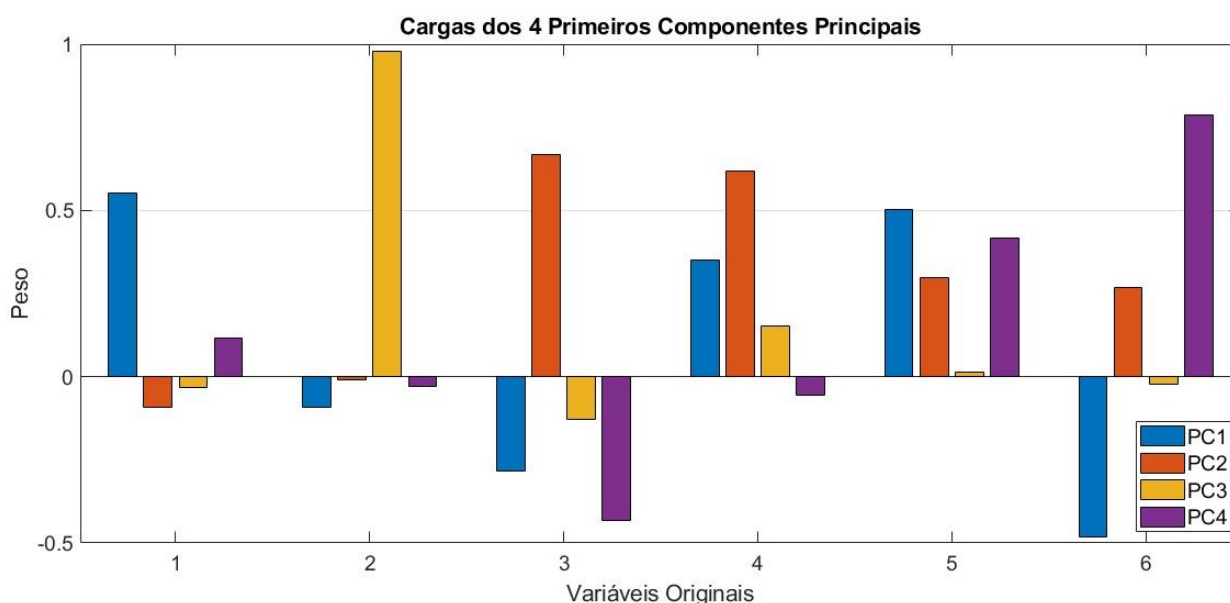
O gráfico indica uma separação parcial entre os dois tipos de diesel, com o diesel S10 tendendo a se concentrar em valores mais negativos de PC1 e o diesel S500 deslocado para valores mais positivos de PC1. Assim como no trabalho de Faria (2019), observa-se no gráfico de PC1xPC2, uma sobreposição das amostras de um mesmo tipo óleo diesel comum (S10 ou S500), o que mostra a similaridade na composição química das amostras.

A partir da Figura 6, também é possível observar, um *outlier* de diesel S10 que está visível no canto superior esquerdo (PC1 negativo e PC2 positivo). Esse *outlier* corresponde a uma amostra de diesel S10, do ano de 2023, que quando comparada com as outras amostras, apresentou um valor alto para a variável (ensaio) destilação 10%, que foi de 835° C. Sendo que, a média dos valores observados para essa mesma variável nas demais amostras é de 200 °C, e segundo a Resolução ANP nº 909, de 18 de novembro de 2022, o limite mínimo estipulado para esse ensaio era de 180° C. Isso mostra, que a aplicação de PCA consegue detectar possíveis erros, discrepâncias ou não conformidades entre os valores contidos no banco de dados estudado neste

trabalho.

A Figura 7, mostra um gráfico com os pesos (cargas ou *loadings*), que indicam quanto cada variável original do conjunto de dados analisados, contribui para cada componente. Ou seja, neste caso, esse gráfico possibilita analisar a influência de cada variável original na organização das amostras de óleo diesel em relação à PC1, PC2, PC3 e PC4.

Figura 7 – Gráfico de pesos dos quatro primeiros componentes principais em relação as variáveis originais do conjunto de dados de óleo diesel analisado por PCA.



Fonte: Produzido pela autora.

Sabendo que neste trabalho, o PCA foi aplicado às variáveis (1) cor, (2) teor de biodiesel, (3) destilação - 10%, (4) destilação - 50%, (5) massa específica a 20°C e (6) ponto de fulgor (vide Figura 1), pode indicar que esses parâmetros ajudam a diferenciar os tipos de diesel. O PC1 é o principal responsável pela distinção entre S10 e S500, explicando 37,42% da variabilidade dos dados de óleo diesel comum.

Assim, observando o gráfico de pesos da Figura 7, verifica-se que as variáveis cores, destilação - 50% e massa específica, correspondem aos pesos positivos (valores de 0,55, 0,35 e 0,50, respectivamente) de PC1 e são as principais responsáveis pela distinção entre o diesel S10 e S500. Sendo que, a cor tem mais influência na separação entre esses dois tipos de diesel. Isto pode ser explicado, devido ao fato de o diesel S10 apresentar, na maioria dos casos, a cor amarela; enquanto que, o diesel S500, costuma apresentar a cor vermelha. Desse modo, ao analisar as

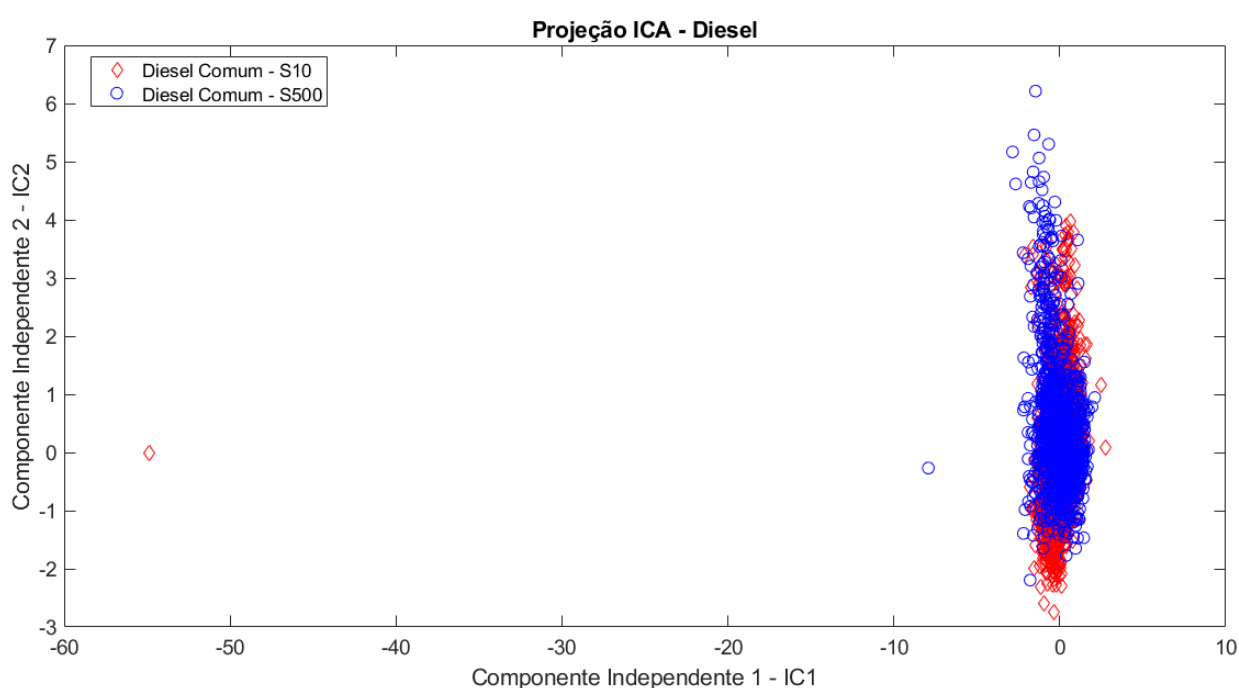
Figuras 6 e 7, observa-se que a cor vermelha (referente ao diesel S500), faz com que PC1 fique mais positivo. Quanto aos pesos negativos de PC1, nota-se que a destilação - 10% e o ponto de fulgor influenciam a parte negativa de PC1. A Figura 6, mostra que a maioria das amostras de diesel S10 está na parte negativa de PC1. Isto indica que, o teor de biodiesel também contribui para a separação das amostras de diesel S10 e S500.

Já para o PC2, que explica 21,97% da variabilidade dos dados de diesel S10 e S500, observou-se que a destilação - 10%, destilação - 50%, massa específica a 20°C e ponto de fulgor são as variáveis que mais influenciam, com pesos positivos que variam entre 0,27 e 0,67, os escores que estão relacionados ao eixo de PC2. Por outro lado, na Figura 3, também pode ser visto que PC3 e PC4, são influenciados majoritariamente pelas variáveis teor de biodiesel e ponto de fulgor, com pesos de 0,98 e 0,79 respectivamente.

5.2 Modelagem por ICA para o Óleo Diesel

Como visto anteriormente, a análise de componentes independentes é uma técnica estatística utilizada para separar um sinal complexo ou misturado em suas componentes independentes originais. A Figura 8, apresenta um gráfico de dispersão gerado a partir da aplicação de ICA, a um conjunto de dados contendo amostras de óleo diesel comum S10 e S500.

Figura 8 – Gráfico bidimensional com os dados das amostras de óleo diesel comum S10 e S500, após a aplicação de ICA.



Fonte: Produzido pela autora.

A partir do gráfico apresentado na Figura 4, observa-se que eixo x representa o primeiro componente independente (IC1), e que, o eixo y representa o segundo componente independente (IC2). Sabe-se também, que os pontos vermelhos em formato de losango representam o diesel comum S10 e os pontos azuis em formato de círculo representam o diesel comum S500.

Analizando a distribuição dos dados na Figura 8, nota-se que maior parte dos dados está concentrada na origem (em $IC1 = 0$), isto indica que a aplicação de ICA a este conjunto de dados, não permitiu a separação de forma clara dos grupos analisados. Além disso, observa-se pontos isolados, fora da distribuição principal, em ambas as direções. Há um ponto vermelho (diesel S10) deslocado para IC1 negativo e alguns pontos azuis (diesel S500) dispersos para valores altos de IC2. Desse modo, infere-se que esses valores extremos indicam que existem amostras de diesel com características muito diferentes do restante do conjunto de dados, possivelmente devido a variações nos ensaios físico-químicos.

Comparando os resultados obtidos com a aplicação de ICA aos obtidos com PCA (vide Figuras 6 e 8), nota-se que diferente do PCA, onde houve uma considerável separação entre os grupos ao longo de PC1, a ICA resultou em uma distribuição mais verticalizada, com sobreposição dos grupos analisados. Considerando que a ICA busca identificar componentes estatisticamente independentes, entende-se que o motivo para isso ter acontecido, é que as variáveis escolhidas não possuem independência estatística forte o suficiente para distinguir claramente em dois grupos, o diesel S10 e o S500.

5.3 Modelagem por PCA para a Gasolina

A Tabela 3 mostra a porcentagem dos valores de variância explicada e variância acumulada, e também, a quantidade de componentes principais (PC) obtidas com aplicação do método PCA no conjunto de dados com amostras de gasolina tipo C comum e aditivada.

Tabela 3 - Valores de variância explicada e variância acumulada obtidos com a aplicação de PCA no conjunto de dados com amostras de gasolina tipo C comum e aditivada.

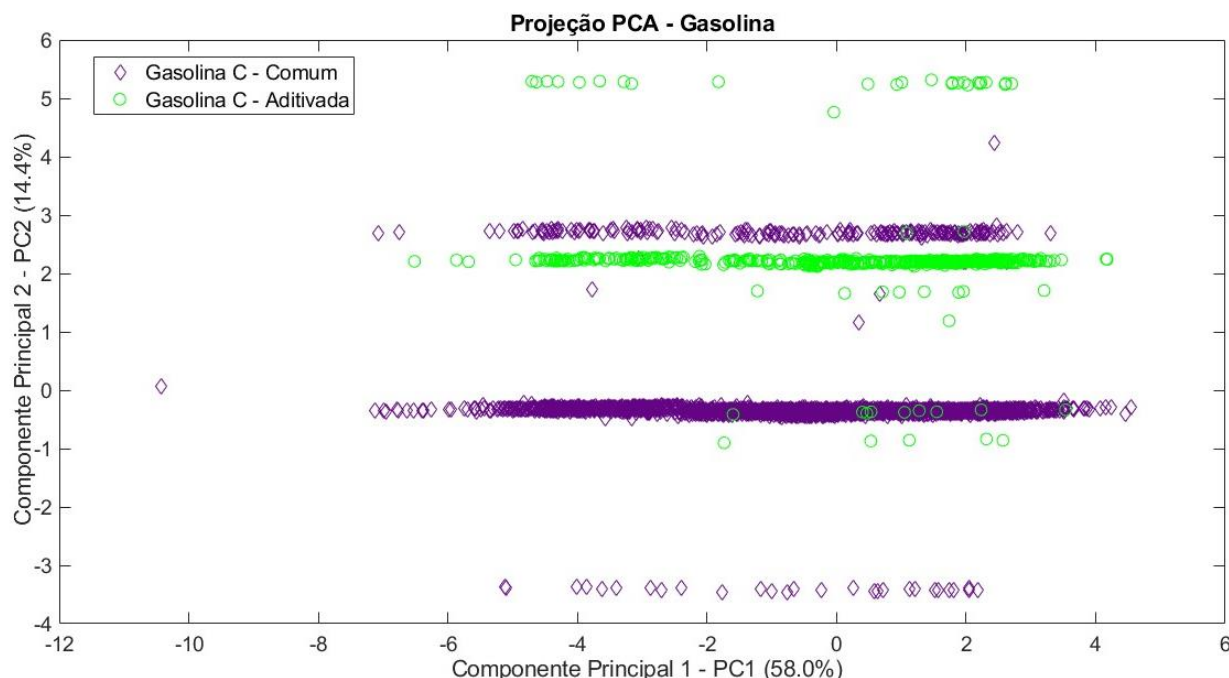
PC	Variância Explicada (%)	Variância Acumulada (%)
1	57,97	57,97
2	14,39	72,36
3	14,18	86,54
4	6,78	93,32
5	3,28	96,60
6	1,88	98,48
7	1,52	100,00

Fonte: Produzido pela autora.

Analisando os resultados apresentados na Tabela 3, nota-se que os dois primeiros componentes principais (PC1 e PC2) representam juntos 72,36% da variabilidade dos dados referentes às amostras de gasolina tipo C comum e aditivada. Sendo que, só PC1 já representa 57,97% da variabilidade desses dados. Assim, ao comparar os resultados obtidos com aplicação de PCA ao conjunto de dados de óleo diesel, com os obtidos a partir do conjunto de dados de gasolina, percebe-se que os dois primeiros PC's encontrados para gasolina, contém mais informação sobre os dados analisados, do que os dois primeiros PC's encontrados para os dados de óleo diesel. Isto pode ser explicado pela quantidade de dados e variáveis consideradas para cada análise. Pois, como já apresentado, neste trabalho foram consideradas 6 variáveis para matriz de dados de óleo diesel S10 e S500 (matriz 4751x7), e sete variáveis para a matriz de dados de gasolina tipo C comum e aditivada (matriz 5532x8).

A Figura 9, mostra um gráfico de dispersão resultante da aplicação da análise de componentes principais a um conjunto de dados contendo amostras de gasolina tipo C comum e gasolina aditivada. A matriz de dados utilizada inclui as seguintes sete variáveis físico-químicas relacionadas à caracterização das amostras de gasolina (vide Figura 5).

Figura 9 – Gráfico bidimensional com os dados das amostras de gasolina tipo C comum e aditivada, após a aplicação de PCA.



Fonte: Produzido pela autora.

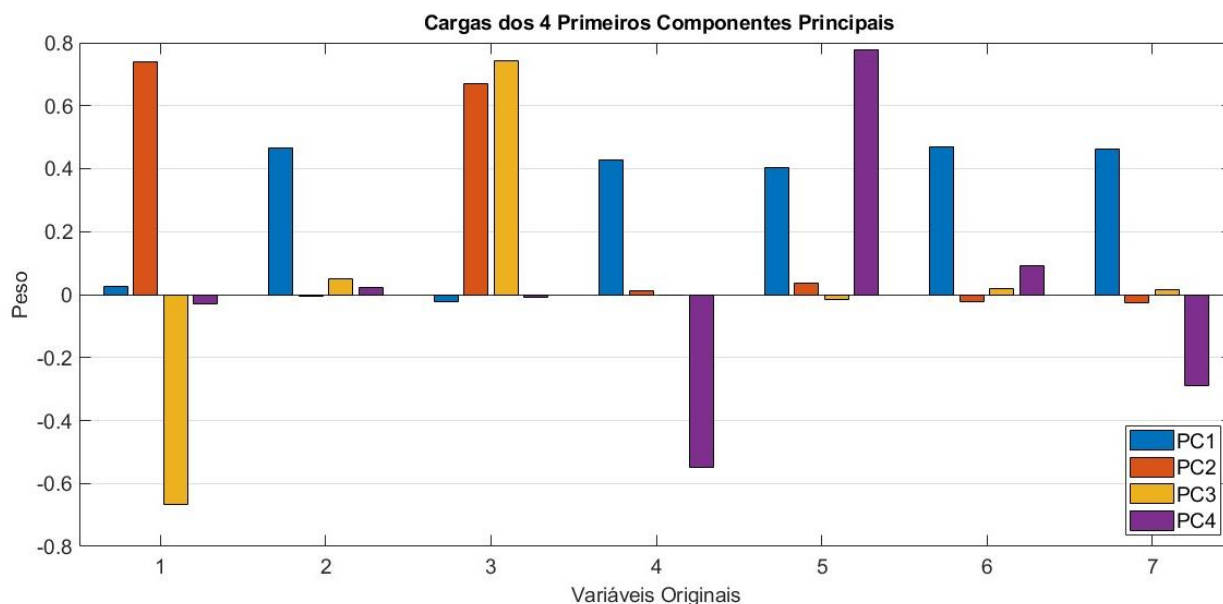
Através do gráfico apresentado na Figura 9, observa-se que os losangos roxos representam amostras de gasolina tipo C comum e os círculos verdes representam amostras de gasolina tipo C aditivada. Além disso, o gráfico também mostra que o eixo x, representa o primeiro componente principal (PC1), que captura 58.0% da variância total dos dados. Já o eixo y, representa o segundo componente principal (PC2), que captura por 14.4% da variância total dos dados.

Analisando a distribuição dos dados da Figura 9, nota-se que as amostras de gasolina comum e aditivada estão distribuídas em boa parte do gráfico de PC1 em relação a PC2. Os pontos roxos (gasolina comum) estão mais concentrados ao longo do eixo de PC1, formando faixas horizontais. Isto pode indicar que as amostras de gasolina comum possuem variação mais significativa em uma variável específica, que estar fortemente associada ao PC1. Por outro lado, os pontos verdes (gasolina aditivada) estão distribuídos de forma mais espalhada no gráfico. Assim, observa-se uma certa separação entre os dois grupos de dados, mostrando que PC1 é um bom discriminante entre a gasolina comum e aditivada.

Seguindo esse contexto, passaremos a interpretar o gráfico de pesos da Figura 10, que viabiliza a análise da influência de cada variável original na organização das amostras de gasolina

em relação à PC1, PC2, PC3 e PC4.

Figura 10 – Gráfico de pesos dos quatro primeiros componentes principais em relação as variáveis originais do conjunto de dados de gasolina analisado por PCA.



Fonte: Produzido pela autora.

Analisado o gráfico da Figura 10, que indica como cada variável original contribui para a formação dos quatro primeiros componentes principais (PC1, PC2, PC3 e PC4), nota-se que PC1 (que explica a maior parte da variabilidade dos dados) está mais fortemente associado às variáveis massa específica a 20°C, teor de etanol anidro, ponto final de ebulição, destilação - 10%, destilação - 50% e destilação - 90% (respectivamente, variáveis 2, 3, 4, 5, 6 e 7). Pois essas variáveis possuem pesos elevados, que variam entre 0,4 a aproximadamente 0,5, como mostrado no gráfico.

Enquanto PC2, possui uma forte relação com as variáveis 1 (cor) e (3) teor de etanol anidro, que apresenta pesos (cargas) altos - de aproximadamente 0,74 e 0,67, respectivamente - indicando que essas variáveis têm um grande impacto na separação das amostras ao longo do eixo PC2 na projeção PCA. Já PC3, apresenta pesos altos e negativos para a variável cor e positivas para a variável teor de etanol anidro (respectivamente, - 0,67 e 0,74), o que significa que ele pode estar capturando variações específicas dessas variáveis. Quanto a PC4, observa-se maior contribuição das variáveis ponto final de ebulição e destilação - 10%, com pesos próximos a - 0,55 e 0,78, respectivamente.

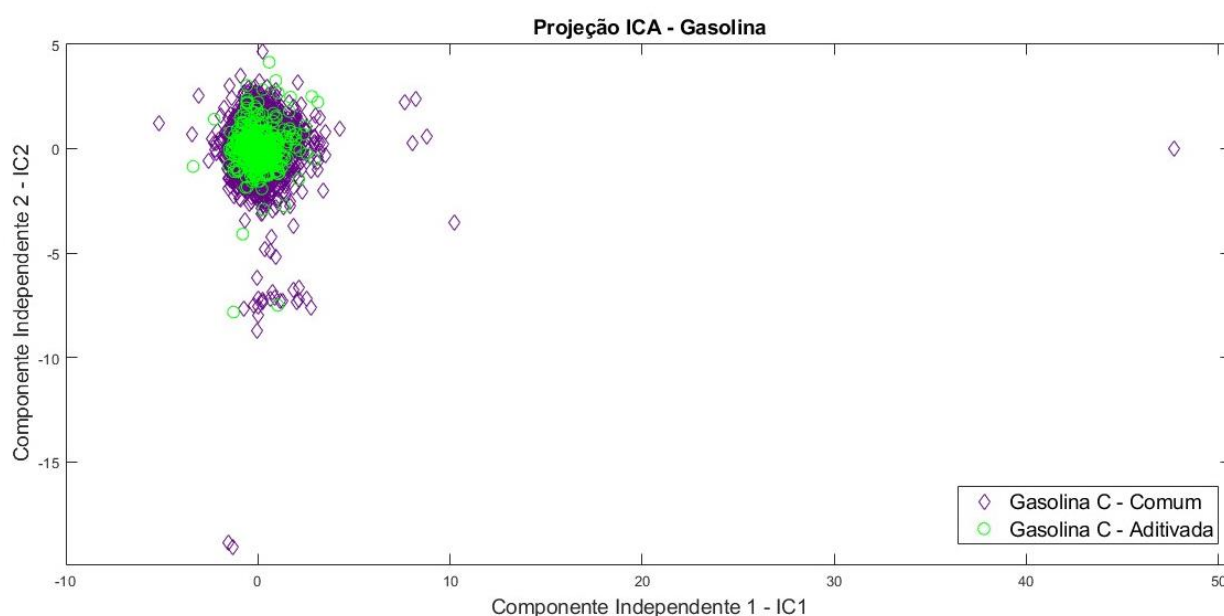
De modo geral, como o PC1 separa bem as amostras, pode-se inferir que a maior diferença

entre gasolina comum e aditivada está relacionada às variáveis com maior peso nesse componente, que provavelmente estão relacionadas às variáveis massa específica a 20°C, ponto final de ebulição, teor de etanol anidro e destilações. Já o PC2, adiciona uma distinção secundária, onde a variável cor tem um papel predominante.

5.4 Modelagem por ICA para a Gasolina

A Figura 11, apresenta um gráfico de dispersão gerado a partir da aplicação de ICA ao conjunto de dados contendo amostras de gasolina tipo C comum e aditivada. Neste gráfico, o eixo x representa o primeiro componente independente (IC1), e o eixo y representa o segundo componente independente (IC2). A gasolina comum é representada por losangos roxos; já a gasolina aditivada é representada por círculos verdes.

Figura 11 – Gráfico bidimensional com os dados das amostras de gasolina tipo C comum e aditivada, após a aplicação de ICA.



Fonte: Produzido pela autora.

Analisando a distribuição das amostras de gasolina comum e aditivada no gráfico da Figura 11, observa-se que a maior parte das amostras de gasolina comum e aditivada está concentrada em uma região próxima ao centro do gráfico. Sendo que, algumas amostras de gasolina comum aparecem dispersas para valores mais elevados de IC1 e IC2, indicando possíveis *outliers* ou maior

variabilidade nas características dessas amostras. Também é possível observar, que existe uma sobreposição significativa entre os grupos, mostrando que os ensaios utilizados podem não ser totalmente eficazes na separação entre os dois tipos de gasolina.

Além disso, sabe-se que a separação das amostras de gasolina comum e aditivada ao longo de IC1 e IC2 depende das variáveis analisadas. Como ICA extrai componentes estatisticamente independentes, a separação ao longo dos eixos IC1 e IC2 pode estar associada a características específicas que diferenciam as amostras, ou até mesmo, a erros referentes aos valores obtidos nos ensaios.

5.5 Modelagem por PCA para o Etanol

A Tabela 4, mostra a porcentagem dos valores de variância explicada e variância acumulada, e também, a quantidade de componentes principais (PC) obtidas com aplicação do método PCA no conjunto de dados de etanol hidratado.

Tabela 4 - Valores de variância explicada e variância acumulada obtidos com a aplicação de PCA no conjunto de dados com amostras etanol hidratado.

PC	Variância Explicada (%)	Variância Acumulada (%)
1	34.13	34.13
2	24.28	58.41
3	16.06	74.47
4	14.65	89.12
5	10.79	99.91
6	0.09	100.00

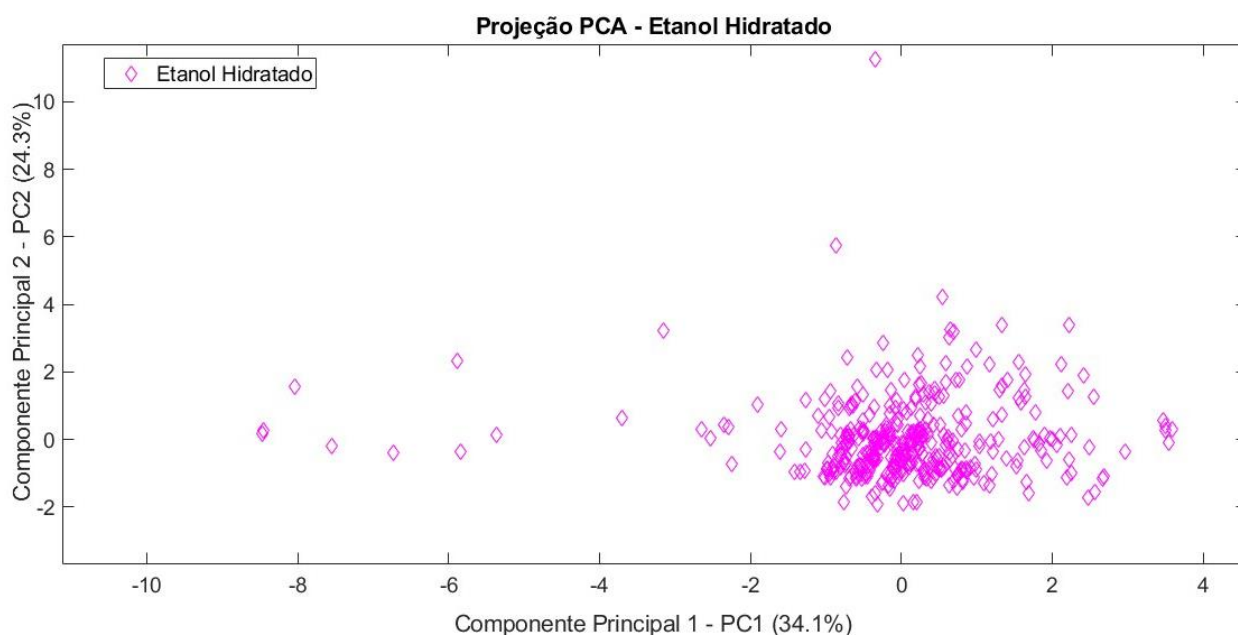
Fonte: Produzido pela autora.

A Tabela 4, mostra que os primeiros componentes principais, PC1 e PC2, explicam juntos 58,41% da variância total dos dados de etanol hidratado. Sendo que, por PC1 explica 34,13% da variância total desses dados, podendo ser considerado como o componente mais importante para capturar as informações contidas nesses dados. Já o PC3 e PC4 representa, respectivamente, 16,06% e 14,65% da variância total dos dados. Mostrando que ao combinar os quatro primeiros

componentes principais, obtêm-se 89,2% de variância acumulada. Por outro lado, PC5 e PC6 apresentam variâncias explicadas pequenas, sendo consideradas pouco relevantes para explicação dos dados de etanol hidratado. Desse modo, observa-se que os quatro componentes (PC1 a PC4) explicam quase 90% da variância total do conjunto de dados de etanol hidratado, sugerindo que apenas esses componentes sejam suficientes para representar bem os dados, permitindo a redução da dimensionalidade, sem perda de muita informação.

A Figura 12, apresenta um gráfico de dispersão referente a aplicação de análise de componentes principais a um conjunto de dados contendo amostras de etanol hidratado. Esse gráfico possibilita a análise sobre a redução da dimensionalidade dos dados e visualização de como as amostras se distribuem em função das variáveis analisadas.

Figura 12 – Gráfico bidimensional com os dados das amostras de etanol hidratado, após a aplicação de PCA.



No gráfico da Figura 12, o eixo x representa o primeiro componente principal (PC1), responsável por 34,1%** da variabilidade total dos dados. Enquanto que, o eixo y representa o segundo componente principal (PC2), responsável por 24,3% da variabilidade total dos dados. Juntos, PC1 e PC2 explicam aproximadamente 58,4% da variabilidade dos dados, sugerindo que uma parte significativa da informação das variáveis originais está contida nesses dois componentes.

Quanto a distribuição dos pontos no gráfico da Figura 12, observa-se uma alta concentração de amostras próximas à origem, indicando similaridade entre muitas delas. Isto pode ser explicado,

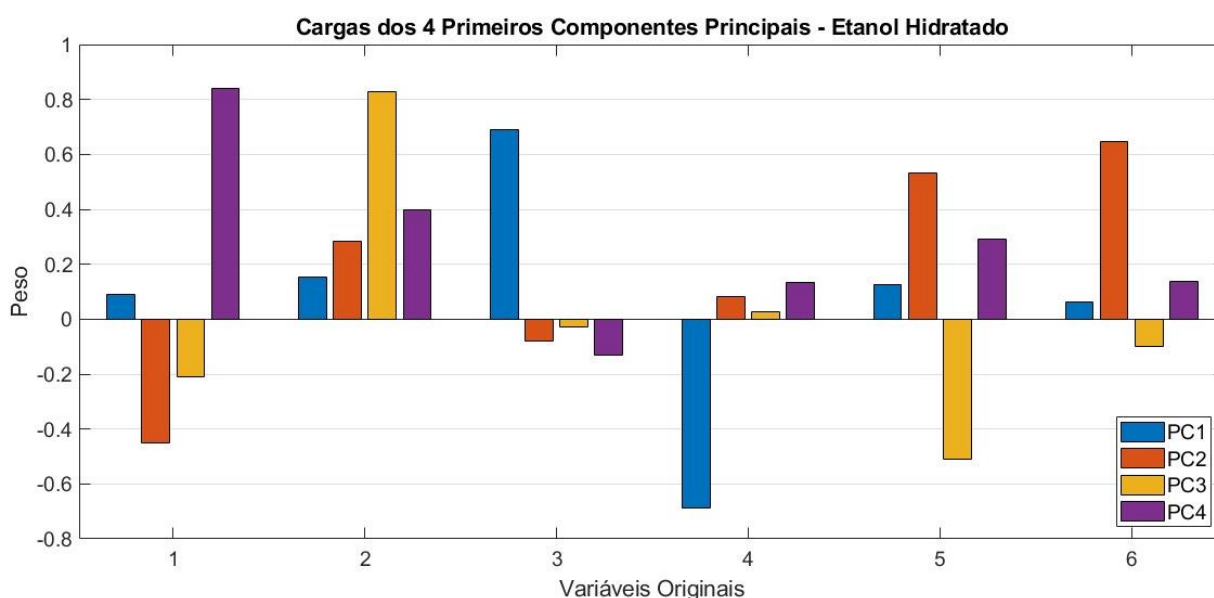
devido todas as amostras analisadas serem de um único tipo de etanol. Algumas amostras estão mais dispersas, sugerindo possíveis *outliers* ou variações em determinadas características.

De modo geral, nota-se pelo gráfico de dispersão referente a aplicação de análise de componentes principais a um conjunto de dados contendo amostras de etanol hidratado, que o primeiro componente principal (PC1) provavelmente representa a combinação linear das variáveis que mais contribuem para a variabilidade dos dados, que são teor alcoólico e massa específica a 20° C, que são parâmetros fundamentais na caracterização do etanol hidratado. Já o segundo componente principal (PC2), pode estar capturando variações relacionadas a fatores secundários, como condutividade elétrica e teor de metanol.

Além disso, existência de pontos mais afastados da origem indica amostras com características distintas, que de acordo com análise mais detalhada do conjunto de dados de amostras de etanol hidratado, foi possível observar que as amostras de etanol hidratado costumam ser incolores. Entretanto, algumas das amostras que são representadas no gráfico como pontos mais afastados, apresentam como resultado para o ensaio cor (variável 1), as colorações amarela e amarelada. Assim, entende-se que a cor foi a responsável por esse afastamento em relação aos demais pontos situados próximo a origem do gráfico.

A Figura 13, apresenta o gráfico de pesos, que viabiliza a análise da influência de cada uma das seis variáveis originais na organização das amostras de etanol hidratado em relação à PC1, PC2, PC3 e PC4, obtidos por PCA.

Figura 13 – Gráfico de pesos dos quatro primeiros componentes principais em relação as variáveis originais do conjunto de dados de etanol hidratado, analisado por PCA.



Fonte: Produzido pela autora.

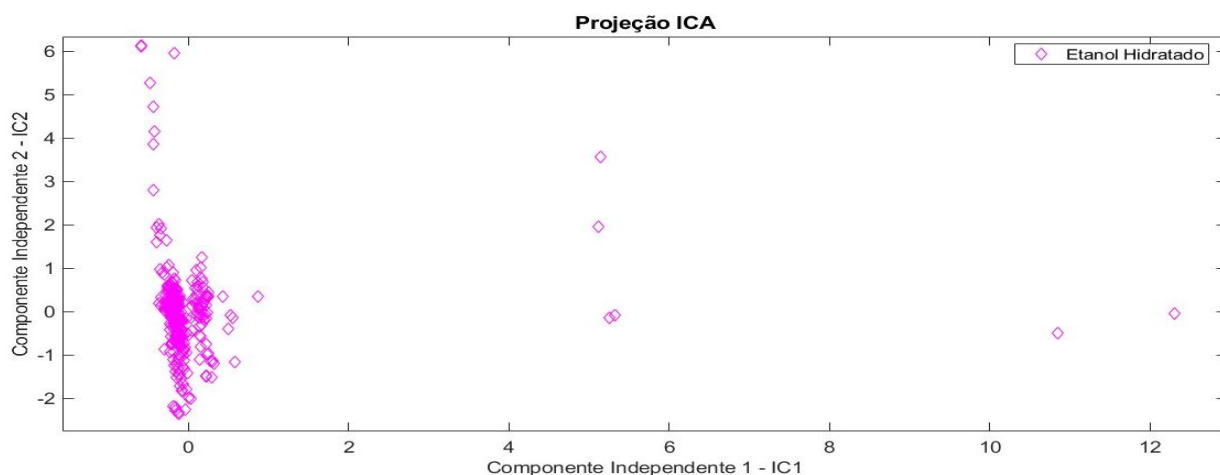
A partir da análise do gráfico da Figura 13, observa-se que os pesos indicam o quanto cada variável ((1) cor, (2) condutividade elétrica, (3) massa específica a 20° C, (4) teor alcoólico, (5) potencial hidrogênico (pH) e (6) teor de metanol), contribui para a formação de cada componente principal. Desse modo, algumas das principais observações que podem ser feitas a partir desse gráfico, é que a PC1 apresenta alto peso negativo no teor alcoólico (aproximadamente - 0,69) e apresenta alto peso positivo na massa específica a 20°C (aproximadamente 0,69), indicando que essas variáveis possuem um papel crucial na formação do primeiro componente.

Analisando as demais componentes contidas no gráfico da Figura 13, observa-se que PC2 apresenta peso negativo para a variável cor (- 0,45), indicando que essa característica é relevante para diferenciar as amostras de etanol hidratado. Além disso, PC2 apresenta peso positivo nas variáveis teor de metanol (0,53) e potencial hidrogênico (0,64), levando a entender que o PC2 está diretamente relacionado a variações na pureza do etanol. Quanto ao PC3, observa-se uma forte influência da variável condutividade elétrica (peso de aproximadamente 0,83), indicando que esse componente pode estar associado à presença de impurezas iônicas. Já PC4, apresenta influência considerável das variáveis cor (peso de aproximadamente 0,84) e condutividade elétrica (peso de aproximadamente 0,34), sugerindo que PC4 possui relação direta com às características visuais e elétricas do etanol hidratado.

5.6 Modelagem por ICA para o Etanol Hidratado

A figura 14, representa a projeção gráfica das amostras de etanol hidratado relacionando os dois primeiros componentes independentes (IC1 e IC2), que foram obtidos através da aplicação da análise de componentes independentes ao conjunto de dados com amostras de etanol hidratado.

Figura 14 – Gráfico bidimensional com os dados das amostras de etanol hidratado, após a aplicação de ICA.



Fonte: Produzido pela autora.

No gráfico da Figura 14, o eixo x representa IC1 e o eixo y representa o IC2. Os losangos em magenta, representam as amostras de etanol hidratado, estando elas projetadas no espaço definido pelos dois primeiros componentes independentes. Analisando a distribuição das amostras no gráfico, observa-se que a maior parte delas está concentrada próximo ao eixo vertical (IC2), mostrando que IC1 pode não ser um fator determinante para a variabilidade dos dados de etanol hidratado. Também é possível observar, que algumas amostras aparecem como pontos distantes (ou *outliers*), indicando que eles se diferem da maioria devido algumas características. Sendo que, foi observado na análise de PCA para as amostras de etanol hidratado, que algumas dessas amostras apresentam variações significativas quanto a variável cor.

Desse modo, a partir da Figura 14, pode-se inferir que a análise de componentes independentes para o conjunto de dados analisados, pode ser útil para identificar impurezas, adulterações ou inconsistências na qualidade do etanol hidratado.

6. Conclusão

Abordando a questão da Análise Multivariada da qualidade dos combustíveis especificamente no Estado do Maranhão, este estudo demonstrou que ao utilizar técnicas de análise estatísticas multivariadas, foi observado que analisando amostras de diesel a Análise de Componentes Principais (PCA) se mostrou mais eficaz do que a Análise de Componentes Independentes (ICA) na distinção dos grupos de diesel, evidenciando que as variáveis analisadas possuem correlações relevantes, capturadas de forma mais eficaz pela PCA. Em contraste, a ICA gerou um gráfico mais compacto ao longo do eixo x, apresentando menor distinção entre os grupos, o que sugere que tal técnica pode não ser a mais adequada para a diferenciação dos tipos de diesel com os ensaios disponíveis. Ademais, a identificação de *outliers* indica que certas amostras possuem características químicas peculiares, possivelmente resultantes de variações na formulação do diesel, contaminação ou erros nas medições.

No que se refere às gasolinas, a PCA evidenciou que, apesar de existir uma certa sobreposição entre a gasolina comum e a aditivada, as amostras apresentam regiões bem demarcadas, o que sugere diferenças substanciais entre os dois tipos de combustível. A separação ao longo do primeiro componente principal (PC1) indica que uma ou mais variáveis físico-químicas desempenham um papel decisivo nessa distinção, enquanto a maior dispersão da gasolina aditivada em PC2 pode refletir características específicas dessa formulação. Já a aplicação de PCA nos dados de etanol hidratado mostrou que os quatro primeiros componentes explicam quase 90% da variância dos dados. Isso indica que, para análises futuras, pode ser possível reduzir o número de variáveis mantendo a maior parte da informação relevante, facilitando a visualização e interpretação dos dados. Enquanto a análise do etanol hidratado por meio da ICA, revelou um padrão homogêneo na maioria das amostras, embora algumas variações significativas possam indicar possíveis impurezas, adulterações ou inconsistências na qualidade do combustível.

Dessa forma, é possível afirmar que a aplicação de técnicas multivariadas para análise da qualidade de combustíveis pode contribuir tanto para a redução da dimensionalidade dos dados, como para a detecção de padrões e agrupamentos de dados que permitem uma análise mais clara sobre as informações contidas nos conjuntos de dados analisados. Por outro lado, é importante ressaltar, que nem todas as técnicas de análise multivariada devem ser aplicadas para análise da qualidade dos combustíveis. Pois, como foi visto neste trabalho, ICA não propiciou resultados relevantes para o estudo realizado, assim, acredita-se que isto se deve aos tipos de dados analisados, que são dados discretos. No entanto, isto não quer dizer que essa técnica não deve ser aplicada em

estudos como este, mas sim, que existem outras técnicas melhores, como por exemplo, PCA.

Ademais, constatado ser uma metodologia eficiente, sendo utilizada a forma de análise estatística mais adequada a cada caso, a análise multivariada da qualidade de combustíveis comerciais é um procedimento estatístico avançado que possibilita avaliar concomitantemente múltiplas variáveis relacionadas às propriedades dos combustíveis. Essa metodologia é de fundamental importância para identificar padrões, verificar adulterações e assegurar a conformidade com os padrões estabelecidos para cada tipo de combustível.

Portanto, estudos que possuem como principal foco a análise multivariada empregada para avaliar a qualidade dos combustíveis comerciais, seja no Estado do Maranhão, como é o caso do presente estudo, ou em qualquer outro estado ou região do país, são de grande valia, pois proporcionam ferramentas eficientes para monitorar e controlar a qualidade e a conformidade dos combustíveis com os padrões regulatórios preestabelecidos.

6.1 Trabalhos futuros

Para trabalhos futuros, com base nos resultados obtidos, as seguintes direções são recomendadas:

1. Análise de variáveis mais relevantes: Realizar um estudo estatístico aprofundado para identificar quais variáveis físico-químicas mais influenciam a separação dos grupos de diesel e gasolina.
2. Otimização de técnicas de análise: Investigar outras metodologias de análise multivariada, como Análise Discriminante Linear (LDA) ou técnicas de aprendizado de máquina, para aprimorar a classificação dos combustíveis.
3. Ampliação do conjunto de dados: Incluir novas amostras de combustíveis de diferentes fornecedores e regiões para verificar a robustez dos modelos aplicados.

Essas direções podem contribuir de maneira significativa para o avanço dos métodos de análise da qualidade dos combustíveis e sua aplicabilidade na indústria.

Referências

AGUIAR, T. C.; CRISPIM, P. T. B.; AZEVEDO, M. S.; MARQUES, C. H. V.; MAGALHÃES, E. O. **Estudo físico-químico da gasolina no estado de Rondônia: uma aplicação das matrizes de correlação e estatística multi-variada**. In: CONGRESSO BRASILEIRO DE QUÍMICA, 2008, Rio de Janeiro. Anais [...]. Rio de Janeiro: CBQ, 2008

ALVES, E. R. et al. **Análise Multivariada de Parâmetros Físico-químicos da Gasolina Comercializada no Agreste da Paraíba**. Disponível em: <https://www.abq.org.br/cbq/2022/trabalhos/4/250-417.html>. Acesso em: 02 de Março de 2025.

ALVES, William Ferreira et al. **Análise Multivariada Dos Parâmetros Físico-Químicos Da Gasolina “Tipo C” Comercializada No Vale Do Juruá-Acre**. South American Journal of Basic Education, Technical and Technological, v. 6, n. 1, 2019.

ANP-AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP (2020a). **Qualidade de Produtos**. Disponível em: <https://www.gov.br/anp/pt-br/assuntos/qualidade-de-produtos>. Acesso em: 28 de Fevereiro de 2025.

ANP-AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP (2020b). **Programa de Monitoramento da Qualidade dos Combustíveis**. Disponível em: <https://www.gov.br/anp/pt-br/assuntos/qualidade-de-produtos/programas-monitoramento/programa-de-monitoramento-da-qualidade-dos-combustiveis>. Acesso em: 14 de Fevereiro de 2025.

ANP-AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP (2020c). **Gasolina**. Disponível em: <https://www.gov.br/anp/pt-br/assuntos/producao-de-derivados-de-petroleo-e-processamento-de-gas-natural/producao-de-derivados-de-petroleo-e-processamento-de-gas-natural/gasolina>. Acesso em: 27 de Fevereiro de 2025.

ANP-AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP (2021). **Boletim de Biocombustíveis e Qualidade de Produtos**. Disponível em: <https://www.gov.br/anp/pt-br/centrais-de-conteudo/publicacoes/boletins-anp/boletim-de-biocombustiveis-e-qualidade-de-produtos>. Acesso em: 14 de Fevereiro de 2025.

ANP-AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP (2024). **Síntese Mensal de Comercialização de Combustíveis**. Disponível em: <https://www.gov.br/anp/pt-br/centrais-de-conteudo/publicacoes/sinteses/scc/2024/sintese-volumes-dezembro.pdf>. Acesso em: 03 de Março de 2025.

BARTARELLI JUNIOR, Admir A. **Introdução à Estatística Multivariada e Análise de Dados** – Notas de Aula. Programa de pós Graduação em Economia Aplicada, Universidade Federal de Juiz de Fora (UFJF). Juiz de Fora, Minas Gerais, 2016.

BORGES, Pamella K. B. **Presença de Etanol no Meio Ambiente: Uma Análise de Estudos e Perspectivas Referente aos Impactos Ambientais e Condições de Contenção para o Armazenamento**. Trabalho de conclusão de curso (Bacharelado - Engenharia Ambiental) - Universidade Estadual Paulista (UNESP), Sorocaba, 2024.

BRASIL (2023). **Anuário Estatístico Brasileiro do Petróleo, Gás Natural e Biocombustíveis: 2023 /**

Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Rio de Janeiro: ANP, 2006.

CHOI, S. et al. Blind Source Separation and Independent Component Analysis: A Review. **Neural Information Processing – Letters and Reviews**, 6, n. 1, January, 2005.

CORREA, Camila. **Metodologias Analíticas para Avaliar a Biodegradabilidade do Diesel, Biodiesel e Blendas B10**. Tese (Doutorado em Química) - Universidade Federal Do Rio Grande Do Sul (UFRGS). Porto Alegre, 2021.

CORRÊA, Paulo José Melo Gomes. **AN ONTOLOGY FOR REPRESENTATION OF KNOWLEDGE OF FIELD ANALYTICAL CHEMISTRY WITH ADDITION OF NEW AGENTS AND FEATURES FOR ANALYSIS AND MONITORING OF FUELS**. 2009. 132 f. Dissertação (Mestrado em Engenharia) - Universidade Federal do Maranhão, São Luis, 2009.

DA ROCHA, Clarice O. *et al.* **INVESTIGAÇÃO DA QUALIDADE DA GASOLINA E DO ETANOL COMERCIALIZADOS EM POSTOS DE SERVIÇOS NO AGRESTE PARAIBANO**. Anais do 9º PdPetro. ABPG, 2017.

DIONYSIO, R. B.; MEIRELLES, F. V. P. **Combustíveis: a Química que move o mundo**. 2016. Cead PUC-RJ. Disponível em: <http://web.ccead.pucRio.br/condigital/mvsl/Sala%20de%20Leitura/conteudos/SL_combustiveis.pdf>. Acesso em: 14 de Fevereiro de 2025.

DO NASCIMENTO, Francisca D. F. et al. **Determinação do Teor de Etanol na Gasolina Comercializada em Postos de Combustíveis do Município de Picos-PI / Determination Of Ethanol Content In Gasoline Marketed In Fuel Stations In The Municipality Of Picos-PI**. Brazilian Journal of Animal and Environmental Research, v. 4, n. 4, p. 5669-5678, 2021.

DUARTE, J. B. **Introdução à Análise de Componentes Principais (Com Aplicação do SAS®-Statistical Analysis System)**. Universidade Federal de Goiás, p. 03-04, 1998.

FERREIRA, D. F. **Análise Multivariada**. Lavras, Minas Gerais, 394 p., 1996.

FOLLI, G. et al. **Tutorial para aplicação didática de quimiometria em software gratuito – Parte I: Análise de Componentes Principais em dados de infravermelho médio e propriedades físico químicas de amostras de petróleo**. Revista Ifes Ciência, v. 9, nº 1, p. 01-14, 2023.

Ge, Z., & Song, Z. (2007). Process Monitoring Based on Independent Component Analysis-Principal Component Analysis (ICA-PCA) and Similarity Factors. *Ind. Eng. Chem. Res*, 46(7), 2054–2063.

GOMES, Maria da Conceição R.; CAVALCANTE, Itabaraci N. **Aplicação da análise estatística multivariada no estudo da qualidade da água subterrânea**. Águas Subterrâneas, v. 31, n. 1, p. 134-149, 2017.

Hyvarinen, A.; Oja, E. Independent component analysis: algorithms and applications. *Neural Network* 2000, 13, 411-430.

IBP (2024) - **Boletim dos Ciclo Diesel 2024**. Disponível em : <https://www.ibp.org.br/personalizado/uploads/2024/03/boletim-ciclo-diesel-i-marco-de-2024-9.pdf>.

Acesso 31 de março de 2024.

MARQUES, Delano. B. **Sistema Integrado de Monitoramento e Controle da Qualidade de Combustível**. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Maranhão (UFMA), São Luís, 2004.

MINGOTI, Sueli A. **Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada**. Belo Horizonte: Editora UFMG, 2005.

MIRANDA, P. E. V. DE. **Combustíveis - materiais essenciais para prover energia à nossa sociedade**. Matéria, v. 18, n. 3, p. 12-15, 2013.

MME – MINISTÉRIO DE MINAS E ENERGIA. **CNPE Passa a ter Competência para Fixar Teor de Etanol Anidro na Gasolina**. Disponível em: <https://www.gov.br/mme/pt-br/assuntos/noticias/cnpe-passa-a-ter-competencia-para-fixar-teor-de-etanol-anidro-na-gasolina#:~:text=O%20etanol%20anidro%20%C3%A9%20misturado,%25%20e%2027%2C5%25>. Acesso em: 03 de Março de 2025.

OLIVEIRA, E. P. et al. Investigação do teor de água no Biodiesel utilizado na composição do Diesel B comercializado por uma distribuidora de combustíveis em Manaus/AM. *Brazilian Journal of Development*, v. 7, n. 9, p. 89663-89680, 2021.

PARENTE, L. E. R. **Análise Exploratória de Perfilagem Acústica para Avaliação da Qualidade de Cimento com Simulações Computacionais**. Relatório – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica, Rio de Janeiro, p. 03 – 04, 2022.

PETROBRAS. **Óleo Diesel: Informações Técnicas** (2023). Disponível em: https://petrobras.com.br/documents/2677942/0/Manual%20de%20Diesel_2023.pdf/#:~:text=O%20%C3%B3leo%20diesel%20%C3%A9%20um,%2C%20nitrog%C3%AAnio%2C%20enxofre%20e%20oxig%C3%AAnio.>. Acesso em: 03 Março 2025.

PMQC-Programa de Monitoramento da Qualidade dos Combustíveis. Painel Dinâmico do Programa de Monitoramento da Qualidade dos Combustíveis. (2025). Disponível em: <https://app.powerbi.com/view?r=eyJrIjoiOTkyOTljYjgtNmQ2Yy00NTU0LTNmYjUtZWNlOG RiNzczMjc2IiwidCI6IjQ0OTlmNGZmLTl0YTtytNGl0Mi1iN2VmLTExNGFmY2FkYzkyMyJ9>. Acesso em: 07 de Março de 2025.

RIBEIRO, Camilo B. et al. **Estudo do Desempenho e das Emissões Gasosas ee Motores Não-Rodoviários de Pequeno Porte Operando com Gasolina e Álcool Etílico Anidro Combustível (Aeac) Misturados em Diferentes Proporções**. Dissertação (Mestrado em Bioenergia) – Universidade Estadual do Centro-Oeste, Paraná, 2017.

RIBEIRO, Camilo Bastos; SCHIRMER, Waldir Nagel. Panorama dos combustíveis e biocombustíveis no Brasil e as emissões gasosas decorrentes do uso da gasolina/etanol. *BIOFIX Scientific Journal*, v. 2, n. 2, 2017.

Roberts, S. and Everson, R. *Independent Component Analysis: principles and practice*. Cambridge University Press. 2001.

SANGUANSAT, P. **Principal Component Analysis**. BoD–Books on Demand, 2012.

SANTOS, Fernando A. et al. **Potencial da palha de cana-de-açúcar para produção de etanol**.

Química nova , v. 35, pág. 1004-1010, 2012.

SANTOS, M. C. S. Dos et al. Classificação de Dados de Produção de Petróleo e Gás. In: Anais do Congresso Nacional de Engenharia de Petróleo, Gás Natural e Biocombustíveis. Anais Even, Aracaju(SE) Centro de Convenções de Sergipe, 2024.

SBQ-Superintendência de Biocombustíveis e Qualidade de Produtos. **Boletim de Biocombustíveis de Qualidade de Produtos (2022)**. Agência Nacional de Petróleo e Biocombustíveis. Rio de Janeiro, 2022.