

ANDRÉ VINICIUS OLIVEIRA MAGGIO

**MODELAGEM DO TORQUE PARA MOTORES DE COMBUSTÃO
FLEX-FUEL APlicando TÉCNICAS DE ANÁLISE DE DADOS E
*MACHINE LEARNING***

SÃO PAULO

2024

ANDRÉ VINICIUS OLIVEIRA MAGGIO

**MODELAGEM DO TORQUE PARA MOTORES DE COMBUSTÃO
FLEX-FUEL APlicando TÉCNICAS DE ANÁLISE DE DADOS E
*MACHINE LEARNING***

Monografia apresentada ao Programa de Educação Continuada da Escola Politécnica da Universidade de São Paulo, para obtenção do título de Especialista, pelo Programa de Pós-Graduação em Engenharia de Dados e Big Data.

Orientador: Prof Dr. Luiz Sérgio de Souza.

SÃO PAULO

2024

ANDRÉ VINICIUS OLIVEIRA MAGGIO

**MODELAGEM DO TORQUE PARA MOTORES DE COMBUSTÃO
FLEX-FUEL APlicando TÉCNICAS DE ANÁLISE DE DADOS E
*MACHINE LEARNING***

Monografia apresentada ao Programa de
Educação Continuada

da Escola Politécnica da Universidade de São
Paulo, para obtenção do título de Especialista,
pelo Programa de Pós-Graduação em
Engenharia de Dados e Big Data.

Área de concentração: Tecnologia da
Informação – Engenharia/ Tecnologia/ Gestão

Orientador: Prof Dr. Luiz Sérgio de Souza.

SÃO PAULO

2024

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catalogação-na-publicação

MAGGIO, ANDRÉ VINICIUS OLIVEIRA

MODELAGEM DO TORQUE PARA MOTORES DE COMBUSTÃO FLEX-FUEL APLICANDO TÉCNICAS DE ANÁLISE DE DADOS E MACHINE LEARNING / A. V. O. MAGGIO -- versão corr. -- São Paulo, 2024.

54 p.

Monografia (Especialização em Engenharia de Dados & Big Data) – Escola Politécnica da Universidade de São Paulo. PECE – Programa de Educação Continuada em Engenharia.

**1.ANÁLISE DE DADOS 2.MOTORES DE COMBUSTÃO INTERNA
3.BIOCOMBUSTÍVEIS 4.SIMULAÇÃO (MODELAGEM COMPUTACIONAL)
I.Universidade de São Paulo. Escola Politécnica. PECE – Programa de Educação Continuada em Engenharia II.t.**

CURSO ENGENHARIA DE DADOS E BIG DATA

Coord.: Prof. Solange N. Alves de Souza

Vice-Coord.: Prof Pedro Luiz Pizzigatti Corrêa

Perspectivas profissionais alcançadas com o curso:

Como engenheiro elétrico de formação, posso afirmar que o curso ampliou significativamente meu leque de conhecimentos e proporcionou uma rica troca de experiências com colegas, professores e especialistas convidados ao longo das aulas. O curso oferece uma visão abrangente sobre diversas áreas de competência, incluindo análise de dados, governança, segurança, regulamentação e ética, temas fundamentais para esta área tão importante da tecnologia.

RESUMO

Neste trabalho, aplicaram-se técnicas de análise de dados e aprendizado de máquina (*machine learning*) a um problema real, com o objetivo de modelar o torque de motores de combustão interna do tipo flex-fuel, a partir de dados de ensaios dinamométricos. Desenvolveu-se a metodologia e a arquitetura de dados, envolvendo as etapas de ingestão, pré-processamento, processamento e análise dos dados. Foi feita a conversão dos arquivos para um formato comum e a padronização dos nomes das variáveis, criando-se um dicionário de dados com o nome, descrição e unidade física de cada variável. Durante o processamento, foi desenvolvido um algoritmo para identificar janelas temporais nos ensaios em regime transitório, nas quais o motor exibiu comportamento estacionário, permitindo combinar os dados dos ensaios em regimes transitório e permanente em um único conjunto de dados para a modelagem do torque. Na modelagem, testaram-se o modelo padrão e duas técnicas de aprendizado de máquina, utilizando redes neurais artificiais (RNAs) e regressão por processos gaussianos (GPR - *Gaussian Process Regression*). O modelo padrão de torque utiliza um mapa (*look-up table*) que representa o torque máximo produzido pelo motor para diferentes condições de rotação e pressão no coletor de admissão, considerando gasolina e etanol comum. A geometria da rede neural foi obtida por meio de um método iterativo, ajustando o número de neurônios nas camadas ocultas e selecionando as quatro melhores arquiteturas. No modelo de regressão por processos gaussianos, foram testados quatro tipos de kernels diferentes. Foram gerados histogramas e gráficos para cada variável do conjunto de dados unificado obtido (*dataset*), permitindo a visualização das características principais do *dataset*, como distribuição e variabilidade dos dados. Realizou-se uma análise de correlação entre as variáveis e o torque, utilizando mapas de calor (*heatmap*) e análises gráficas de correlação de *Spearman*. Por fim, utilizaram-se indicadores de desempenho (KPIs - *Key Performance Indicators*) para avaliar e comparar o desempenho dos modelos obtidos, permitindo selecionar o modelo com melhor desempenho e menor tempo de execução.

Palavras-chave: ANÁLISE DE DADOS, MACHINE LEARNING, MOTORES DE COMBUSTÃO INTERNA, MODELAGEM, BIOCOMBUSTÍVEIS.

LISTA DE FIGURAS

Figura 1 – Variáveis controladas pela ECU	15
Figura 2 – Quatro tempos do motor de combustão interna.	16
Figura 3 – Principais sensores e atuadores do motor de combustão interna.	16
Figura 4 – Modelo de torque padrão.....	18
Figura 5 – Tipos principais de dinamômetros.	19
Figura 6 – Fatores da eficiência indicada dependente da rotação ($\eta\omega m$), relação ar-combustível ($\eta\lambda$) e ângulo de ignição ($\eta\zeta$).	22
Figura 7 – Exemplo de arquitetura de rede neural para modelagem do torque.....	24
Figura 8 – Regressão baseada em Processos Gaussianos.....	25
Figura 9 – Arquitetura da GPR para modelagem do torque.	26
Figura 10 – Modelo de processo de dados.....	27
Figura 11 – Fonte de dados.	28
Figura 12 – Pontos de interesse.	33
Figura 13 – Mapa de torque para gasolina E27	35
Figura 14 – Mapa de torque para etanol E100.	35
Figura 15 – Histogramas do mainDataset.	38
Figura 16 – Gráficos para cada variável do mainDataset.....	39
Figura 17 – Mapa de calor (<i>heatmap</i>).	40
Figura 18 – Análises gráficas de correlação entre cada variável e o torque.....	41
Figura 19 – Gráficos de regressão para modelo padrão.	42
Figura 20 – Gráficos de desempenho das quatro melhores arquiteturas de rede neural.....	43
Figura 21 – Gráficos de regressão das quatro melhores arquiteturas de rede neural.	44
Figura 22 – Gráficos de regressão das quatro tipo de kernel utilizados.	45
Figura 23 – Gráficos de dispersão com erro absoluto percentual para cada ponto do mainDataset	47

LISTA DE TABELAS

Tabela 1 – Nomes padronizados, descrição e unidade física de cada variável.....	30
Tabela 2 – Indicadores de desempenho e tempo de execução para cada modelo...	47
Tabela 3 – Indicadores de desempenho para os principais artigos de referência (<i>Benchmark</i>).....	48
Tabela 4 – Dados do Laboratório de Combustíveis Líquidos.....	54

LISTA DE ABREVIATURAS E SIGLAS

APE	Absolute Percentage Error (Erro percentual absoluto).
°APMS	graus Antes do Ponto Morto Superior
ARD	Automatic Relevance Determination
CO	monóxido de carbono
CO ₂	dióxido de carbono
dat	data
ECM	Engine Control Module
ECU	Engine Control Unit
GPR	Gaussian Process Regression regressão por processos gaussianos
IEMA	Instituto de Energia e Meio Ambiente
KPIs	Key Performance Indicators
MAF	Mass Air Flow sensor
MAPE	Mean Absolute Percentage Error (Erro percentual absoluto médio).
mat	MATLAB daTa file
MBT	Maximum Break Torque
mdf4	Measurement Data Format v.4
MLE	Maximum Likelihood Estimation
MLP	Multilayer Perceptron
MSE	Mean Squared Error (erro medio quadrático).
NOx	óxidos de nitrogênio
NRMSE	Normalized Root Mean Squared Error (Erro quadrático médio normalizado).
O ₂	oxigênio
PMI	Ponto Morto Inferior
PMS	Ponto Morto Superior
RMSE	Root Mean Squared Error (Raiz do erro quadrático médio).
RNAs	Redes Neurais Artificiais
THC	Total HydroCarbon (hidrocarbonetos totais).
TMAP	Manifold Pressure and Temperature sensor
TPS	(Throttle Position Sensor).
xls	eXcel Spreadsheet
xlsx	eXcel Spreadsheet

SUMÁRIO

1.	Introdução	10
1.1.	Objetivos.	12
1.2.	Justificativa.....	12
2.	Modelagem do torque para motores de combustão interna.	14
2.1.	Motor de combustão interna.....	14
2.2.	Modelagem do torque.....	18
2.2.1.	Modelagem analítica do torque.....	20
2.3.	Modelagem do torque com <i>machine learning</i>	23
2.3.1.	Redes neurais artificiais (RNAs).	23
2.3.2.	Regressão baseada em processos gaussianos (GPR).	24
3.	Desenvolvimento dos modelos de torque a partir dos dados experimentais.....	27
3.1.	Ingestão de dados.	27
3.2.	Pré-processamento dos dados.	30
3.3.	Processamento dos dados.	31
3.4.	Análise de dados.....	34
3.4.1.	Modelos de torque obtidos.....	34
4.	Analise e resultados obtidos.....	38
4.1.	Indicadores de desempenho dos modelos.....	45
5.	Conclusão	49
	Referências	50
	Anexo A – equipamentos utilizados.	53

1. INTRODUÇÃO.

Ao longo de mais de um século de história, os veículos automotores revolucionaram a humanidade, oferecendo um meio de transporte seguro, confortável e rápido. A maioria dos veículos modernos utiliza um sistema de propulsão (*powertrain*) equipado com um motor de combustão interna [1], [2]. O uso dos combustíveis derivados de petróleo e o crescimento em larga escala da frota mundial de veículos gerou uma série de impactos, sobretudo ambientais. Dados do Instituto de Energia e Meio Ambiente (IEMA) indicam que, na cidade de São Paulo, os automóveis são responsáveis por mais de 73% das emissões de gases que contribuem para o aquecimento global.

Atender às demandas de conforto, segurança, desempenho, eficiência energética, ao mesmo tempo que se busca reduzir as emissões de gases, representa um desafio complexo de engenharia, incentivando pesquisas científicas em diversas áreas [3], [4], [5]. Nas últimas décadas, surgiram novas tecnologias, como conversores catalíticos, comando de válvulas variável (*Variable Valve Timing*, VVT), injeção direta de combustível e sistemas de sobrealimentação. Surgiram também, combustíveis alternativos, como o etanol, veículos elétricos e híbridos. Essas tecnologias, em conjunto com o gerenciamento eletrônico dos motores de combustão, têm contribuído para a redução do consumo de combustível e das emissões em comparação aos veículos convencionais [1].

Para otimizar seu funcionamento, os motores de combustão interna modernos são equipados com sistemas de controle eletrônicos embarcados [1] [2]. O hardware utilizado no controle de motor é chamado de unidade de controle do motor ou ECU (*Engine Control Unit*), também conhecido como ECM (*Engine Control Module*) [4]. A exemplo do Brasil, as emissões de monóxido de carbono (CO) de um veículo leve nacional, que eram cerca de 54 g/km em 1986, atualmente são inferiores a 0,4 g/km, com o uso dos sistemas de controle eletrônico nos motores de combustão modernos.

Uma das principais variáveis de saída de um motor de combustão interna é o torque, que é uma medida da força rotacional gerada pelo motor no eixo do virabrequim. O torque produzido pelo motor, é aplicado ao eixo de transmissão, permitindo que o veículo ganhe velocidade [1].

O processo de calibração da ECU requer uma série de testes dinamométricos e ensaios com o veículo. Para melhorar a resposta e a dirigibilidade, a ECU do motor, deve controlar o torque produzido [5]. Para medir o torque e a potência do motor, utiliza-se um equipamento chamado dinamômetro, que pode ser de dois tipos: de bancada e de chassi. No dinamômetro de bancada, o motor é montado diretamente no equipamento, sem o veículo, permitindo medir o torque e a potência no próprio eixo do motor. Já no dinamômetro de chassi, o veículo é fixado sobre rolos, permitindo a medição do torque e da potência de saída diretamente nas rodas [1].

O dinamômetro é um equipamento de custo elevado e dimensões incompatíveis com o uso em veículos durante condições reais de condução. Portanto, a medição direta do torque no veículo em condições reais de condução acaba sendo inviável, tornando necessário o uso de modelos para estimar o torque. Esse modelo funciona como um "sensor virtual" ou estimador de torque, permitindo que a ECU estime o valor do torque produzido em cada ponto de operação do motor [3], [4], [5].

Modelar com precisão o torque em motores de combustão *flex-fuel* é uma tarefa complexa, pois envolve diversas variáveis, como o tipo de combustível, o ângulo de ignição e a proporção da mistura ar-combustível. Os modelos tradicionais utilizam mapas tridimensionais, nos quais um dos eixos representa a rotação do motor e o outro representa a abertura da válvula borboleta ou a pressão no coletor de admissão de ar [3], [6]. Os mapas são obtidos por meio de ensaios dinamométricos que coletam dados do motor em condições estacionárias para cada combinação de rotação e pressão na admissão [3].

1.1. Objetivos.

O objetivo geral deste trabalho é desenvolver e validar modelos de estimativa de torque para motores de combustão interna do tipo *flex-fuel*, utilizando técnicas de *machine learning* e análise de dados, com base em dados disponíveis de ensaios previamente realizados em regimes transitórios e estacionários do motor de combustão no dinamômetro.

Dos os objetivos específicos, destacam-se:

- Aplicar técnicas de análise aos dados disponíveis dos diversos ensaios em regime transitório do motor, que contêm séries temporais das variáveis, para identificar e extrair automaticamente as regiões de operação do motor em regime estacionário.
- Incorporar os dados extraídos de ensaios em regime transitório aos dados disponíveis em condições estacionárias, gerando um único *dataset* integrado para a modelagem do torque.
- Realizar a ingestão dos dados a partir dos diversos arquivos disponíveis contendo os dados brutos (*raw data*) dos ensaios em regime estacionário e transitório do motor.
- Converter os diferentes arquivos para um formato comum, padronizando os nomes das variáveis presentes nos dados dos ensaios.
- Construir um dicionário de dados padronizado para facilitar a manipulação e análise dos dados.

1.2. Justificativa.

A indústria automotiva enfrenta o desafio constante de reduzir emissões e melhorar a eficiência energética dos veículos. Em motores de combustão interna, o torque produzido influencia diretamente o desempenho, a dirigibilidade e as emissões do veículo [6], [7], [10], [11].

A unidade de controle do motor (ECU) controla as diversas variáveis do motor em função da demanda de torque gerada pelo condutor, garantindo uma resposta

eficiente do veículo. No entanto, a medição direta do torque em condições reais de condução é inviável devido ao custo e às dimensões dos equipamentos, tornando necessário o uso de modelos para estimar o torque produzido pelo motor. Modelar com precisão o torque do motor é uma tarefa complexa, pois envolve a interação de diversas variáveis do motor [6], [7], [10], [11].

A utilização das técnicas de análise de dados e *machine learning* permite aprimorar os modelos tradicionais, oferecendo maior precisão e uso de um número maior de variáveis. Os modelos obtidos, permitem estimar o torque para diferentes tipos de combustíveis e diversas condições de operação do motor, tornando-os adequados para uso em motores do tipo *flex-fuel* utilizados pela indústria brasileira.

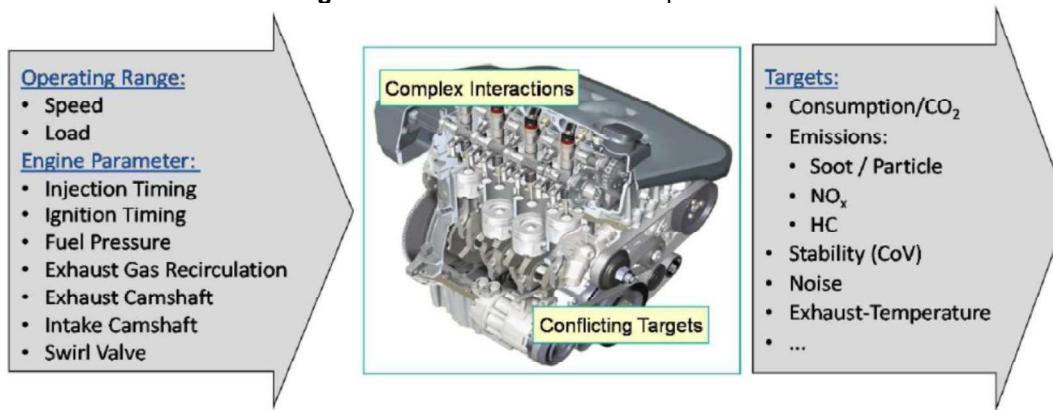
2. MODELAGEM DO TORQUE PARA MOTORES DE COMBUSTÃO INTERNA.

A eletrônica automotiva está cada vez mais presente nos veículos modernos, especialmente em comparação com os veículos desenvolvidos no início do século passado. Nos primeiros motores de combustão, o controle da proporção ar-combustível e da ignição era realizado inteiramente por sistemas eletromecânicos. Com o avanço da tecnologia e a crescente necessidade de otimizar o funcionamento dos motores de combustão interna, os veículos modernos passaram a utilizar sistemas de controle eletrônico embarcado. Esses sistemas são conhecidos como ECU (*Electronic Control Unit*). A ECU que controla o motor de combustão interna é também chamada de unidade de controle do motor (*Engine Control Unit*) e utiliza a mesma sigla. É um dos sistemas mais importantes do veículo, pois realiza o controle e gerenciamento do motor de combustão, influenciando diretamente o consumo de combustível, o desempenho e níveis de emissões de poluentes [3] [4] [5]. A ECU controla o torque produzido pelo motor em função da demanda gerada pelo condutor por meio do pedal do acelerador. A ECU interpreta o sinal da posição do pedal do acelerador como um indicador do nível de torque desejado pelo condutor e ajusta automaticamente os sistemas do motor para fornecer a quantidade de torque necessária, garantindo uma resposta eficiente ao condutor [5].

2.1. MOTOR DE COMBUSTÃO INTERNA.

Uma ECU moderna possui entre 50 a 120 mapas de correção e consulta, podendo chegar a 30 variáveis medidas e 15 manipuladas [14], e possui diferentes estratégias (estados) de operação, tais como partida, marcha lenta, operação normal e operação priorizando a potência (modo *boost*) [14] figura 1.

Figura 1 – Variáveis controladas pela ECU.

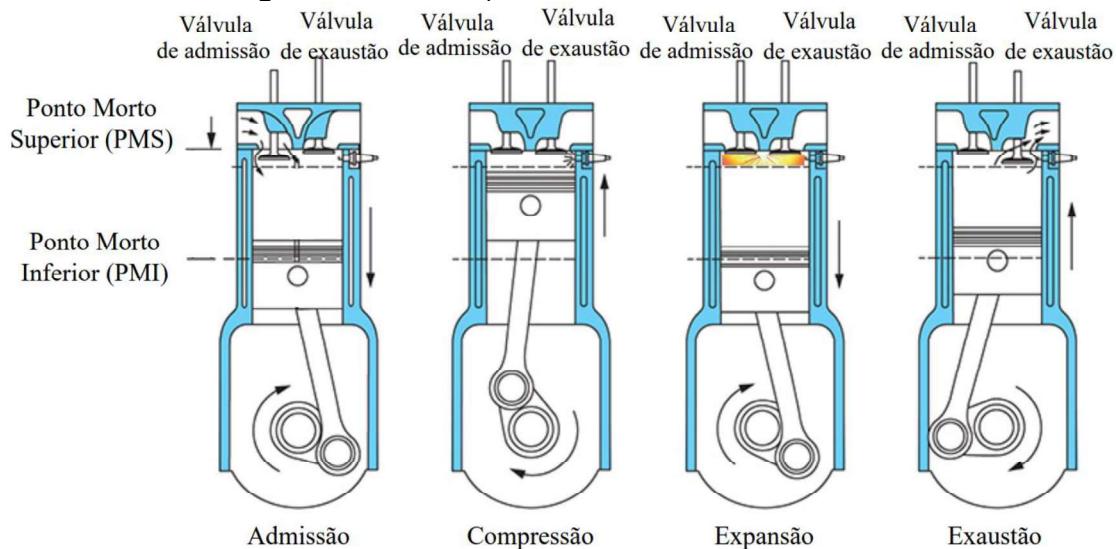


Fonte: [14].

O motor de combustão é uma máquina térmica que converte a energia química do combustível em energia mecânica através de quatro ciclos: admissão, compressão, expansão e exaustão (figura 2). No ciclo de admissão, a válvula de admissão se abre, e o pistão se move do ponto morto superior (PMS) para o ponto morto inferior (PMI), aspirando a mistura de ar-combustível e preenchendo o cilindro com essa mistura. Em seguida, no ciclo de compressão, a válvula de admissão se fecha, e o pistão começa a se mover para cima, em direção ao PMS, comprimindo a mistura de ar-combustível. Quando o pistão está próximo do PMS, a vela de ignição gera uma centelha, iniciando a reação de combustão e produzindo uma frente de chama que se propaga pela mistura[1], [2]

Durante o ciclo de expansão, essa frente de chama continua a se propagar, consumindo completamente a mistura de ar-combustível e fazendo a temperatura e a pressão dos gases dentro do cilindro aumentarem. Os gases quentes empurram o pistão para baixo, em direção ao PMI, convertendo a energia térmica em trabalho mecânico. Por fim, no ciclo de exaustão, a válvula de exaustão se abre, e o pistão se move para cima, expelindo os gases de combustão para o coletor de escape e completando o ciclo [1], [2].

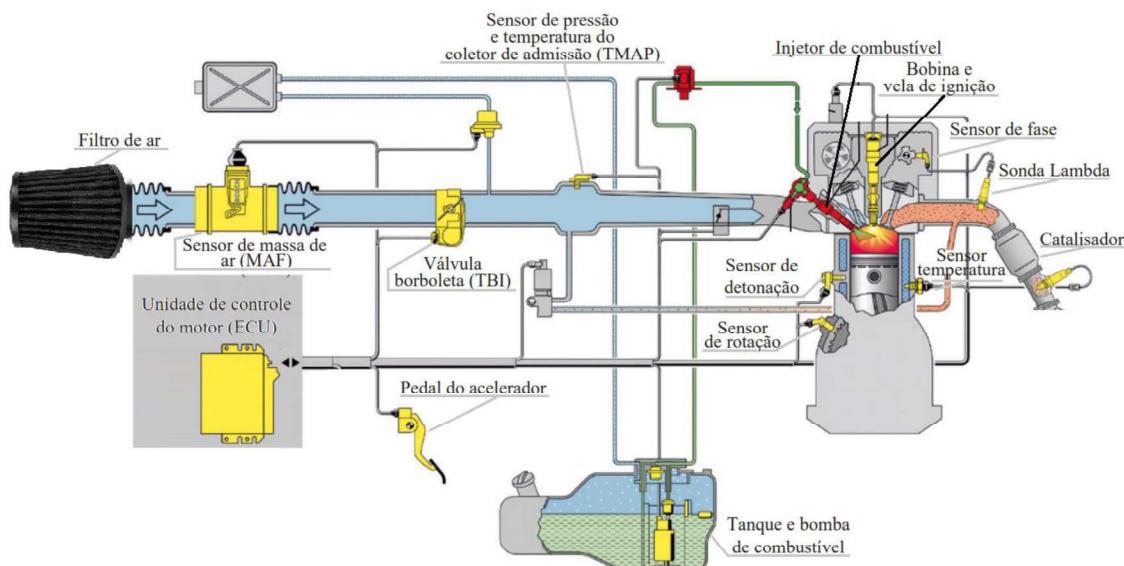
Figura 2 – Quatro tempos do motor de combustão interna.



Fonte: adaptado de [1].

A ECU utiliza diversos sensores e atuadores para monitorar e ajustar as condições de operação do motor, controlando a pressão no ciclo de admissão, a proporção da mistura ar-combustível e o instante em que a vela de ignição gera a centelha que inicia a combustão,[5] figura 3.

Figura 3 – Principais sensores e atuadores do motor de combustão interna.



Fonte: adaptado de [5].

Para controlar a quantidade de ar admitida, o motor utiliza uma válvula chamada válvula borboleta, que regula o fluxo de ar proveniente do filtro de ar (figura 3). A válvula borboleta possui um prato que, dependendo da sua abertura, oferece maior ou menor restrição à passagem do ar, controlando o fluxo. Antes da válvula, existe o sensor de massa de ar (MAF - *Mass Air Flow sensor*), que monitora a quantidade de ar admitida pelo motor. Após passar pela válvula borboleta, o ar segue para o coletor de admissão, onde um sensor monitora a pressão e a temperatura (TMAP - *Manifold Pressure and Temperature sensor*), permitindo à ECU monitorar a pressão e a densidade do ar admitido [2], [5].

Em seguida, o fluxo de ar segue para a válvula de admissão. Próximo a ela, encontra-se o injetor de combustível, que recebe o combustível pressurizado vindo da bomba. Quando acionado, o injetor permite a passagem do combustível por um orifício muito fino, atomizando-o. Essa atomização permite a mistura do combustível com o ar, formando a mistura ar-combustível. Através do acionamento do injetor, a ECU controla a massa de combustível admitida pelo motor em cada ciclo. No coletor de escape, há um sensor de concentração de oxigênio, chamado sonda lambda. A sonda permite que o sistema de controle estime a concentração de oxigênio nos gases de escape e ajuste a proporção da mistura ar-combustível, conforme o regime de operação do motor [2], [5].

A ECU controla o instante angular em que ocorre a centelha na vela de ignição, acionando a bobina de ignição. A bobina converte a tensão elétrica da bateria em uma tensão elevada, necessária para gerar a centelha. O ângulo de ignição é geralmente expresso em °APMS, ou graus antes do ponto morto superior (PMS), indicando o momento da centelha em relação à posição angular do pistão [2].

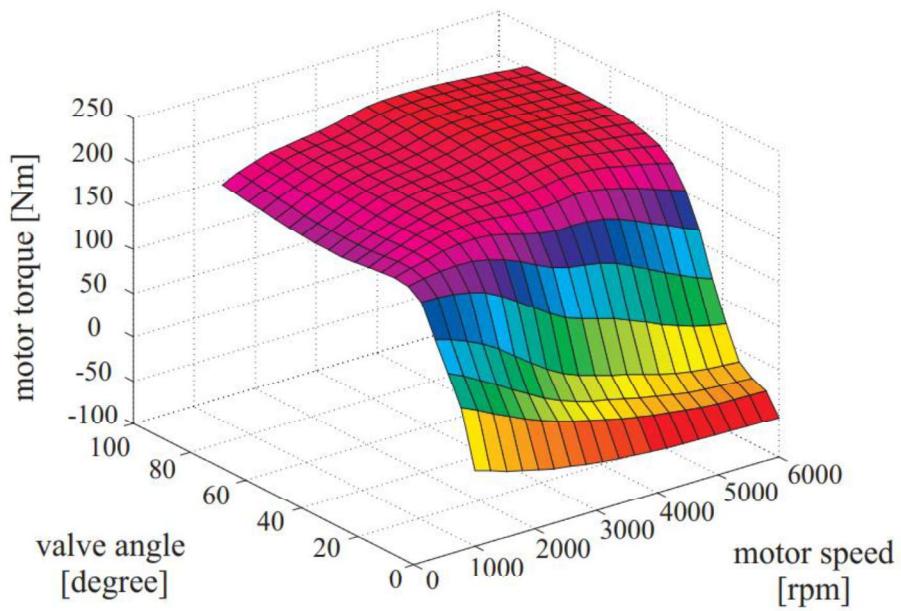
O sensor de rotação permite que a ECU monitore a posição angular do motor e calcule sua velocidade angular. O sensor de fase, informa o momento exato em que a válvula de admissão está se abrindo, permitindo o acionamento síncrono do injetor de combustível. Por fim, o sensor de temperatura do motor permite à ECU monitorar

a temperatura do motor e ajustar seu funcionamento conforme necessário, protegendo o motor [2].

2.2. MODELAGEM DO TORQUE.

O modelo de torque padrão utilizado na ECU é composto por um mapa (*look-up table*). Um dos eixos representa a rotação do motor, o outro eixo representa a abertura da válvula borboleta ou a pressão no coletor de admissão, e o terceiro eixo indica o torque produzido pelo motor. Esses mapas são construídos a partir de ensaios dinamométricos, que coletam dados do motor em condições estacionárias, nos pontos ideais de operação para cada combinação de rotação e pressão de admissão, figura 4 [3], [5]. No caso de veículos *flex-fuel*, normalmente existem dois mapas: um para o torque produzido pelo motor utilizando gasolina e outro mapa para o torque com etanol. Quando o motor está operando com uma mistura de etanol/gasolina, o sistema interpola os valores desses dois mapas conforme a proporção de etanol na mistura [2], [3].

Figura 4 – Modelo de torque padrão.



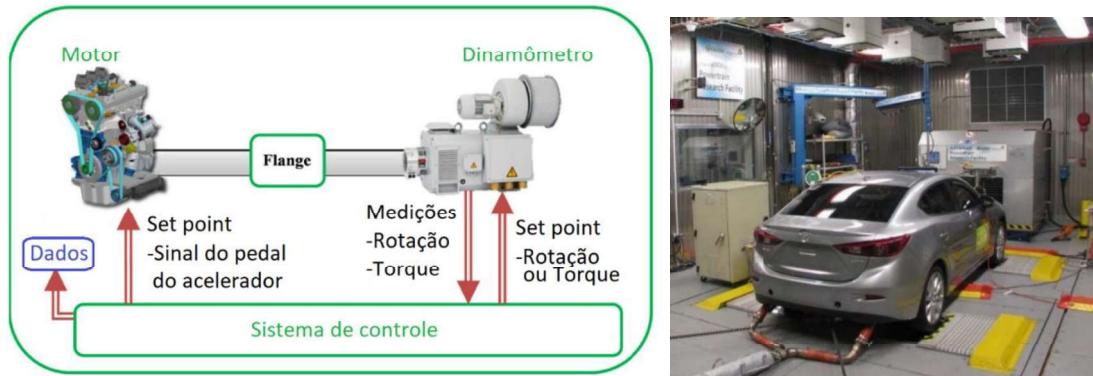
Fonte: adaptado de [5].

O dinamômetro é utilizado para medir a potência e o torque de motores, nas etapas de desenvolvimento, calibração e otimização dos veículos. O dinamômetro funciona como uma carga que consome a energia gerada pelo motor. Ele pode operar

mantendo a rotação do motor constante para uma determinada condição de operação, permitindo que o torque produzido varie, ou controlando o torque de carga aplicado no motor, permitindo que a rotação do motor varie conforme a condição de operação. Ao fixar o torque ou a rotação, durante os ensaios dinamométricos, os engenheiros podem calibrar, validar e otimizar todo o software da unidade de controle do motor (ECU) [2], [3], [5].

Existem dois tipos principais de dinamômetros: o dinamômetro de bancada e o dinamômetro de chassi, figura 5. No dinamômetro de bancada, o motor é testado isoladamente em um ambiente de laboratório, o que permite operá-lo em diferentes condições e medir diversas variáveis. Esse tipo de dinamômetro permite medir o torque diretamente no eixo do motor, sem a influência da massa do veículo ou do sistema de transmissão. Os testes típicos realizados em dinamômetros de bancada avaliam a eficiência, durabilidade, emissões de poluentes e características de desempenho e consumo do motor [1], [2].

Figura 5 – Tipos principais de dinamômetros.



Fonte: adaptado de [15], [16].

No dinamômetro de chassi, o motor é instalado no veículo, com as rodas acionando diretamente o dinamômetro, permitindo avaliar o desempenho do veículo como um todo. Ele simula condições de diferentes velocidades e cargas, medindo o comportamento do motor em situações próximas à condução real. O dinamômetro de chassi é amplamente utilizado em testes de consumo, emissões, aceleração e desempenho geral do veículo [1], [2].

2.2.1. Modelagem analítica do torque.

O modelo de torque padrão utilizado nas ECUs não considera a influência de diversos parâmetros de operação do motor. Para incluir a eficiência desses fatores, são utilizados modelos mais complexos, com base em princípios físicos, conforme a equação (3) [3].

$$P_m = \frac{2 \cdot \pi}{30} \cdot \omega_m \cdot T = 2 \cdot \pi \cdot \omega_m \cdot (T_e - T_{perdas}(T_m, p_{adm})) = \dot{Q} \cdot \eta_i \quad (1)$$

$$\dot{Q} = \dot{m}_{comb}(t_{inj}) \cdot H_l(P_{etan}) \quad (2)$$

$$T = \frac{60}{2 \cdot \pi} \cdot \frac{\dot{m}_{comb}(t_{inj}) \cdot H_l(P_{etan})}{\omega_m} \cdot \eta_{\omega_m} \cdot \eta_{\lambda} \cdot \eta_{\zeta} + T_{perdas}(T_m, p_{adm}) \quad (3)$$

H_l :	Poder calorífico da mistura de combustível utilizada (J/kg);
\dot{m}_{comb} :	Vazão mássica admitida de combustível (kg/s);
p_{adm} :	Pressão no coletor de admissão (kPa);
P_{etan} :	Porcentagem de etanol no combustível (%);
P_m :	Potência produzida pelo motor (W);
\dot{Q} :	Potência térmica gerada pela combustão (W);
T :	Torque produzido pelo motor (Nm);
T_e :	Torque efetivo produzido pelo motor (Nm);
T_{perdas} :	Perdas (Nm);
T_m :	Temperatura do motor (°C);
t_{inj} :	Tempo de injeção em (s);
ω_m :	Rotação do motor em (RPM);
η :	Eficiência indicada, adimensional;
η_{ω_m} :	Fator da eficiência indicada dependente da rotação do motor;
η_{λ} :	Fator da eficiência indicada dependente da relação ar-combustível;
η_{ζ} :	Fator da eficiência indicada dependente do ângulo de ignição;
ζ :	Ângulo ótimo de ignição em °APMS;
λ :	Fator lambda, adimensional.

O torque produzido pelo motor é gerado a partir da energia química fornecida pelo combustível injetado, considerando seu poder calorífico e o rendimento térmico do motor, descontando as perdas, equação (1). O torque de saída do motor, medido no dinamômetro, corresponde ao torque efetivo produzido, subtraídas as perdas. Essas perdas incluem o atrito entre as partes mecânicas do motor e o trabalho necessário para o bombeamento de gases entre o coletor de admissão e o de escape. A massa de combustível injetada depende do tempo de injeção, que corresponde ao período de acionamento da válvula injetora. O poder calorífico depende diretamente da composição do combustível utilizado, equação (2). A eficiência do motor é um

parâmetro complexo, pois varia em função de diversos fatores e do ponto de operação. Para reduzir a complexidade, a eficiência é frequentemente decomposta em fatores de correção, sendo que cada um considera os efeitos de diferentes parâmetros de operação, como o ângulo de ignição, o fator lambda e a velocidade angular (rotação) do motor [3].

O fator η_{ω_m} infere a influência da rotação do motor na eficiência indicada. Quando o motor se encontra em regime de baixa rotação, aumenta o tempo que os gases da combustão têm para trocar calor com as paredes do cilindro, aumentando as perdas e reduzindo a eficiência térmica. Já quando o motor está em regime de alta rotação, o tempo necessário para que a combustão ocorra por completo limita a quantidade de calor liberada no ciclo de expansão. Existe uma faixa ótima de rotação para o motor onde se observa um maior rendimento [3] [5]. O fator η_{ζ} estabelece a influência do instante em que ocorre a centelha na vela de ignição, chamado de ângulo de ignição (ζ), sobre a eficiência do motor. O ângulo de ignição é comumente expressado em graus antes do ponto morto superior (°APMS). Existe um ângulo ótimo em que a ignição deve ocorrer, o ângulo ótimo é chamado de ângulo de máximo torque de saída o MBT (*Maximum Break Torque*) ou (ζ_0) [3], [5].

O fator η_{λ} relaciona a influência da proporção ar-combustível (fator λ), na reação de combustão, devido à falta ou excesso de oxigênio disponível. Existe uma quantidade mínima de ar necessária para consumir completamente uma determinada quantidade de combustível na reação de combustão, chamada de estequiométrica. A relação estequiométrica é definida como a razão mínima de massa de ar necessária por unidade de massa de combustível. Seu valor varia de 14,7:1, no caso da gasolina pura, a 9:1, no caso do etanol. Ou seja, para consumir completamente 1 kg de gasolina na reação de combustão, são necessários 14,7 kg de ar. O fator lambda (λ) relaciona a proporção ar-combustível na qual o motor está operando com a relação estequiométrica, conforme a equação (4) [3], [4], [5]. Na condição estequiométrica da mistura, há oxigênio suficiente para que a reação de combustão ocorra de forma completa. Em condições de excesso de oxigênio na mistura ($\lambda > 1$), considera-se que

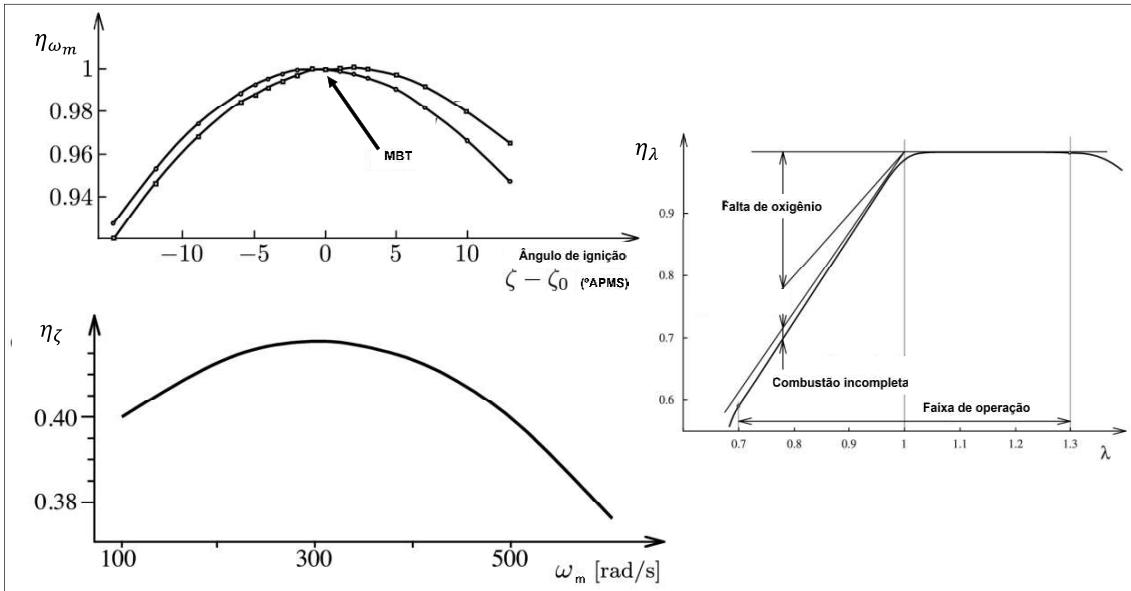
todo o combustível é consumido, e o fator η_λ vale 1. Em condições de excesso de combustível ($\lambda < 1$), fator η_λ diminui proporcionalmente ao valor de λ [3].

$$\lambda = \frac{\dot{m}_{ar}}{\dot{m}_{comb} \cdot \sigma_0(P_{etan})} \quad (4)$$

- \dot{m}_{ar} : Vazão mássica admitida de ar admitida (kg/s);
- \dot{m}_{comb} : Vazão mássica admitida de combustível (kg/s);
- P_{etan} : Porcentagem de etanol no combustível (%);
- σ_0 : Relação estequiométrica, adimensional;
- λ : Fator lambda, adimensional.

Os fatores de eficiência, como o de ignição, lambda e rotação, não seguem um comportamento linear e apresentam pontos ótimos em função das condições de operação do motor, como observado na figura 6 [3].

Figura 6 – Fatores da eficiência indicada dependente da rotação (η_{ω_m}), relação ar-combustível (η_λ) e ângulo de ignição (η_ζ).



Fonte: adaptado de [3].

É possível observar que as principais variáveis de interesse, por serem fatores diretamente relacionados ao desempenho do motor e à geração de torque, são: ângulo de ignição (1), composição do combustível (2), fator lambda (3), pressão no coletor de admissão (4), rotação do motor (5), temperatura do motor (6) e tempo de injeção (7). Essas sete variáveis foram utilizadas neste trabalho como base para o desenvolvimento e validação dos modelos.

2.3. MODELAGEM DO TORQUE COM *MACHINE LEARNING*.

Para estimar o torque produzido pelo motor a partir de seu ponto de operação, selecionou-se duas técnicas de *machine learning*: redes neurais artificiais (RNAs) e regressão por processos gaussianos (*Gaussian Process Regression* – GPR). Essas técnicas foram escolhidas por serem amplamente utilizadas nas referências.

2.3.1. Redes neurais artificiais (RNAs).

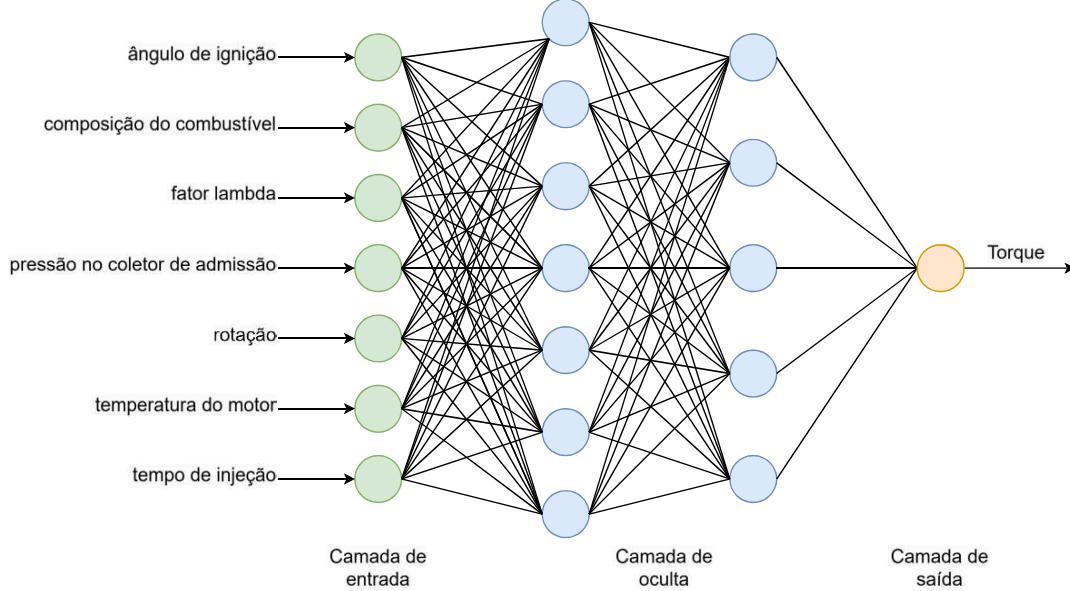
As redes neurais artificiais (RNAs) são modelos computacionais capazes de reconhecer padrões, aproximar funções e possuem ampla aplicação em predição e classificação de variáveis. Inspiradas no cérebro humano, as RNAs são compostas por um sistema de neurônios interconectados formando camadas (*layers*), permitindo diversas variáveis de entrada e saída. As variáveis de entrada alimentam os neurônios na primeira camada, chamada de camada de entrada. A saída dos neurônios da camada de entrada, alimentam os neurônios das camadas intermediárias (ou ocultas), chegando ao(s) neurônio(s) da camada de saída (figura 7) [6], [7], [8], [9].

Utilizando um número adequado de neurônios e camadas, as redes neurais podem aproximar funções complexas, combinando os sinais gerados pelos diversos neurônios da rede. Os parâmetros de cada neurônio são obtidos durante o treinamento da rede, utilizando algoritmos de otimização que estimam os valores ideais de cada parâmetro, com o objetivo de minimizar o erro.

O torque produzido por motores de combustão interna depende de interações altamente não lineares entre diversas variáveis. As RNAs são uma solução eficaz para modelar essas relações complexas sem a necessidade de desenvolver modelos analíticos baseados em fenômenos físicos. Possuem também a capacidade de lidar com ruídos de medição presentes em dados experimentais, garantindo maior robustez, [6], [7], [8], [9].

Neste trabalho, utilizou-se uma rede neural do tipo *feedforward*, treinada para predição de variáveis ou regressão. Em uma rede *feedforward*, as informações fluem de forma unidirecional, da camada de entrada para a camada de saída, sem loops ou retroalimentações. Conectando as camadas de entrada e de saída, existem camadas intermediárias, chamadas de camadas ocultas (ou *hidden layers*). Essa rede pertence à categoria MLP (*Multilayer Perceptron*), contendo uma ou mais camadas ocultas com funções de ativação não lineares (figura 7) [9]. A Figura 7 apresenta uma rede neural com as variáveis selecionadas para modelar o torque na camada de entrada, duas camadas intermediárias e uma camada de saída, onde é estimado o valor do torque.

Figura 7 – Exemplo de arquitetura de rede neural para modelagem do torque.



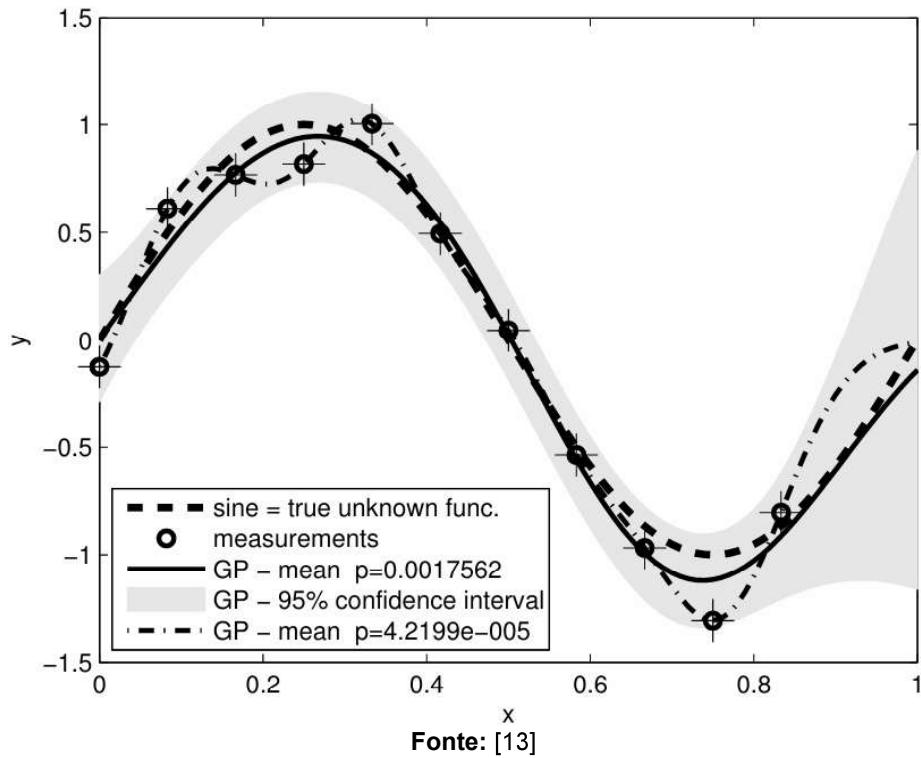
Fonte: O autor.

2.3.2. Regressão baseada em Processos Gaussianos (GPR).

A regressão baseada em Processos Gaussianos (*Gaussian Process Regression – GPR*) é uma técnica estatística não paramétrica de aprendizado supervisionado, utilizada para prever e classificar variáveis. Permite também realizar previsões com intervalos de confiança (figura 8), o que é especialmente útil na modelagem de sistemas onde o comportamento tem natureza ruidosa ou incerta. A GPR é uma

solução para modelar torque produzido por motores de combustão interna depende de interações altamente não lineares entre diversas variáveis. [10], [11], [12], [13], [17] [18].

Figura 8 – Regressão baseada em Processos Gaussianos.



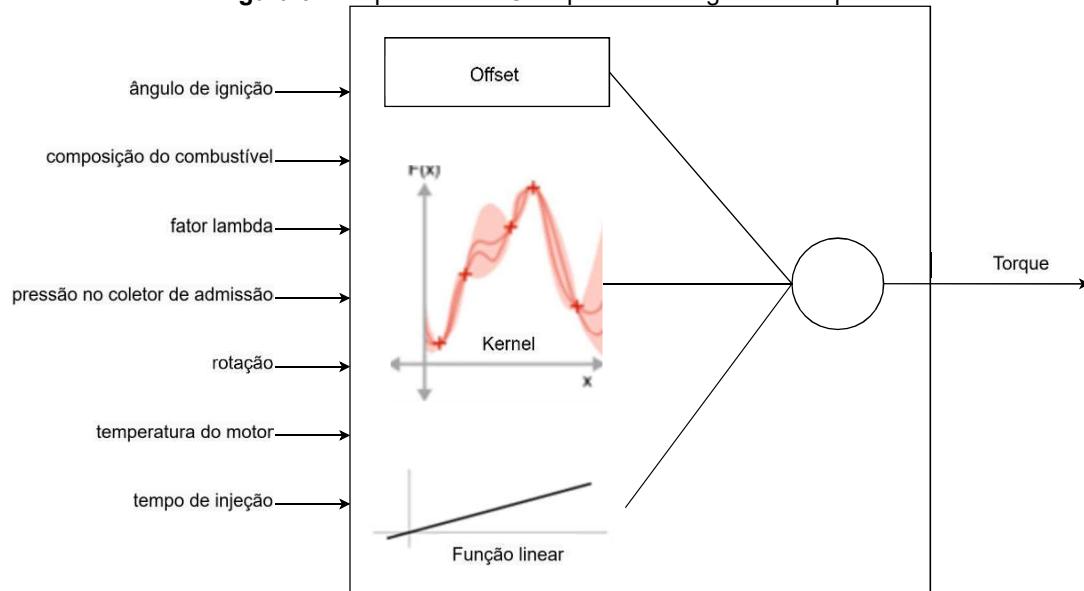
Fonte: [13]

A regressão baseada em Processos Gaussianos utiliza o conceito de que um dado conjunto finito de variáveis aleatórias segue uma distribuição do tipo Gaussiana. Ela estima a relação entre entradas e saídas por meio de uma função de média e funções de covariância, que descrevem a semelhança entre os pontos no espaço de entrada e de saída (figura 8). Essas relações são feitas utilizando uma ou mais funções de kernel, de forma semelhante aos neurônios presentes em redes neurais [13], [14], [19].

Para ajustar os parâmetros do modelo, são utilizados algoritmos de otimização, como a base na estimativa da máxima probabilidade da semelhança (MLE, *Maximum Likelihood Estimation*), buscando os valores ótimos para os

parâmetros do kernel, minimizando o erro nas previsões. Neste trabalho, utilizou-se uma GPR, projetada para aproximação de funções ou regressão, Figura 9 [19].

Figura 9 – Arquitetura da GPR para modelagem do torque.



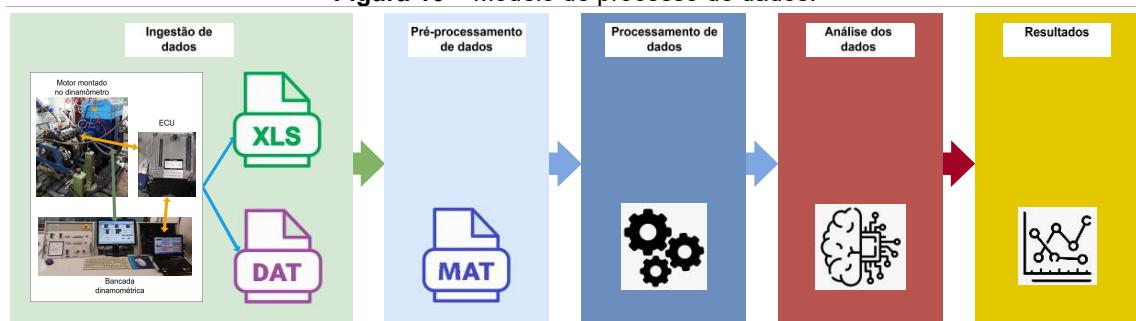
Fonte: O autor.

3. DESENVOLVIMENTO DOS MODELOS DE TORQUE A PARTIR DOS DADOS EXPERIMENTAIS.

Para este trabalho, utilizou-se o MATLAB, um software amplamente utilizado pela indústria automotiva, que oferece diversas ferramentas de análise e processamento de dados. Este software também é utilizado pelo grupo de pesquisa em eletrônica automotiva da Poli-USP, permitindo que os códigos e a metodologia desenvolvidos ao longo deste trabalho sejam reutilizados em projetos futuros.

Para o modelo de processo de dados (figura 10), utilizou-se a metodologia de análise e processamento conforme as etapas da referência [20]. A arquitetura utilizada possui as etapas de: ingestão, pré-processamento, processamento, análise dos dados e apresentação dos resultados.

Figura 10 – Modelo de processo de dados.



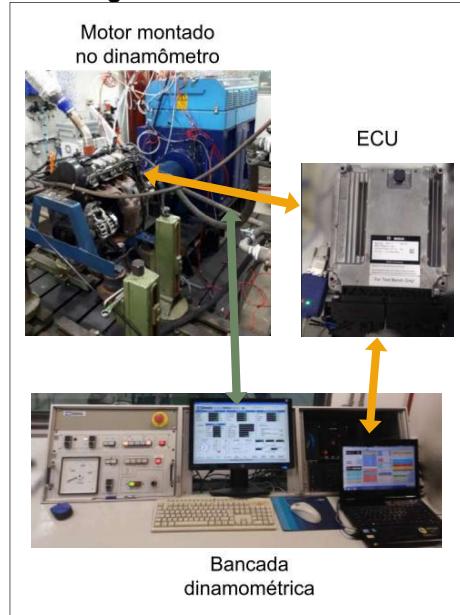
Fonte: O autor.

3.1. Ingestão de Dados.

A ingestão dos dados é um processo complexo devido à variedade de formatos de arquivo das fontes de dados. Os dados dos ensaios estavam em diversos arquivos com extensão “dat” (data), “xls” e “xlsx” (eXceL Spreadsheet). Esses arquivos foram convertidos para um formato de arquivo compatível com o MATLAB, o “mat” (MATLAB *data file*). Após a conversão, identificou-se uma inconsistência nos nomes das variáveis, sendo necessária a padronização desses nomes. Realizou-se o mapeamento do nome das variáveis nos arquivos para padronizá-las, garantindo a uniformidade e consistência nos dados extraídos dos arquivos disponíveis, contendo os ensaios realizados no motor.

A principal fonte de dados utilizada neste trabalho é composta pelos arquivos provenientes dos ensaios realizados no motor (figura 11) [21]. Esses ensaios são classificados em dois tipos principais: regime permanente e regime transitório.

Figura 11 – Fonte de dados.



Fonte: O autor.

Os arquivos dos ensaios em regime transitório contêm séries temporais, com o registro ao longo do tempo das variáveis geradas pela ECU (Unidade de Controle do Motor). Já os ensaios em regime permanente consistem em planilhas, geradas pelo operador do dinamômetro, com dados coletados em condições estacionárias do motor.

Armazenaram-se os arquivos dos ensaios realizados em uma pasta chamada RAWData. Desenvolveu-se um algoritmo para fazer a varredura da pasta RAWData, carregando e convertendo os arquivos disponíveis para o formato padrão (“mat”).

O registro das variáveis da ECU é feito em arquivos no formato “dat”. Esse arquivo contém os valores das variáveis ao longo do tempo, compactadas no formato “mdf4” (*Measurement Data Format v.4*). O formato “mdf4” é um formato de arquivo utilizado na indústria automotiva e em outras áreas de engenharia. A ECU registra as variáveis do motor juntamente com suas respectivas bases de tempo. Como a ECU

controla diversos processos que ocorrem em tempos de execução diferentes, a gravação das variáveis é realizada em bases de tempo distintas. Cada variável registrada está associada a uma base de tempo específica, refletindo a frequência com que os dados são atualizados. Sendo necessário identificar a base de tempo correspondente para cada variável. O algoritmo utiliza o número identificador no nome e o número de amostras de cada variável com o da base de tempo selecionada pelo número de identificação. Cada variável e sua respectiva base de tempo são armazenadas em uma única variável do tipo *timeseries* do MATLAB, permitindo sincronizar os dados das variáveis no tempo.

Os arquivos convertidos foram armazenados em uma pasta chamada de convData. Durante o processo de conversão, os nomes originais dos arquivos foram preservados, alterando apenas seus formatos, de “dat” e “xls” para “mat”. O algoritmo desenvolvido, detecta a existência de arquivos com o mesmo nome na pasta convData evitando que arquivos de ensaios diferentes sejam sobreescritos.

Ao carregar os arquivos de cada ensaio, observou-se uma grande variedade nos nomes das variáveis. Sendo necessário mapear os nomes das variáveis para padronizá-las na etapa de processamento, garantindo que todos os arquivos fossem processados de maneira uniforme. Foi possível observar a falta de uma cultura de dados e governança. Se práticas como essas tivessem sido adotadas durante a elaboração dos ensaios, a etapa de ingestão de dados teria sido muito mais simples, sem a necessidade de padronizar e converter os nomes das variáveis. Além disso, foi reforçada para as equipes responsáveis pela realização dos ensaios a importância da cultura de dados, uma vez que esses arquivos podem ser utilizados em trabalhos futuros. A ausência de padronização e metadados dificulta o uso desses dados em futuras aplicações.

3.2. Pré-processamento dos dados.

Antes da etapa de processamento dos dados, realizou-se o pré-processamento para padronizar os nomes das variáveis e dos arquivos que contêm os dados dos ensaios. Os arquivos, previamente convertidos para o formato “mat” durante a fase de ingestão de dados, foram carregados, e os nomes das variáveis presentes nesses arquivos foram renomeados seguindo o mapeamento do nome das variáveis. Após essa padronização, foi gerado um novo arquivo, com as variáveis devidamente renomeadas, e o nome de cada arquivo também é padronizado. Esse novo arquivo foi armazenado em uma pasta chamada `processData`. Elaborou-se o dicionário de dados com o nome, descrição e unidade física de cada uma das variáveis.

Desenvolveu-se um algoritmo para realizar a varredura dos arquivos da pasta `convData`. O algoritmo renomeia as variáveis e os arquivos seguindo um padrão com o nome do motor, o tipo de ensaio, o número do arquivo lido e a composição do combustível. O tipo de ensaio pode ser do tipo dinâmico, que corresponde às gravações feitas em regime transitório do motor, ou estático, que corresponde aos dados obtidos nos ensaios do motor em regime estacionário. Na Tabela 1, estão os nomes padronizados das variáveis, suas descrições e unidades físicas. Para os dados em regime transitório, foi utilizada uma notação específica para as variáveis em formato de séries temporais, seguindo o nome padrão apresentado na Tabela 1, porém com todas as letras em maiúsculas.

O mapeamento das variáveis foi feito de forma manual, utilizando as ferramentas de busca do Excel. Essa etapa foi feita de forma manual, pois é uma etapa crítica influenciando diretamente a confiabilidade dos dados.

Tabela 1 – Nomes padronizados, descrição e unidade física de cada variável.

Variável	Descrição	Unidade
<code>Etpc</code>	Porcentagem de etanol no combustível.	(%) Percentual
<code>exh_co</code>	Nível de emissão de monóxido de carbono (CO) nos gases de escape.	(ppm) Partes por milhão
<code>exh_co2</code>	Nível de emissão de dióxido de carbono (CO ₂) nos gases de escape.	(ppm) Partes por milhão
<code>exh_nox</code>	Nível de emissão de óxidos de nitrogênio (NOx) nos gases de escape.	(ppm) Partes por milhão

exh_o2	Nível de oxigênio nos gases de escape.	(ppm) Partes por milhão
exh_thc	Nível de emissão de hidrocarbonetos totais (THC) nos gases de escape.	(ppm) Partes por milhão
ign	Ângulo de ignição.	(°APMS) Graus antes do ponto morto superior
lamb	Fator lambda, que indica a relação ar-combustível na combustão.	(.) Adimensional
maf	Fluxo de massa de ar.	(kg/h) Quilograma por hora
padm	Pressão no coletor de admissão.	(kPa) Quilopascal
pamb	Pressão ambiente.	(kPa) Quilopascal
pexh	Pressão no sistema de escapamento.	(kPa) Quilopascal
poleo	Pressão do óleo do motor.	(kPa) Quilopascal
rpm	Rotação do motor.	(RPM) Rotações por minuto
tadm	Temperatura no coletor de admissão.	(°C) Graus Celsius
tamb	Temperatura ambiente.	(°C) Graus Celsius
texh	Temperatura no sistema de escapamento.	(°C) Graus Celsius
th2o	Temperatura do fluido de arrefecimento do motor.	(°C) Graus Celsius
th2o_in	Temperatura do fluido de arrefecimento na entrada do motor.	(°C) Graus Celsius
th2o_out	Temperatura do fluido de arrefecimento na saída do motor.	(°C) Graus Celsius
tinj	Tempo de injeção de combustível.	(ms) Milissegundos
toleo	Temperatura do óleo do motor.	(°C) Graus Celsius
tps	Abertura da válvula borboleta (<i>Throttle Position Sensor</i>).	(%) Percentual
trq	Torque gerado pelo motor.	(Nm) Newton-metro

Fonte: O autor.

O dicionário de dados foi elaborado a partir dos nomes padronizados. Ele serve como referência para o uso das, permitindo que os dados dos ensaios possam ser utilizados em simulações diversas, utilizando as gravações no tempo com os nomes padronizados e os dados convertidos do motor em regime estacionário e transitório.

3.3. Processamento dos Dados.

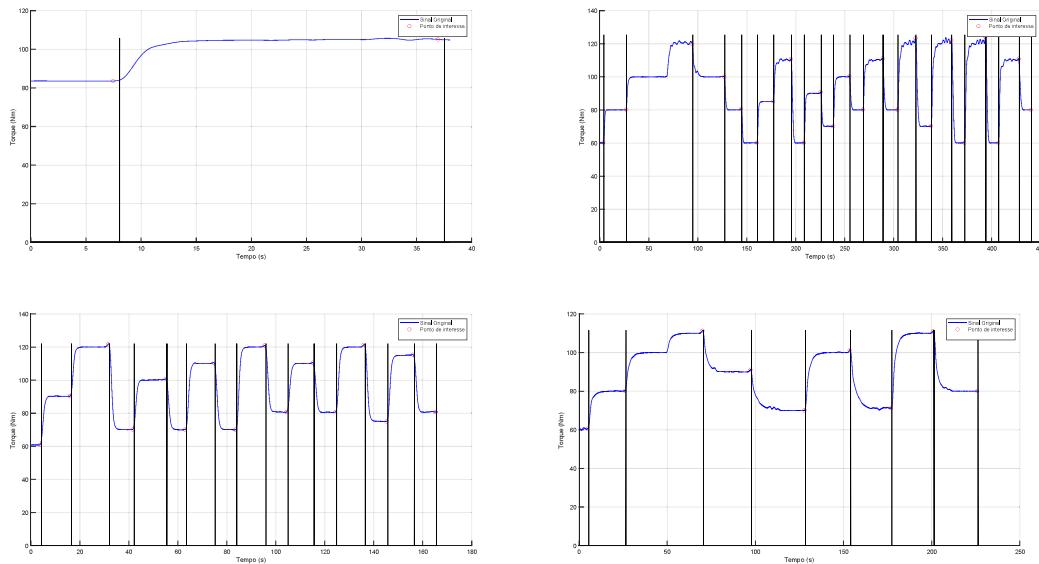
Nesse processo, os dados foram submetidos a uma etapa de limpeza, descartando-se os arquivos de ensaios que não continham as variáveis essenciais para a modelagem do torque. Foram aplicadas técnicas de análise de dados para extrair janelas de operação do motor em regime estacionário a partir dos ensaios em regime transitório do motor. Essa análise permitiu a unificação dos dados provenientes

de ensaios em regimes transitório e estacionário em um único dataset, para ser utilizado nas etapas de análise e modelagem do torque do motor.

Durante a etapa de limpeza, foi gerado um contendo informações como o nome do arquivo do ensaio, o arquivo original antes da conversão e o motivo da eliminação (arquivo de log). Em alguns ensaios, a temperatura do óleo não estava disponível, sendo seu valor aproximado como igual ao valor da temperatura do líquido de arrefecimento do motor.

Após a limpeza, o algoritmo realizou a identificação de janelas de tempo em que o motor apresentava comportamento em regime estacionário. Para isso, foi feita uma busca ao longo do sinal de torque na série temporal, utilizando a derivada do sinal para detectar variações. O algoritmo identificou os picos de variação (figura 12), classificando-os como pontos de interesse, e criou janelas de 5 segundos antes e depois desses pontos. Em seguida, verificou-se se o valor do torque permanecia estável dentro dessas janelas de tempo. Caso fosse constatada estabilidade, extraíndo os valores médios das variáveis: ângulo de ignição (1), composição do combustível (2), fator lambda (3), pressão no coletor de admissão (4), rotação do motor (5), temperatura do motor (6) e tempo de injeção (7).

Após o processo de janelamento, os dados extraídos foram integrados aos dados provenientes dos ensaios em regime estacionário do motor, gerando o dataset principal denominado mainDataset. Esse mainDataset foi organizado no formato de tabela, contendo o nome das variáveis, suas descrições e as respectivas unidades físicas.

Figura 12 – Pontos de interesse.

Fonte: O autor.

As variáveis contínuas foram discretizadas em intervalos predefinidos, com o objetivo de reduzir a dispersão nos dados e melhorar a qualidade da análise. Por exemplo, considerando dois pontos no dataset em que a temperatura do motor é de 70°C e 71°C, a discretização permite agrupá-los em uma única classe, reduzindo a variabilidade e facilitando a obtenção dos modelos.

A pressão no coletor de admissão foi discretizada no intervalo de 30 a 90 kPa (quilopascais), com incrementos de 15 kPa. A rotação do motor foi discretizada no intervalo de 1500 a 5000 RPM (rotações por minuto), com incrementos de 500 RPM. A composição do combustível foi discretizada em intervalos fixos correspondentes a valores de [27, 45, 64, 82, 100] (%). O fator lambda foi discretizado no intervalo de 0,85 a 1,15, com incrementos de 0,15 (adimensional). As demais variáveis foram arredondadas para seus valores inteiros. Para realizar a discretização, o algoritmo percorreu todos os pontos do mainDataset e ajustou os valores ao intervalo discretizado mais próximo de cada ponto.

3.4. Análise de dados.

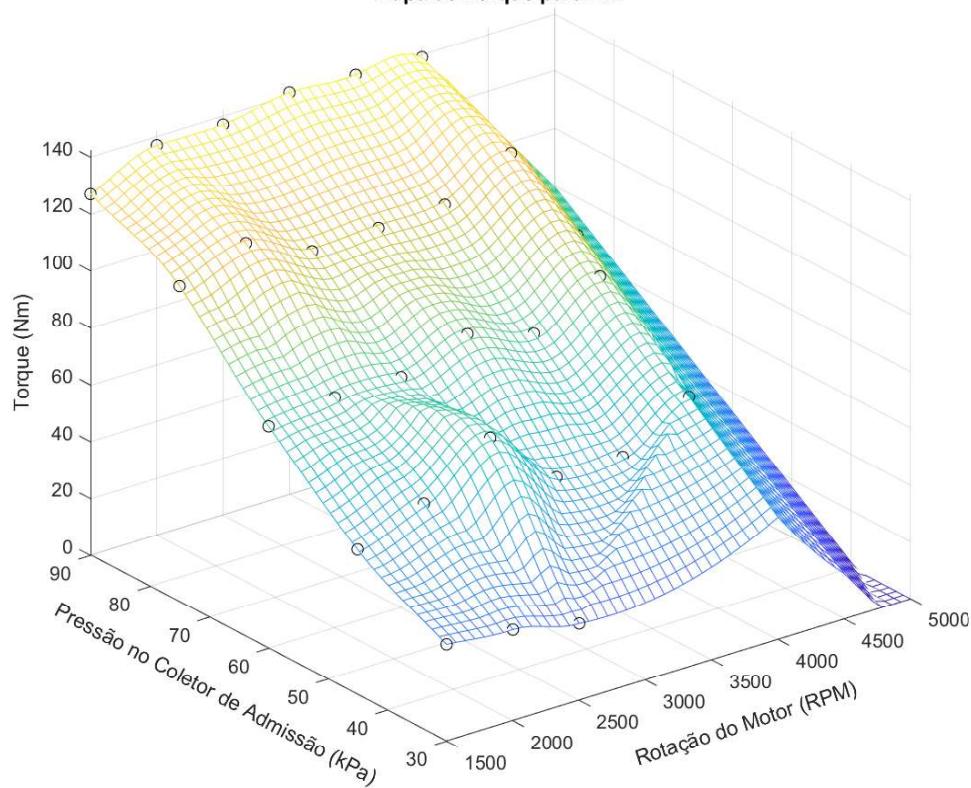
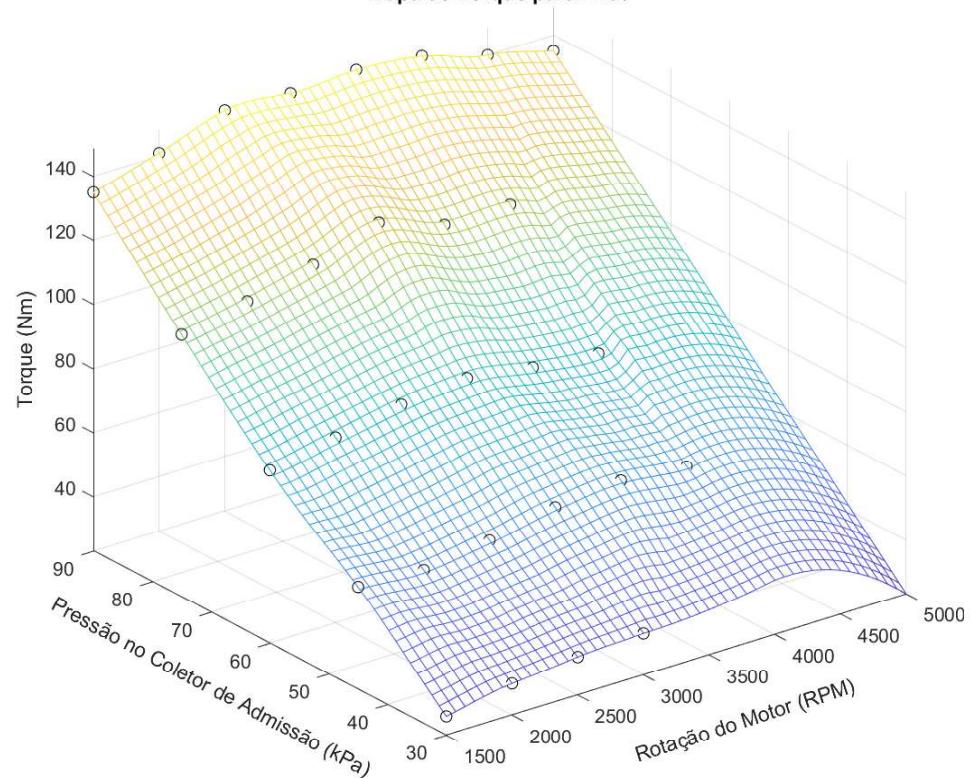
Após a obtenção do dataset, foi realizada a etapa de análise de dados para desenvolver os modelos de torque produzido pelo motor. Para o treinamento e validação dos modelos, o mainDataset foi dividido em três subconjuntos: treino (*Training*), validação (*Validation*) e teste (*Testing*). O desempenho final dos modelos foi avaliado por meio de gráficos gerados nas etapas de treino, validação e teste, bem como pela análise de Indicadores de Desempenho (KPIs). Além disso, foi calculado o tempo gasto por cada modelo para predizer os valores de torque com base nos pontos do mainDataset.

3.4.1. Modelos de Torque obtidos.

Utilizou-se três modelos para o torque: o modelo padrão, modelo baseado em Redes Neurais Artificiais (RNAs) e Regressão por Processos Gaussianos (GPR).

- **Modelo padrão.**

O modelo de torque padrão consiste em um mapa (look-up table) que representa o torque máximo produzido pelo motor para uma determinada condição de rotação do motor e pressão no coletor de admissão, considerando os combustíveis gasolina e etanol comum. Para gerar o modelo padrão, utilizou-se as duas composições padrão de combustível: gasolina comum E27 (com 27% de etanol) e etanol comum E100 (100% de etanol). O algoritmo identificou os pontos no mainDataset correspondentes a essas composições e selecionou o maior valor de torque para cada combinação de pressão no coletor de admissão e rotação do motor. Esse processo resultou na criação do conjunto de pontos de treino. Os pontos restantes, que não foram utilizados no treinamento, foram divididos igualmente entre os conjuntos de teste e validação. Obteve-se dois mapas de torque, um para cada composição padrão de combustível, figura 13 e figura 14.

Figura 13 – Mapa de torque para gasolina E27.**Mapa de Torque para E27****Fonte:** O autor.**Figura 14 – Mapa de torque para etanol E100.****Mapa de Torque para E100****Fonte:** O autor.

O valor de torque estimado pelo modelo para composições intermediárias de combustível, bem como para outras condições de operação, é calculado por meio da interpolação entre os dois mapas.

- **Redes neurais artificiais (RNAs).**

Para os modelos de torque baseados em Redes Neurais Artificiais (RNAs) e Regressão por Processos Gaussianos (GPR), o mainDataset foi dividido em subconjuntos de treino, validação e teste, seguindo a proporção de 70%, 15% e 15%, respectivamente. Realizou-se essa divisão de forma aleatória, utilizando essa proporção para cada valor de composição de combustível presente no mainDataset ([27, 45, 64, 82, 100]). Essa estratégia foi adotada para reduzir o risco de *overfitting*. Caso os pontos fossem selecionados aleatoriamente, diretamente do mainDataset, os dados de treino poderiam se concentrar em composições específicas, como 27%, comprometendo o desempenho geral dos modelos.

Para a geometria da rede neural, foi utilizada uma topologia com duas camadas ocultas e um método interativo em *loop* para ajustar o número de neurônios nas camadas. Testou-se geometrias com o número de neurônios em cada camada variando de 1 a 10. Sendo selecionadas as quatro arquiteturas que apresentaram os menores erros.

- **Regressão baseada em Processos Gaussianos (GPR).**

Para o modelo GPR (Regressão por Processos Gaussianos), testando-se quatro diferentes tipos de kernels [14], [19]:

1. *Squared Exponential:*

$$k(x_i, x_j | \theta) = \sigma_f^2 \exp\left(-\frac{(x_i - x_j)^T (x_i - x_j)}{2\sigma_l^2}\right) \quad (5)$$

2. *Exponential:*

$$k(x_i, x_j | \theta) = \sigma_f^2 \exp\left(-\frac{r}{\sigma_l}\right) \quad (6)$$

$$r = \sqrt{(x_i - x_j)^T (x_i - x_j)} \quad (7)$$

3. *Rational Quadratic:*

$$k(x_i, x_j | \theta) = \sigma_f^2 \left(1 + \frac{r^2}{2\alpha\sigma_l^2}\right)^{-\alpha} \quad (8)$$

$$r = \sqrt{(x_i - x_j)^T (x_i - x_j)} \quad (9)$$

4. *(Automatic Relevance Determination) ARD Squared Exponential:*

$$k(x_i, x_j | \theta) = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}\right) \quad (10)$$

$k(x_i, x_j | \theta)$: Função de covariância (kernel).

σ_f^2 : Variância do sinal.

σ_l : Comprimento de escala (*length scale*), distância em que as entradas (x_i, x_j) permanecem correlacionadas.

x_i, x_j : Vetores de variáveis de entrada e de saída.

r : Distância euclidiana entre as entradas (x_i, x_j) .

α : Parâmetro de transição entre comprimentos de escala.

σ_m^2 : Comprimento de escala individual para a cada dimensão do vetor de entrada.

d : Dimensão do vetor de entrada (número de variáveis).

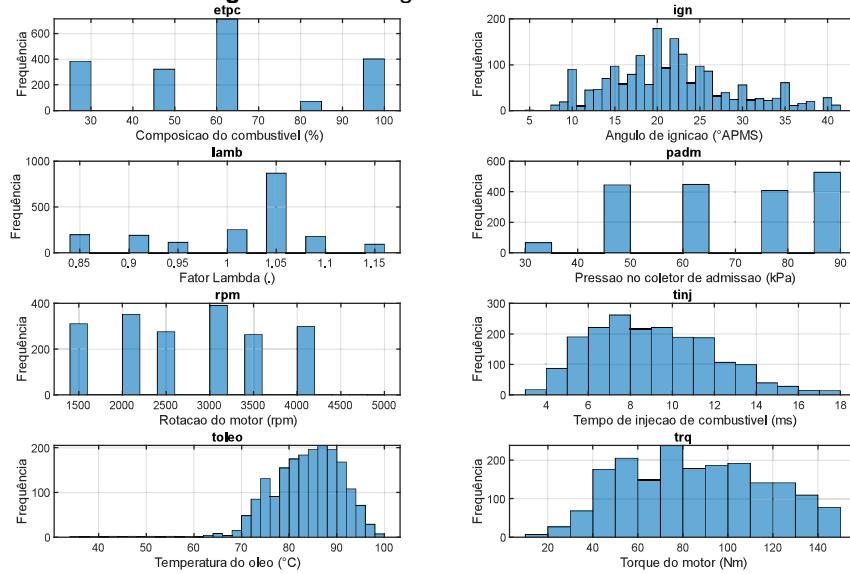
θ : vetor de parâmetros transformados, $(\theta_1 = \log \sigma_l, \theta_2 = \log \sigma_f)$.

4. ANALISE E RESULTADOS OBTIDOS.

A partir do mainDataset, foram gerados histogramas e gráficos para cada variável, permitindo a visualização das características principais, como distribuição e variabilidade dos dados (figura 15 e figura 16). Observou-se que a maioria dos ensaios foi realizada com o fator lambda entre 1,0 e 1,05, caracterizando uma operação em regime estequiométrico do motor.

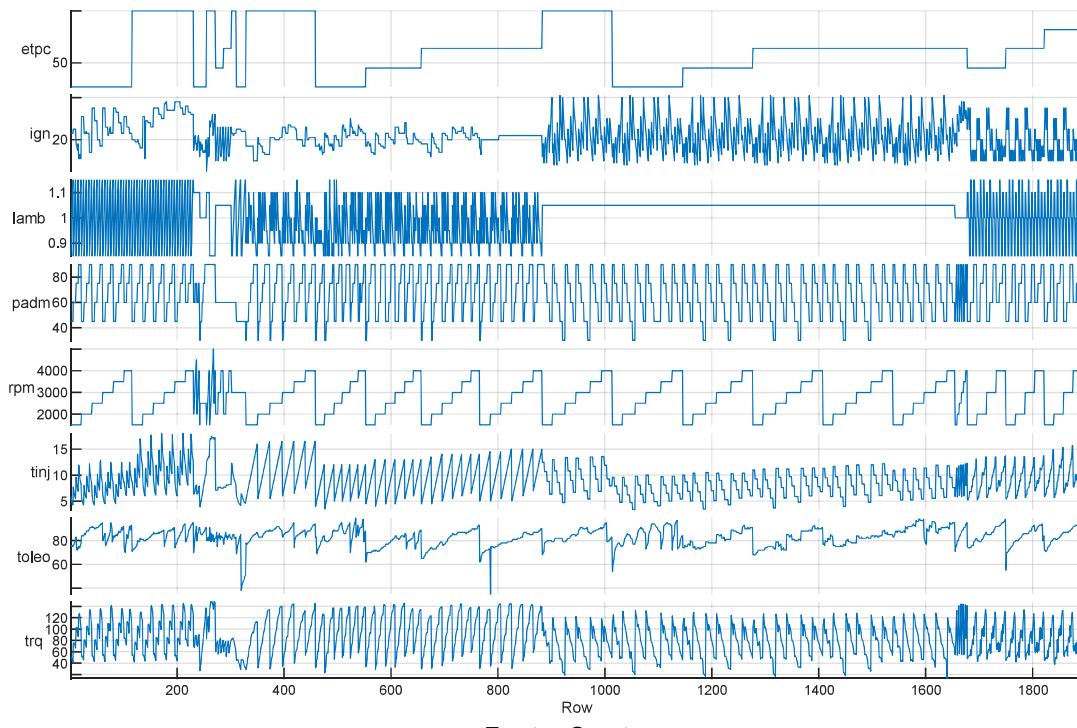
Os Dados estão bem distribuídos em relação à pressão no coletor de admissão, com exceção dos ensaios realizados com pressões mais baixas, próximas a 30 kPa (figura 15). O torque apresentou uma distribuição aproximadamente uniforme, entre 40 Nm e 120 Nm (figura 15). É possível observar que a maioria dos ensaios foi realizada com o motor quente (acima de 70 °C) (figura 15). A rotação do motor também está bem distribuída ao longo dos valores analisados, exceto por um número reduzido de ensaios registrados em 4.500 RPM (figura 15).

Figura 15 – Histogramas do mainDataset.



Fonte: O autor.

Figura 16 – Gráficos para cada variável do mainDataset.



Fonte: O autor.

Realizou-se uma análise de correlação entre as variáveis e o torque, utilizando mapas de calor (*heatmap*) e análises gráficas de correlação de *Spearman* (figura 17 e figura 18).

Os resultados indicaram forte correlação (próxima de 1) entre o torque e as variáveis pressão no coletor de admissão, tempo de injeção, ângulo de ignição e fator lambda A pressão no coletor e tempo de injeção tiveram as correlações mais fortes, refletindo sua influência direta na energia química admitida pelo motor, como previsto pela literatura.

Nas variáveis ângulo de ignição, fator lambda e rotação, observa-se uma correlação menos intensa devido ao comportamento não linear da influência dessas variáveis no torque produzido. Na revisão da literatura, observa-se que a influência dessas variáveis sobre o torque produzido é não linear e apresenta um valor ótimo. O torque produzido diminui à medida que o ponto de operação do motor se afasta desse valor ideal. Isso torna a análise de correlação dessas variáveis mais complexa do que

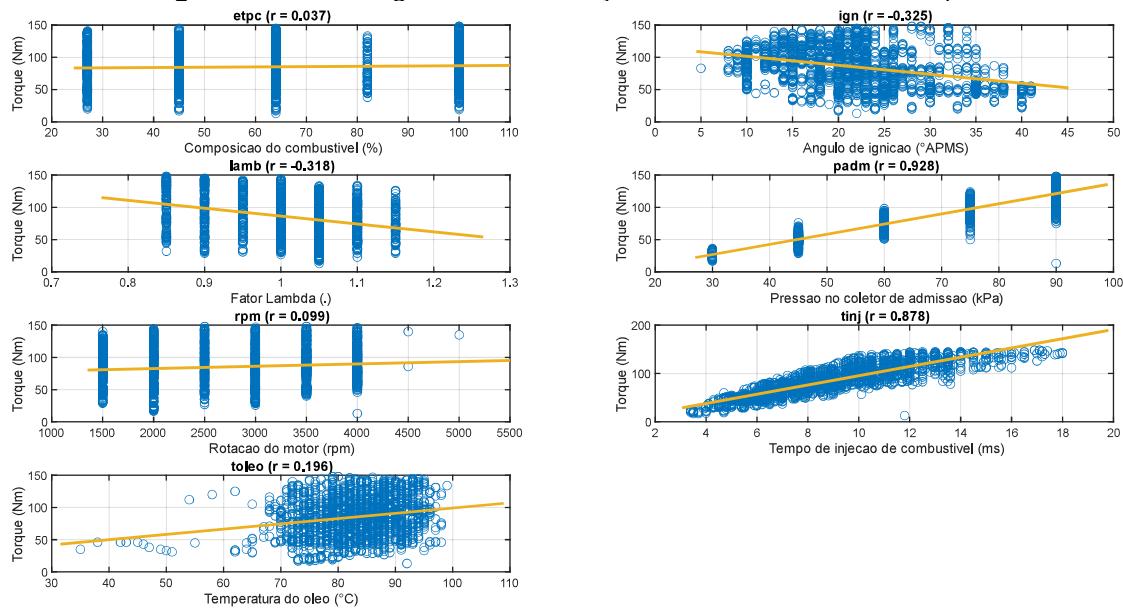
aquela obtida utilizando a correlação de *Spearman*, uma vez que a relação entre elas é não linear. É possível observar também uma leve correlação positiva entre o torque e a concentração de etanol no combustível, como esperado. Observou-se ainda que o torque apresenta uma correlação positiva com o aumento da temperatura do motor. Esse comportamento ocorre porque temperaturas mais altas reduzem a viscosidade do óleo, minimizando as perdas por atrito no motor.

Figura 17 – Mapa de calor (heatmap).



Fonte: O autor.

Figura 18 – Análises gráficas de correlação entre cada variável e o torque.



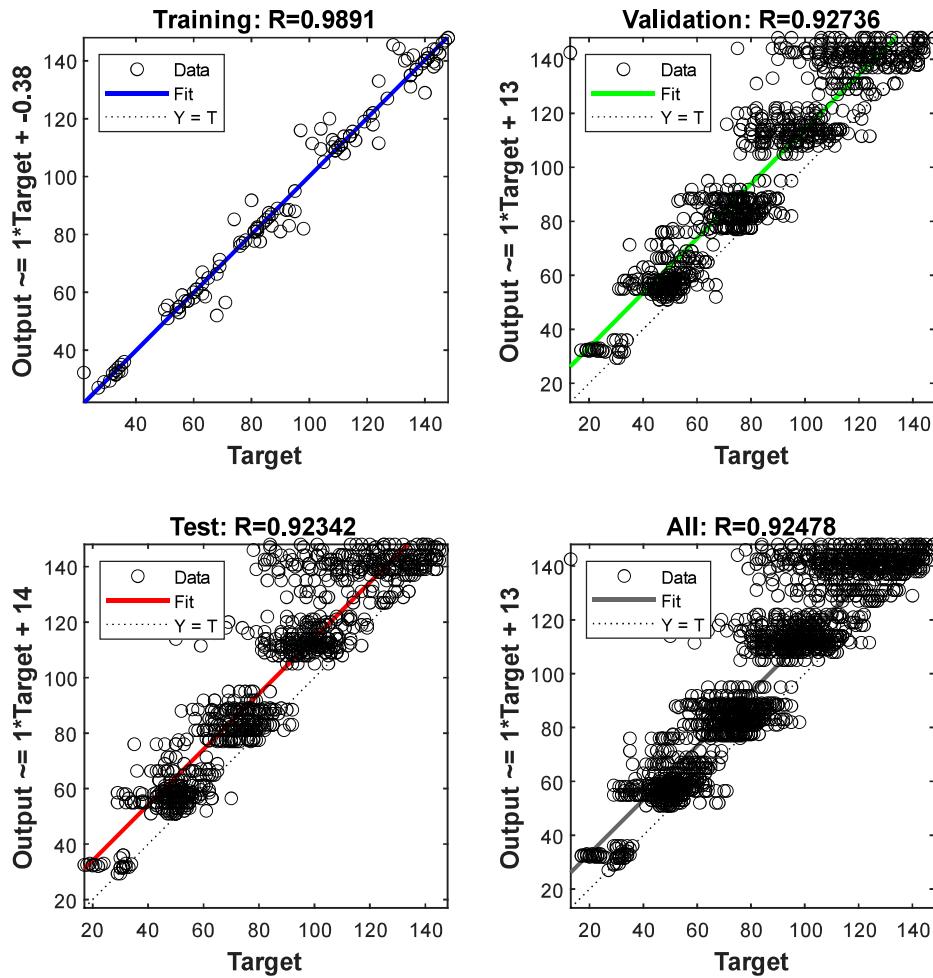
Fonte: O autor.

- **Modelo padrão.**

Gerou-se gráficos de regressão para cada um dos modelos obtidos. No modelo padrão, observou-se um *overfitting* em relação aos dados de treino. Os pontos desse conjunto estão bem alinhados com a linha de identidade, indicando um bom ajuste em relação aos dados de treino (figura 19).

É observada uma grande dispersão nos dados de validação e teste, evidenciando o baixo desempenho do modelo. Esse comportamento era esperado, conforme a revisão da literatura, já que o modelo é construído com base nos valores de torque máximo para cada combinação de rotação e pressão, resultando em um desempenho ruim para os dados de validação e teste.

Figura 19 – Gráficos de regressão para modelo padrão.

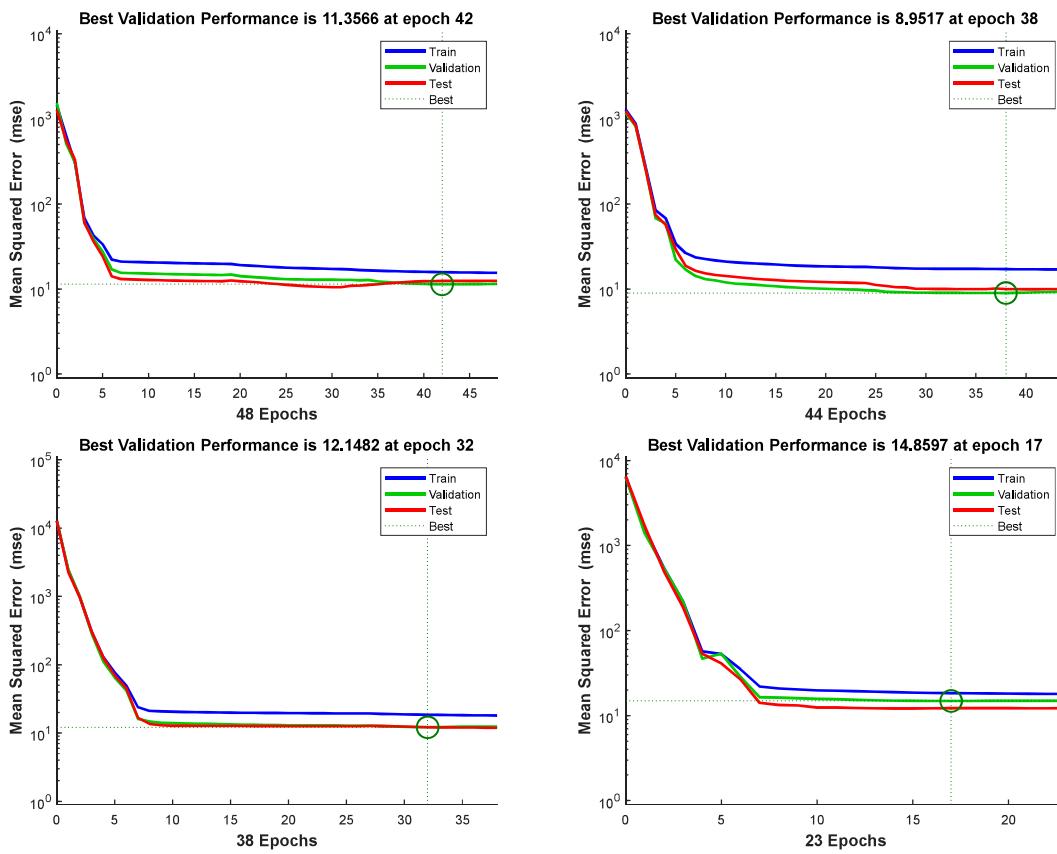


Fonte: O autor.

- **Redes neurais artificiais (RNAs).**

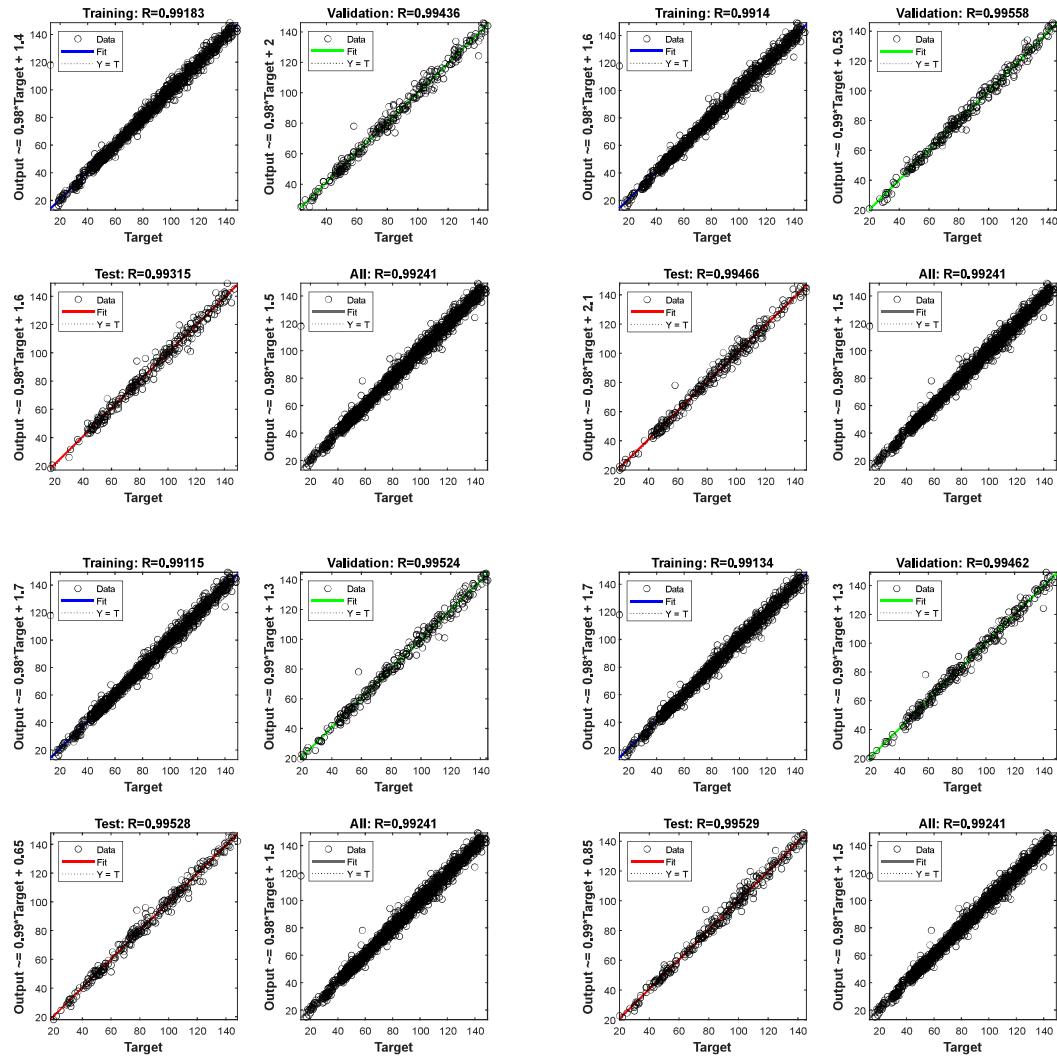
Selecionou-se as quatro melhores topologias de redes neurais que apresentaram o menor erro quadrático médio (MSE, *Mean Squared Error*). Para essas RNAs, gerou-se os gráficos de regressão e de desempenho. Os gráficos de desempenho permitem visualizar as etapas de treinamento, validação e teste da rede neural em cada iteração do algoritmo, possibilitando identificar a iteração em que o algoritmo converge (figura 20). Nos gráficos de regressão, os pontos estão bem localizados próximos à linha de identidade (figura 21), e o algoritmo demonstrou rápida convergência (figura 20). Os resultados evidenciaram um bom ajuste das redes neurais na predição do torque, sem *overfitting* ou *underfitting* (figura 21).

Figura 20 – Gráficos de desempenho das quatro melhores arquiteturas de rede neural.



Fonte: O autor.

Figura 21 – Gráficos de regressão das quatro melhores arquiteturas de rede neural.

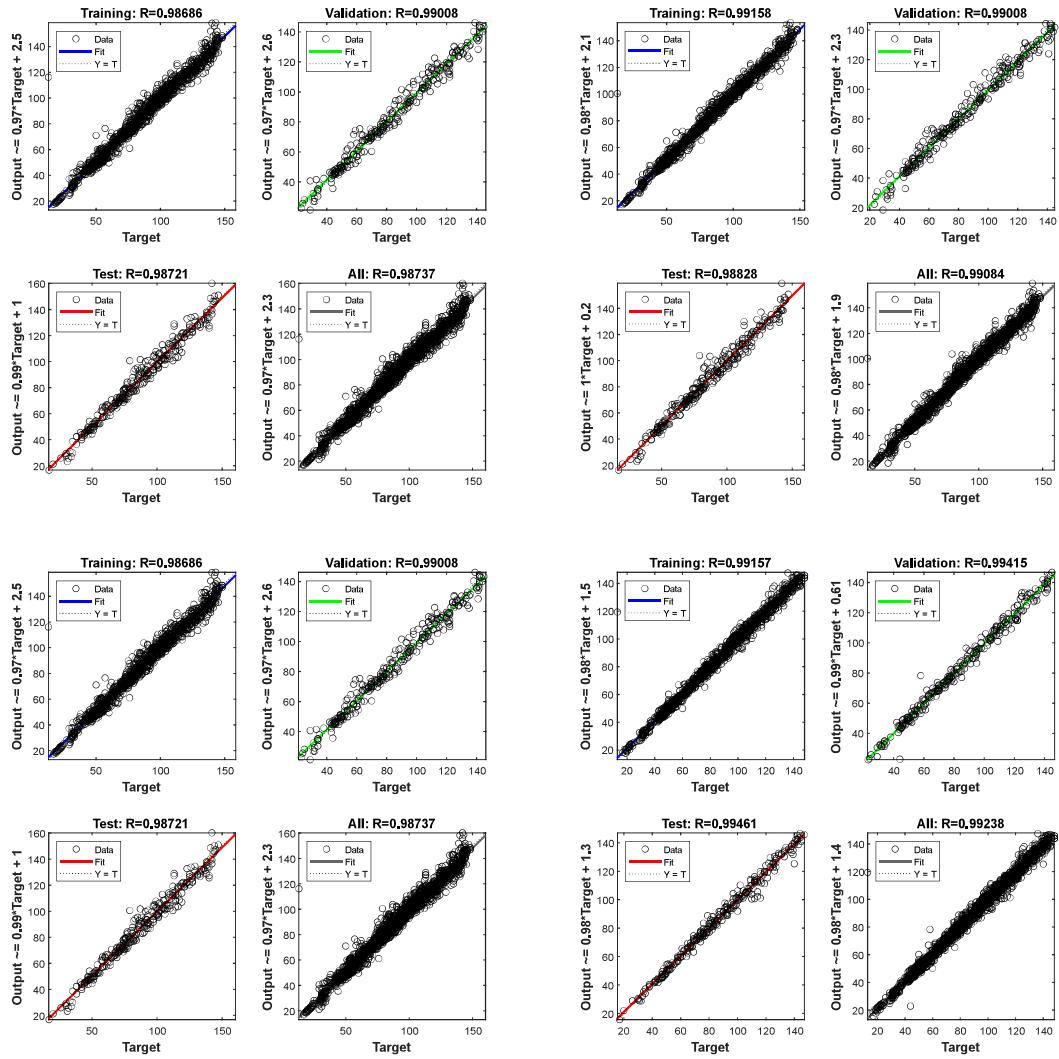


Fonte: O autor.

- **Regressão baseada em Processos Gaussianos (GPR).**

Para o modelo GPR, testou-se quatro diferentes tipos de kernels: *squared exponential*, *exponential*, *rational quadratic*, (*Automatic Relevance Determination*) *ARD squared exponential*. É possível observar, nos gráficos de regressão, que os pontos estão bem localizados próximos à linha de identidade, indicando um bom ajuste na predição do torque. É possível identificar a ausência de *overfitting* e *underfitting* (figura 22).

Figura 22 – Gráficos de regressão das quatro tipo de kernel utilizados.



Fonte: O autor.

4.1. Indicadores de Desempenho dos modelos.

Além dos gráficos de desempenho, foram utilizadas métricas indicadoras de desempenho (KPIs, *Key Performance Indicators*) para avaliar os modelos obtidos:

1. Erro percentual absoluto (APE, *Absolute Percentage Error*)

$$APE = \left| \frac{y_i - y_{mdl_i}}{y_i} \right| \cdot 100\% \quad (11)$$

2. Erro percentual absoluto médio (MAPE, *Mean Absolute Percentage Error*)

$$\text{MAPE} = \frac{1}{N} \cdot \sum_{i=1}^N \left| \frac{y_i - y_{mdl_i}}{y_i} \right| \cdot 100\% \quad (12)$$

3. Raiz do erro quadrático médio (RMSE, *Root Mean Squared Error*)

$$\text{RMSE} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (y_i - y_{mdl_i})^2} \quad (13)$$

4. Erro quadrático médio normalizado (NRMSE, *Normalized Root Mean Squared Error*)

$$\text{NRMSE} = \frac{1}{\bar{y}} \cdot \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (y_i - y_{mdl_i})^2} \quad (14)$$

5. Coeficiente de determinação (R^2 , *Coefficient of Determination*)

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - y_{mdl_i})^2}{\sum_{i=1}^N (\bar{y} - y_i)^2} \quad (15)$$

N : Número total de amostras.

y_i : valor da amostra.

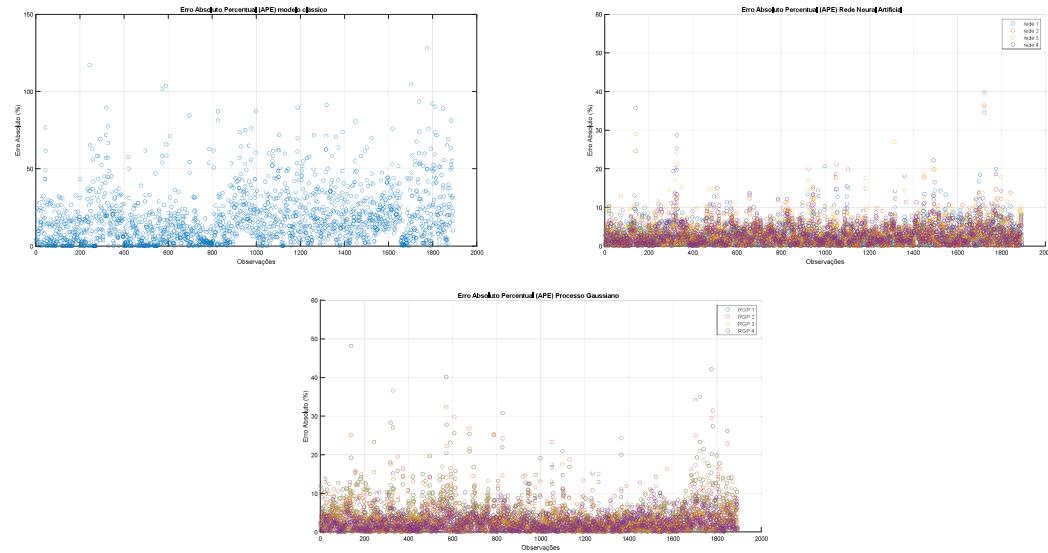
y_{mdl_i} : valor predito pelo modelo.

\bar{y} : valor médio das amostras.

É elaborado um gráfico de dispersão com o erro percentual absoluto para cada ponto do mainDataset e para cada tipo de modelo (figura 23). Observou-se que

o modelo padrão apresentou valores de erro significativamente mais elevados quando comparado às redes neurais artificiais (RNAs) e à regressão por processos gaussianos (GPR). Além disso, o erro no modelo padrão demonstrou uma dispersão bem maior em relação aos demais modelos, indicando um desempenho inferior na predição.

Figura 23 – Gráficos de dispersão com erro absoluto percentual para cada ponto do mainDataset.



Fonte: O autor.

Na Tabela 2, são apresentados os KPIs para o modelo padrão, as quatro melhores topologias de rede neural e cada um dos quatro kernels selecionados para a regressão por processos gaussianos (GPR). Observou-se que o erro absoluto médio (MAPE), o erro quadrático médio (RMSE) e o erro médio quadrático normalizado (NRMSE) foram, em média, cinco vezes menores nos modelos baseados em técnicas de *machine learning* em comparação ao modelo padrão.

Tabela 2 – Indicadores de desempenho e tempo de execução para cada modelo.

Modelo	MAPE	RMSE	NRMSE	R2	tempo execução (ms)
Modelo Padrão	18,69	18,33	0,21	0,65	13,71
Rede Neural Artificial	3,12	3,75	0,04	0,99	7,33
Rede Neural Artificial	3,24	3,87	0,05	0,98	6,49
Rede Neural Artificial	3,36	3,99	0,05	0,98	6,08
Rede Neural Artificial	3,67	4,04	0,05	0,98	6,29
Processo Gaussiano	4,38	4,92	0,06	0,97	14,79
Processo Gaussiano	3,78	4,20	0,05	0,98	20,60

Processo Gaussiano	4,38	4,92	0,06	0,97	20,47
Processo Gaussiano	3,14	3,83	0,04	0,98	38,94

Fonte: O autor.

Na Tabela 3, são apresentados os KPIs encontrados nos principais artigos de referência, permitindo um comparativo com os resultados obtidos (benchmark).

Tabela 3 – Indicadores de desempenho para os principais artigos de referência (Benchmark).

Referência	Método	Aplicação	Métricas de Erro
[6]	RNA	Torque de saída	MAPE: 3,33% (médio)
[7]	RNA	Torque de saída	
[8]	RNA	Potência de saída	RMSE: 0.0046 R ² : 0.9952 MAE: 1,51%
[9]	RNA Convolucional	Torque de saída em série temporal	RMSE: 0.0018 MAPE: 0,56%.
[10]	GPR	Emissões de gases	RMSE: 0.1907 MAE: 0.1364 R ² : 0.9518 (médio)
[11]	GPR	Sistema de admissão de ar	RMSE: 128,01 NRMSE: 0,06 R ² : 0,92 (médio)
[12]	GPR	Emissões de gases	NRMSE: 4,80% (médio)
[13]	GPR	Emissões de gases	NRMSE: 0,392% (médio)
[17]	GPR	Torque de saída e emissões de gases	RMSE: 6.1125 MSE: 37.362 MAE: 2.8717
[18]	GPR	Massa de ar estimada	MAPE: 3.3% RMSE: 7.86

Fonte: O autor.

Os modelos obtidos com a Rede Neural Artificial (RNA) e a regressão com Processo Gaussiano (GPR) apresentam excelente desempenho em comparação aos *benchmarks*, com MAPE de 3,12% e 3,14%, RMSE de 3,75 e 3,83, e R² de 0,99 e 0,98, respectivamente. O coeficiente de determinação (R²) apresentou valores próximos de 1 para ambos os modelos. O tempo de execução das redes neurais foi inferior à metade do tempo dos modelos baseados em GPR. Com base nos resultados, a RNA destacou-se como a melhor solução, apresentando erros comparáveis ao GPR, porém com menor tempo de execução.

5. CONCLUSÃO.

No decorrer deste trabalho, aplicaram-se técnicas de análise de dados e aprendizado de máquina (*Machine Learning*) a um problema real, com o objetivo de modelar o torque em motores de combustão interna.

As técnicas de análise de dados permitiram identificar janelas temporais dentro dos ensaios em regime transitório, nas quais o motor apresentou comportamento em regime permanente. Isso possibilitou a integração dos dados dos ensaios realizados em regimes transitório e permanente em um único conjunto de dados.

Os modelos que utilizaram aprendizado de máquina apresentaram uma capacidade de previsão do torque com precisão superior ao modelo padrão, com um erro aproximadamente cinco vezes menor. Foram testadas duas técnicas de aprendizado de máquina: redes neurais artificiais e regressão por processos gaussianos. O resultado final indicou que a rede neural foi a melhor solução, apresentando um erro semelhante ao da regressão por processos gaussianos, porém com um tempo de execução inferior à metade do tempo necessário para executar o modelo de regressão por processos gaussianos.

Com base nos resultados obtidos, o trabalho desenvolvido pode ser ampliado para a modelagem de emissões de gases poluentes, como óxidos de nitrogênio (NO_x), hidrocarbonetos totais (THC - *Total Hydrocarbons*), dióxido de carbono (CO₂) e monóxido de carbono (CO). Outra possibilidade de continuidade deste trabalho é a inclusão de novos dados, considerando motores de modelos e tecnologias diferentes, como motores turboalimentados, incluindo os modelos de injeção direta (TSI - *Turbocharged Stratified Injection*).

REFERÊNCIAS

- [1] HEYWOOD, J. B. **Internal combustion engine fundamentals**. McGraw-Hill Education, 2018.
- [2] BRUNETTI, F. **Motores de combustão interna**. São Paulo: Blucher, 2012.
- [3] GUZZELLA, L.; ONDER, C. **Introduction to modeling and control of internal combustion engine systems**. ed. 2. Springer Science & Business Media, 2010.
- [4] ERIKSSON, L.; NIELSEN, L. **Modeling and control of engines and drivelines**. John Wiley & Sons, 2014.
- [5] ISERMANN, Rolf. **Engine modeling and control**. Berlin: Springer Berlin Heidelberg, v. 1017, 2014.
- [6] Ge, Y., Huang, Y., Hao, D., Li, G., & Li, H. An indicated torque estimation method based on the Elman neural network for a turbocharged diesel engine. **Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering**, v. 230, n. 10, p. 1299-1313, 2016. DOI:10.1177/0954407015606271
- [7] Kaisheng, H., Shuaiyu, W., Zhenhua, J., & Dinan, J. Feedforward method of engine torque estimation. In: **2006 IEEE international conference on vehicular electronics and safety**. IEEE, 2006. p. 246-249. DOI: 10.1109/ICVES.2006.371592
- [8] HOFNY, M. S., GHAZALY, N. M., SHMROUKH, A. N., & ABOUELSOUD, M. Comparative Study of ANN and SVM Model Network Performance for Predicting Brake Power in SI Engines Using E15 Fuel. **International Journal of Robotics and Control Systems**, v. 4, n. 3, p. 979-999, 2024. DOI:10.31763/ijrcs.v4i3.1429

[9] HOFNY, M. S., GHAZALY, N. M., SHMROUKH, A. N., & ABOUELSOUD, M. Predicting SI Engine Performance Using Deep Learning with CNNs on Time-Series Data. **Journal of Robotics and Control (JRC)**, v. 5, n. 5, p. 1457-1469, 2024. DOI: 10.18196/jrc.v5i5.22558

[10] Pan T, Cai Y, Chen S (2021) "Development of an Engine Calibration Model Using Gaussian Process Regression", **International Journal of Automotive Technology**. DOI: 10.1007/s12239-021-0031-5

[11] Aran V, Unel M (2018) "Gaussian Process Regression Feedforward Controller for Diesel Engine Airpath", **International Journal of Automotive Technology**. DOI: 10.1007/s12239-018-0060-x

[12] Berger B, Rauscher F (2012) "Robust Gaussian Process Modelling For Engine Calibration" **IFAC Proceedings Volumes**. DOI: 10.3182/20120215-3-AT-3016.00028

[13] Berger B, Rauscher F, Lohmann B (2011) "Analysing Gaussian processes for stationary black-box combustion engine modelling", **IFAC Proceedings Volumes**. DOI: 10.3182/20110828-6-IT-1002.01160

[14] Gutjahr, T., Kruse, T., Huber, T. (2017) "Advanced modeling and optimization for virtual calibration of internal combustion engines", **NDIA ground vehicle systems engineering and technology symposium**

[15] Klein, S., Savelsberg, R., Xia, F., Guse, D., Andert, J., Blochwitz, T., & Amringer, N. Engine in the loop: Closed loop test bench control with real-time simulation. **SAE International journal of commercial vehicles**, v. 10, n. 2017-01-0219, p. 95-105, 2017. DOI: 10.4271/2017-01-0219.

- [16] CARLSON, R. B., WISHART, J., & STUTENBERG, K. On-road and dynamometer evaluation of vehicle auxiliary loads. **SAE International Journal of Fuels and Lubricants**, v. 9, n. 1, p. 260-268, 2016. DOI: 10.4271/2016-01-0901
- [17] Shi X, Jiang D, Qian W, Liang Y (2022) "Application of the Gaussian Process Regression Method Based on a Combined Kernel Function in Engine Performance Prediction", **ACS Omega**. DOI: 10.1021/acsomega.2c05952
- [18] Kolewe B, Haghani A, Beckmann R, Jeinsch T (2015) "Gaussian mixture regression and local linear network model for data-driven estimation of air mass", **IET Control Theory and Applications**. DOI: 10.1049/iet-cta.2014.0213
- [19] Rasmussen, C. E. and C. K. I. Williams. **Gaussian Processes for Machine Learning**. MIT Press. Cambridge, Massachusetts, 2006.
- [20] DAMA INTERNATIONAL. **DAMA-DMBOK: Data Management Body of Knowledge**. 2nd ed. Technics Publications, 2017.
- [21] Maggio, A. V. O., Laganá, A. A. M., & Silva, M. H. C. (2024). **Dataset para Estimação de Torque de Motores de Combustão Interna Flex-Fuel [Data set]**. Zenodo. 10.5281/zenodo.14251864.

ANEXO A – Equipamentos utilizados.

Dinamômetros
<p>Fabricante: SCHENCK</p> <p>Modelo: D 360 1E automatizado, <i>software</i> ECAT</p> <p>Capacidade: 700Nm e 5000RPM</p> <p>Tipo: hidráulico, passivo.</p>
<p>Fabricante: Antriebstechnik</p> <p>Modelo: INDY 33/4P</p> <p>Capacidade: 330kW e 8000RPM</p> <p>Tipo: Elétrico, ativo.</p>
Motor
<p>Fabricante: Volkswagen</p> <p>Modelo: EA 111 VHT (very high torque) 1.6L Total Flex Ano 2008.</p> <p>Especificações: Alimentação: aspirado; 4 cilindros com 8 válvulas; taxa de compressão: 12,1:1; diâmetro/curso do pistão: 76,5/86,9mm; cilindrada: 1.598 cm³; Potência máxima a 5250RPM: 101,0cv (74 kW) com gasolina e 104,0cv (76 kW) com etanol; Torque máximo a 2500 RPM: 151Nm com gasolina, 153Nm com etanol;</p>
ECU
<p>Fabricante: BOSCH/ETAS</p> <p>Modelo: Flex ECU MED17ETAS-2.41</p> <p>Software interno: FlexECU-G1 para gasolina</p>
Software de calibração e aquisição da ECU
<p>Fabricante: ETAS</p> <p>Modelo: INCA V7 .1</p>
Sensor Lambda
<p>Fabricante: BOSCH</p> <p>Modelo: LSU4.9</p> <p>Tipo: Planar de banda larga (<i>Planar wide band</i>)</p>
Medidor Lambda
<p>Fabricante: ETAS</p> <p>Modelo: LA4</p>

Sensor de Composição			
Fabricante: Continental			
Modelo: Flex-fuel sensor generation II			
Sensor de fluxo de massa de ar (MAF)			
Fabricante: MTE			
Modelo: 7113			
Sensor de pressão do escape			
Fabricante: Wika			
Modelo: P# 9013512.			
Sensor de temperatura do escape, líquido de refrigeração e do óleo			
Fabricante: ETAS			
Modelo: ES650			
Tipo: termopar tipo K para a temperatura do escape, e termopar tipo J para a temperatura líquido de refrigeração e do óleo.			
Hardware de simulação			
Fabricante: ETAS			
Modelo: Core i7 3,44GHz –8GB RAM; placa de interface: ES5340.2 –ICE board			
Software de controle da simulação			
Fabricante: ETAS			
Modelo: Experiment Environment e LabCar–IP V5.4			

Características dos combustíveis utilizados nos ensaios. Os ensaios foram realizados no Laboratório de Combustíveis Líquidos do Instituto de Pesquisas Tecnológicas do Estado de São Paulo. Os dados foram obtidos pelos métodos: ASTM D2699, D2700, D4052 e ABNT: 13992, 15639 (E100) Tabela 4.

Tabela 4 – Dados do Laboratório de Combustíveis Líquidos

Mistura	E27	E65	E100
Densidade a 20°C (Kg/m ³)	753,2	778,4	805
Etanol anidro [Vol.%]	27	65	94,7
Água [Vol. %]	0,5 – 0	2,3	5,3
Índice antidetonante [IAD]	88	97	100
Relação estequiométrica	13,1	10	9