DOTTORATO DI RICERCA IN

AUTOMOTIVE ENGINEERING FOR INTELLIGENT MOBILITY

Ciclo 37

Settore Concorsuale: 09/C1 - MACCHINE E SISTEMI PER L'ENERGIA E L'AMBIENTE

Settore Scientifico Disciplinare: ING-IND/08 - MACCHINE A FLUIDO

DATA-DRIVEN APPROACHES FOR ENHANCED ON-BOARD FAULT DIAGNOSIS AND EMISSIONS MONITORING TO SUPPORT EURO 7 STANDARD IMPLEMENTATION

Presentata da: Stella Canè

Coordinatore Dottorato	**Supervisore**
Davide Moro	Nicolò Cavina

Esame finale anno 2025

Abstract

The European Commission has recently published the new Euro 7 standard, with the target of reducing the impact on pollutant emissions due to the transport sector. Besides forcing internal combustion engines to operate cleaner for an extended useful life, the new regulation points out the role of On-Board Monitoring (OBM) as a key enabler to ensure limited pollutant emissions over the whole vehicle lifetime, necessarily taking into account the natural ageing of involved systems, as well as possible faults and malfunctions. In this scenario, the research activity described in this dissertation aims at investigating the potential of data-driven approaches in detecting emission-relevant engine faults, supporting standard On-Board Diagnostics (OBD), and allowing real-time monitoring of pollutant emissions both in nominal and non-nominal conditions, which is one of the main challenges introduced by Euro7 OBM requirements.

As a preliminary step, a selection of common emission-relevant engine faults has been simulated by means of a detailed and validated 0-D model representative of a Diesel Plug-in Hybrid Electric Vehicle (PHEV). The aim is to define the effects of the selected faults on engine-out NOx emissions and to identify on-board available signals that contain useful fault-related information to be used as input for the OBM-oriented data-driven models. In this regard, a detailed performance evaluation of different classifiers has been carried out to identify the most suitable data-driven methods to be applied to engine fault diagnosis. The analysis considers not only model accuracy, but also training time and resulting prediction speed, which are key parameters for real-time model implementation. The results demonstrate that Tree, Ensemble, and Neural Network classifiers are the best-performing ones in terms of accuracy, exhibiting also a good trade-off between classification performance and computational complexity. Based on this analysis, the mentioned classification models have been optimized through Bayesian optimization to further improve their accuracy. The same methodology has been applied to develop OBM-oriented data-driven regression models for NOx emission estimation. More in detail, the same input dataset considered for the fault classification model has been used to train selected regression models, based on a preliminary performance analysis of the most common regression algorithms. The model output consists of a correction factor to be applied to the NOx emission estimated by the reference ECU model, which is not sufficiently sensitive to induced faults, thus resulting in a significant residual error on the final cumulated emission when a non-nominal condition is considered. The developed regression models, taking advantage of easily available on-board signals that contain fault-related

information, allow the calculation of the required correction factor to be applied to the reference NOx estimation in order to reduce the deviation between actual and estimated NOx emission. To assess the robustness of the developed models, as well as their interpolation and extrapolation capability, different driving cycles and fault conditions have been simulated. Based on these testing results, it can be concluded that the proposed data-driven classification approach allows to correctly detect and identify selected faults or combination of them with high overall accuracy, showing satisfactory interpolation and extrapolation capabilities when tested on selected driving cycles and specific fault conditions not included in the corresponding training dataset. Regarding the on-board NOx emission monitoring task, the developed data-driven regression models show promising results when tested on different testing cycles: the correction factor calculated by the models allows to significantly reduce the gap between actual and estimated NOx emission compared to the reference map-based controller model, even if highly emission-relevant faults are introduced. For both classification and regression tasks, the Neural Network model is preferred over the other considered approaches, since providing the best trade-off between accuracy, generalisation capability, and model complexity.

In view of vehicle on-board implementation, the developed models have been deployed on a commercial real-time hardware, namely a Raspberry Pi computer, and tested at the HiL to evaluate their real-time capability and related computational load. In addition, the proposed hardware setup offers the possibility to perform on-board vehicle testing of the developed models at low cost, enabling communication via CAN-bus with the ECU and on-board sensors, taking advantage of information that is easily available on a standard vehicle without the need of modifying the already existing hardware/software architecture. As an added value to this research project, following the promising initial outcomes achieved in simulation, the same validated methodology has been applied to a real-world dataset obtained from test bench measurements, including both nominal conditions and deliberately introduced faults. The results in terms of effectiveness, robustness, and real-world applicability provide a foundation for further investigations in this field, offering a promising solution to assist manufacturers in meeting the requirements introduced by Euro 7 regulation.

All the activities presented in this doctoral dissertation have been carried out at the *Green Mobility Research Lab* (GMRL), a research center resulting from the partnership between the University of Bologna and FEV Italia s.r.l., which represents the industrial partner of the research project.

Keywords

Internal combustion engines · Fault Detection and Identification (FDI) · On-Board Monitoring (OBM) · Euro 7 · Machine Learning (ML) · Neural networks · Nitrogen oxides (NOx) · Mean value Engine Model (MVEM) · Real-time implementation

Abstract in lingua italiana

La Commissione Europea ha recentemente pubblicato la nuova normativa Euro 7, con l'obiettivo di ridurre l'impatto del settore dei trasporti sulle emissioni inquinanti. Oltre ad imporre ai motori a combustione interna di operare in modo più "pulito" per una vita utile prolungata, le nuove disposizioni sottolineano il ruolo dell' *On-Board Monitoring* (OBM) come elemento chiave per garantire emissioni inquinanti limitate per l'intera durata di vita del veicolo, tenendo necessariamente conto del naturale invecchiamento dei sistemi coinvolti, nonché di possibili guasti e malfunzionamenti. In questo scenario, l'attività di ricerca descritta in questa tesi mira a studiare il potenziale degli approcci *data-driven* nel rilevare eventuali guasti del motore rilevanti dal punto di vista delle emissioni, supportando la diagnostica di bordo (*On-Board Diagnostics*, OBD) e consentendo il monitoraggio in tempo reale delle emissioni inquinanti sia in condizioni nominali che non nominali, trattandosi di una delle principali sfide introdotte dai requisiti Euro7.

Come primo passo, è stata simulata una selezione di guasti e malfunzionamenti che coinvolgono il motore e incidono sulle principali emissioni inquinanti. Per le simulazioni è stato utilizzato un modello 0-D dettagliato e validato su dati sperimentali, rappresentativo di un veicolo ibrido elettrico plug-in diesel. L'obiettivo è definire gli effetti dei guasti selezionati sulle emissioni di ossidi di azoto (NOx) del motore e identificare i segnali disponibili a bordo che contengono informazioni utili relative ai guasti da utilizzare come input per i modelli *data-driven*. A tal scopo, è stata effettuata una valutazione dettagliata delle prestazioni di diversi classificatori per individuare i metodi *data-drive*n più adatti per la diagnosi dei guasti del motore. L'analisi considera non solo l'accuratezza del modello, ma anche il tempo di allenamento e la velocità di previsione, parametri fondamentali per l'implementazione *real-time* del modello. I risultati dimostrano che i classificatori di tipo *tree*, *ensemble* e reti neurali sono i più performanti in termini di accuratezza, mostrando anche un buon compromesso tra prestazioni di classificazione e carico computazionale. Sulla base di questa analisi, i modelli di classificazione menzionati sono stati ottimizzati tramite l'ottimizzazione bayesiana per migliorare ulteriormente la loro accuratezza. La stessa metodologia è stata applicata per sviluppare modelli di regressione *data-driven* per la stima delle emissioni di NOx. Più in dettaglio, lo stesso set di dati di input considerato per il modello di classificazione dei guasti è stato utilizzato per allenare una selezione di modelli di regressione, scelti sulla base di un'analisi preliminare delle prestazioni degli algoritmi di regressione più comuni. L'output del

modello consiste in un fattore correttivo da applicare all'emissione di NOx stimata dal modello centralina di riferimento, che si dimostra non sufficientemente sensibile ai guasti indotti, portando così ad un errore residuo significativo sull'emissione cumulata a fine ciclo guida ogni volta che si considera una condizione non nominale. I modelli di regressione sviluppati, sfruttando alcuni segnali facilmente reperibili a bordo e che contengono informazioni rilevanti relative ai guasti, consentono di calcolare il fattore correttivo necessario da applicare alla stima degli NOx di riferimento per ridurre la deviazione tra le emissioni di NOx effettive e quelle stimate. Per valutare la robustezza dei modelli sviluppati e la loro capacità di interpolazione ed estrapolazione, sono stati simulati diversi cicli di guida e condizioni di guasto. Sulla base dei risultati di questi test, si può concludere che l'approccio di classificazione *data-driven* proposto consente di rilevare e identificare correttamente i singoli guasti considerati, così come una combinazione di essi, con un'elevata accuratezza complessiva, mostrando capacità di interpolazione ed estrapolazione soddisfacenti quando testate su cicli di guida e specifiche condizioni di guasto non incluse nel corrispondente set di dati di allenamento. Per quanto riguarda il monitoraggio delle emissioni di NOx, i modelli di regressione sviluppati mostrano risultati promettenti se testati su diversi cicli guida: il fattore di correzione calcolato dai modelli consente di ridurre significativamente il divario tra le emissioni di NOx effettive e quelle stimate rispetto al modello in centralina basato su mappe di riferimento, anche se vengono introdotti guasti altamente impattanti sulle emissioni. Sia nel caso della classificazione che in quello della regressione, i modelli basati sulle reti neurali sono preferibili rispetto agli altri approcci considerati, in quanto offrono il miglior compromesso tra accuratezza, capacità di generalizzazione e complessità del modello.

In vista dell'implementazione a bordo veicolo, i modelli sviluppati sono stati implementati su un hardware commerciale, in particolare un computer Raspberry Pi, per poi essere testati all'Hardware-in-the-Loop (HiL) per valutarne la capacità di calcolo *real-time* e il relativo carico computazionale. Inoltre, la configurazione hardware proposta offre la possibilità di eseguire test a bordo veicolo a basso costo, consentendo la comunicazione via CAN-bus con la centralina motore e i sensori di bordo, sfruttando le informazioni facilmente disponibili su un veicolo commerciale senza la necessità di modificare l'architettura hardware e software già esistente. Come valore aggiunto a questo progetto di ricerca, visti i promettenti risultati iniziali ottenuti in simulazione, la stessa metodologia è stata applicata a un set di dati reali ottenuti da misurazioni a banco, che includono sia condizioni nominali che guasti appositamente introdotti. I risultati in termini di efficacia, robustezza ed effettiva applicabilità della metodologia proposta costituiscono una base per ulteriori indagini in questo campo, offrendo una soluzione promettente per supportare i produttori nel soddisfare i requisiti introdotti dalla normativa Euro 7.

Tutte le attività presentate in questa tesi di dottorato sono state svolte presso il *Green Mobility Research Lab* (GMRL), un centro di ricerca nato dalla collaborazione tra l'Università di Bologna e FEV Italia s.r.l., che rappresenta il partner industriale del progetto di ricerca.

Contents

x

List of figures

xiv

List of tables

Nomenclature

Acronyms

CH_4	Methane
CO_2	Carbon dioxide
CO	Carbon monoxide
N_2O	Nitrous oxide
NH_3	Ammonia
NO_2	Nitrogen dioxide
NOx	Nitrogen oxides
NO	Nitrogen monoxide
SO_2	Sulfur oxide
AFR	Air-Fuel Ratio
ANN	Artificial Neural Network
AT	Automated Transmission
BCM	Body Control Module
BEV	Battery Electric Vehicle
BMEP	Brake Mean Effective Pressure
BSG	Belt-driven Starter Generator
CAC	Charge Air Cooler
CAN	Controller Area Network
CARB	California Air Resources Board
CF	Conformity Factor
CFD	Computational Fluid Dynamics
CI	Compression Ignition
CLOVE	Consortium for ultra LOw Vehicle Emissions
CNN	Convolutional Neural Network
DL	Deep Learning
DOC	Diesel Oxidation Catalyst
DPF	Diesel Particulate Filter
DT	Decision Tree

EATS	Exhaust After-Treatment System
EC	European Commission
ECU	Electronic / Engine Control Unit
EEA	European Environmental Agency
EEDWS	Excess Emission Driver Warning System
EGC	EGR Cooler
EGR	Exhaust Gas Recirculation
EM	Electric Motor
EPA	Environmental Protection Agency
EU	European Union
FDI	Fault Detection and Identification
FN(R)	False Negative (Rate)
FP(R)	False Positive (Rate)
FTP	Federal Test Procedure
GHG	Greenhouse gas
GPR	Gaussian Process Regression
GWP	Global Warming Potential
HC	Unburned hydrocarbons
HCU	Hybrid Control Unit
HD	Heavy-Duty (vehicles)
HEV	Hybrid Electric Vehicles
HiL	Hardware-in-the-Loop
HP EGR	High-Pressure EGR
HV	High-Voltage
ICE	Internal Combustion Engine
ISC	In-Service Conformity
ISO	International Organization for Standardization
IUPR	In-Use Performance Ratio
kNN	k-Nearest Neighbours
LCV	Light Commercial Vehicle
LD	Light-Duty (vehicles)
LP EGR	Low-Pressure EGR
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAF	Mass Air Flow
MAP	Manifold Absolute Pressure
MaS	Market Surveillance
MIL	Malfunction Indicator Light
MIQ	Mutual Information Quotient

ML	Machine Learning
MRMR	Minimum Redundancy Maximum Relevance
MVEM	Mean Value Engine Model
NEDC	New European Driving Cycle
NMVOC	Non-Methane Volatile Organic Compounds
OBD	On-Board Diagnostics
OBFCM	On-Board Fuel and Energy Consumption Monitoring
OBM	On-Board Monitoring
OEM	Original Equipment Manufacturers
OTA	Over-The-Air
PCA	Principal Component Analysis
PEMS	Portable Emission Measuring Systems
PHEV	Plug-in Hybrid Electric Vehicle
PM	Particulate matter
PN	Particulate number
RDE	Real Driving Emissions
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SAE	Society of Automotive Engineers
SCR	Selective Catalytic Reduction
SDG	Signed Directed Graphs
SI	Spark Ignition
SoC	State of Charge
SoH	State of Health
SVM	Support Vector Machines
TN(R)	True Negative (Rate)
TP(R)	True Positive (Rate)
UNECE	United Nations Economic Commission for Europe
VGT	Variable Geometry Turbine
VVT	Variable Valve Timing
WG	Waste Gate
WHO	World Health Organization
WLTC	Worldwide harmonized Light vehicles Test Cycle
WLTP	Worldwide harmonized Light vehicles Test Procedure

Roman symbols

c_p	Specific heat capacity (at constant pressure) $[J \cdot kg^{-1} \cdot K^{-1}]$

$C_{bat}(\cdot)$	Battery nominal capacity [Ah]
d_{rk}	Euclidean distance between two observations r and k [-]
f'_{NOx}	Correction factor on engine-out NOx concentration [-]
$F(\cdot)$	Force [N]
f_0, f_1, f_2	Coast-down parameters [N], [N/(km/h)], [N/(km/h)^2]
$F_t(\cdot)$	Propulsion force [N]
$F_a(\cdot)$	Aerodynamic friction force [N]
$F_g(\cdot)$	Gravitational force [N]
f_{NOx}	Correction factor on engine-out NOx mass flow [-]
$f_{P,acc}, f_{I,acc}$	Proportional and integral factors of the PI controller for traction torque request
$f_{P,brk}, f_{I,brk}$	Proportional and integral factors of the PI controller for braking torque request
$F_{res}(\cdot)$	Resistant force acting on the vehicle [N]
$F_r(\cdot)$	Rolling friction force [N]
g	Gravitational acceleration [m/s^2]
$I(X;Y)$	Mutual information between variable X and Y [-]
$I_{bat}(\cdot)$	Battery current [A]
$k_{T_{InMfd}}, k_{T_{Co}}$	Correction parameters for cylinder external charge calculation [-]
$\dot{m}_{cpr}(\cdot)$	Mass flow entering the compressor [kg/s]
$\dot{m}_{exh}(\cdot)$	Exhaust gas mass flow entering the turbine [kg/s]
$\dot{m}_f(\cdot)$	Injected fuel mass flow [kg/s]
$\dot{m}_{HpEgr}(\cdot)$	High-pressure EGR mass flow [kg/s]
$\dot{m}_{NOx}(\cdot)$	Engine-out NOx mass flow [g/s]
$m_{cyl,ext}$	Cylinder external charge [kg]
$m_{eq,rot}(\cdot)$	Equivalent mass of rotative components [kg]
m_v	Vehicle mass [kg]
$n_{eng}(\cdot)$	Engine revolution speed [rpm]
$n_{mot,act}(\cdot)$	Actual electric motor revolution speed [rpm]
$n_{mot,act}(\cdot)$	Revolution speed [rpm]
n_p, n_s	Number of parallel and series cells [-]
$P(\cdot)$	Power [W]
	Pressure [kPa]
$p(\cdot)$	Probability [-]
$p(X = x_i)$	Marginal probability of an event occurring independently [-]
$p(X = x_i, Y = y_j)$	Combined probability of two events occurring together [-]
P_{amb}	Ambient pressure [kPa]
$P_{bat,loss}(\cdot)$	Battery power losses [W]
$P_{el}(\cdot)$	Electrical power [W]
$P_{ExMfd}(\cdot)$	Exhaust manifold pressure [kPa]

$P_{InMfd}(\cdot)$	Intake manifold pressure [kPa]
$P_m(\cdot)$	Mechanical power [W]
$R(\cdot)$	Ohmic resistance [Ω]
r_w	Dynamic wheel radius [m]
r_{air}	Air specific gas constant [$J \cdot kg^1 \cdot K^{-1}$]
$R_{cel,int}(\cdot)$	Battery cell internal resistance [Ω]
$V_{cyl,efc}$	Effective cylinder volume [m^3]
t	Time [s]
$T(\cdot)$	Temperature [°C]
T_{amb}	Ambient temperature [°C]
$T_{c,in}, T_{c,out}$	Input / output coolant temperature [°C]
$T_{co,in}$	Temperature of the coolant fluid at the cooler inlet [°C]
$T_{ExMfd}(\cdot)$	Exhaust manifold temperature [°C]
$T_{InMfd}(\cdot)$	Intake manifold temperature [°C]
T_{us}, T_{ds}	Upstream / downstream temperature [°C]
$Tq(\cdot)$	Torque [Nm]
$Tq_{brk,ct}(\cdot)$	Braking torque requested by the controller [Nm]
$Tq_{eng}(\cdot)$	Engine torque [Nm]
$Tq_{mot,act}(\cdot)$	Actual electric motor torque [Nm]
$Tq_{req,ct}(\cdot)$	Torque requested at the wheels after PI controller intervention [Nm]
$Tq_{req}(\cdot)$	Torque requested by the driver for traction [Nm]
$V(\cdot)$	Voltage [V]
$v(\cdot)$	Vehicle longitudinal speed [km/h]
V_x	Relevance of a feature x [-]
$V_{bat,int}(\cdot)$	Battery internal voltage drop [V]
$V_{bat}(\cdot)$	Battery voltage [V]
$V_{cal,OC}(\cdot)$	Battery cell open circuit voltage [V]
$v_{req}(\cdot)$	Vehicle speed requested by the driver [km/h]
W_x	Redundancy of a feature x [-]
$y_{NOx}(\cdot)$	Engine-out NOx concentration [ppm]

Greek symbols

$\alpha(\cdot)$	Angle of slope [rad]
β	Compression ratio [-]
$\beta_{LpCpr}, \beta_{HpCpr}$	Compression ratios of low-pressure / high-pressure compressors [-]
$\beta_{LpTrb}, \beta_{HpTrb}$	Pressure ratios of low-pressure / high-pressure turbines [-]
Δh	Specific enthalpy difference [$J \cdot kg^{-1}$]

ΔP	Differential pressure / Pressure loss [kPa]
ΔP_{AirFt}	Pressure losses through air filter [kPa]
ΔP_{IntCo}	Pressure losses through intercooler [kPa]
ΔT	Differential temperature [°C]
ΔT_{AirFt}	Temperature difference upstream and downstream air filter [°C]
ΔT_{IntCo}	Temperature difference upstream and downstream intercooler [°C]
$\Delta T_{\text{LpCpr}}, \Delta T_{\text{HpCpr}}$	Temperature difference upstream and downstream low-pressure / high-pressure compressors [°C]
$\Delta T_{\text{LpTrb}}, \Delta T_{\text{HpTrb}}$	Temperature difference upstream and downstream low-pressure / high-pressure turbines [°C]
Δ	Difference [-]
η	Efficiency [-]
η_{co}	Cooling efficiency [-]
η_c	Inverter efficiency [-]
η_m	Mechanical efficiency [-]
η_v	Volumetric efficiency [-]
τ_{fd}	Final drive ratio [-]
τ_{gr}	Gear ratio [-]

Chapter 1

Introduction

In this chapter, an overview of the current automotive background is given in order to practically comprehend the reasons and motivations lying behind the research project and how the global regulatory targets are driving the automotive industry. In particular, the latest European regulations in terms of pollutant emissions are presented, focusing on imposed thresholds and monitoring requirements as well as the technical challenges associated with meeting them.

1.1 Motivation

Over the past decades, the growing level of industrialization and urbanization across several developing countries has led to a marked increase in the global atmospheric concentration of human-made greenhouse gases (GHG) and pollutants. As well known, this trend is closely linked to the main concerns of global warming, climate change, air quality and environmental degradation, together with all their severe implications.

In this context,researchers have highlighted that the primary greenhouse gases requiring global attention today are carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), and sulfur dioxide (SO_2). Even though water vapor is the most naturally abundant greenhouse gas in the atmosphere, CO_2 is the most frequently emitted by human activities [1]. More in detail, recent studies indicate that the transportation sector plays a significant role in GHG emissions, contributing for more than 20% of the global amount of equivalent CO_2[2]. With regard to pollutant emissions, recent advancements in combustion processes and aftertreatment technologies have shifted the focus primarily to nitrogen oxides (NOx), a group of highly reactive gases containing nitrogen and oxygen in varying proportions. Among these, nitrous oxide (N_2O) is a powerful greenhouse gas with a global warming potential (GWP) approximately three hundreds times that of CO_2. On the other hand, NOx (as the sum of NO and NO_2 is commonly referred to) has direct adverse effects on human health and the environment, including protein oxidation, cellular membrane damage, immune system disruption, respiratory and cardiovascular diseases, accelerated ageing of green

leaves, and reduced photosynthetic activity and biomass production [3, 4]. The situation becomes even more critical considering the increasing rate of urbanization that led to a massive migration of people from the rural areas to cities Fig. 1.1 [5]. In these urban centres, the concentration of pollutants, especially NOx and particulate matter (PM), tends to be much higher (red dots on the map in Fig. 1.2) and may exceed the safe limits set by the EU air quality standards and the World Health Organization (WHO) [6], as graphically shown in Fig. 1.3.
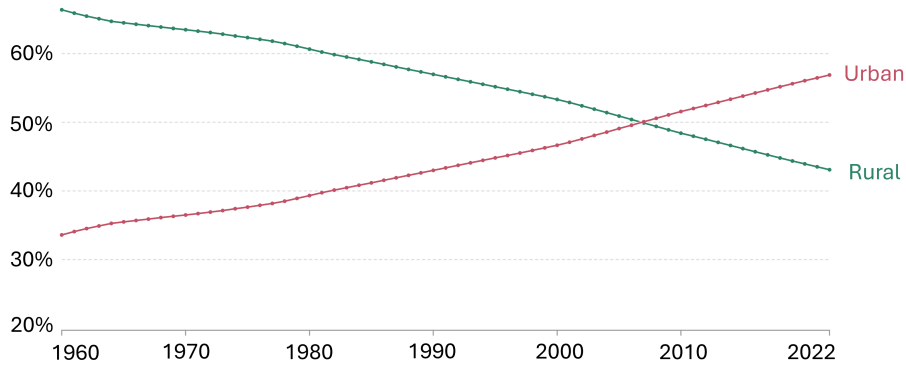
Figure 1.1: Global trend of the share of people living in urban and rural areas between 1960 and 2022. [7]
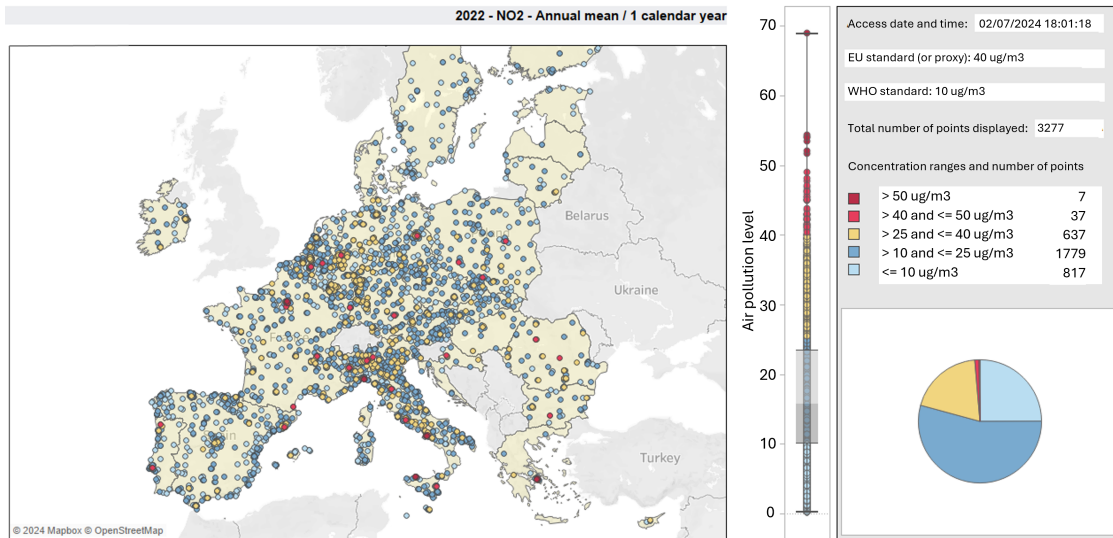
Figure 1.2: Average NO$_2$ concentration in Europe in 2022 [8]

As shown in Fig. 1.4, road transport is a leading source of NOx emissions in Europe, accounting for more than 36% of the total emissions in 2022; it also significantly contributes to PM2.5 emission (fine particular matter consisting of particles with a diameter of 2.5 micrometers or less), reaching the 10% of the total share [9]. As mentioned before, these percentages are much higher in cities, where the transport sector is regularly the main contributor to air pollution [10]. This, combined
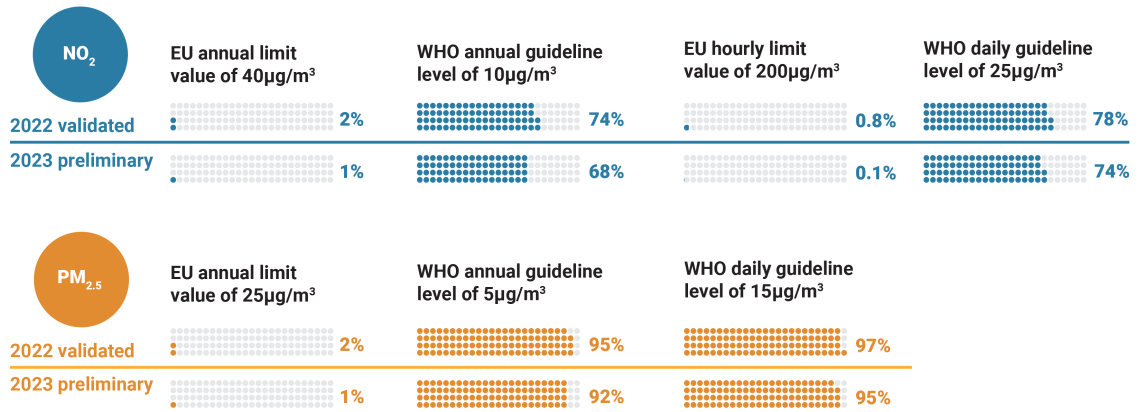
Figure 1.3: Percentage of monitoring stations registering NO_2 and PM2.5 concentrations above the EU annual limit value and the WHO guideline levels in 2022 and 2023

with the significant contribution to CO_2 emissions, underscores the necessity for cleaner and smarter transportation solutions.
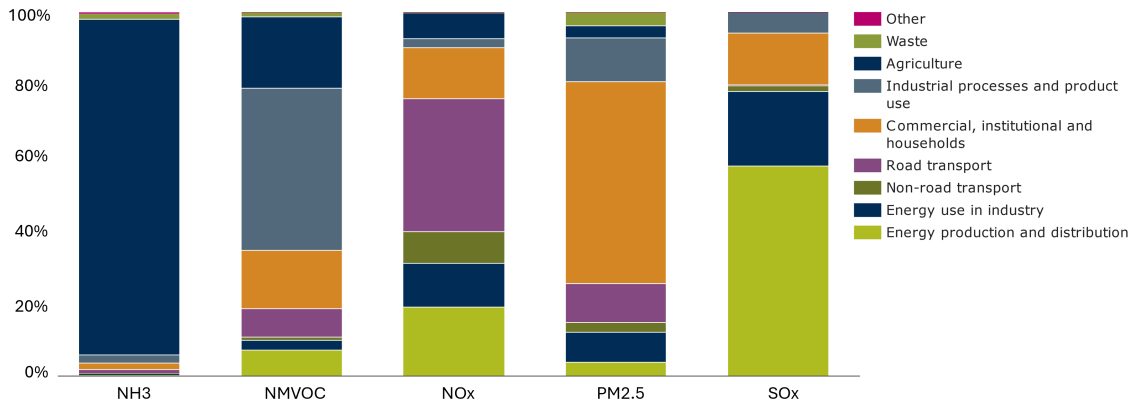
Figure 1.4: Emission of the main air pollutants by sector group in Europe in 2022 [11]

To address these issues, in recent years European regulations on CO_2 and pollutant emissions have become increasingly stringent for nearly all vehicle categories, from passenger cars to heavy-duty vehicles. In particular, in 2019, a target level of 118 [g/km] of CO_2 has been set by the European Commission for new passengers cars starting from 2021. Then, a 15% reduction of the mentioned threshold has been set as new target by 2025, followed by a stronger reduction of 55% by 2030 [12, 13]. Starting from 2019, all these CO_2 targets are assessed using the Worldwide harmonized Light vehicles Test Procedure (WLTP) [14], which replaced the New European Driving Cycle (NEDC) to reduce the observed gap between on-road emissions and laboratory testing [15].

Besides these demanding CO_2 standards, vehicle homologating procedures have also evolved

in the last years with regard to pollutant emissions: the tightening of emission limits and the introduction of real driving emissions (RDE) tests alongside the standard WLTP [16], has been a significant driver for innovation and the development of new powertrain and after-treatment technologies, forcing internal combustion engines to perform more efficiently and cleanly across a broader range of operating conditions. Looking forward, new challenges will be introduced with the upcoming Euro 7 regulation: emission compliance will have to be ensured throughout a significantly extended vehicle lifetime and to be confirmed by real-world testing and continuous emission On-Board Monitoring (OBM) on all trips [17].

Due to the multitude of factors affecting vehicle emissions as well as the measuring accuracy, meeting the intended continuous emissions monitoring standards represents the dominating challenge to comply with the Euro 7 standards [18]. In fact, recent studies prove that already existing technologies, eventually integrated into larger exhaust after-treatment systems (EATS) and combined with optimized thermal management, can address current weaknesses in terms of emission performance, such as during cold start phases, harsh accelerations, and uphill driving, allowing to comply with Euro 7 emissions thresholds [19, 20]. The same can not be said about current emission monitoring systems: while there are available and effective solutions for some pollutant species, including NOx, NH_3, and partially PM (although requiring further development for continuous monitoring), there are still no ready-to-use on-board sensors for other important pollutants like HC, CO, and PN, and future technical viability of such sensors has yet to be fully assessed [18]. Moreover, even considering the most mature technologies, there are intrinsic limitations for continuous and accurate Euro 7-compliant monitoring: focusing on NOx sensors, besides suffering from a non-negligible ammonia cross-sensitivity [21], they are characterized by a significant activation time, during which they are kept switched-off to avoid the risk of damage due to water condensation [22]. As shown in [23], the sensor can be kept switched off for several minutes, which can cover a significant part of the driving cycle, especially during short, cold-start tests. This has a substantial impact on data quality and integrity in view of a Euro 7-oriented real-time emission monitoring, mainly considering that cold starts are among the most critical driving events from an emission point of view due to the reduced efficiency of the EATS [24].

To face the request of a continuous monitoring of pollutant emissions, covering all the possible situations in which direct information from a physical sensor is not available or reliable enough, a model-based approach is needed in combination with on-board measurements. To date, accurate and robust physics-based emissions modelling is already possible considering vehicle nominal conditions [25–27]. The main challenge comes when continuous emission monitoring is requested in any possible operating condition, meaning that emission models need to be adapted to correctly predict emissions even when faulty conditions of one or more components occur.

In this context, this research activity is aimed at exploring possible approaches to face Euro 7 OBM requirements, especially focusing on the development of models that are capable of predicting emissions both in nominal and faulty conditions. For this purpose, data-driven methods have been investigated, taking advantage of available real-time on-board measurements to perform

engine fault detection and to correctly predict emissions in any possible operating condition, supporting standard on-board diagnostics to meet the challenging Euro 7 requirements.

1.2 Pollutant Emission Regulation

As already introduced, this work explores possible approaches to face Euro 7 OBM standards, mainly focusing on the development of different data-driven models that are capable of performing robust engine fault detection and predicting pollutant emissions, especially NOx, both in nominal and faulty conditions. Therefore, a brief introduction to the in-force emission regulations, with a focus on European laws, is provided in this section. Then, being significantly relevant for the topic of the dissertation, the newly approved Euro 7 regulation will be discussed in Section 1.2.3, with a deeper focus on emission OBM requirements and related challenges.

1.2.1 Regulated Pollutants

The European Environmental Agency (EEA) has defined a list of air pollutants that damage human health and the environment, including ammonia (NH_3), carbon monoxide (CO), non-methane volatile organic compounds (NMVOCs), nitrogen oxides (NOx), particulate matter (PM2.5, PM10), sulphur oxides (SOx), and unburned hydrocarbons (HC) [28]. Among these, the most critical pollutants associated with the road transport sector are:

- **Carbon monoxide (CO)**: formed during combustion as an intermediate oxidation product [29]:

$$2\,C + O_2 \longrightarrow 2\,CO \tag{1.1}$$

Under ideal conditions all CO is oxidized into CO_2, according to the reaction:

$$CO + OH \underset{-1}{\overset{+1}{\rightleftharpoons}} CO_2 + H \tag{1.2}$$

CO formation can be predominantly controlled by the air-fuel mixture in the cylinder, independently of fuel type [30]. With a rich mixture, more fuel than the stoichiometric amount can react with the air in the cylinder. Therefore, not all of the fuel can be completely burned. Conversely, with lean mixture, the probability of complete combustion increases and the CO decreases.

- **Unburned hydrocarbons (HC)**: primarily formed from the hydrocarbons present in the fuel that partially reacts, recombines, or remains unburned. Even the partial oxidation and evaporation of the lubricant oil could generate a small amount of HC. During the exhaust

5

phase, in the presence of oxygen, a large part of the HC is oxidized; the process can continue in the exhaust manifold (especially if secondary air is introduced)[29].

- **Nitrogen oxides (NO$_x$)**: they include both nitrogen monoxide (NO) and nitrogen dioxide (NO$_2$) species. NO can be formed from three primary mechanisms during combustion: Prompt NO is formed from the reaction of atmospheric nitrogen with combustion radicals, and this occurs only in the earliest stages of combustion; Fuel NO comes from the nitrogen chemically bonded to the fuel, which is usually removed in actual refinery processes; Thermal NO is the predominant NOx formation mechanism in internal combustion engines and can be described by the well-known Zeldovich mechanism [29], including three reactions for near-stoichiometric air-fuel mixtures:

$$N_2 + O \leftrightarrow NO + N \tag{1.3}$$

$$N + O_2 \leftrightarrow NO + O \tag{1.4}$$

$$N + OH \leftrightarrow NO + H \tag{1.5}$$

These reactions are highly temperature dependent, so the hotter the combustion, the more NOx is formed, independently from the type of fuel. This makes reducing combustion temperature the key strategy to low engine-out NOx. On the other hand, NO$_2$ is formed in colder regions from NO being further oxidized according to the forward reaction:

$$NO + O \leftrightarrow NO_2 \tag{1.6}$$

However, since reaction rates for NO oxidation reaction tend to be low, NO concentration in the exhaust gases is much higher than NO$_2$ concentration [30].

- **Particulate matter (PM) and number (PN)**: composed mainly of coal particles (soot) carried in suspension by exhaust gases. It is generated in a rich environment at high pressure and temperatures. The particulate matter is grouped with respect to the particles diameter (e.g. PM10 and PM2.5 comprise particles that are 10 or 2.5 micrometers or less in diameter respectively); both the mass and the number of particles are regulated.

1.2.2 In-force regulation: global trend and EU standards

Given their significant threat to human health and the increasing number of passenger cars and transport vehicles on the roads, pollutant emissions are more widely regulated compared to greenhouse gases, with over 70 countries implementing restrictions on vehicle emissions. The United States initiated these efforts with the EPA (Environmental Protection Agency) and CARB (California Air Resources Board) emission standards in the 1970s, while Europe introduced the

Euro emission standards in the 1990s [30]. The primary differences between the two systems lie in the testing procedures (Euro regulations using the WLTP, whereas the US employing the FTP-75) and in whether particulate number (PN) is regulated [31]. Apart from China, Japan and India, most countries follow one of these two systems, often with a time lag. Starting from 2017 to 2020, there is almost universal agreement on the allowed pollution levels for vehicles, with most pollutants being restricted to less than 5% of the original legislated limits. As an example, Fig. 1.5 shows the trend of NOx emission limits according to seven different legislations in the last decades.

Figure 1.5: Global light-duty NOx emissions limits showing at least a ten-fold decrease since the first limits and a global confluence at around 0.05 g/km in 2020 [31].

Focusing on the European legislation, EU emission standards for new light duty vehicles, including passenger cars and light commercial vehicles (LCV), were once specified in Directive 70/220/EEC [32], with a number of amendments adopted through 2004, defining the emission standards from Euro 1 to Euro 4. In 2007, this Directive was repealed and replaced by Regulation 715/2007 [33], defining the Euro 5 and Euro 6 standards. Tab. 1.1 summarizes the emission limits for the regulated pollutants from Euro 1 to Euro 6 standards. In general, EU regulations introduced different emission limits for compression ignition (CI) and spark ignition (SI) vehicles. More in detail, CI engines have more stringent CO standards but are allowed higher NOx emissions, while SI vehicles were exempted from PM standards through the Euro 4 stage. Then Euro 5 and Euro 6 regulations introduced PM mass emission limits, equal to those for diesels, for SI vehicles with direct injection engines. In addition to the reduced emission thresholds, especially regarding NOx, among the main changes introduced by the Euro 6 there is the PN limitation introduction, together with the introduction of the Worldwide harmonized Light vehicles Test Cycle (WLTC) chassis dynamometer procedure, which replaced the earlier NEDC test [14]. Since the Euro 6d-TEMP stage, in addition to laboratory testing, vehicles must be tested while operating on the road over their normal driving patterns, conditions, and payloads [16]. This type-approval test, identified as Real Driving Emission (RDE), is performed during vehicle operation using a portable

emissions monitoring system (PEMS). The test must last from 90 to 120 minutes and the route must include three segments in this order: urban (< 60 km/h), rural (60-90 km/h) and motorway (> 90 km/h). Each segment must cover one third of the total time share and a distance of at least 16 km. During RDE tests, NOx emissions must be measured on all Euro 6 passenger cars and LCVs, while PN emissions must be measured on all Euro 6 vehicles which have a PN limit set. CO emissions have to be measured and recorded as well, even if no not-to-exceed (NTE) limits are imposed for this pollutant [34]. More in detail, NOx and PN RDE emission limits are defined by multiplying the respective Euro 6 emission limit by a conformity factor (CF) for a given emission:

$$\text{NTE}_{\text{pollutant}} = \text{CF}_{\text{pollutant}} \times \text{EURO6} \tag{1.7}$$

So a conformity factor consists in the cap by which the Euro 6 emission limit would be allowed to be exceeded in real world driving. This margin of error exists mainly because PEMS equipment may suffer of significant inaccuracies, so that RDE tests will not show the same level of repeatable accuracy as a laboratory test like the WLTP. The conformity factors adopted by the Euro 6 standard are listed in Tab. 1.2.

Table 1.1: Emission standards for SI (Gasoline) and CI (Diesel) vehicles according to EU regulation.

Stage	Date	CO	HC	HC+NOx	NOx	PM	PN
		[g/km]					[#/km]
Spark Ignition (Gasoline)							
Euro 1	1992/07	2.72	-	0.97	-	-	-
Euro 2	1996/01	2.2	-	0.5	-	-	-
Euro 3	2000/01	2.3	0.20	-	0.15	-	-
Euro 4	2005/01	1.0	0.10	-	0.08	-	-
Euro 5	2009/09	1.0	0.10	-	0.06	0.005[a]	-
Euro 6	2014/09	1.0	0.10	-	0.06	0.005[a]	6.0×10^{11}[a]
Compression Ignition (Diesel)							
Euro 1	1992/07	2.72	-	0.97	-	0.14	-
Euro 2, IDI	1996/01	1.0	-	0.70	-	0.08	-
Euro 2, DI	1996/01	1.0	-	0.90	-	0.10	-
Euro 3	2000/01	0.64	-	0.56	0.50	0.05	-
Euro 4	2005/01	0.50	-	0.30	0.25	0.025	-
Euro 5a	2009/09	0.50	-	0.23	0.18	0.005	-
Euro 5b	2011/09	0.50	-	0.23	0.18	0.005	6.0×10^{11}
Euro 6	2014/09	0.50	-	0.17	0.08	0.005	6.0×10^{11}

[a] Applicable only to vehicles equipped with DI engines

Table 1.2: RDE conformity factors for different Euro 6 stages.

Stage	NOx CF	PN CF	CO
Euro 6d-TEMP	2.1	1.5	Measured and
Euro 6d	1.43	1.5	recorded,
Euro 6e	1.10	1.34	no limits

1.2.3   Euro 7: introduction and overview of the new emission standard

Aims and proposal   In November 2022 the European Commission presented the official Euro 7 proposal [10], with the aim of reducing air pollution from new vehicles sold in the EU to meet the stricter air quality standards of the European Green Deal proposed in October 2022 [35], while keeping vehicles affordable for consumers and promoting Europe's competitiveness. The idea behind the new Euro 7 emission standard was to replace and simplify previously separate emission rules for cars and vans, bringing emission limits for all vehicles under a single set of fuel- and technology-neutral rules. The final target was to ensure that cars, vans, lorries and buses were significantly cleaner, complying to much stricter emission thresholds under real driving conditions that more accurately represent urban environments, where air pollution is most severe. Furthermore, the new standards would have extended the compliance period beyond that of current regulations, also tackling emissions from tailpipe as well as from brakes and tyres for the first time. More in detail, according to [36], the first draft of the regulation proposal would have required that the emission limits, besides being much lower than Euro 6 ones (up to 90% NOx reduction was required for the most demanding scenario), were enforced in nearly any possible driving condition, with no more speed, acceleration and road grade limitations during the homologation procedure; moreover, a wider range of ambient temperatures and altitudes was considered as well. As evidenced in [37], all these requirements would have forced car manufacturers to develop and adopt extremely advanced technologies to fulfil such demanding emission standards in a very short time, leading to increased and nearly unaffordable development and manufacturing costs. After further evaluations from the European Commission supported by CLOVE (Consortium for ultra LOw Vehicle Emissions), the emission thresholds became much less strict in the final version of the regulation proposal published in 2022, set equal to the lowest values already observed in Euro 6, but applied to all cars and vans regardless of fuel type and technologies.

Legislation introduction   In May 2024, regulation (EU) 2024/1257 [17], namely the base Euro 7 legislation, came into force, superseding the Euro 6 regulation (EC) 715/2007 for passenger cars and vans and the Euro VI regulation for trucks and buses. As shown in Fig. 1.6, Euro 7 will first apply to the type-approval of new Light-Duty (LD) vehicle models and their brake systems more than two years after the regulation came into force. One year later, all newly registered vehicles will need to comply with the new rules. For Heavy-Duty (HD) vehicles, Euro 7 will apply to new vehicle models four years after the entry into force and after five years to all new vehicles [38]. Despite being a single regulation for both vehicle categories, it defines different requirements and limits for LD and HD vehicles. However, according to the scope of this dissertation, only LD regulation will be further explored.

Tailpipe emissions   Euro 7 tailpipe emission limits for LD vehicles are basically identical to the Euro 6 limits. Contrary to what stated in the Euro 7 proposal, different limits continue to apply for

November 2022
EU Commission proposes new Euro 7 standards

May 2024
Base legislation comes into force

May 2025
Last LD implementation acts are finished

November 2026
New type approvals

November 2027
All vehicle registrations

2022 — 2023 — 2024 — 2025 — 2026 — 2027 — 2028 — 2029

Final appendixes likely published from May 2025; important rules published earlier

November 2026
Last HD implementation acts are finished

May 2028
New type approvals

May 2029
All vehicle registrations

Figure 1.6: Euro 7 timeline: key steps in the introduction and implementation of the new emission standard.

CI and SI engine vehicles as well as for cars and vans. The main differences introduced by Euro 7 compared to Euro 6 in terms of tailpipe pollutant emissions are:

- SI vehicles, with both direct and indirect injection, are required to comply with PN and PM limits, while under Euro 6 only direct injection engines were subjected to particulate limits;

- PN10 count towards the particle number limit, while the Euro 6 limit was only considering particles as small as 23 micrometers (PM23).

Brake and tire emissions   In addition to tailpipe emissions, Euro 7 introduces emission limits for particle emissions from brakes and tires, which were not addressed by Euro 6 legislation. Brake particle emissions are generated from abrasion of brake pads and discs and their amount is not negligible, being of the same order of magnitude as tailpipe PM emissions. They can be reduced by optimizing brake system components, introducing brake dust suction devices, or by using regenerative braking in the case of electrified vehicles. This is the reason why a lower brake PM limit of 3 mg/km has been set for LD battery electric vehicles compared to other powertrain types, for which a higher limit of 7 mg/km is provided. On the other hand, tire particle emissions stem from tires abrasion. The test procedure and limits are under development at the United Nations Economic Commission for Europe (UNECE) and will amend the Euro 7 regulation [38]. In the first stage, Euro 7 applies to new tire models that are homologated for the first time, meaning that only tires with abrasion rates below the Euro 7 limits will receive type approval. One year later, new vehicles put on the market must be equipped with Euro 7 type-approved tires, and another year later, all tires put on the market must comply with Euro 7 requirements.

Vehicle lifetime   As for Euro 6, vehicles must comply with emissions standards for a specified vehicle lifetime. During this period, emission compliance can be verified through In-Service Conformity (ISC) and Market Surveillance (MaS) testing. Compared to Euro 6, Euro 7 extends vehicle useful life and introduces an "additional lifetime" beyond the "main lifetime", during which gaseous emissions are allowed to be 20% higher than Euro 7 limits. The lifetime can be expressed

11

in terms of calendar age in years and vehicle mileage in kilometers, as shown in Fig. 1.7a, and it is exceeded when one of the two thresholds is passed.

Battery durability   For the first time, Euro 7 sets durability requirements for high-voltage batteries of LD Battery Electric Vehicles (BEV) and Plug-in Hybrid Electric Vehicles (PHEV). As shown in Fig. 1.7b, after 5 years or 100000 km, whichever is reached sooner, batteries of BEV and PHEV passenger cars must retain an energy storage capability of at least 80%. The minimum energy storage capacity becomes 72% after 8 years or 160000 km. For LCVs, the durability thresholds are set at 75% and 67% for the same usage periods.

(a)   (b)
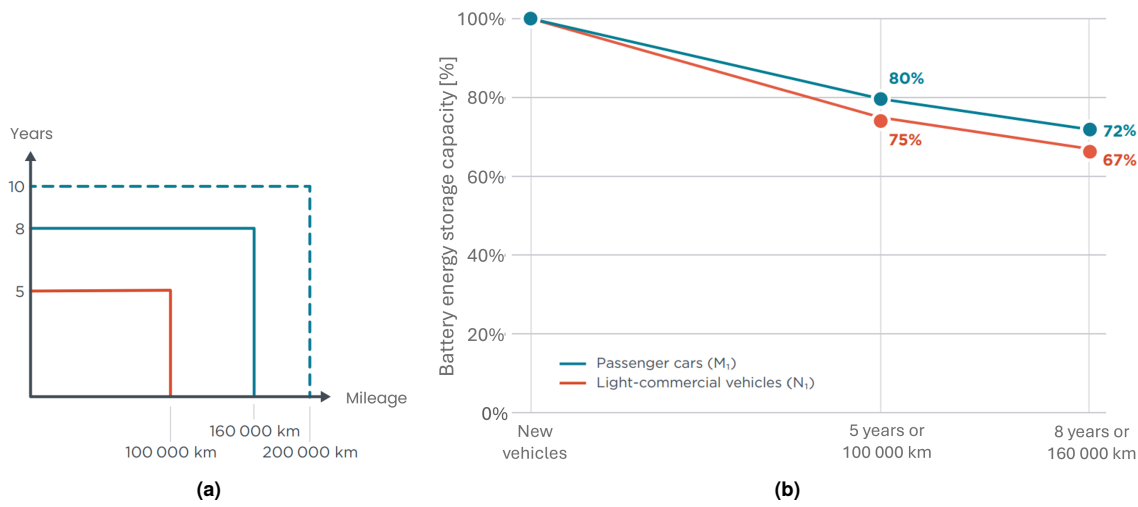
Figure 1.7: Euro 7 lifetime requirements for LD vehicles (passenger cars and LCVs) compared to Euro 6 (a); Euro 7 battery durability requirements for LD BEVs and PHEVs (b)[38]

On-board monitoring   On-Board Fuel and Energy Consumption Monitoring (OBFCM) devices were progressively implemented in the EU for all new cars and vans with internal combustion engines beginning in 2020. The Euro 7 standard will expand the application of OBFCM to include all vehicle categories and powertrain types. While OBFCM data from Euro 6 vehicles were accessible only through the vehicle's diagnostic interface, Euro 7 vehicles will be required to display this information to their users, with the possibility to transmit the data Over-The-Air (OTA). Besides OBFCM, On-Board emission Monitoring, simply referred to as OBM, is introduced by Euro 7 as a new compliance verification element to monitor the levels of NOx and PM emissions and eventually triggering a driving warning system to ensure an early intervention. OBM will be discussed more deeply in section Section 1.3.2.

- Same limits set for **EURO6** LD vehicles
- PN_{10} emission replacing PN_{23}

Revised emission limits

- Continuous **emission on-board monitoring** with focus on NO_x and PN
- Data available for **all trips**

Introduction of OBM

EURO7

- **Main lifetime**: 160000 km / 8 years
- **Additional lifetime**: 200000 km / 10 years with increased limits

Lifetime emission compliance

- PM emission from **brakes and tires** considered
- **Battery minimum SoH** at 160000 km / 8 years set to 72%

Brakes, tires and battery

Figure 1.8: Main updates/novelties introduced by the Euro 7 regulation

1.3 On-board diagnostics and emission monitoring

1.3.1 On-Board Diagnostic: European standards

For the sake of brevity, only European On-Board Diagnostics (EOBD) standards will be discussed in this section, tracing their evolution within European regulations. OBD as it is currently intended, was firstly introduced in Europe with the Euro 3 regulation [39, 40] in 2000 for gasoline engines and four year later for diesel engines. Its role is to identify malfunctions and deterioration of monitored components that cause emissions to exceed set thresholds [41]. To do that, OBD uses information from on-board sensors to directly and indirectly monitor the performance of emission controls. The driver is notified upon a malfunction detection by means of a warning light appearing on the vehicle instrument panel. This way it is possible to mitigate risks and enforce stricter safety policies before the situation becomes critical, enhancing safety and potentially saving money by means of an early diagnosis. Moreover, when a malfunction is detected, information about the malfunctioning component is stored and technicians can download the information, communicated in a standardized format (SAE and ISO standards), with a scan tool to support vehicle repair. In this regard, the standard ISO 15031-3 [42], specifies the type of OBD diagnostic connector used in communication between vehicle and external test equipment for emissions-related diagnostics, together with the electrical signalling protocols and the messaging format. It also details the vehicle parameters to be monitored and the encoding methods for transmitting and storing the associated data. The communication protocol used for the OBD is the CAN (Controller Area Network) that can communicate real-time with the Body Control Module (BCM) to which all the devices of a vehicle are connected [43]. The final aim of this standardization is to promote the use of a common diagnostic connector and standardized messaging throughout the automotive industry allowing for easier access to the OBD data.
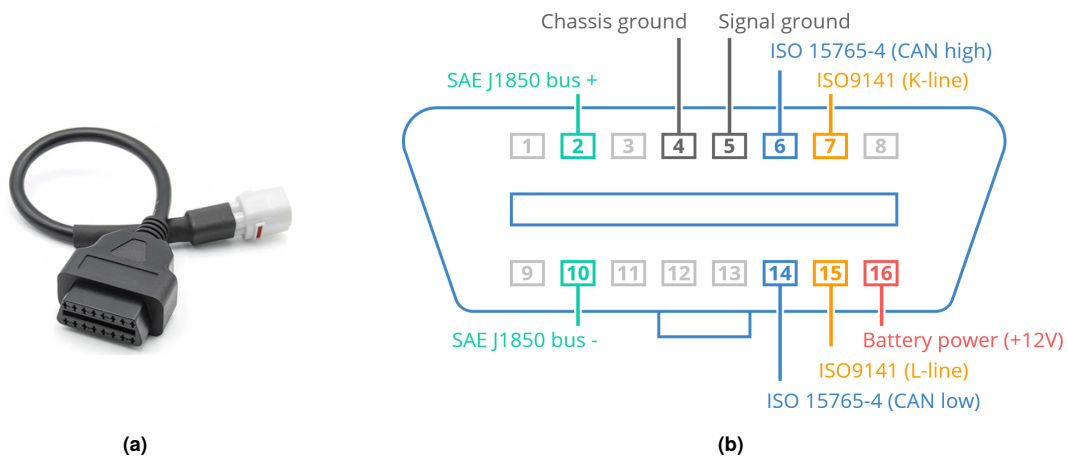
Figure 1.9: OBD connector (a) and standard OBD port pinout (b) [44]. Selected pins use SAE/ISO-defined standard protocols; where not specified, pins are left to the manufacturer's discretion.

Malfunctions are reported if they cause emissions to exceed specific thresholds set by the regulation. As shown in Tab. 1.3, no updates on OBD emission thresholds were introduced by Euro 4 compared to Euro 3, while much lower thresholds have been set by Euro 5, introducing PM threshold also for gasoline DI engines. Accordingly with the updated pollutant emission standards, OBD thresholds were further lowered with the introduction of the first and second stage of Euro 6 OBD standards, especially for NOx and PM. Moreover, as summarized in Tab. 1.4, the initial monitoring areas were significantly expanded from EURO 3-4 to EURO 5, with a particular focus on Exhaust Gas Recirculation (EGR) and NOx abatement systems. Starting from the last stage of EURO 5, called EURO 5b+, additional requirements in terms of In-Use Performance Ratio (IUPR) monitoring have been introduced. IUPR indicates how often a specific monitoring function is running relative to the overall vehicle operation time, and it can be calculated using the following expression:

Table 1.3: Pollutant emission EOBD thresholds according to Euro 3-4 and Euro 5 regulation

EOBD thresholds	Date	CO		HC		NOx		PM	
[g/km]		SI	CI	SI	CI	SI	CI	SI	CI
Euro 3-4	2000/01	3.20	3.20	0.40	0.40	0.60	1.20	-	0.18
Euro 5	2009/09	1.90	1.90	0.25	0.32	0.30	0.54	0.050[a]	0.050
Euro 6-1	2014/09	1.90	1.75	0.17	0.29	0.15	0.18	0.025	0.025
Euro 6-2	2017/09	1.90	1.75	0.17	0.29	0.09	0.14	0.012	0.012

[a] Applicable only to vehicles equipped with DI engines

$$IUPR = \frac{N_m}{N_c} \tag{1.8}$$

where:

- N_m is the number of times a specific monitoring function has operated and a malfunction could have been detected;

- N_c is the number of times the vehicle has operated under conditions suitable for monitoring.

Tab. 1.5 shows the IUPR values from Euro 5b+ until the last stage of Euro 6. An increasing trend of IUPR is observed according to Euro 6 regulation for both SI and CI engines compared to when it was firstly introduced, meaning that a more frequent monitoring is requested for all the monitored systems listed in the table.

1.3.2 On-Board Monitoring: overview and challenges

Concept and aims   As already mentioned in Section 1.2.3, on-board monitoring of pollutant emissions, or briefly OBM, is one of the novel elements introduced by the Euro 7 regulation, with

Table 1.4: Monitored components and areas according to Euro 3-4 and additional areas introduced by Euro 5 EOBD

Euro 3-4 monitor areas	SI	CI
Catalyst converter (for gasoline HC only)	x	x
Engine misfire	x	
Oxygen sensor deterioration (front sensor)	x	
Particulate trap		x
Fuel injection system	x	x
Circuit continuity of all emission related powertrain components	x	x
Any other emissions related components or systems (air flow, EGR, etc) if malfunction causes increase above thresholds	x	x
Euro 5 additional monitor areas	SI	CI
EGR system efficiency monitoring	x	x
EGR flow and cooler monitoring	x	x
Catalyst against HC and NOx	x	x
NOx aftertreatment device with or without reagent efficiency monitoring	x	x
All O_2 Sensors to monitor catalyst (in addition to front sensor)	x	x
PM monitoring	x	x

Table 1.5: IUPR values introduced by EURO 5b+ and EURO 6

IUPR	Euro 5b+	Euro 6b		Euro 6c/d	
System/component	SI/CI	SI	CI	SI	CI
Catalyst	0.1	0.336	0.336	0.336	0.336
EGR system	0.1	0.336	0.336	0.336	0.336
O_2 sensors	0.1	0.336	0.336	0.336	0.336
NOx sensors	0.1	0.336	0.336	0.336	0.336
NOx aftertreatment system	0.1	0.1	0.336	0.336	0.336
Secondary air	0.1	0.26	n.a.	0.26	n.a.
Cold start diagnostics[a]	-	0.26	0.26	0.26	0.26
VVT system	-	0.336	0.26	0.26	0.26
Boost pressure control[b]	0.1[c]	0.336	0.336	-	0.336
EVAP system	0.1	0.52	n.a.	0.52	n.a.
Diesel oxidation catalyst	0.1	0.336	0.336[d]	0.336	0.336[d]
Particulate filter	0.1[c]	0.336	0.336[d]	-	0.336[d]

[a] Incremented only after cold start < 35°C coolant
[b] Boost control active > 15 sec
[c] Only for CI engines
[d] Additional monitoring requirement of total failure or removal

17

the primary target of ensuring that all vehicles on the road are fulfilling the emission standards over vehicle life time. According to the definition included in the legislation [17],"'on-board monitoring system' or 'OBM system' means a system on board a vehicle that is capable of monitoring exhaust emissions, detecting exhaust emission exceedances and capable of communicating that information together with the state-of-health information off-board". Based on this definition, OBM has two primary purposes: the first consists in detecting whether emissions exceed allowed thresholds, at the individual vehicle level; on the other hand, the second function ensures compliance with emission standards at the fleet level, transmitting relevant data to authorities for comparison with regulatory limits. Therefore, the effective implementation of the OBM concept relies on two key factors: the accuracy and reliability of available emission sensors and models in providing meaningful data on vehicle emissions, and the capability to transmit information Over-The-Air (OTA) for market surveillance, taking advantage of the growing vehicular connectivity and cloud-based analysis [45].

OBM is essentially born to address the shortcomings of current OBD systems [46]. While OBD is effective in monitoring failures of systems and components, it does not provide a comprehensive assessment of total vehicle emissions. Consequently, as pointed out by [47] and [22], large gaps between in-use and certified emissions can be observed, with a significant fraction of vehicles on the road emitting higher levels of air pollutants than allowed, potentially due to aged or lightly damaged EATS and engine components that are not identified by standard OBD. On the other hand, OBM is not just monitoring the functional status of a single, faulty component which may result in an exceedance of the given OBD emission limits, but directly and continuously monitors tailpipe emissions on real road trips during all-day vehicle operation, to accurately and promptly detect high-emitting vehicles [18]. However, it is important to clarify that OBM is not replacing OBD, but it is instead an integration to it, taking advantage of many shared hardware elements such as embedded sensors and the standard OBD port [45]. Moreover, based on its definition, OBM will only detect emission exceedance and not directly the root cause. Although this pinpointing is not explicitly required by the legislation, an extended OBD or a new system for identifying faulty components would allow quick and cost-effective repairs, in the full interest of vehicle manufacturers and users.

Euro 7 requirements   According to Euro 7 regulation, in its first implementation, OBM covers the on-board monitoring of NOx, NH_3 and PM emissions from LD vehicles. OBM systems will be technology-neutral, meaning that the same requirements apply to all vehicles of the same category equipped with an internal combustion engine, regardless of the powertrain (with limited exceptions such as vehicles used by emergency services). The following functions have to be performed by the OBM system by means of appropriate hardware and software installed in each vehicle:

- manage the Excess Emission Driver Warning System (EEDWS) for NOx, NH_3, and PM emissions, including driver inducement for repairs;

18

- provide time-resolved (at 1Hz frequency) NOx and NH$_3$ emissions data and auxiliary data for emissions testing via the OBD port (not stored or broadcast) (Tab. 1.6);

- evaluate distance-specific NOx and NH$_3$ emissions for each trip;

- store summaries and metadata of recent trips and a random sample of trips in non-volatile memory, accessible via the OBD port;

- periodically transmit these summaries and metadata for random trips to the manufacturer's servers using OTA data transmission.

Distance-specific emission data, evaluated at the end of each trip, are then processed according to the regulation for ISC tests and eventually corrected or excluded from the calculation if driving conditions fall outside defined boundaries. The ISC-processed distance-specific emissions for a trip are one of the key building blocks of OBM, as these results are used both to verify the accuracy of OBM and to determine the emissions conformity of vehicle types based on a large number of randomly selected trips from different vehicles [48]. Even if the first stage of Euro 7 regulation is focusing mainly on NOx and NH$_3$, PM emissions are still within OBM scope: the system must set an EEDWS monitoring status for PM and trigger inducement if the vehicle is unlikely to pass an ISC emissions test at less than 2.5 times the PM emission limit, as for NOx and NH$_3$.

At the end of each trip, the so called "lifetime" and "long-term" data values are calculated and updated as well. However, unlike distance-specific emission data, these data are not used to verify emissions or OBM system conformity, but to support the activities of MaS authorities and consumers information (e.g. for second-hand vehicles buyers). A list of the OBM lifetime values to be updated by the OBM system at the end of each trip is given in Tab. 1.7. While lifetime data refer to the average values calculated on the last trip, regardless to its length, long-term data represent the approximate vehicle emissions performance over the last 1000 km, in terms of NOx, NH3 and EEDWS states.

Table 1.6: Time-resolved OBM signals to be made available from the standard OBD port [48]

OBM signal	Unit
NOx emission rate	[mg/s]
NH$_3$ emission rate	[mg/s]
NOx concentration	[ppm]
NH$_3$ concentration	[ppm]
Vehicle speed	[m/s]
Exhaust mass flow	[g/s]
Data flags (vectors of binary values)	[-]

Table 1.7: Time-resolved OBM signals to be made available from the standard OBD port [48]

OBM lifetime data	Unit
Estimated lifetime emissions (unprocessed; one value per pollutant for NOx and NH$_3$)	[mg/km]
Estimated lifetime emissions (ISC-processed; one value per pollutant for NOx and NH$_3$)	[mg/km]
Share of lifetime distance driven with EEDWS status "normal" (one value per pollutant for NOx, NH$_3$ and PM)	[%]
Share of lifetime distance driven with EEDWS status "intermediate" (one value per pollutant for NOx, NH$_3$ and PM)	[%]
Share of lifetime distance driven with EEDWS status "error" (one value per pollutant for NOx, NH$_3$ and PM)	[%]
Share of lifetime distance driven with data flag active (one value per data flag)	[%]

Excess Emission Driver Warning System The EEDWS is designed and installed on EURO 7 vehicles to inform users about excess emissions and enforce repairs. The system monitors emissions of NOx, NH_3, and PM, setting a status for each single pollutant that can be updated at the end of each drive cycle. The EEDWS works like a "traffic" light system, linking a vehicle emissions status to its eligibility for ISC testing. More in detail, as summarized in Tab. 1.8, the monitoring status of each pollutant can be in one of the following:

- normal state, meaning that the vehicle emissions control systems are operating properly and the OBM system is confident in the accuracy of the measured emissions trip values. Vehicles in this state may be selected for ISC testing, subject to other selection criteria like modifications or maintenance status;

- intermediate state, meaning that the monitoring systems cannot conclusively assess the emissions control systems, resulting in increased uncertainty in OBM emissions evaluations. Vehicles in this state may still be selected for ISC testing, with possible special conditions like enhanced preconditioning before testing;

- error state, indicating a risk of consistently high emissions or failures in emissions control that require repair. Vehicles in this state must begin the inducement procedure, displaying a standard indication through the Malfunction Indicator Light (MIL). Moreover, the system may eventually prevent engine start if not repaired. Vehicles in this state are unfit for ISC emissions testing, though they may still be tested to verify correct inducement application without resulting in a fail for the ISC procedure.

According to the legislation, the rule for setting the error status is whether the vehicle, based on its own evaluation, can pass a regulated emissions ISC test at less than 2.5 times the applicable emission limit for a specific pollutant. This threshold is aligned with the "extreme outlier" definition in the ISC statistical procedure for Euro 6 vehicles [49]. When implementing the EEDWS, manufacturers may use any available data to determine vehicle monitoring status based on the expected tailpipe emissions or performance of emissions control systems.

Table 1.8: EEDWS monitoring status. Adapted from [50]

EEDWS status	Traffic light	Emission status	ISC status
Normal	(○○●)	< threshold	Fit for ISC
Intermediate	(○●○)	Indeterminate	Fit for ISC
Error	(●○○)	> threshold	Unfit for ISC, check for inducement triggering

Technical challenges   As extensively introduced, OBM will require continuous and accurate monitoring of selected pollutant species. According to the latest Euro 6 applications, the state of the art of on-board sensors is not mature enough to ensure robust and accurate measurements for all pollutant species meeting Euro 7 standards. Tab. 1.9 summarizes currently available on-board sensor technologies for the main engine pollutants [18]. Diesel NOx sensors are the only ones readily available on the market and routinely used for CI engines. Depending on installation, positioning and package constraints, the accuracy of current sensors can reach +/-10 ppm for NOx concentrations below 100 ppm and +/-15% for higher NOx concentrations, potentially enhanced down to +/- 7 ppm and +/-8% respectively for next generation sensors . These performance of NOx sensors may be degraded over time due to ageing and exposure to substances in the exhaust gas, such as hydrocarbons, soot, urea, or NH_3. This makes the estimation of tailpipe NOx very challenging during regeneration events of the particulate filter. Moreover, NOx sensors are cross-sensitive to NH_3 [21], requiring additional algorithms, sensors, or models to distinguish NH_3 contributions from the overall sensor acquisition and accurately estimate NOx concentration, presenting an extra technical challenge, especially in situations with concurrent NOx breakthrough and NH_3 slip [50]. Moreoveer, NOx sensors are designed for lean or stoichiometric exhaust, and measurements are generally not reliable under rich exhaust conditions. As a result, NOx sensors may need further development for use with gasoline engines. Besides suffering from the mentioned limitations affecting measurement accuracy, NOx sensors are characterized by a very long activation time, during which they are kept switched-off to avoid the risk of damage of the sensing element due to water condensation [22]. In cold start conditions, this phase, also known as the "dew point preparation phase", may take from a few minutes up to more than 20 minutes depending on the considered engine, sensor position, driving cycle, and ambient conditions [51]. As shown in [23], on a cold WLTC, a standard NOx sensor can be kept switched off for more than 50% of the cycle duration. Some newer generation NOx sensors are dew point free, which reduces the delay until the NOx sensor measurement is reliable. However, there is still a time delay to reach sensor light off before a valid measurement is provided, which becomes longer the further downstream of the engine the sensor is located. To face this problem, quick light-off control strategies have been developed, like the one presented in [52], which allows the sensor light off time to be halved. Despite these improvements, due to intrinsic sensors limitations,an accurate measurement of NOx emission during cold-start phases is not currently possible.

Strong technical limitations are observed for NH_3 and PM sensors as well. Besides being characterized by a significant activation time for dew point detection like NOx sensors, measurement capability of available NH_3 sensors is even more limited than NOx sensors and decreases significantly with ageing. Error on a new NH_3 sensor is approximately 25%, compared to the approximately 15% error on NOx sensors, and could reach 40% for aged sensors. Therefore, even if theoretically NH_3 sensors could be used to estimate NH_3 emissions and solve NOx sensors cross sensitivity, this is not realistic with the limited accuracy of such sensors [50]. A reliable measurement of PM rate is even more challenging. Commonly used resistive sensors are inherently not

capable of providing a readily available and continuous measurement: once the sensor accumulates enough soot to generate a current signal, the current increases almost exponentially, making these sensors unsuitable for reliably measuring soot at resolutions below DPF failure thresholds. Electrostatic sensors may offer improved capabilities, but they are not yet commercially available and still present technical challenges due to signal conditioning and post-processing [53].

In this context, sensor-based Euro 7-compliant emission monitoring is currently challenging, particularly during cold starts, which are among the most critical driving events in terms of emissions. Therefore, to face the request of a continuous monitoring of pollutant emissions, covering all the possible situations in which direct information from a physical sensor is not available or reliable enough, a model-based approach is needed in combination with on-board measurements. As already introduced in Section 1.1, to date, accurate and robust physical-based modelling of engine-out emissions is already possible in nominal conditions, including fast transient events and a multitude of boundary conditions [25–27]. However, even the most complex models can not comprehensively capture the effects of degraded or faulty components, which is crucial for OBM, requiring continuous emission monitoring in any possible operating condition, independently from the State of Health (SoH) of system components. A more detailed review of the state of the art of emission models, related limitations and consequent improvements needed to fulfil Euro 7 OBM requirements will be presented in Section 2.2.

Table 1.9: Overview of existing on-board emission sensors

Pollutant	SI engines	CI engines
NOx	Not in production	Available on the market Long activation time Tolerance for OBM borderline
NH3	Not in production	In production for off high-way engines Durability and tolerance issues
PM	Not in production	Available on the market Long cumulation time Tolerance for OBM borderline
PN	Not in production	Not in production
CO/HC	Not in production	Not in production

Chapter 2

Literature review

Chapter 1 depicted the new regulatory scenario introduced by Euro 7 legislation, highlighting the crucial role of OBM in ensuring that all vehicles on the road are fulfilling the emission standards over vehicle life time, identifying high-emitters due to degradation or failures of systems and components, or to the deliberate manipulation by the vehicle owners. Besides the unquestionable improvements in terms of in-service conformity verification, market surveillance and reduced time-to-repair enabled by an Euro 7-compliant OBM, demanding technological challenges have been introduced with the new regulation, forcing OEMs to find innovative approaches to fulfil the requirements, which must be at the same time effective, reliable, and both economically and technologically viable.

In this context, integrating Machine Learning (ML) approaches with standard OBD and conventional physics-based models for pinpointing damaged components and continuous emission monitoring is seen as a promising solution for OEMs. Thus, this research project focuses on investigating the potential of ML methods applied to on-board fault detection and identification of damaged component, in parallel with engine-out emission estimation in case emission-relevant faults occur. To understand the weaknesses and shortcomings of state-of-the-art fault detection and emission models in view of OBM applications, a comprehensive literature review on such topics is here presented. Firstly, an overview of possible approaches to fault diagnosis is given, with a deeper focus on data-driven fault detection. Then, the more relevant studies and applications of such methods to mechanical systems are presented as well. In parallel, an overview on both physics-based and data-driven emission models is given, focusing on relevant applications involving engine-out emission modelling. Finally, the already mentioned issue of non-nominal emission modelling will be discussed separately, providing references of already proposed methodologies to face the problem and highlighting the novel contribution of this research activity.

2.1 Fault detection and identification

2.1.1 Concept and methods

Fault diagnosis is a crucial aspect to increase safety, reliability, and efficiency of complex dynamic systems, avoiding as much as possible severe consequences on humans and environment, limiting repair costs, and shortening downtime periods which can cause inconveniences and significant economic losses. Therefore, many studies have been carried out, all referable to the subfield of control engineering known as Fault Detection and Isolation (FDI) [54], focused on developing monitoring systems that can recognize when a fault occurs and identify its type and location. The issue of FDI can be divided into two complementary steps [55]:

- detection, intended to identify the presence of potential faults within the system;

- diagnosis, aimed at determining the root causes of the detected fault (isolation), its type, and eventually its size and time profile (identification).

If only the first of the two steps is supplied, the procedure more accurately aligns with the concept of "anomaly detection". The target of this process is limited to finding patterns or instances in a dataset that deviate significantly from the expected or normal behaviour, generically classifying them as anomalies, with no further distinction between them.

FDI approaches can be divided into three macro-categories, namely model-based, knowledge-based and data-driven FDI [56]. A brief introduction to each of them is provided in the next paragraphs.

2.1.1.1 Model-based and knowledge-based approaches

Analytical model-based FDI relies on the knowledge of the physics of the process to be monitored, which is used to build a mathematical model of the system to determine if the system is failing. Consistency checks between physical measurements and model outputs are performed through the application of a selected decision rule on calculated residuals in order to detect a fault. Different methods can be used to generate residuals [55]:

- **state observer**, used to process inputs, measurements and models to reconstruct some observable states of the process. The errors between the model and process outputs are fed back to recalibrate the reconstructed states and the residuals generated from the reconstructed states are compared with fixed or adaptive thresholds. If such limits are exceeded, a fault is detected. Handling of modelling uncertainties and measurement noise are among the main critical aspects, and many efforts have been made to address these shortcomings [57];

- **parity space**, transforming the process model to parity relations which describe the inputs and outputs relationships independently of the states of the system. These relations, represented by algebraic equations or difference equations, are used to generate residuals to check

the consistency between the model and process outputs. In presence of a fault, they are no longer satisfied, thereby generating non-zero residuals as fault indication. The same issues highlighted for observer-based FDI must be addressed also for this approach [58];

- **parameter estimation**, based on the assumption that a fault can cause changes in some unmeasurable physical parameters of the system, and these changes will be further reflected in variations of some model parameters. When the calculated parameter set exceeds the space of allowed solutions, a fault is detected. This method requires an accurate parametric model and the solution of a complex non-linear optimisation problem, computationally demanding for online application [59];

- **Bond graph**, a graphical approach able to represent energy exchanges in mixed systems and that can be used to generate residuals. This methodology is usually implemented in dedicated software and integrates multiple types of information, leveraging its causal and structural characteristics to directly derive a set of fault indicators [60].

As model-based FDI, deep-knowledge-based FDI relies on the fundamental knowledge of the underlying physics of the process, but, differently from model-based methods, it employs a non-analytical type of knowledge. In fact, in many situations it could be difficult to obtain detailed mathematical relationships describing the monitored process; however different forms of process knowledge could be available, such as qualitative behaviour of the system, if-then rules, probabilistic conditional relationships, causal effect relations and process connectivity. Typical techniques include [61]:

- **qualitative simulation**, providing an abstract description of the behaviour of a system, ignoring much of its quantitative details. It is easy to use, but fundamentally limited to algebraic equations and hence to describe steady-state behaviour;

- **expert system**, based on the fact that human operators accumulate theoretical and practical expertise over years, allowing them to reason out the cause of a potential fault and suggest corresponding corrective actions. The expert system is an organized knowledge system designed to automate this process, relying on four essential components, namely knowledge base, inference machine, knowledge management and user interface [62];

- **Signed Directed Graphs** (SDGs), using nodes and directed arcs to represent the events or variables of the system, and the causal relationships among them. A sign is attached to each node to qualitatively represent its status, while the directed arcs among the nodes are drawn from the cause to the effect nodes. Differently from Bond graphs, that are built based on mathematical models relying on algebraic and differential equations, SDGs are obtained from expert knowledge on process connectivity and causality. Despite their intrinsic limitations, they can be used in combination with other approaches to improve diagnosis accuracy [63];

- **fault tree**, a graphical and hierarchical model that propagates primal events, namely fault origins, to the top event, namely a hazard. The root cause of an undesired event can be deduced by a top down analysis, whilst the consequences of a basic event can be inferred in a down-to-up manner, by means of different logic operators. Establishing the correct logical links between all the involved events is the critical aspect, being a time-consuming and error-prone process;

- **bayesian network**, employing a graphical model which consists of qualitative and quantitative parts. As for SGDs, nodes and their links are used to represent a set of variables and their causal relationships. However, in this case conditional probabilities are assigned to each node for a better knowledge representation under uncertainty. The quantification of this probabilities is the critical requirement of this method [64].

A key strength of analytical model-based methods lies their foundation in comprehensive system or process knowledge and their utilization of empirically validated correlations. However, physics-based models are inherently limited by their inability to handle infinite complexity, leading to inevitable simplifications to the advantage of a reasonable computational effort. Moreover, a deep understanding of system physical, mechanical and electrical data flow is required to develop and refine an accurate model, making the process time-consuming and costly. As the complexity of the monitored system increases, inevitable modelling errors and uncertainties, which cannot be easily quantified, may results into false alarms from the monitoring system [55]. Similarly, deep knowledge-based fault detection, although it can be highly effective in certain scenarios, suffers from limitations such as impossibility to encode all relevant events into the model as complexity of the monitored system increases, dependency on expert knowledge of the monitored system, complexity in development and maintenance, and therefore high related costs. Moreover, they are characterized by limited flexibility in their application, as they rely heavily on predefined rules or models that might not generalize well to different scenarios or systems, thus requiring frequent model updates and integrations. These factors make this approach less suitable for dynamic, complex, or rapidly changing environments where more adaptive approaches might be preferable.

In this context, data-driven approaches offer a valid alternative to model-based and knowledge-based methods. On the negative side, they necessitate vast amounts of data to achieve remarkable results on the entire application range, due to their limited capacity to extrapolate beyond the data they were trained on and neglecting basic physical principles that determine the behaviour of real-world systems. However, on the positive side, these approaches can provide exceptional flexibility and accuracy in modelling complex dynamic systems with limited effort, making them particularly effective for solving high-dimensional problems that would be difficult to analyse analytically or schematically.

2.1.1.2 Data driven approaches

Nowadays, due to the increased complexity and automation degree of the monitored systems, the development of model-based diagnostic approaches capable of considering discrete and continuous system dynamics as well as the interactions between the large number of interconnected components included in the system, has become a challenging task [65]. Moreover, enhanced by the advancements involving Internet of Things, wireless communications, and development of mobile devices, the amount of available data that could be used for fault diagnosis purposes has grown exponentially. In this context, as evidenced by literature [66], data-driven FDI has gained significant popularity, proving to be a promising, feasible, and cost-effective tool to detect and classify the occurrence of faults in complex dynamic systems.

In this regard, data-driven approaches applied to fault diagnosis can be broadly categorized into the following types:

- **signal processing**, involving time-frequency analysis or wavelet transforms of signals to detect anomalies. They are particularly effective in applications where faults manifest as changes in the signals frequency content or other signal characteristics;

- **statistical methods**, using statistical analysis to detect deviations from normal system behaviour. Common techniques include control charts, hypothesis testing, and Principal Component Analysis (PCA) [67]. They work by analyzing historical data to establish a baseline, and then monitoring real-time data to eventually find anomalies;

- **machine learning methods**, including several algorithms like Support Vector Machines (SVM), k-Nearest Neighbours (k-NN), Decision Trees (DT), and ensemble methods, like Bagged Trees or Boosted Trees, to classify data as normal or faulty. Differently from previous categories, these methods can learn from labeled datasets not only to detect, but also to identify known faults by recognizing specific patterns in the data [68];

- deep learning methods, employing different types of Neural Networks (NNs), such as Artificial Neural Networks (ANNs) or more complex Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which can automatically extract features and detect faults from raw data, often providing high accuracy but requiring large datasets for training;

- hybrid methods, combining multiple data-driven techniques, such as statistical analysis or signal processing with machine learning or deep learning methods, to enhance fault detection and isolation capabilities. The combination of model-based or knowledge-based methods with data-driven techniques is also possible, as shown in [63] where PCA is combined with knowledge-based SDGs with the aim of leveraging the strengths of each method to improve accuracy and robustness.

Another possible classification of data-driven approaches applied to fault detection can be made according to whether responses to certain input datasets are included in the training or not:

- **unsupervised learning methods**, including PCA, Independent Component Analysis (ICA), K-means and other clustering methods [69], aimed to automatically draw pattern characteristics of a dataset without any knowledge about the individuals belonging to different classes or specific guidance on expected outputs;

- **supervised learning methods**, including ML and DL classification algorithms like SVM, DT, k-NN, and ANN-based classifiers, using historical data to construct a learning model, which is used for the fault detection and diagnosis of the new data. More in detail, they are capable of inferring a mapping function between a pair of input object and target value based on a fully labelled training dataset, meaning that each input has a corresponding known output;

- **semi-supervised-learning methods**, combining aspects of supervised and unsupervised learning to improve model accuracy using a small amount of labelled data to draw more information about the pattern characteristics of the dataset [68].

Unsupervised learning methods have the significant advantage of not requiring labelled data, which is particularly beneficial where data labelling is time-consuming, expensive, or infeasible due to rare and unpredictable faults. However, they are generally binary classifiers, distinguishing only between nominal and faulty conditions, thus limited to anomaly detection tasks. On the other hand, despite the challenge of requiring well-labelled data, supervised learning approaches can achieve high diagnostic accuracy, being uniquely effective when the objective is not only to detect faults but also to pinpoint the specific root cause and distinguish between different fault types. As reviewed in [70], supervised learning is the most common approach used for fault diagnosis and it has been already explored within several different application fields, including power plants [71], photovoltaic systems [72], rotating machinery [73, 74] (e.g. bearings [75], automotive transmissions [76], etc.), industrial machinery [77], automotive electrical components (e.g. inverters [78], electric motors [79], batteries [80]), unmanned aerial systems [81], nuclear power plants [82], aircraft turbines [83] and shipboard systems [84].

Focusing on the automotive sector, with the introduction of the new Euro 7 standard, widely discussed in 1.2.3, vehicle emission compliance must be ensured throughout the entire vehicle lifetime and confirmed by real-world testing and on-board monitoring of defined pollutants. Thus, fast and reliable detection and identification of emission-relevant faults of engine sensors and components plays an important role in ensuring the nominal expected emission behaviour of the monitored vehicle. In this context, the first part of the present research activity aims to study the potential of data-driven approaches in detecting emission-relevant engine faults, thereby assisting standard OBD systems in pinpointing faulty or damaged components and addressing one of the primary challenges introduced by Euro 7 OBM requirements. Focusing on the target of this dissertation, a comprehensive bibliography research has been carried out to give a deeper overview on the state of the art of data-driven fault detection methods applied to automotive engines. The most relevant results are discussed in the following section, emphasizing the most interesting aspects, valuable innovations, and limitations of each work to establish a strong foundation for

this research activity and identify the gaps that need to be addressed. A general overview of the cited previous works on this topic is given in Tab. 2.1.

2.1.2   Data-driven FDI for automotive engines

In the automotive industry, most vehicle components are continuously monitored to ensure proper functioning. More in detail, almost every potential source of excessive emissions is under control, including catalyst efficiency, misfire detection, evaporative systems, and fuel systems. This comprehensive monitoring must be designed to prevent emissions issues and reduce environmental impact, while protecting the vehicle components until they can be serviced. To achieve this, different methods are employed for condition monitoring, including data-driven approaches, that have gained increasing importance in recent times as part of broader efforts to enhance emissions control and improve overall vehicle performance. In this context, a selection of related works from the field of fault and anomaly detection applied to automotive engines is presented in this section.

In [85], a multi-layer ANN model for engine failure classification using features extracted from sound intensity analysis is presented. While the proposed approach can successfully detect different fault classes, including malfunctions of injector nozzles, electric motor, throttle orientation potentiometer, or ECU-sensors communication, it requires a dedicated set-up for the acquisition of high quality acoustic data, making it unsuitable for on-board vehicle implementation. In [86], an alternative method is proposed where engine vibration data is used as the input for an ANN-based classification model. This fault diagnostic system effectively detects known engine faults with varying severity levels. However, testing must be conducted in a semi-anechoic chamber to prevent external noise from affecting the system vibration response, which makes it unfit for OBM purposes. Similarly, in [87], the authors propose using vibration signals to identify various faults, including malfunctions of manifold absolute pressure sensor, knock sensor, and misfire. Their multi-layered ANN is capable of accurately classifying different fault scenarios, having been trained on both individual faults and their combinations. A significant drawback of this approach lies in the need to acquire all potential combinations of faults and provide them as input to the models as distinct classes. The number of combinations increases exponentially as the number of individual faults increases, thus requiring a large data collection.

A data-driven approach to detect knock phenomena using ECU signals instead of vibration data is presented in [88]. In this case, different unsupervised ML methods, including Autoencoder, Support Vector Machines (SVM), and Isolated Forest, have been investigated to perform anomaly detection through binary classification. The proposed anomaly detection method successfully identifies engine knock issues with good precision, demonstrating the potential of already available ECU data for improving engine diagnostics. However, this method supports only binary classification, which is not sufficient if the goal is to identify and distinguish between various engine faults. Similarly, the work described in [89] presents a Deep Neural Network (DNN)-based

31

pre-ignition detection starting from all available ECU signals, showcasing the potential of deep learning for advanced diagnostics. Nevertheless, the same lack is present in terms of handling multiple types of faults. Analogously, in [90] the authors conduct a comparative analysis of various ML algorithms, demonstrating the effectiveness of Deep Neural Networks in detecting anomalies through CAN messages. For this purpose, both synthetic and real CAN data have been processed to identify anomalies for a wheel speed control system, even if no classification of different types of faults is supported, as for [88] and [89].

In [91], a Probabilistic Neural Network (PNN) is designed to detect different engine faults, including failures of different sensors, like oxygen, water temperature, boost pressure, camshaft position and throttle position sensors, as well as injection and ignition failures, based on the information extracted from exhaust gases, namely HC, CO, NOx, CO_2, and O_2 concentrations. High average accuracy is reached with low computational effort, but measurements obtained with a dedicated exhaust gas analyser are required, precluding the possibility of on-board implementation. The work in [92] presents a fault diagnosis system based on Sparse Bayesian Extreme Learning Machines (SBELM) which is able to detect both single and simultaneous engine faults while the system is trained by single-fault patterns only, solving the main drawback of the approach proposed by [87]. However, an inductive pickup clamp, five gas analysers, and a microphone are used to acquire the air-to-fuel ratio, ignition, and acoustic wave patterns respectively, which makes on-board implementation of such a system not feasible. Moreover, an extensive evaluation of the computational cost is not provided, thus requiring additional considerations on computational efficiency and model complexity in view of real-world applications.

In [93], the authors propose a fuel consumption classification system for heavy-duty vehicles based on bagged decision trees. The aim is to detect high fuel consumption patterns which are independent of vehicle load and slope, to identify anomalous driving behaviours or system failures, based on engine-related and vehicle dynamics-related features. As for [88], the proposed approach does not address the root cause of the anomaly, and does not even distinguish between a fault and an unusual driver behaviour. An alternative data-driven approach based on two-class and one-class classifiers trained and validated on a dataset of vehicle recordings is proposed in [94]. As for the last mentioned research, the aim is to classify data as either normal or anomalous, regardless of the cause of the anomaly. Although the developed anomaly detector yields good results for both known and unknown faults on different driving scenarios, it was designed for off-line analysis of on-road trial recordings, with no further investigation for real-time application.

In [95], an innovative hybrid methodology utilizing both model-based residuals generation together with a Long Short-Term Memory (LSTM) neural network in proposed to predict two selected engine faults, namely manifold pressure sensor failure and air mass leakage into the intake manifold. However, along with the lack of evidence of real-time implementability, only one residual signal is used, namely intake manifold pressure, and a very limited number of faults is considered if compared to other reviewed researches. Similarly, the work in [96] proposes a hybrid approach that integrates data-driven and model-based techniques for fault detection

32

and isolation, aiming to leverage the strengths of both methodologies to enhance diagnostic accuracy, at the cost of increased complexity in system design and implementation. In [97], the authors propose a probabilistic fault classification algorithm for classification of real residual data computed from sensor data and model analysis of an internal combustion engine. One-class support vector machines for fault class modelling combined with Bayesian filtering for time-series analysis have been implemented, ranking different fault classes and identifying unknown faults not present in the training dataset. The idea is to perform fault diagnosis in two steps, starting with a fault detection step followed by a fault classification step when a fault is detected. Even if the effectiveness of the proposed approach has been demonstrated, real-time capability of the proposed approach still needs to be assessed. Moreover, a high-precision mathematical model is required, as for [95] and [96], which in turn increases the cost and complexity.

In [98], an artificial neural network bank is used to detect and replace damaged engine sensors with estimated signals. More in detail, the system compares ANN-generated signals to engine sensor readings to isolate and replace faulty sensors, including intake manifold pressure sensor, MAF sensor, throttle position sensor and their sequential combinations. Despite showing promising results on both single and multiple faults detection, the approach is only applicable to sensors fault detection and requires one calibrated model for each monitored sensor. The work described in [99] focuses on identifying single and simultaneous faults in automotive software systems using an ensemble approach that combines LSTM networks and Random Forest classifiers, addressing challenges posed by noisy and imbalanced data in fault diagnosis. As an additional contribution compared to other reviewed studies, real-time validation using a Hardware-in-the-Loop (HiL) platform is provided. However, as for [98], the approach is designed to handle only sensor faults, such as drift, delay, or gain. Moreover, as observed for [87], the training dataset includes all single faults and their possible combinations, with the resulting drawbacks.

In [100], the authors present an FDI scheme using ANNs for isolating damaged injectors in Electronic Fuel Injection (EFI) systems, based on time and frequency analysis of the measured rail pressure signal. Among the remarkable results, the proposed method exhibits high classification accuracy in offline testing, which has been then validated in real-time. Nonetheless, the objective of this research is to identify faulty injectors without addressing the root cause of the fault. Moreover, the model has been trained on all possible scenarios, which necessitates a large amount of data for broader applications. In [101], several classifiers, including Naïve Bayes, k -Nearest Neighbours, Support Vector Machine, Random Forest, and Decision Tree and an ANN, have been trained on a real-world dataset to evaluate the effectiveness of a set of selected features as potential engine malfunction indicators. As a result, the decision tree and random forest models achieved exemplary training performance, while ANN is capable of better capturing the variability and complex patterns of the considered dataset. Unquestionably, emissions data, including CO, HC, and CO_2, provide insights into the combustion efficiency and thus potential faults in the engine or exhaust system. However, as observed for [91], these measurements can not be acquired on-board for real-time processing, thus making this approach not applicable for on-board FDI purposes.

33

Table 2.1: Overview of remarkable studies on data-driven FDI in the automotive sector

Reference	Approach	Data Source	Input Features	Detected Faults	Performance	On-board Implementation
Y.S. Wang et al. (2020) [85]	ANN	Test bench (idle operation), Nominal + single faults (labelled data)	Sound intensity analysis	Injector nozzle, electric motor, throttle orientation potentiometer, Hall sensor, ECU not receiving selected signals (knock, oxygen sensor)	Accuracy: 100%	Unlikely
R. Ahmed et al. (2015) [86]	ANN	Test bench in semi-anechoic chamber (fixed steady-state condition), Nominal + single faults (labelled data)	Vibration signals	Missing bearing, piston chirp, chain tensioner, lash adjuster (loose or collapsed), chain sprocket.	Accuracy: 97%	Not feasible
M.N. Khajavi et al. (2014) [87]	ANN	Test bench (fixed steady-state condition), Nominal + single faults and their combinations (labelled data)	Vibration signals	Manifold pressure sensor, knock sensor, misfire and their combinations	Accuracy: 100%	Unlikely
L.T. Francis et al. (2022) [88]	Autoencoder, SVM, Isolated Forest	Real-driving ECU acquisitions (unlabelled data)	Knock signal, Mean knock noise MfdT, MfdP, MFF, AmbT, ECT, EngTq, Rpm, MAF	Knock (anomaly detection)	Precision: 80%, F1 score: 81%	Feasible
P. Wolf et al. (2018) [89]	CNN, LSTM	Real-driving ECU acquisitions (unlabelled data)	50% of all ECU signals based on statistical relevance	Pre-ignition (anomaly detection)	F1 score: 90%	Feasible
A. Maruf et al. (2024)	DNN, Naïve Bayes, SVM	Real-driving ECU acquisitions (unlabelled data)	CAN messages	Wheel speed control system (anomaly detection)	Accuracy: 99.8%	Feasible
S. Zhu et al. (2021) [91]	PNN	Test bench (idle operation), Nominal + single faults (labelled data)	HC, CO, NOx, CO_2 (from exhaust gas analyzer)	Sensors (oxygen, water temperature, boost pressure, camshaft and throttle position), injection and ignition failures	Accuracy: 100%	Not feasible
P.K. Wong et al. (2016) [92]	SBELM	Test bench (fixed steady-state condition), Nominal + single faults	AFR, ignition, acoustic wave patterns	15 single faults (e.g. misfire, knock, leakage in the manifold, clogged air filter, defective spark plug of injector etc.), and 5 combinations of them	Accuracy: 92% (single faults), 81% (simultaneous faults)	Not feasible
M. Mumcuoglu et al. (2023) [93]	Bagged DT	Real-driving records (unlabelled data)	EngTq, VehS, RoG, VehM, AccP, BrkP, Rpm, EOT, ECT, pitch angle	High fuel consumption (anomaly detection)	Accuracy: 99.2%, F1 score: 78%	Unlikely

Reference	Approach	Data Source	Input Features	Detected Faults	Performance	On-board Implementation
A. Theissler et al. (2017) [94]	One-class + two-class classifiers (Naïve Bayes, RF, SVM, etc.)	Real-driving records (labelled and unlabelled data)	ECT, EngTq, Rpm, MfdP, VehS, ignition timing, injection pulse width, ThrP	Defected injection or ignition, unavailable ECT	Accuracy: 84%	To be tested
M. Youssef et al. (2021) [95]	Model-based residuals + LSTM	MVEM + stochastic fault model, Nominal + single faults (labelled data)	Intake manifold pressure residual	Manifold pressure sensor failure, leakage into the intake manifold	Accuracy: 96%	To be tested
D. Jung et al. (2019) [96]	Model-based residuals + t-student stochastic nearest embedding (t-SNE)	Test bench (FTP-75) + intake air mass flow model, Nominal + single faults (labelled data)	Set of residuals	MAF sensor, intercooler pressure and temperature sensors, intake manifold pressure sensor	False alarm rate: 2%	To be tested
D. Jung et al. (2020) [97]	One-class SVMs + Bayesian filtering	Test bench + intake air mass flow model, Nominal + single faults (labelled and unlabelled data)	Set of residuals	MAF sensor, intercooler pressure and temperature sensors, intake manifold pressure sensor, air filter clogging and leakage, throttle clogging and leakage	Accuracy: 92%	To be tested
M. Cervantes et al. (2023) [98]	ANNs bank	Test bench (steady-state and step variations), Nominal + single and sequential faults (labelled data)	MAF, CKP, AFR, MAP, ThrP	Single manifold pressure sensor, MAF sensor, throttle position sensor faults and combinations	StdDev: 0.1008 (single faults), 0.112 (sequential faults)	To be tested
M. Abboush et al. (2023) [99]	LSTM-RF ensemble	HiL platform (real driving cycles), Nominal + single faults and combinations	EngTq, Rpm, VehS, ThrP, ECT, MfdP, RailP	Single sensor faults and their combinations	Accuracy: 91.2%	Likely (Real-time validated)
A. Namigtle-Jimenez et al. (2023) [100]	ANN	Test bench (idle operation), Nominal + single faults (labelled data)	Injector rail pressure (time and frequency domain)	Injector faults (anomaly detection)	Accuracy: 100% (offline), 85% (online)	Likely (Real-time validated)
M. Vergara et al. (2023) [101]	NB, k-NN, SVM, RF, DT, ANN	Chassis dynamometer, Nominal + single faults (labelled data)	MfdP, ThrP, EngT, EngPwr, MFF, VehS, AFR, lambda, O_2, CO, HC, CO_2	Rich mixture, lean mixture, and low voltage	Accuracy:75%, F1 score: 75%	Not feasible

35

2.2   Engine performance and emission modelling

Modelling of internal combustion engines is a complex field that involves a variety of approaches designed for different purposes such as diagnostics, data analysis, optimization, and control to improve performance and reduce emissions. Engine models can generally be categorized into three main groups: physics-based "white-box" models, empirical "black-box" models and semi-empirical "grey-box" models [102]. These approaches will be deeply discussed in the next sections, with a particular focus on pollutant emission modelling, being the relevant topic for the present research activity.

2.2.1   Physics-based models: overview and applications

Physics-based modelling approach, also known as "white-box" modelling, is based on prior knowledge of the modelled phenomena and the existence of universal equations that can be applied to build the model [103]. Thus, physics-based models simulate engine characteristics by applying physical principles such as laws of thermodynamics, mass conservation, and thermal kinetic relations to replicate the phenomena occurring during an engine cycle. These may include chemical species transport, heat transfer, fuel injection and atomization, fuel-air mixing, and combustion, which significantly impact emissions and performance. Various physics-based models can be selected depending on the desired outputs, level of accuracy, and computational performance. The main features and common applications of each group of models are summarized in this section.

Zero-Dimensional (0-D) models   These models calculate engine parameters on a crank-angle basis by solving the mass and energy conservation equations, along with the gas state equation, in their ordinary differential form without spatial information [104]. This approach allows for the calculation of key parameters such as pressure, temperature, and gas composition within the engine cylinders and manifolds. The physical combustion process is not modelled, but it is simulated using phenomenological models, like the Wiebe function-based model [105], that can be either single-zone, which provide a balance between process representation and accuracy, or multi-zone, which offer a more detailed depiction of the combustion process and enhanced prediction of exhaust gas emissions. These models use simplified, lumped-parameter representations of one or several engine cycles, focusing on average conditions rather than spatial variations, offering quick predictions of engine performance and emissions with limited computational cost. 0-D models will be further discussed in Section 2.2.3.1, where related examples will be included as well.

Mean value models   A Mean Value Engine Model (MVEM) is a 0-D model that dynamically predicts the mean values of major external engine variables, like crankshaft speed and manifold pressure, and the gross internal engine variables, like thermal and volumetric efficiency, neglecting the discrete engine cycles and assuming that all processes and effects are spread out over the engine

cycle [106, 107]. The time scale for this mean value description is much longer than that required for a single engine cycle, but sufficiently shorter than that required for describing longitudinal vehicle motion. Positioned between large cyclic simulation models and basic transfer function models, MVEMs strike a balance between accuracy and computational efficiency, which makes them suitable for longitudinal vehicle control applications. Several examples of MVEM models for engine performance and emission prediction will be provided in Section 2.2.3.1.

One-Dimensional (1-D) models   These models simulate gas dynamics, mass flow, and heat transfer along the engine pathways using one-dimensional flow equations. This means that the relevant properties of the flow (air, fuel or exhaust gases), such as velocity, pressure, and density, only vary along a single spatial dimension, typically the direction of the flow, which may coincide with the axis of pipes or ducts. Propagation of pressure waves and pulse effects within the engine air pathways can be effectively captured. 1-D models are particularly useful for analysing the flow through pipes, ducts, or nozzles where the cross-sectional area perpendicular to the flow direction is constant or for system-level interactions and transient response in components like manifolds, heat exchangers, turbochargers, and EATS components [102]. As a further step, quasi-dimensional models can be adopted when flow properties are assumed to vary in the direction of the flow as well as to some degree in the cross-sectional directions, but the latter is small compared to the stream-wise variation. Quasi-one-dimensional flow is often used to model flow through components with gradually changing cross-sectional area, such as diffusers, nozzles, or ducts with slight curvature. As a meaningful example, in [108] a novel quasi dimensional-model for the simulation of the combustion process in CI engines is presented. The presented 1-D control-volume-based approach discretizes the in-cylinder air-fuel mass on fixed values of local equivalence ratio based on the charge stratification determined from a 2-D reconstruction. The model has been validated against experimental data and detailed 3-D simulation results, showing an almost negligible underestimation in terms of predicted fuel consumption and NOx emission.

Three-Dimensional (3-D) models   These models provide detailed simulations of fluid flow, heat transfer, and chemical reactions within the engine, using numerical methods to solve complex fluid dynamics problems by dividing the system into sufficiently small cells, and solving the governing partial differential equations for each cell. These fluid-dynamic-based models, referred to as Computational Fluid Dynamics (CFD) models, are capable of simulating the behaviour of a fluid system by considering a wide range of submodels including fuel spray and breakup, turbulence, chemical reactions, heat transfer, fluid-surface interactions, multi-phase fluid interactions [29]. In addition, due to their cell-by-cell simulation structure, CFD models can adapt to complex geometries, and provide useful information about the system at different spatial locations in three-dimensional space, including temperature and pressure gradients, concentration gradient of species throughout the simulated region, and others. A large number of CFD models have been developed by the engine researchers to study the combustion phenomena of different types of

ICEs and to predict engine performance and emissions. CFD models are used for in-depth studies of combustion dynamics, turbulence, and emission characteristics, but their computational time is definitely too high for online optimization and control purposes. The works presented in [109] and [110] show two possible applications of CFD models to internal combustion engines. In the first one, an experimentally validated CFD model is used to study the effect of water injection on knock mitigation, flame propagation and pollutant emission on a hydrogen-enriched natural gas engine; in the latter, the effects of increasing turbulence intensity on the performance and pollutant emission of a HD diesel engine has been investigated using a validated CFD model. Thanks to their accuracy and robustness, CFD models can be used also to calibrate and train simpler and faster models. In this regard, in [111], an example of this application is given, where several ML algorithms are trained on data from CFD simulations to predict combustion parameters and NOx emission of a gasoline CI engine.

Thermo-Kinetics Models (TKM) Thermo-kinetics or chemical kinetics models explain engine emission and performance using thermodynamic laws and focusing on chemical reactions involved in combustion and emission formation. These models are generally used to study the kinetics of pollutant formation. A large number of chemical reactions take place during the combustion process, mainly dependent on the in-cylinder gas temperature and pressure [102]. High-order TKMs try to capture as many as thermo-kinetic reactions to improve the model accuracy, while reduced order TKMs only include the most dominant reactions to reduce the computation efforts at the expense of a slightly reduced model accuracy [112]. Depending on the complexity of the phenomena or the required accuracy, TKMs can be used in zero-dimensional or multi-dimensional modelling approaches. As an example, the work in[113] introduces a modified chemical kinetics mechanism to simulate Diesel fuel oxidation, coupling NOx related ionic reactions to the main formation mechanisms, namely thermal, prompt and NO oxidation mechanisms. The results show that, although most of the engine exhaust NOx is produced by thermal mechanism, the proposed approach considering ionic reactions allows to increase prediction accuracy of engine exhaust NOx.

Different types of models can be combined together to obtain the required target of accuracy, depending on the specific application, for each variable of interest. As an example, a robust numerical framework based on combination of TKM, 1-D and 3-D tools for CFD analyses of internal combustion engines fuelled with ammonia-hydrogen mixtures is presented in [114]. More in detail, the 1-D engine model provides boundary conditions for the multi-dimensional investigations and estimates the overall engine performance; 3-D in-cylinder analyses are considered to predict combustion efficiency, heat transfer, knock tendency, and emissions, with a detailed chemistry based approach; 0-D/1-D TKM models support 3-D analyses and allow insights on aspects such as NOx formation and identification of mixture qualities able to reduce the emissions.

While the level of accuracy reached by 3-D CFD models and TKMs is necessary for precisely modelling the combustion process, these models are too complex and detailed to be adopted for

real-time control or monitoring. When the target is to develop a control-oriented model, a good balance between reasonable accuracy and computational effort is needed, which can be provided by 0-D models [115]. According to the target of the model and thus the time scale of interest, whether it is a single or several engine cycles, the simulation step can vary from fractions of a crank angle (in a crank angle-based domain), to hundredths or tenths of a second (in a time-based domain). 0-D crank-angle based models like Wiebe function-based and simplified reaction-based models are able to provide a good description of the combustion process, with the evolution of all relevant combustion parameters (pressure, temperature, mass of fuel burnt, etc.) through the single engine cycles and the associated pollutants generated. This level of detail is required for combustion control and calibration. On the other hand, the already introduced MVEMs is a simplified time-based model that relies on a general understanding of the combustion dynamics to capture the main combustion characteristics averaged over one or several engine cycles, without modelling the physical combustion process. Thanks to the short execution time, but still good accuracy in the calculation of crucial parameters like engine torque, MVEM approach is often used in combination with other modelling methodologies (e.g. empirical equations or maps) and integrated with other subsystems models to investigate vehicle-level performance, including acceleration, fuel consumption, and pollutant emissions [105]. Being the core of the simulation model presented within the first part of this dissertation, a more detailed overview on mean value modelling approach is given in Section 3.2.3, together with a comprehensive description of the developed MVEM.

2.2.2 Empirical and semi-empirical models: overview and applications

Empirical and semi-empirical methods typically implement simple correlations to predict target output variables, such as engine performance parameters or emission levels. The main difference is that empirical methods use input variables that can be directly measured, while semi-empirical methods can include input variables estimated by sub-models [116]. In this respect, semi-empirical methods, also referred to as grey-box models, lie between physics-based white-box models and fully empirical black-box models.

Black-box engine models provide a data-driven representations of the system input-output relations to help understanding the system behaviour when physical models are either unavailable, too complicated or computationally prohibitive for the target application. Many different types of empirical models can be found in literature. The main families of black-box models applied to engine end emission modelling are listed below.

Statistical models   They use mathematical frameworks to identify patterns in the data and make predictions. They can be used for capturing linear and non-linear relationships between inputs and outputs of the system [117]. Common statistical techniques include:

- **regression analysis**, employing mathematical models like linear regression, polynomial

regression, and logistic regression to describe a continuous response variable as a linear or non-linear function of one or more predictor variables;

- **Principal Component Analysis** (PCA), together with Pearson Correlation Coefficient (PCC), used for dimensionality reduction to simplify complex datasets;

- **Time series analysis**, employing techniques like AutoRegressive Integrated Moving Average (ARIMA), Multiple Linear Regression (MLR), and Regression error with Autoregressive Moving Average (RegARMA) for modelling data that varies over time.

Several applications of the mentioned statistical techniques can be found in literature. More in detail, PCA and PCC have been extensively used to select the most significant input features for different data-driven models, recognizing the degree of correlation between input and output variables [111, 118–121]. On the other hand, in [122] different regression techniques are used to develop speed-based emission models aimed at quantifying the actual pollutant emissions and defining the emission factors for diesel cars using on-board emissions measurements. To provide an additional example, the study described in [123] focuses on different intelligent time series modelling techniques, including ARIMA, MLR and RegARMA, to predict NOx emission and fuel consumption of a medium duty truck diesel engine based on on-board measured inputs, namely exhaust gas pressure and temperature, engine speed and load, engine coolant temperature, common rail pressure, intake manifold air temperature and pressure, accelerator pedal percentage, VGT position, and SCR outlet temperature.

As a main advantage, classical statistical models are well-suited for interpreting data with clear linear relationships and quantifying uncertainty with limited complexity, thus resulting in good model explainability and reasonable computational requirements. On the other hand, these methods often lack the ability to capture highly non-linear dynamics compared to types of models such as ML-based and DL-based ones [124].

Machine learning models   ML regression models are data-driven approaches that automatically learn patterns from a training dataset and can be used to make predictions for unknown given datasets. They can handle complex, non-linear relationships between input and output variables. As already introduced in Section 2.1.1 for ML classification approaches, two types of ML regression models can be identified, namely supervised and unsupervised learning methods. However, only supervised regression models are considered of interest if the goal is to reconstruct the trend of specific quantities, like engine emissions for this research activity. So, more specifically, the algorithm takes a known set of input data and known responses to these data and trains the ML model to generate reasonable predictions for the response to new data [125]. According to literature [126], the most widely used ML models in the field of engine and emission modelling are:

- **Support Vector Machines** (SVM), aiming at finding a function that deviates from measured

data by a minimum value (usually not greater than a predefined ε) for each training point, and at the same time is as flat as possible to reduce sensitivity to errors;

- **Decision Trees** (DT) and **Random Forests** (RF), according to which responses are predicted following a flow chart-like structure from the root through branching conditions, where predictor values are compared to trained weights, until reaching a leaf node denoting the final results of the algorithm. RF is an ensemble method where multiple decision trees runs in parallel and then the single outputs are averaged to provide one single output. As the number of trees increases, the accuracy and robustness increase as well, at the cost of increased model complexity and harder interpretability;

- **K-Nearest Neighbors** (KNN), predicting the output for a new data point by averaging the values of its k nearest neighbours in the feature space, based on a specified distance metric. The value of k must be optimized according to the input dataset and expected performance;

- **Gradient Boosting Machines** (GBM), an ensemble method achieving high accuracy in prediction tasks by boosting weaker models, for example decision trees. In this case they differ from RF since trees are not independently trained to run in parallel, but they are combined and trained in a fixed sequence so that each corrects the errors of the previous ones. This method can be very accurate, but it requires a longer training phase and it is more sensible to overfitting compared to simpler models.

In [127], an RF model which extracts information from OBD data streams transmitted by a remote emission management vehicle terminal is proposed to provide a novel solution for the online screening of anomaly NOx emissions. The results show that RF outperforms statistical models like logistic regression, due to the fact that the relationship of NOx emissions with OBD features is strongly non-linear. Moreover, compared to other ML methods like SVM or GBM, when the target is to identify anomalies in the input data, RF still shows appreciable flexibility, but with lower computational cost and less tuning work. Besides emission estimation, RF technique is efficiently used also to estimate the relative importance of input variables for other data-driven models predicting emission trends [128, 129]. In [130], a comparison between different ML methods is presented, including SVM e KNN, to predict emission (NOx and CO) and performance of a CI engine fuelled with various metal-oxide nanoparticles, based on engine speed, type of fuel, and fuel mass flow rate. However, even if both SVM and KNN models show promising performance, neural networks provide the best results in the prediction of engine responses. The work in [131], offers an example of application of the Extreme Gradient Boosting (XGBoost) ML algorithm to develop a virtual sensor for NOx monitoring in diesel engines, achieving excellent results in steady state conditions and still meaningful results in transient conditions.

Compared to simpler statistical models, the mentioned ML models show higher flexibility, with the ability to model non-linear interactions even with medium-sized datasets. However, this comes at the cost of increased complexity and computational power required by these models,

especially for KNN, SVM and GBM, together with the need of accurate, and thus time-consuming, tuning of model hyperparameters.

Deep Learning (DL) models   Deep learning models are a subset of machine learning that uses neural networks with multiple layers to learn hierarchical representations of data. They excel at capturing intricate patterns and correlations in complex datasets. Neural networks can be grouped into different types:

- **Artificial Neural Networks** (ANNs), basic networks predicting continuous values by propagating input data through interconnected layers of neurons that learn weights and biases to relate inputs to desired outputs. As deeply reviewed by [126], in the last years ANNs have become increasingly popular in addressing problems like engine emission estimation, thanks to their flexibility, potential accuracy and limited computational requirements enabling real-time applications;

- **Convolutional Neural Networks** (CNNs), employing convolutional layers to capture spatial features from input data, followed by fully connected layers that predict continuous target values, making them suitable for tasks involving structured data like images. They are not so commonly applied in the field of emission estimation, but some examples can still be found in literature [129];

- **Recurrent Neural Networks** (RNNs), performing regression tasks by processing sequential data with feedback loops, which allows them to model dependencies and predict continuous outputs based on time series information. Long Short-Term Memory (LSTMs) and Gate Recurrent Unit (GRU) are now the standard architectures for RNNs, showing improved performance compared to previously developed RNNs [132].

As an example of application of different DL methods, the work presented in [121] compares several ML and DL methods, including ANN, LSTM, GRU, applied to transient emission prediction of a diesel engine. The results show that there is no single model outperforming the others in all emission characteristic prediction. For example GRU achieves the best performance in NOx prediction, while LSTM in CO_2 prediction. So, the final solution proposed by the authors is combining the best algorithms corresponding to each emission parameter to generate a more complex model achieving the overall best performance in all emission characteristic prediction. As a main advantage, DL techniques can model highly non-linear relationships and patterns, effectively handling larger datasets compared to other ML models with better performance in terms of prediction speed compared to the most complex ML algorithms, such as ensemble models like GBM. On the other hand, one drawback that must be highlighted is that DL models are computationally demanding compared to simpler ML models, especially in the training phase and hyperparameters optimization of the most complex networks (CNNs, RNNs, LSTMs), requiring also a large amount of training data. However, the great potential of DL methods, in parallel with

recent progress in computing power and the availability of extremely large sets of information, makes them attractive for challenging tasks like engine emission modelling, where high accuracy is required in capturing complex trends and relationships between a large number of variables, while allowing real-time implementation of trained models. This is confirmed by the extended amount of works found in literature dealing with the application of neural networks for engine emission prediction. A detailed review of the most relevant related works can be found in Section 2.2.3.2, emphasizing the most interesting solutions for on-board monitoring purposes.

Probabilistic Models   Probabilistic models incorporate uncertainty directly into the modelling process, providing probabilistic predictions rather than deterministic outputs. Due to their nature, they are rarely used for engine performance and emission prediction, even if they find application when stochastic processes are involved (e.g. system anomalies or knocking). Common probabilistic approaches include:

- **Bayesian Networks**, consisting in a probabilistic graphical model that measures the conditional dependence structure of a set of random variables based on the Bayes theorem [133]. It can be used for modelling probabilistic relationships between engine parameters. Within this field, an example is provided by [134], employing a Bayesian network for the development of a stochastic controller for engine knock;

- **Hidden Markov Models** (HMM), statistical models that can be used to describe the evolution of observable events that depend on internal factors, which are not directly observable [135], showing effectiveness for sequence modelling where state transitions are uncertain. As an example, [136] presents an implementation of an HMM for engine misfire detection through processing of crankshaft speed fluctuations;

- **Gaussian Processes** (GP), used for regression tasks with uncertainty quantification; it is a non-parametric data-driven model used for predicting the value of a continuous response variable, modelled as a random Gaussian process using covariances of each input variable and modelling the uncertainty of the response as well.

According to literature, among probabilistic models, only GPs find significant applications in the field of engine emission modelling. As an example, in [119] the most relevant information related to actuators, environmental status and engine operating status have been selected through PCA as input features for a regression model estimating torque and raw NOx emission based on the Gaussian process regression principle, achieving satisfactory results in terms of accuracy. In [137], another example of application of GP-based model to predict NOx emission of a Diesel engine is given. Probabilistic models show significant advantages compared to other models if the target is to predict stochastic phenomena, naturally incorporating uncertainty and providing predictions within a certain confidence interval. However, they are computationally demanding, especially for high-dimensional data, which does not make them a suitable choice when applied to real-time emission modelling.

Grey-box models provide benefits of both groups of physics-based models (white-box models) and black-box models, by combining these two model types: physics-based models are characterized by higher robustness and good extrapolation capability with a minimum calibration effort, but simplified mean value approaches are required for real-time application, which in turn limits model accuracy; on the other hand, black-box models can provide high accuracy, but they may show a poor predictive robustness in regions not extensively included in the training dataset [138]. The idea behind grey-box models is to improve the accuracy of physics-based models by an additional data-driven model, maintaining the same robustness. Grey-box models can be generally classified according to different approaches [139]:

- **Semi-empirical models**, based on simplified physical representations including empirically calibrated parameters used to fine-tune the model, providing a trade-off between computational efficiency, model interpretability, and prediction accuracy;

- **Hybrid serial approach**, integrating a sequence of both white-box and black-box models where the outputs of one model can be the inputs for another model, with the final target of improving prediction accuracy;

- **Hybrid parallel approach**, aiming to use both white-box and black box models together to estimate the same target variable, exploiting the advantages of the two types of models in terms of accuracy and robustness;

- **Physics-informed ML models**, embedding known physical principles directly into ML architectures, guiding the learning process by limiting the space of admissible solutions to improve generalization, flexibility and interpretability, and also overcoming the issue of low data availability. The works in [140] and [141] show possible applications of this approach for NOx and soot prediction respectively, where information from simplified physics-based models are used to generate more easily explainable ML models architectures. However, according to literature, only few applications of physics-informed ML models can be found in the field of engine performance and emission modelling.

As an example, the work in [142] presents a cyclic peak pressure-based semi-empirical NOx prediction model, calibrated using NOx emission measurements during transient operation. However, in-cylinder peak pressure estimation is needed as main input parameter for the simplified chemical kinematics model, together with adiabatic flame temperature and other combustion-related variables, which need to be estimated using an external software. An additional calibration is performed based on high-resolution NOx measurements to tune a parameter related to a simplified Zeldovich NOx formation mechanism. The so developed semi-empirical model is able to effectively capture transient cyclic NOx emission trends with better accuracy than a reference ANN model using the same input features. However, no optimization of the mentioned ANN model has been performed, which would probably increase model accuracy, making the comparison more

fair. Moreover, the developed model shows lower accuracy at operating points where non-thermal NOx production is relevant (since only Zeldovich mechanism is modelled), which on the contrary is well captured by the simple ANN model without any added calibration effort. In [143], a grey-box model is developed to estimate combustion phasing, engine torque, exhaust gas temperature and engine emission in a HCCI engine for control purposes. The proposed model presents a serial architecture consisting of physical models to predict CA50, IMEP, adiabatic flame temperature, and exhaust gas temperature and an ANN black-box model to predict emission concentration. The inputs to the white-box model are fuel equivalence ratio (φ), intake manifold pressure and temperature, and engine speed. The same quantities, together with the mentioned outputs of the physical models, are used as input to the ANN to predict HC, CO and NOx emission. The results show that the emission grey-box model can predict pollutants concentrations with three times better accuracy compared to those from the reference black-box model. The limited computation time allows the developed grey-box model to be used as a virtual engine test-bed for model-based controller design. However, further investigation is required to evaluate the feasibility of on-board implementation on a standard ECU. In [138], an hybrid modelling approach for NOx estimation is proposed. More in detail, two different control-oriented engine-out NOx models are considered, namely a simplified physics-based model and a purely data-driven model based on GP algorithm, which has been trained on a limited dataset including only warm engine conditions. The two models are running in parallel, while a calibrated factor is applied to weight the related outputs. In the presented use case, the weighting term is a function of engine temperature, allowing to rely on the physics-based model rather than on the data-driven model in all the conditions where the latter may not be able to provide appropriate results, namely cold engine condition in the specific use case. The proposed hybrid engine-out NOx emission model is tested on a cold-started WLTC, where the resulting cumulated prediction error is significantly reduced compared to both physics-based and black-box models.

As already mentioned, grey-box models show the advantages characterizing both white-box and black-box models in terms of interpretability, robustness and accuracy. On the other hand, they also show the critical issues peculiar to the two types of models, namely the need for knowledge of the underlying physical phenomena and a reasonable simplification of them, together with the definition and tuning of empirical parameters.

2.2.3   OBM-oriented emission estimation

Meeting the latest emission and OBM standards necessitates the implementation of efficient engine control strategies and accurate monitoring approaches that can operate in real-time. As already introduced in Section 2.2.1, most of the physics-based engine and emission models are too complex and computationally demanding for on-board implementation. Thus only simplified physics-based models or faster black-box models can be taken into consideration for this purpose, as an alternative to conventional simpler methods, like map-based approach. In this regard, a

deeper review on OBM-oriented emission models is presented in the following sections, to frame the state of the art and identify key areas for improvement.

2.2.3.1 Conventional and model-based approaches

In combination with available on-board measurements from OBD systems [144, 145], conventional ECU control-oriented emission models usually employ look-up tables or correlations developed from engine calibration testing. They are characterized by fast response rates, but they require a significant experimental effort and may perform poorly under off-design conditions [142]. On the other hand, several advantages may be obtained from adopting a model-based controller instead of a conventional map-based one, such as the possibility of realizing a real-time optimization of engine calibration parameters [115]. In addition, thanks to its physical consistency, a model-based approach is potentially capable of taking into account the effects of the variability of the environmental conditions or the effects related to the engine transient operation, without the need of implementing further correction maps that must be experimentally derived [146]. For these reasons, the development of model-based approaches against the conventional map-based ones has become of great interest, further boosted by the increasing computational performance of modern processors that are now available for mobility applications. The already deeply introduced multi-dimensional combustion models are capable of estimating the in-cylinder gradients of temperature and chemical composition with high accuracy and predictive strength (also in off-design conditions), and may be coupled with pollutant formation submodels to accurately predict their concentration. However, they require a computational time that is not compatible with real-time combustion control applications, at least considering the computational performance of available ECUs [147]. On the other hand, as already highlighted in Section 2.2.1, 0-D MVEMs are capable of simulating the combustion and emission formation processes with a good predictive capability, even outside the calibration range, in both steady-state and transient engine operating conditions, while requiring limited computational load [148].

In this regard, [25] proposes a mean value NOx model deriving from a simplified 0-D thermodynamic model, namely an averaged NOx formation kinetic model based on the Zeldovich mechanism, thus including only the most relevant phenomena contributing to NOx formation. The model shows high accuracy and flexibility at low computational cost, even if online application still needs to be tested. Moreover, the mathematical formulation of the model is simple, since it is based on a single algebraic relation, and also calibration requires limited effort and few experimental data. However, the key input for the model is the maximum burned gas temperature, which is not measurable and needs to be estimated by an additional model, in the specific case a semi-empirical model, requiring measured in-cylinder pressure among the inputs and a non-negligible calibration effort. A similar approach has been previously followed by the authors in [149], presenting a real-time semi-empirical model for predicting NOx emissions per engine cycle based on an alternative mathematical expression compared to [25], but still requiring maximum

burned gas temperature as main input and the calibration of different parameters.

In [150], a cycle-averaged engine-out NOx model is developed, taking into account all main diesel engine control variables. A good prediction accuracy on cumulated NOx emission at the end of a standard homologation driving cycle is reached by the developed model. Moreover, differently from other solutions found in literature, including [25], the model structure is developed such that it can be embedded in an engine control unit without any need for an in-cylinder pressure sensor. However, additional calibrated models to estimate the combustion parameters required by the NOx model are needed also in this case, while feasibility of on-board implementation needs to be tested. Similarly, in [146] a model-based approach to control Brake Mean Effective Pressure (BMEP) and NOx emissions has been developed, based on the inversion of a zero-dimensional real-time combustion model, which is capable of predicting the main combustion parameters as well as engine-out NOx emissions. Real-time capability has been assessed through the implementation of the developed model on a rapid prototyping device, which requires an incrementation of the calculation step, while the effectiveness of the developed controller still needs to be tested.

Th work presented in [151] and [152] is focused on the development of ECU-oriented models for the on-board estimation of NOx emission. In the first part of the paper, the proposed modelling approach employs look-up tables for the estimation of nominal NOx emission and the effects of external parameters such as environmental conditions, following a conventional map-based approach. Then physics-based models are used in parallel to calculate correction factors to be applied to the mapped nominal NOx, as a function of the intake oxygen rate and engine thermal load. In the second part, the authors propose a data-fusion approach based on Kalman filter to combine the output of the ECU-oriented NOx model with measurements coming from the on-board NOx sensor. Thanks to this approach, online updating of look-up tables for modelling NOx can be performed, as well as calibration of the physics-based models, eventually identifying significant biases. A complication is definitely added due to the need of taking into consideration the delay of NOx sensor measurements, mainly due to the transport and hardware delay of the sensor itself. However, the proposed solution becomes interesting from an OBM perspective, as it could serve as a foundation for handling model drifts resulting from inevitable system ageing. Such drifts may lead to the original control calibration failing to meet emission targets after a certain mileage and causing predicted NOx emissions to deviate from actual levels, which is a significant issue that the Euro 7 standard must address.

2.2.3.2 Data-driven approaches

As already introduced in the previous sections, data-driven approaches, and especially ML and DL-based approaches, have gained significant popularity among researchers in the automotive field [153]. In fact, these types of black-box models stand out for their ability to model complex non-linear system behaviours without requiring detailed physical knowledge of the investigated processes. Through relatively simple mathematical operations, they can provide an

accurate representation of system dynamics, including engine performance and emission characteristics. Moreover, recent progress in ML techniques, computing power, and the ever-increasing availability of large sets of data significantly enhances the applicability and effectiveness of these models [154], allowing them to learn and adapt more efficiently to the behaviour of the modelled systems. Another significant advantage of these models is their limited computational time, making them highly suitable for implementation in standard control units for on-board monitoring tasks [147]. Focusing on the target of this dissertation, a comprehensive bibliography research has been carried out to give a deeper overview on the state of the art of data-driven emission modelling. The most relevant results are discussed in this section, highlighting peculiarities, valuable novelties and drawbacks of each work, in order to provide a solid basis to this research activity and point out the shortcomings that need to be covered. A general overview of the cited previous works on this topic is given in Tab. 2.2.

In [155], an integrated neural network based on different types of RNNs, including LSTM and GRU, has been developed to predict CO_2, CO, HC, and NOx emission, taking as input speed profile, engine family, engine manufacturer, engine model year, vehicle inertia, odometer reading, number of cylinders, and fuel type. The network has been trained on chassys dynamometer test profiles performed for several LDVs, showing good performance in terms of average accuracy compared to simpler networks. However, the model is missing some relevant emission peaks, and moreover this set of input features would never allow to take into account the effects of non-nominal conditions, since most of the inputs are not directly physically related to emission formation.

In [156] an RF model was used as a regression method to integrate various characteristics to improve the predictive effect of a reference emission model (MOVES, provided by EPA) and to study and quantify the corresponding contribution rates of various characteristics on CO_2 and NOx emissions. The integration of ML methods leads to better results in terms of accuracy on average CO2 and NOx emission compared to the reference model. However, while this approach is interesting for offline emission analysis, it is not suitable for OBM-oriented emission estimation, due to the fact that only average emissions per road segments are calculated. A similar output is provided by the already cited speed-based emission model developed in [122].

In [157] an ANN-based approach to predict tailpipe CO_2, NOx, and CO emissions of diesel vehicles is presented. More in detail, six operating parameters, including vehicle speed, engine speed and torque, engine coolant temperature, AFR ratio, and intake air mass flow, collected through OBD interface, are used as predictors for exhaust emissions. Response variables are measured through PEMS on real driving tests from four diesel passenger vehicles complying with different emission standards. From an evaluation of the importance of each input parameter to the emission predictions, intake air mass flow rate and AFR ratio are shown to be the inputs with the highest relative importance for the prediction of NOx and CO emission. Significant differences were observed in prediction performance depending on vehicle type, with very limited accuracy

for the vehicles equipped with the most advanced EATS. This result was expected since the set of predictor variables was not sufficient to characterize so complex systems and behaviours, so the authors suggest considering additional variables to improve accuracy.

In [120], a super-learner approach generated from the combination of simpler models (RF, XGBoost, LightGBM, and CatBoost) is used to build CO_2 and tailpipe NOx transient emission models. On-road test data from different HDVs are used to train and test the model. Despite differences in the emission behaviour between the vehicles, the proposed model shows robust prediction performance. However, similarly to [155], this approach would not be suitable for OBM purposes due to the chosen input features, not allowing online implementation and modelling of non-nominal conditions: most of them are related to driving cycle characteristics and statistics (e.g. average vehicle speed and acceleration, acceleration and speed variance, etc.) or to vehicle specifications (i.e. Vehicle Specific Power (VSP), engine displacement, etc.). Analogous comments are valid also for [158], where a similar input dataset is used to predict CO_2 and pollutant emissions from a CNG-fueled bus employing a combined XGBoost-GRU algorithm.

The study in [159] proposes a GRU network combined with genetic algorithm to be embedded as a soft sensor in the OBD system to monitor real-time emissions of NOx and PN. The model is trained on data from PEMS measurements collected from Diesel HDVs in real driving conditions, showing excellent results and outperforming other conventional ML models thanks to its accuracy and generalization ability. Most of the required input features are conventional on-board available signals, including engine coolant temperature, vehicle speed, air mass flow rate, ambient air temperature, engine oil temperature, engine fuel rate, actual engine torque, catalyst temperature, DPF differential pressure, DPF outlet pressure, DPF inlet and outlet temperature. Compared to [157], highest accuracy in the tailpipe emission estimation is reached also thanks to additional input features related to EATS operating conditions. The final target of the authors is to monitor real-time pollutant emissions during actual driving, effectively minimizing the need for physical sensors. However, in view of an OBM-oriented application, the feasibility of implementation on a standard real-time hardware still needs to be assessed, as well as the sensibility to engine and EATS faults.

Similarly, the authors in [160] employ another kind of RNN, namely LSTM, to build soft sensor for instantaneous NOx emission from diesel vehicles. The input features for the model are vehicle speed, engine speed and torque, instantaneous fuel consumption, and accelerator pedal position. So the model has the great advantage of using signals that are readily available from a standard OBD port. Signal analysis and decomposition is used to improve network performance, which shows more reliable and accurate prediction compared with other standard ML models. Despite these promising results, as for the other cited works, the computational efficiency of the proposed model needs to be quantified in view of on-board implementation. Moreover, although the limited number of selected input signals makes the implementation easier, it also reduces the capability of the model to adapt and predict emission correctly in non-nominal conditions.

In [161], ANN and LSTM networks are used to estimate engine-out CO_2, HC, CO, and NOx

49

from a gasoline engine. Real-world driving profiles were collected from a test vehicle and then repeated on a transient engine test bench to generate the training dataset. Despite the small dataset and the use of just three input parameters, namely engine torque, speed, and lambda, the models are capable of delivering acceptable quantitative results in most of the test cases. However, comparison with other ML methods, shows that RF and SVM models can produce much better results on a small training data set like the one considered in this application. Given the considered input features, the capability of the model to predict emission in non-nominal conditions is even more unlikely than in the previous work.

In [162], a control-oriented diesel engine modelling approach is investigated to predict both engine performance and emissions in transient and steady state operation. As a starting point, thermodynamic and chemical reaction models altogether with 1-D gas dynamic model are employed, requiring a time-consuming iterative solution method. As already discussed in 2.2.2, the adoption of ML-based methods allows to decrease the run-time of the model while still achieving good accuracy. So, the authors introduce an ANN to mimic the thermodynamic model and another ANN to predict soot and NOx emission, taking as input inlet air pressure and temperature, engine speed, rail pressure, mass of injected fuel, mass of pilot injection, start of main injection, pilot retard, EGR rate, exhaust pressure. The comparison of experimental data with model results shows satisfactory agreement in both performance and emission prediction capabilities. However, only steady-state tests or step variation of different actuations (e.g. throttle, VGT) have been considered, so performance assessment on driving cycles is missing in perspective of OBM application. A similar set of input features is considered in [163], where an ANN is combined to a fuzzy inference system to predict emissions, namely soot and NOx, and performance of a Reactivity-Controlled compression Ignition (RCCI) engine. An experimentally validated CFD model is used to generate the training and testing dataset for the neural-fuzzy system. Intake pressure, injected fuel mass, EGR rate and SOI are used as input features, achieving high correlation factors. However, also in this case, only steady-state tests have been considered and application on a real-world dataset is needed to confirm the model robustness. Similarly, in [164], engine speed, intake manifold pressure, ignition timing, fuel consumption rates are used as input variables to predict output torque and emission performance of a gasoline oxygen-enriched combustion engine for different steady-state operating conditions. At this stage, this specific ANN application could be an effective tool to assist test bench studies, not implementable for OBM purposes. The same ANN-based modelling approach is investigated also by [165] to predict NOx emission from a diesel engine, using the same inputs as in [162]. As for [162] and [163], the training and testing data are generated at the test bench, but in this case a transient driving cycle is performed, showing high accuracy of the model in predicting transient NOx emissions.

The study in [129] provides one of the very few applications of CNNs for emission estimation tasks. Data from a transient driving cycle were employed to train and validate the model. The six most relevant input features for the prediction of each pollutant are selected among engine speed, torque, intake pressure and temperature, intake flow, oil pressure and temperature, fuel

rate, and water temperature. The developed model shows excellent robustness and generalization capability, achieving higher accuracy and also reduced computational effort compared to other advanced algorithms, including XGBoost, RNN, and LSTM. Considering the selected input variables, on-board implementation would be possible, but no investigations on real-time capability and sensibility to non-nominal conditions have been performed.

In [166], a GBM-based model is used to predict NOx emission, CO_2 emission, and fuel consumption of real driving diesel vehicles in urban, suburban and highway scenarios. To this aim, a PEMS system is used to collect data. The input features for the NOx model include both engine and ambient-related variables, including intake air mass flow, exhaust gas flow rate, CO_2 concentration, engine speed, vehicle acceleration, tailpipe exhaust temperature, vehicle speed, ambient humidity, ambient temperature, and GPS altitude. Taking into account measurements that are physically related to engine operation and emission formation, this approach would be likely sensible to system faults if trained properly. However, some of the considered input features may not be readily available from conventional on-board sensors: a direct on-board measurement of exhaust gas flow rate, for example, is not usually provided, while it would be one of the most important features for the developed model according to the authors, as well as CO_2 concentration.

The study presented in [167] proposes an interesting application of DL methods for OBM purposes. More in detail, an algorithm based on the combination of RF and ANN is proposed to perform the post correction of NOx measurements from on-board sensors installed on diesel HDVs, to limit the issues of dynamic measurement delay and low measurement accuracy. The model is trained on the statistics data describing the relationship between NOx sensor measurements and PEMS acquisitions, considered as the actual emission. Thanks to the proposed correction applied to the on-board measurement, the final error on NOx emission is halved. The main issue of this approach is that it relies on on-board NOx measurement, which is not always available, as discussed in Section 1.3.2. However, the promising results achieved by the proposed approach would make it worth considering further investigations, including non-nominal conditions of the monitored system.

In [168], a DL-based approach to calculate engine-out NOx concentration is proposed, to be used as a reference for fault diagnosis of the on-board NOx sensor and to consequently reconstruct the correct signal. The input for the ANN model are engine speed, injected fuel, intake manifold pressure and temperature, intake air mass flow, mapped reference NOx concentration, and intake oxygen concentration correction factor. The proposed approach is computationally efficient (even if no quantitative data on that are provided) and effective in detecting sensor abnormal behaviours like drifts or deviations; however, actual intake oxygen concentration (needed to calculate the corresponding correction factor), is usually not measurable on-board, thus making it necessary to rely on models that may be significantly inaccurate in non-nominal conditions, affecting the performance of the proposed approach. Further investigations in this field would be interesting. Similarly, [169] proposes a NOx emission model also based on ANN, but which is fed with variables that are likely available from conventional OBD systems, including engine speed and

torque, exhaust temperature, coolant temperature, intake air mass flow, fuel mass flow, intake temperature and pressure. Moreover, both engine-out and tailpipe NOx emission are modelled, thus allowing to perform fault detection on SCR systems. Given the promising results, also in this case a demonstration of the on-board implementability and the evaluation of model sensibility to faults could be an interesting further step.

In summary, this bibliography research highlights the effectiveness of data-driven methods for engine emission modelling, particularly emphasizing the increasing popularity of ANNs in this field compared to other ML methods. This is mainly due to their capability of handling large amounts of data and accurately modelling complex systems behaviours with limited calibration effort and relatively low computational time. Moreover, it stresses the importance of key dataset features and careful parameters selection to optimize model performance both in terms of accuracy and robustness. According to the results obtained in literature, on-board available signals provide valuable information to be used for engine emission prediction, which is particularly interesting in an Euro7-oriented OBM perspective. In this respect, several studies demonstrate the potential of data-driven methods employed as soft sensors to perform diagnosis of physical sensors or to be used in combination with them. However, despite these promising results, the possibility to implement such models on standard vehicle control units needs to be further investigated, both in terms of robustness and computational efficiency. Moreover, the application of ML and DL approaches to correctly predict engine emissions also in non-nominal conditions, due to aged, damaged or faulty system components, has not been explored in literature. As extensively mentioned, non-nominal emission modelling is a critical requirement introduced by Euro 7 OBM standard, which would be difficult to fulfil by means of conventional approaches. In this context, data-driven approaches may offer an innovative approach to handle this issue, taking advantage of the large amount of emission-relevant data made available by standard OBD.

Table 2.2: Overview of remarkable studies on OBM-oriented data-driven emission modelling

Reference	Approach	Data source	Input features	Predicted variables	Performance	On-board implementation	Sensitivity to faults
A.M.Howlader et al. (2023) [155]	Integrated RNNs	Chassis dynamometer, Diesel LDVs	SP, EngF, EngM, EngV, VehV, VehD, NCyl, FuT	CO_2, HC, CO, NOx; Type: instantaneous tailpipe emission	nRMSE: 8.8%, 2.5%, 5.1%, 4.4%	Not interesting	No
R.Liu et al. (2023) [156]	RF	RDE PEMS acquisition, gasoline LDVs	AmbP, AmbT, Vehs, Mfp, Mft, ECT, Exht, Rpm, AccP, RoG, RoT	CO_2, NOx; Type: average tailpipe emission per road segment	R^2: 0.93, 0.75 RMSE: 40.3, 3.8	Not interesting	Not investigated, unlikely
C.Chandrashekar et al. (2024) [122]	LR, SVM	RDE PEMS acquisition, passenger car	SP (intervals)	CO_2, HC, CO, NOx; Type: average tailpipe emission per road segment	R^2: $0.6 - 0.8$	Not interesting	No
J.Seo et al. (2023) [157]	ANNs	RDE PEMS acquisition, Diesel passenger cars	VehS, Rpm, EngTq, ECT, AFR, MAF	CO_2, CO, NOx; Type: instantaneous tailpipe emission	R^2: 0.98, 0.45, 0.62.	Not tested, feasible	Not investigated, unlikely
N.Wei et al. (2022) [120]	RF-XGBoost-LightGBM	Previous RDE test campaign, HDVs	RP, EngV, LP, AveV, VarV, AveA, VarA, VSP, RPA, VehV+VehA, TID, AmbT	CO_2, NOx; Type: instantaneous tailpipe emission	R^2: 0.82, 0.74	Not interesting	No
L.Hu et al. (2021) [158]	Combined XGBoost-GRU	RDE PEMS acquisition, CNG-fueled bus	LP, Vehs, VehA, VSP, RoG, AmbT, AmbH + historical vehicle emissions	CO_2, HC, CO, NOx; Type: average tailpipe emission per road segment	Avg R^2: 0.98 Avg MAPE: 9%	Not interesting	No
L.Yang et al. (2024) [159]	GA-GRU network	RDE PEMS acquisition, Diesel HDVs	ETC, VehS, MAF, AmbT, EOT, MFF, EngTq, CaTt, DpfDp, DpfOp, DpfInt, DpfOutT	NOx, PN; Type: instantaneous tailpipe emission	R^2: 0.98, 0.90	Not tested, feasible	Not investigated, possible
Y.Yu et al. (2021) [160]	CEEDMAN - LSTM	OBD acquisition + NOx sensor	VehS, Rpm, EngTq, MFF, AccP	NOx; Type: instantaneous engine-out emission	R^2: 0.98 nRMSE: 2.71%	Not tested, feasible	Unlikely
G.Sundaram et al. (2023) [161]	RF, GB, SVM, ANN, LSTM	RDE + test bench	Rpm, EngTq, lambda	CO_2, HC, CO, NOx; Type: instantaneous engine-out emission	-	Not tested, feasible	Unlikely

Reference	Approach	Data source	Input features	Predicted variables	Performance	On-board implementation	Sensitivity to faults
K.Nikzadfar et al. (2015) [162]	ANN	Test bench, diesel engine (steady-state and step variations)	MfdP, MfdT, Rpm, RailP, MFF, SOI, Pilot Retard, EGR Rate, ExhP	PM, NOx Type: instantaneous engine-out emission	nRMSE: 5.8%, 5.4%	Not tested, feasible	Not investigated, possible
X.Chen et al. (2024) [165]	ANN	Test bench (WHTC), Diesel engine	MFF, Rpm, EGR rate, SOI, RailP, MfdP, MfdT, ExhT	NOx Type: instantaneous engine-out emission	R^2: 0.96 MAE: 2.8%	Not tested, feasible	Not investigated, possible
M.Koç et al. (2021) [163]	ANN – fuzzy inference	CFD simulation	MfdP, MFF, EGR rate, SOI	NOx Type: steady-state engine-out emission	R^2: 0.96	Not tested, feasible	Not investigated, unlikely
J.Liao et al. (2024) [129]	CNN	Test bench (WHTC), Diesel engine	Rpm, EngTq, MfdP, MfdT, MAF, EOT, EOP, MFF, ECT	CO_2, HC, CO, NOx Type: instantaneous engine-out emission	R^2: 0.95, 0.92, 0.93, 0.93	Not tested, feasible	Not investigated, unlikely
H.T.Wen et al. (2021) [166] 54	GB	RDE PEMS acquisition, Diesel passenger car	MAF, ExhT, CO2, Rpm, VehA, ExhT, VehS, AmbP, AmbT, Altitude	NOx, CO_2 Type: instantaneous tailpipe emission	Avg R^2: 0.92, 0.81	Not tested, unlikely	Not investigated, unlikely
C.Wu et al. (2023) [167]	Combined RF - ANN	RDE PEMS + on-board NOx sensor acquisition	Statistics data between PEMS and on-board measurements	NOx Type: instantaneous tailpipe emission	R^2: 0.92	Not tested, feasible	Not investigated, possible
W.Zhang et al. (2022) [168]	DNN	Test bench (WHTC), Diesel engine	Rpm, MFF, MfdP, MdfT, MAF, NOx conc. ref., O2 conc.	NOx Type: instantaneous tailpipe emission	Accuracy: 98% (on cumulated emission)	Not tested, feasible	Not investigated, likely
R.Pillai et al. (2023) [169]	DNN	Test bench and chassis dynamometer (FTP), Diesel HDV	EngTq, Rpm, ExhT, ECT, MAF, MFF, MdfT	NOx Type: instantaneous engine-out and tailpipe emission	R^2: 0.95	Not tested, feasible	Not investigated, likely

2.3   Innovative contribution of the dissertation

The present dissertation is focused on the investigation of possible solutions based on data-driven approaches to help manufacturers meet the requirements introduced by Euro 7 regulation, developing OBM-oriented models to perform engine fault detection and NOx emission monitoring taking advantage of on-board available signals. The first part of the manuscript is focused on the definition of a clear methodology for the development of the introduced data-driven models, that has been validated in simulation before being applied to real experimental data in the second part of the dissertation. More in detail, as a preliminary step, a selection of common emission-relevant engine faults have been simulated by means of a comprehensive validated 0-D vehicle model to define their effects on engine-out NOx emissions, namely the ones for which a continuous monitoring is required according to Euro 7 OBM. Based on a combination of physical and analytical considerations, a significant set of input features has been selected among the available signals. It is worth emphasising that, differently from many studies discussed within the literature review, all the considered signals would be available on-board according to the real sensor layout of the modelled vehicle. Then, a comprehensive performance assessment of different classifiers and regression models has been performed, in order to identify the most suitable data-driven methods to be applied to engine FDI and NOx monitoring tasks. The analysis takes into account not only model accuracy, but also computational complexity, that may affect not only the training time, which could be performed offline, but also the model size and prediction speed, thus strongly limiting the feasibility of real-time model implementation. In this regard, since great attention is placed on the possibility of effectively implementing the developed models on-board, they have been deployed on a conventional real-time hardware, namely a commercial Raspberry Pi computer, and tested at the HiL to evaluate their real-time capability. As an additional contribution, the proposed hardware setup offers the possibility to perform on-board vehicle testing at low cost, enabling the communication via CAN-bus with the ECU and on-board sensors to retrieve the input signals required by the models, without the need of modifying the already existing vehicle architecture.

Based on the simulation results of this first part of the dissertation, it can be concluded that the proposed data-driven classification approach allows to correctly detect and identify single-fault conditions with high overall accuracy, regardless of the considered driving cycle or fault severity; Tree, Ensemble and Neural Network classifiers have proved to be the best performing ones, showing satisfactory interpolation and extrapolation capabilities when tested on selected driving cycles and specific fault conditions not included in the corresponding training dataset. When a combination of two different faults is considered, the Neural Network classifier shows the best behaviour: in most of the tests considered, it is able to correctly detect both induced faults without false fault detections, even if trained only on single-fault conditions, which is instead an important limitation highlighted in some of the publications cited above. Regarding the on-board NOx emission monitoring task, the developed data-driven regression models show

promising results when tested on different unseen driving cycles: the correction factor calculated by the models allows to significantly reduce the gap between actual and estimated NOx emission compared to the reference map-based controller model, even if highly emission-relevant faults are introduced. As for the classification task, the Neural Network is preferred over the other considered regression models, since providing the best trade-off between accuracy and model complexity. As an additional result, all the developed classification and regression models, deployed on the Raspberry Pi, proved to be real-time capable, with a maximum task execution time of less than 3 ms in the worst case, which is far below the 10 ms threshold corresponding to the application time-step of standard control units.

Based on the promising results achieved in simulation, the second part of the dissertation is focused on the application of the same methodology on real experimental data acquired at the test bench. As a first step, selected fault conditions, that would not permanently damage the engine, have been introduced. The most relevant measured quantities from both standard on-board sensors and test bench equipment have been analysed, to study the effect of the considered faults on NOx emissions and identify significant signals to be used as input for the data-driven models. Even in this case, both physical and analytical considerations have been applied to select a proper set of input features, starting from the signals that would be available on-board and neglecting the ones measured only at the test bench. The same classifiers and regression models selected in the simulation phase have been trained on a portion of the available experimental data and then tested on the remaining data. Taking into account the limited amount of training data, the results obtained from both fault detection and NOx correction models demonstrate the potential of the proposed approach: the considered engine faults are detected and identified with limited misclassification errors by the developed models, while the applied data-driven NOx emission correction allows to strongly reduce the deviation between actual and estimated emissions compared to the reference ECU model, with a good matching of the instantaneous emission trend and a residual error on the cumulated emission that is below 5% in most of the test cases.

2.4   Organization of the dissertation

Following this introduction to the research topics and the presented novel contributions, the dissertation is organized as follows:

- in Chapter 3, the vehicle and powertrain model of the LCV PHEV under study is analysed, with a deeper focus on the Mean Value Engine Model (MVEM), being the core of the considered control-oriented model for the specific application described in this dissertation. Then, a simplified HiL simulation environment is described, with the main target of assessing the real-time capability of the developed data-driven models;

- in Chapter 4, the effects of selected emission-relevant faults are investigated by means of a set

of simulations. Then, the methodology followed to develop several FDI and OBM-oriented data-driven models based on simulation data is comprehensively described, together with the implementation of the developed models on a real-time hardware;

- in Chapter 5, the results obtained applying the optimized FDI and NOx emission models on selected test cases are presented and analysed in terms of accuracy and real-time capability;

- in Chapter 6, the experimental setup used for real-world data acquisition is described. According to the same approach followed in simulation, the effects of the induced emission-relevant faults on s set of on-board-measurable quantities are investigated, together with the performance of built-in ECU emisssion models to be used as reference;

- in Chapter 7, selected data-driven models, based on the results obtained in simulation, are trained on the experimental dataset acquired at the test bench. The training and optimization process of both fault classification and NOx regession models is extensively described, investigating different approaches in order to select the most effective one for this specific application;

- in Chapter 8, the results obtained applying the developed FDI and NOx emission models to a real-world testing dataset are presented and deeply discussed;

- in Chapter 9, the achievements of this work are summarized and analysed, together with the possible future works to overcome the limits of this Ph.D. research project.

Part I

Simulation

Chapter 3

Simulation environment

In this chapter, the focus is on the simulation platform that has been employed to evaluate the effects of selected faults on engine-out emissions, and then to generate the training dataset and assess the performance of the developed data-driven fault detection and NOx emission models. To this end, an overview of the vehicle under test is given, together with the adopted modelling approach. Moreover, the models of vehicle dynamics and the main powertrain components are introduced and analytically described, even if a more detailed and complete analysis can be found in [170]. Particular attention is paid to the description and validation of the Mean Value Engine Model (MVEM), which plays a crucial role in estimating engine emissions and the effects that each fault has on them. An overview of model validation on experimental data is provided as well. Finally, the architecture of the Hardware-in-the-Loop (HiL) environment used to test the real-time capability of the developed data-driven models is presented and shortly explained.

3.1   Modelled vehicle

The modelled vehicle is a prototype Diesel plug-in hybrid Light Commercial Vehicle (LCV) property of FEV, which is currently used for research purposes. The vehicle is equipped with a 2-liter compression ignition engine providing a rated power of 120 kW, with a double-stage turbocharger and High-Pressure (HP) EGR. The engine is coupled with a Diesel Euro 6-compliant ATS, which includes a Diesel Oxidizing Catalyst (DOC), a Diesel Particulate Filter (DPF), and a Selective Catalytic Reduction (SCR) catalyst. The considered hybrid architecture, represented Fig. 3.1, provides high flexibility in terms of possible operating modes and torque split strategies. The location of the electric machines, thanks to the separation clutch, allows to perform effective regenerative braking as well as pure electric drive, while the combination of a 15 kWh High-Voltage (HV) battery and a 50 kW/350 Nm Electric Motor (EM) ensures a significant all-electric range. All the main vehicle specifications are listed in Table 1 Tab. 3.1. The operating mode between pure-electric drive and hybrid drive is managed through a calibrated strategy based on vehicle speed,

requested torque and battery State of Charge (SoC). Focusing on hybrid operating mode, different configurations are possible thanks to the flexible architecture: the EM can provide additional torque to support the Internal Combustion Engine (ICE) when the driver request is particularly relevant or, on the contrary, it can operate as a generator to charge the battery with the engine providing additional torque with respect to the driver request, allowing to follow the best trade-off between fuel economy and pollutant emissions. A so-called range-extender mode is also possible when the vehicle is at stand-still and the engine is running for the sole purpose of charging the battery. Since out of the scope of this research activity, the same hybrid strategy calibrated and optimized within previous projects has been considered.
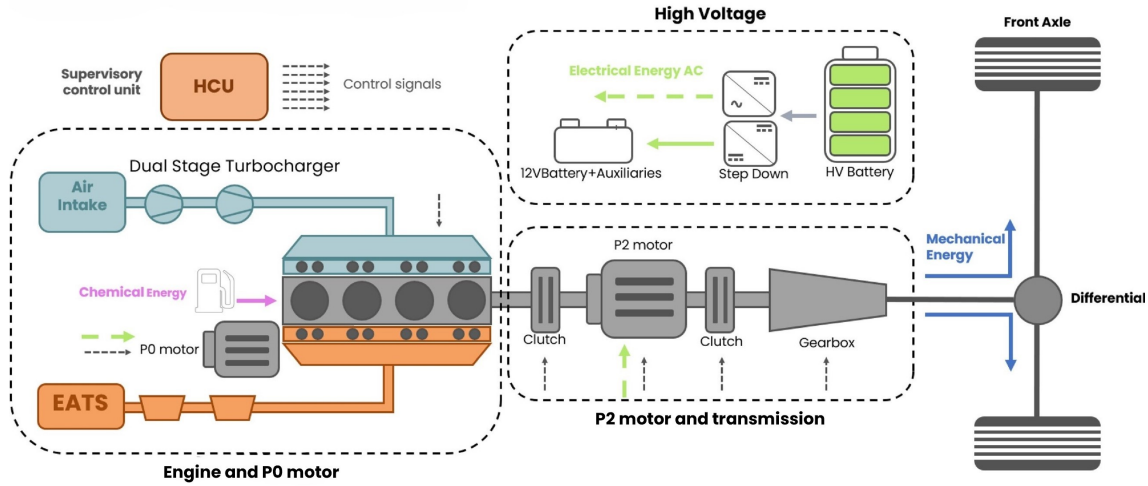
Figure 3.1: Prototype PHEV: schematic representation of vehicle hybrid powertrain architecture

From the simulation side, the prototype is entirely modelled in *Matlab* & *Simulink*. The models of the physical components are described in Section 3.2.1, while the models of the vehicle controllers in Section 3.2.2. All of them have been validated over experimental data and the results of the validation are shown in Section 3.3.

3.2 *Simulink* model

Nowadays, model-based development in the automotive industry is widely adopted since this approach allows to reduce the so-called time to market by simplifying the validation process, in order to be competitive in the fast changing automotive market. The availability of a reliable simulation tool is a key enabler for the design and the optimization of complex systems such as Hybrid Electric Vehicles (HEVs). The adoption of a 0-D quasi-static vehicle model developed in *Matlab* & *Simulink* environment is a common practice in the automotive field when the main focus is to simulate fuel consumption, pollutant emissions or vehicle performance over long duration driving cycles, providing a good trade-off between simulation accuracy and computational time. In

Table 3.1: Prototype PHEV: Vehicle powertrain specifications

Vehicle	Category	Plug-in LCV
	Curb weight	3330 kg
	Configuration	P0P2, front-wheel drive
Engine	Type	Diesel, turbocharged
	Displacement	2.0 l
	Rated power	120 kW @ 3600 rpm
	Rated torque	400 Nm @ 1800 rpm
	Gearbox	8-AT
Electric machine	Type	PM synchronous
	Continuous power	50 kW
	Continuous/peak torque	250/350 Nm
	Maximum speed	7000 rpm
Battery	Type	Lithium-ion polymer
	Capacity	15 kWh / 50 Ah
	Nominal voltage	300 V

this regard, a comprehensive 0-D control-oriented vehicle model has been developed in *Simulink*, starting from already existing and previously validated simulation platforms made available by FEV Europe GmbH.

A "forward-backward" approach has been chosen for vehicle modelling. In this approach, the powertrain is dynamically modelled with a "forward" approach (also known as "dynamic" approach), meaning that the energy flows from the propellers to the wheels and the vehicle speed is the result of the control chain. The latter is based on a "backward" model, and so on the inverted path of the energy flow inside the vehicle, used by the driver model to compute the torque set-point. On one hand, the accuracy of the speed-following control is increased due to the feedback of the effective vehicle speed, as it would happen for a real driver. On the other hand, the dynamic powertrain model ensures the power limitations of each component, introduced by the forward-facing energy flow. A more detailed explanation of this approach can be found in [171].

From an architectural point of view, the model is structured in modular blocks that are organized in different levels, so that a block can contain further sub-blocks modelling specific components or performing more specific functions. Each block has its own input and output signals and all blocks are connected to each other by means of a bus creator that transmits all the signals. The model top layer is organized in three different sub-systems: physical, controller and scope subsystem. The first two subsystems contain respectively all vehicle components physical models and corresponding control units, while the third one allows to visualize all the main signals that can be useful for monitoring the model. A schematic representation of model architecture is given by Fig. 3.2. An overview of the main functions and equations modelled in each block inside physical and controller subsystems is given in the next paragraphs.
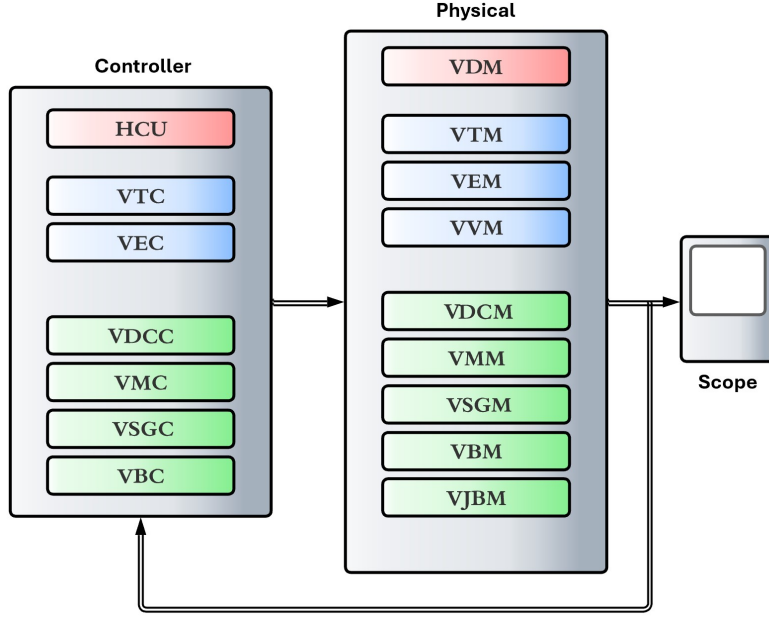
Figure 3.2: Schematic representation of model architecture including main physical and controller subsystems

3.2.1 Vehicle physical subsystems

Vehicle model (VVM) Given the objective of the present work, the vehicle performance in terms of drivability (vehicle stability, handling, etc.) and comfort (noise, harshness, and vibrations) are not investigated. For this reason, lateral and horizontal vehicle dynamics are not taken into account. Hence, vehicle longitudinal dynamics is described by the following fundamental equations:

$$m_v \frac{d}{dt} v(t) = F_t(t) - F_{res}(t) \tag{3.1}$$

$$F_{res}(t) = F_a(t) + F_r(t) + F_g(t) \tag{3.2}$$

where m_v is the vehicle mass, v is the actual vehicle speed, F_t is the propulsion force and F_{res} is the resistance force acting on the vehicle, calculated as the sum of aerodynamic, rolling and road gradient related resistance. Aerodynamic and rolling resistance forces depend on vehicle speed and they are modelled through experimental coast-down coefficients f_0, f_1 and f_2, so that the total resistance force becomes:

$$F_{res}(t) = f_0 + f_1 v(t) + f_2 v^2(t) + m_v g \cdot sin \alpha(t) \tag{3.3}$$

where α is the road slope. Another important output of this block, together with vehicle actual speed and acceleration, is the maximum wheel torque request, which is calculated according to

grip limitations.
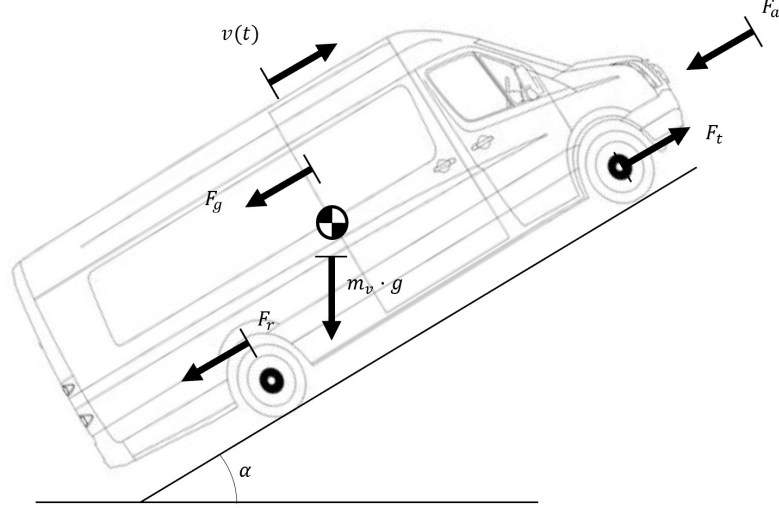
Figure 3.3: Longitudinal forces acting on a vehicle in motion

Driver model (VDM)   It contains information about the considered driving cycle, in terms of speed profile and road gradient over time, and simulates the driver behaviour. The torque requested at the wheels Tq_{req} in order to follow the speed profile is calculated through the equations:

$$F_{req}(t) = F_{res}(t) + (m_v + m_{eq,rot}) \cdot \frac{d}{dt} v_{req}(t) \tag{3.4}$$

$$Tq_{req}(t) = F_{req}(t) \cdot r_w \tag{3.5}$$

where F_{req} is the total force requested at the wheels, $m_{eq,rot}$ is the equivalent mass of rotative components, r_w is the dynamic wheel radius and v_{req} is the vehicle speed imposed by the drive cycle. Then a PI controller simulates the driver action in terms of positive or negative torque request to perform an acceleration or a braking action, based on the difference between requested speed and actual vehicle speed feedback. More in detail, braking torque request $Tq_{brk,ct}$ after controller intervention is calculated using the following equation:

$$Tq_{brk,ct} = (v_{req} - v_{act}) \cdot f_{P,brk} + f_{I,brk} \cdot \int (v_{req} - v_{act}) \cdot dt \tag{3.6}$$

where $f_{P,brk}$ and $f_{I,brk}$ are the proportional and integral factors for the PI controller, while v_{req} and v_{act} are the vehicle speed requested by the drive cycle and the actual speed respectively. The integral in Eq. (3.6) is reset whenever actual vehicle speed drops below a specific threshold. On the propulsion side, the PI controller adjusts the torque requested at the wheels, calculated as in

Eq. (3.5), based on the error between requested and actual vehicle speed:

$$Tq_{req,ct} = (v_{req} - v_{act}) \cdot f_{P,acc} + f_{I,acc} \cdot \int (v_{req} - v_{act}) \cdot dt \tag{3.7}$$

being $Tq_{req,ct}$ the torque requested at the wheels after the PI controller intervention and $f_{P,acc}$ and $f_{I,acc}$ the proportional and integral factors of the controller as seen in Eq. (3.6).

Engine model (VEM)   It provides engine actual torque and speed based on different operating conditions and engine torque request coming from the Engine Control Unit (ECU), taking into account inertia effects; a simplified engine thermal model simulates the heat-up of oil, coolant and engine block, while fuel consumption and engine-out emissions of HC, CO, soot and NO_x are calculated by means of an accurate MVEM, described in detail in Section 3.2.3.

Transmission model (VTM)   A physical transmission model calculates the torque transmitted through the launch clutch, between the electric motor and the gearbox, and the separation clutch, between the engine and the electric motor, according to clutch states (locked, unlocked, slip control). A second subsystems represents the transmission thermal model, in which also transmission efficiency is calculated using a map-based approach depending on temperature, torque and active gear. A state machine performs the gear change when requested.

EM model (VMM)   The model accounts for the calculation of motor speed, torque and actual electric power consumption, based on torque request coming from the EM controller, considering inertia losses and eventually engine drag torque. The core is made by experimental efficiency maps, which are used to take into account the EM losses both in motor and generator mode as function of speed and load. Motor efficiency is then used to calculate the electric energy requested to the HV battery, based on the mechanical and electrical power balance, as follows:

$$\begin{cases} P_m = 2\pi \cdot Tq_{mot,act} \cdot \frac{n_{mot,act}}{60} \\ P_{el} = P_m \cdot \eta_c^k \cdot \eta_m^k \end{cases} \tag{3.8}$$

where P_m is the mechanical power, $Tq_{mot,act}$ is the actual motor torque and $n_{mot,act}$ is the actual motor speed. On the electric side, η_c is the inverter efficiency, η_m is the mechanical efficiency, k is a constant, equal to -1 when the EM operates as a motor and equal to 1 when it operates as a generator), and P_{el} is the electric power request sent to the HV battery. Finally, the electric motor rotational speed is computed starting from actual vehicle speed v_{act} through the following equation:

$$n_{mot,act} = \frac{v_{act}}{r_w} \cdot \tau_{gr} \cdot \tau_{fd} \tag{3.9}$$

being r_w the dynamic wheel radius, τ_{gr} the current gear ratio, and τ_{fd} the final drive ratio.

Belt-driven Starter Generator model (VSGM)   The physical representation is analogous to that of the already introduced VMM.

HV battery model (VBM)   The physical component is modelled through an equivalent RC circuit that allows to calculate battery voltage V_{bat} and current I_{bat} through the following system of equations:

$$V_{bat}(t) = (V_{cel,OC}(T,SoC) - R_{cel,int}(T,SoC) \cdot I_{bat}(t)/n_p) \cdot n_s \tag{3.10}$$

$$P_{bat}(t) = V_{bat}(t) \cdot I_{bat}(t) \tag{3.11}$$

where n_p, n_s are the number of parallel and series cells, P_{bat} is the battery power request coming from the electric motor and the DC/DC converter, $R_{cel,int}$ and $V_{cel,OC}$ are respectively cell internal resistance and open circuit voltage, which are obtained from experimental maps as function of temperature and battery SoC. Power losses due to resistive heating are then calculated as follows:

$$P_{bat,loss}(t) = V_{bat,int} \cdot I_{bat}(t) \tag{3.12}$$

$$V_{bat,int} = \frac{n_s}{n_p} R_{cel,int}(T,SoC) \cdot I_{bat}(t) \tag{3.13}$$

where $V_{bat,int}$ is the internal voltage drop due to battery internal resistance. Battery actual state of charge SoC is determined through the following equation:

$$SoC(t) = SoC_0 - \frac{\int_0^t I_{bat}(t)\,dt}{C_{bat}} \tag{3.14}$$

where SoC_0 is the initial state of charge and C_{bat} is the nominal battery capacity. A thermal model calculates battery temperature depending on power losses and thermal exchange with air and coolant.

DC/DC converter model (VDCM)   It translates the power requested to the 12 V battery into power requested to the HV battery to maintain a fixed SoC of the 12 V battery, taking into account its efficiency.

12V battery model   The physical representation is analogous to that of the HV battery.

Junction box model (VJBM)   It represents the physical connection between all HV consumers, in this case the electric motor and the DC/DC converter.

3.2.2   Vehicle controller subsystems

Hybrid Control Unit (HCU)   It controls powertrain operation and acts as a supervisor for all the other slave controllers. First of all, a specific function determines the driving mode between pure

electric drive (e-drive) or hybrid mode, according to the implemented hybrid strategy. Hybrid mode is further divided into different powertrain operating modes including engine start phase, engine load point shift, range extender mode and engine stop phase. Another function evaluates the negative torque requested to the electric motor according to the driver request, and eventually sends the additional brake torque request to the vehicle model, so that regenerative braking is always maximized according to electric motor and battery limits. Engine and motor torque request are determined according to the operating mode and sent to the corresponding controllers. Another block calculates the requested gear according to the adopted shifting strategy based on speed thresholds maps depending on vehicle speed and requested torque.

Engine controller model (VEC)   The first of the different functions included in the ECU model gives information about the current engine status depending on operating conditions (e.g. start, stop, idling, fuel cut-off etc.); another function acts on engine torque request according to HCU request taking into account engine operating mode, idle speed control and torque limitations; the final torque request includes also engine friction losses and cold start measures (e.g. cat-heating phase). All the input signal needed by the physical engine model, and more specifically by the MVEM, are calculated in a dedicated subsystem that has been developed following an architecture inspired by real ECUs, which is also very similar to the one of the physical MVEM (Section 3.2.3). Therefore, all the main physical quantities are calculated within the VEC for control purposes, thanks to the same semi-empirical models implemented in the engine physical subsystem. Moreover, the engine controller receives all the signals coming from the physical engine model, that on a real vehicle would correspond to the measurements from on-board sensors.

Transmission controller (VTC)   It controls through a state machine the operating conditions of both launch and separation clutch, according to requested torque, engine and motor states and speed, shifting request, and transmission torque limits.

Electric motor controller (VMC)   It contains limitations on the torque requested to the electric motor, obtained from motor continuous and peak torque curves. A state machine determines motor operating modes preventing it to work in peak operation for too long, which in reality can cause overheating issues and consequent power limitation;

Battery controller (VBC)   It contains map-based power limitations according to battery peak or continuous operation modes.

3.2.3   Mean Value Engine Model

As anticipated, considering the focus of this research activity, the Mean Value Engine Model (MVEM) can be considered as the core of the simulation environment, which includes all the

main engine sub-models, namely the air system that determines how much air is introduced into the cylinder; the fuel system that determines how much fuel is injected into the cylinder; the torque generation system that determines how much torque is produced by the air and fuel in the cylinder as determined by the first two parts; the engine thermal system that determines the dynamic thermal behaviour of the engine; the pollutants formation system that models the engine-out emission. As discussed in 2.2.1, MVEMs are control-oriented models by definition, meaning that they model the input-output behaviour of the systems with reasonable precision but low computational complexity. The reciprocating behaviour of the engine is not considered, neglecting the discrete engine cycles and assuming that all processes and effects are spread out over the engine cycle, still including all relevant transient effects [107]. The MVEM considered for this application combines physics-based models, which are based on physical principles and on a set of experiments necessary to identify some key parameters, with a model-based ECU control structure, leading to a flexible simulation software that can be eventually adapted to be representative of different hardware architectures [172].

As already introduced in Section 3.1, the modelled engine is a 2-liter 4-cylinder compression ignition engine equipped with a high-pressure EGR system and a double-stage turbocharger, including Variable Geometry Turbine (VGT) in the high-pressure stage. The engine is coupled with a Diesel Euro 6-compliant ATS, which includes DOC, DPF, and SCR catalyst. A schematic representation of engine layout and main related components, installed sensors, and actuators, is given in Fig. 3.4, highlighting the most significant signals sent from the on-board sensors to the engine controller (solid line), listed in Tab. 3.2, and from the controller to the actuators (dashed line). The approach that has been followed to calculate the most relevant physical quantities within the adopted MVEM is presented in the next paragraphs, together with exemplary related equations.

Mass flows calculation It is based on the principle of mass conservation and on generic flow equations. More in detail, starting from the cylinder filling model, the cylinder external charge $m_{cyl,ext}$ is calculated as:

$$m_{cyl,ext} = \eta_v \cdot \frac{P_{\text{InMfd}}}{T_{\text{InMfd}} \cdot r_{\text{air}}} \cdot V_{cyl,efc} \cdot k_{T_{\text{InMfd}}} \cdot k_{T_{Co}} \tag{3.15}$$

where η_v is the volumetric efficiency, P_{InMfd} and T_{InMfd} are the pressure and temperature in the intake manifold respectively, r_{air} is the specific gas constant for the air, $V_{cyl,efc}$ is the effective volume that can be filled with fresh charge considering the volume occupied by residual gas at exhaust valve closure and the expansion of residual gas from exhaust to intake manifold pressure. $k_{T_{\text{InMfd}}}$ and $k_{T_{Co}}$ are two experimentally-derived correction parameters based on intake manifold temperature and engine coolant temperature respectively. For the calibration of the cylinder filling model, the determination of several parameters is needed, including the compression volume profile, base maps of volumetric efficiency, reference maps for intake manifold temperature and
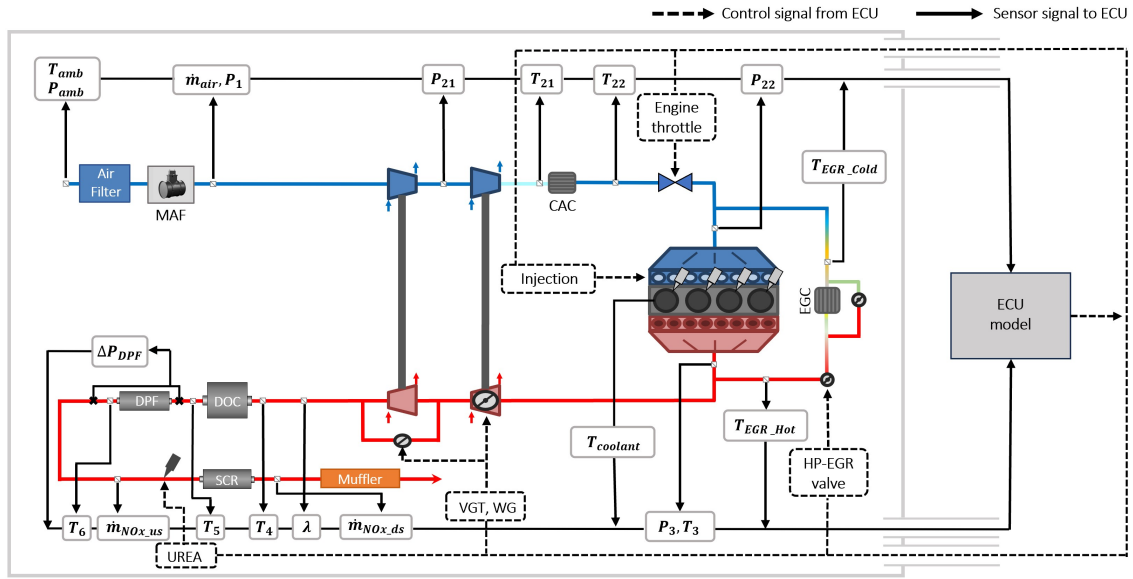
Figure 3.4: Modelled engine architecture including sensors/actuators layout

Table 3.2: List of available on-board sensors signals according to real vehicle layout

Abbreviation	Description
T_{amb}	Ambient temperature
P_{amb}	Ambient pressure
$\dot{m}_{air}$	Intake air mass flow
P_1	Pressure upstream low-pressure compressor
P_{21}	Pressure upstream high-pressure compressor
P_{22}	Intake manifold pressure
P_3	Exhaust manifold pressure
T_{21}	Temperature downstream high-pressure compressor
T_{22}	Temperature downstream Charge Air Cooler (CAC)
$T_{EGR-Hot}$	Temperature upstream EGR valve
$T_{EGR-Cold}$	Temperature downstream EGR cooler
T_3	Exhaust manifold temperature
T_4	Temperature downstream low-pressure turbine
T_5	Temperature downstream DOC
T_6	Temperature downstream DPF
ΔP_{DPF}	Differential pressure across DPF
λ	Oxygen concentration after combustion
$T_{coolant}$	Engine coolant temperature
$\dot{m}_{NOx_{us}}$	NOx mass flow upstream SCR
$\dot{m}_{NOx_{ds}}$	NOx mass flow downstream SCR

engine coolant temperature as a function of engine speed and load, as well as the related correction parameters for each engine operating point. If no low-pressure EGR system is present, as for this specific application, the air mass flow $\dot{m}_{cpr}$ going through the two stages of compression and the exhaust gas mass flow $\dot{m}_{exh}$ entering the high-pressure turbine can be defined by the following equations:

$$\dot{m}_{cpr} = \dot{m}_{cyl,ext} - \dot{m}_{HpEgr} \tag{3.16}$$

$$\dot{m}_{exh} = \dot{m}_{cpr} + \dot{m}_f \tag{3.17}$$

where $\dot{m}_{HpEgr}$ is the high-pressure EGR mass flow, calculated as a function of EGR valve position, pressure and temperature upstream the EGR valve and the ratio between upstream and downstream pressure, while $\dot{m}_f$ is the injected fuel mass flow.

Pressures calculation   Two different approaches are adopted for the calculation of the pressure levels along the air path. More in detail, as schematically represented in Fig. 3.5, pressures in the intake side are calculated starting from ambient pressure according to a forward calculation approach, taking into account pressure losses in the air filter (ΔP_{AirFt}) and in the intercooler (ΔP_{IntCo}), modelled as function of the product between mass flow and volume flow, as well as the compression ratios of the two compression stages (β_{HpCpr} and β_{LpCpr}). On the other hand, backward calculation is employed for the determination of pressure levels in the exhaust side, starting from the pressure downstream low-pressure turbine and considering the pressure ratios of the two expansion stages (β_{HpTrb} and β_{LpTrb}). As for the cylinder filling model, there are some parameters that need to be experimentally determined, including the curves of pressure losses across filters and heat exchangers, and the time constants of applied filters to match pressure signals behaviour during transients.
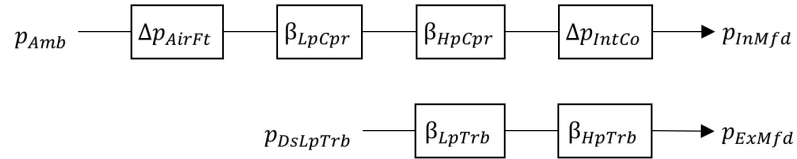
Figure 3.5: Forward (top) and backward (bottom) approach for air path pressures calculation

Temperatures calculation   Differently from pressure estimation, a forward calculation approach is adopted for the determination of temperature levels in the intake and exhaust sides starting from ambient temperature until turbine outlet temperature, as schematized in Fig. 3.6.

More in detail, the temperature downstream of the intercooler ($T_{dsIntCo}$) is calculated from the upstream temperature ($T_{usIntCo}$), considering cooler efficiency η and temperature of the coolant
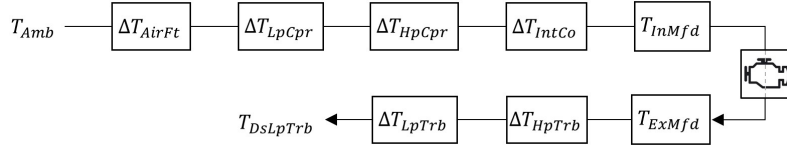
Figure 3.6: Forward approach for air path temperatures calculation

fluid at the cooler inlet $T_{co,in}$, according to the following equation:

$$T_{dsIntCo} = T_{usIntCo} + \eta_{co} \left(T_{usIntCo} - T_{co,in} \right) \tag{3.18}$$

Cooler efficiency is defined based on gas temperature upstream of the cooler and mass flow across the cooler by means of an experimentally calibrated map. In addition to the efficiency map, as seen for pressure calculation, the time constants of applied filters to account for the thermal inertia of involved components need to be calibrated as well through experimental measurements, in order to match gas temperature dynamics. Intake manifold temperature (T_{InMfd}) is calculated considering an adiabatic mixing between the air coming from the Charge Air Cooler (CAC) and high-pressure EGR gases, described by the following equation:

$$T_{\mathrm{InMfd}} = \frac{T_{dsIntCo} \cdot c_{p,\mathrm{dsIntCo}} \cdot \dot{m}_{\mathrm{IntCo}} + T_{HpEgr} \cdot c_{p,HpEgr} \cdot \dot{m}_{HpEgr}}{c_{p,\mathrm{dsIntCo}} \cdot \dot{m}_{\mathrm{IntCo}} + c_{p,HpEgr} \cdot \dot{m}_{HpEgr}} \tag{3.19}$$

where $c_{p,\mathrm{dsIntCo}}$ and $c_{p,HpEgr}$ are the specific heat capacity of the air downstream the intercooler and of the high-pressure EGR gases respectively, while T_{HpEgr} is the temperature of the high-pressure EGR gases. A further extension of this model has been considered to account for the impact of residual gas back flow, based on difference between in-cylinder pressure at Intake Valve Closing (IVO) and intake manifold pressure, assuming an instantaneous back flow of residual gas into intake manifold and no variation of pressure in the intake manifold due to residual gas back flow. Exhaust manifold temperature is calculated as follows based on the converted share of fuel energy content:

$$T_{\mathrm{ExMfd}} = T_{\mathrm{InMfd}} + \frac{m_f \cdot \Delta h}{m_{cyl} \cdot c_p} \tag{3.20}$$

where T_{InMfd} is the intake manifold temperature, m_f and m_{cyl} are the mass of injected fuel and the air mass in the cylinder respectively, c_p is the air specific heat capacity, and Δh is the specific enthalpy difference. The latter is modelled considering the impact of air-fuel ratio, combustion phasing, cylinder charge, exhaust valve timing and engine coolant temperature. A heat transfer model is employed to calculate proper corrections to account for convective heat exchange between the exhaust gas and convective and radiative heat exchange between the manifold and the ambient air. As seen for the other models previously introduced, also in this case there are some parameters that need to be experimentally mapped and calibrated, including the specific enthalpy difference

for each engine operating point and the time constants of implemented filters to emulate gas temperature and sensors dynamics.

Pollutant emissions calculation   The most relevant engine-out raw emissions, namely NOx, HC, CO, and soot, are properly predicted within the engine model with a semi-physical approach, reaching accurate results also for dynamic responses thanks to a set of corrections based on oxygen concentration, temperatures, pressures, or combustion properties (center of combustion, rail pressure, etc.) [173]. Focusing on NOx, the semi-empirical engine-out emission model is based on the main relationship between in-cylinder oxygen concentration before combustion to which the above-mentioned corrections are applied, accounting for the main parameters impacting the reaction temperature. More in detail, the impact of oxygen concentration, as well as the one of the other considered parameters, is treated as a variation from reference conditions. This means that reference maps for each involved parameter need to be generated as a function of engine speed and load, as well as related correction factors maps. However, this approach provides significant benefits, as high accuracy in transient cycles and good extrapolation capability.
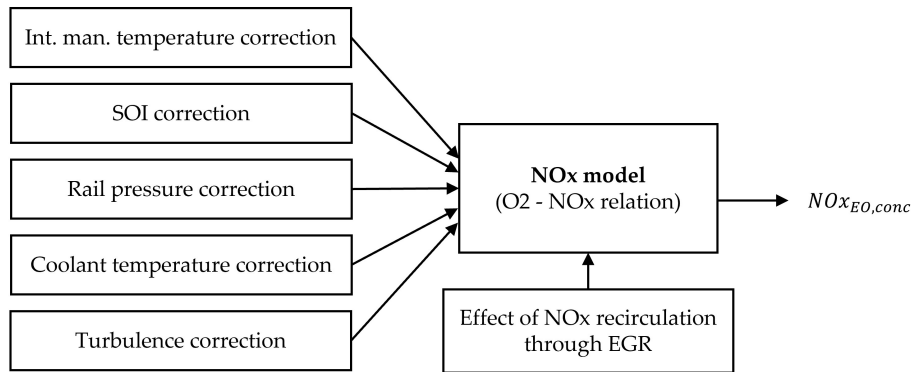
Figure 3.7: Modelling approach for engine-out NOx concentration

The resulting tailpipe pollutant emissions after oxidizing/reducing reactions that take place in the ATS are modelled with a map-based approach, by first calculating the conversion efficiency of each ATS component, as a function of gas average temperature, space velocity, and NO_2/NOx ratio (in the case of NOx conversion efficiency). Since out of the scope of the dissertation, for a comprehensive description of the ATS modelling approach, the reader is referred to [174]. All the mentioned empirical maps and calibrations for the engine and ATS models have been obtained from experimental tests previously carried out within FEV and the related data are confidential. Further considerations on the modelling approach and involved physical equations can be found in [27].

3.3 Engine model validation

As part of a previous FEV project, the presented mean value engine model has been firstly calibrated on steady-state test bench measurements for what concerns all the empirical maps function of engine operating point. Then, transient measurements, namely load steps and cycles, have been used to calibrate all the parameters and time constants to refine gas pressure and temperature dynamics models, as well as modelled sensors dynamics. Finally, the calibrated model has been validated under steady-state conditions and on different driving cycles, such as the one shown in Fig. 3.8. For the sake of brevity, a selection of the most significant results of the validation process is presented, including intake air mass flow (Fig. 3.9), intake manifold pressure and temperature (Fig. 3.10), exhaust manifold pressure and temperature (Fig. 3.11), injected fuel mass (Fig. 3.12) and EGR flow rate (Fig. 3.13). Considering the main focus of this dissertation, a reliable NOx estimation is a preliminary requirement. In this respect, Fig. 3.14 shows the experimental and simulated NOx emission trends. For reasons of confidentiality, all the considered signals have been normalized with respect to the maximum value observed in the corresponding plot. For the sake of clarity, it is worth specifying that although the corresponding models were calibrated and validated under both steady-state and transient conditions, the emissions of the other main pollutants besides NOx (i.e. HC, CO, and soot) were not validated on experimental data from real driving cycles, which, on the other hand, is considered to be outside the scope of this research activity.
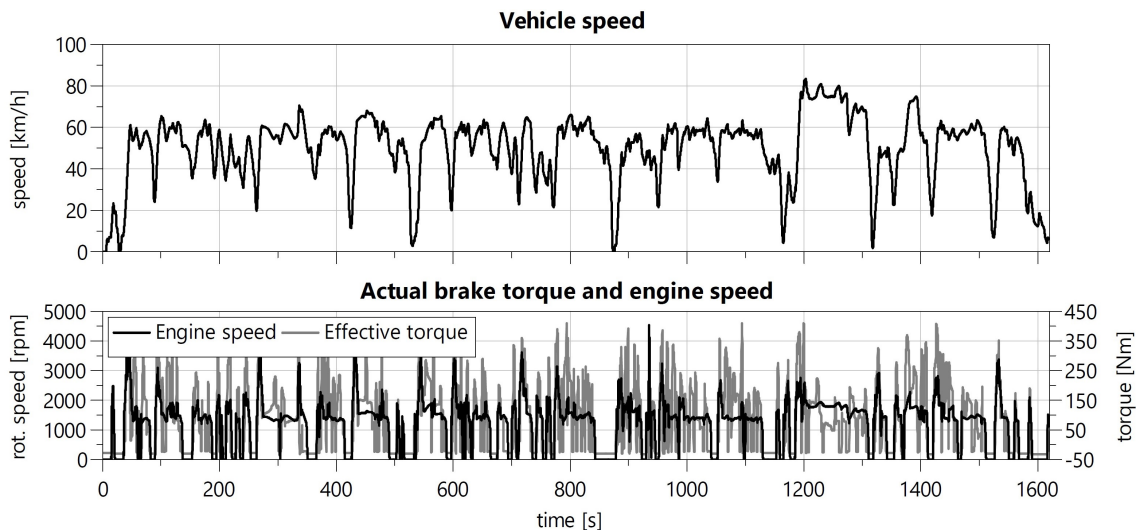

Figure 3.8: Validation real driving cycle

The comparison of modelled and measured signals demonstrates the capability of the model of capturing the correct trends of the mentioned physical quantities. Fig. 3.15 shows a comparison of the overall measured and simulated cumulated air mass, injected fuel mass, and NOx emission

Figure 3.9: Air mass flow model validation: normalized experimental (black) and simulated (red) results over a real driving cycle

on the same driving cycle with related percentage deviations. As noticed also in Fig. 3.9, for the same engine torque and speed, an almost perfect matching between measured and simulated air mass is achieved, while an overestimation of 5% is observed for the total injected fuel mass, as well as an underestimation of 7% for the cumulated NOx emission, which is still considered a good level of accuracy for a control-oriented modelling approach. Therefore, with the adopted simulation approach, a realistic estimation of engine-out NOx emissions, fuel consumption, as well as pressure and temperature levels along the air path is given. Model accuracy is considered satisfactory for the purpose of investigating the potential of data-driven methods applied to on-board fault diagnosis and emission monitoring, as a preliminary step before considering the application of the same approach to real-world data, as presented in Part II of this dissertation.

3.4   HiL architecture

As a final step of the simulation process, the developed fault classification and NOx emission regression models have been tested at HiL to evaluate the feasibility of real on-board implementation in terms of requested computational load. In this regard, the considered HiL architecture includes the following components:
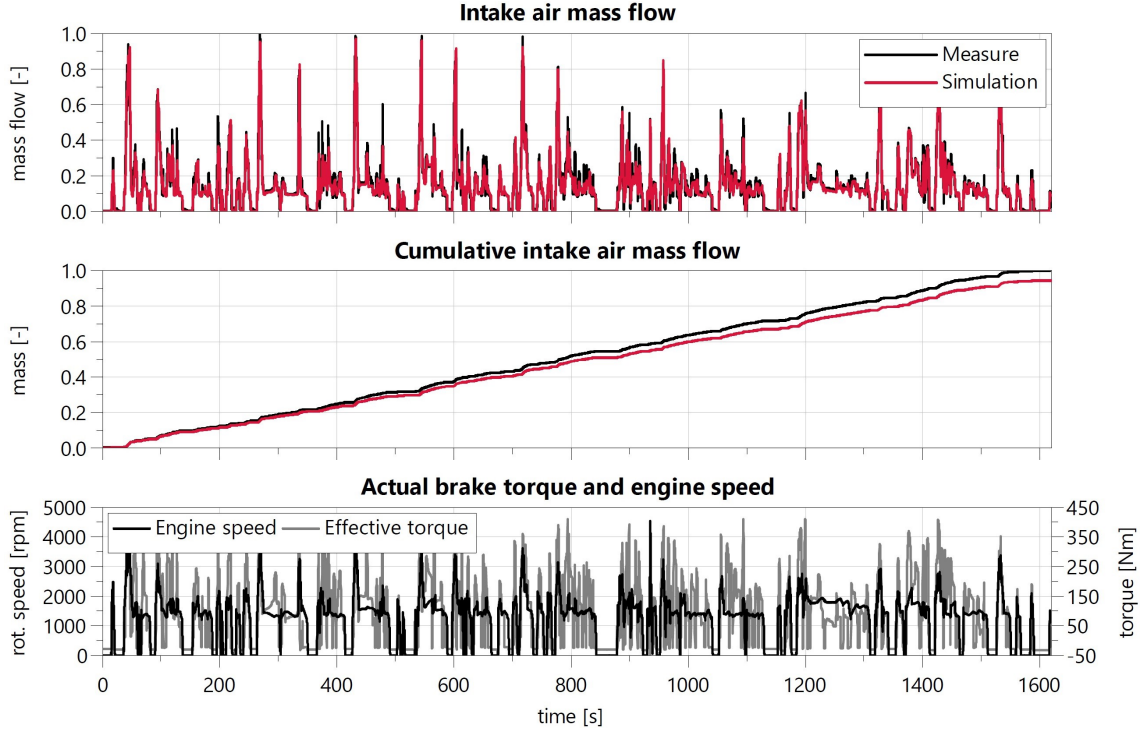
Figure 3.10: Intake manifold model validation: normalized experimental (black) and simulated (red) results over a real driving cycle

- a real-time PC, where the detailed vehicle and engine physical models described in Section 3.1 are implemented. More specifically, a SCALEXIO platform from dSPACE GmbH [175] has been used to simulate the environment in which a real ECU would operate;

- a rapid-prototyping control unit, specifically a MicroAutoBox II by dSPACE GmbH [176], where the ECU software is running;

- a single-board computer, more specifically a Raspberry Pi 4 [177], where the developed data-driven models to be tested have been deployed.

All the software, including the data-driven regression and classification models, have been written in *MATLAB & Simulink* and then deployed on the corresponding hardware. The engine controller software implemented in the MicroAutoBox, as well as the complete vehicle model, is proprietary of FEV Europe GmbH and fully open to modifications. The real-time PC is connected via CAN-bus to the rapid-prototyping ECU, allowing to feed the ECU control functions with realistic data sent through a real communication protocol. This way, all the useful physical signals can be sent from the real-time PC to the rapid-prototyping ECU, as it would happen with on-board sensors sending the measured signals to the real vehicle ECU. The same applies for the control signals sent from the ECU to the actuators, namely from the MicroAutoBox to the real-time PC.
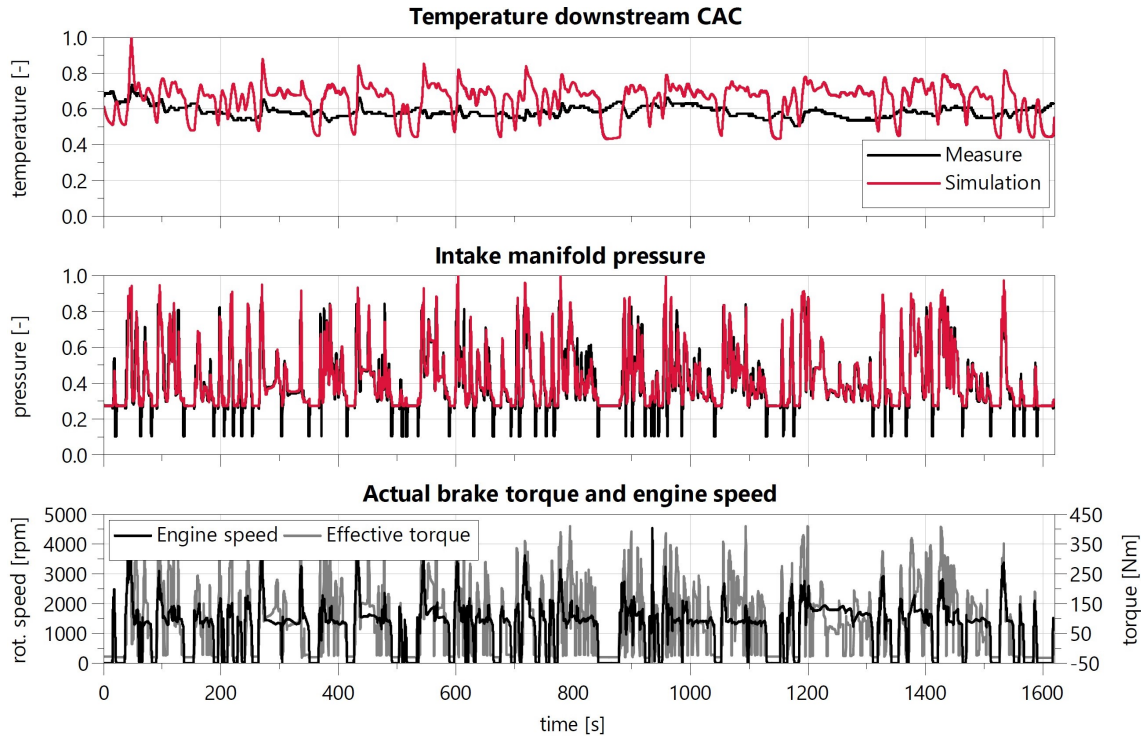
Figure 3.11: Exhaust manifold model validation: normalized experimental (black) and simulated (red) results over a real driving cycle

Finally, the Raspberry Pi is coupled with a PiCAN2 board [178] to manage the automotive CAN-bus standard communication and interact with the other hardware components, receiving all the signals needed by the implemented data-driven models. A detailed explanation of the complete HiL architecture, including hardware components that have not been used within this research activity, can be found in [179].

The main advantage of the described real-time hardware implementation, is that the Raspberry Pi is able to communicate via CAN-bus, meaning that once tested at the HiL, it would be ready to be connected to a real vehicle control unit, thus receiving all the available signals transmitted by the ECU and the on-board sensors. Taking advantage of information that is readily available on a standard vehicle, the proposed solution based on a Raspberry Pi computer ensures high modularity at low cost, without the need of modifying the already existing vehicle software.
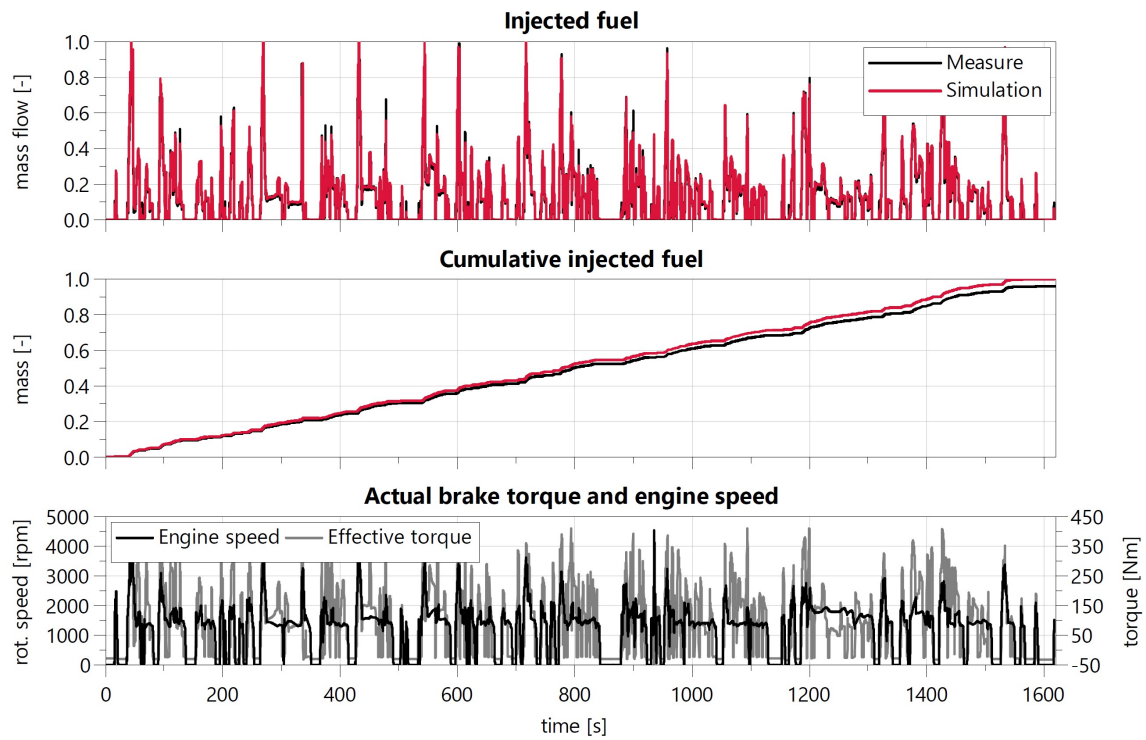
Figure 3.12: Injected fuel quantity validation: normalized experimental (black) and simulated (red) results over a real driving cycle

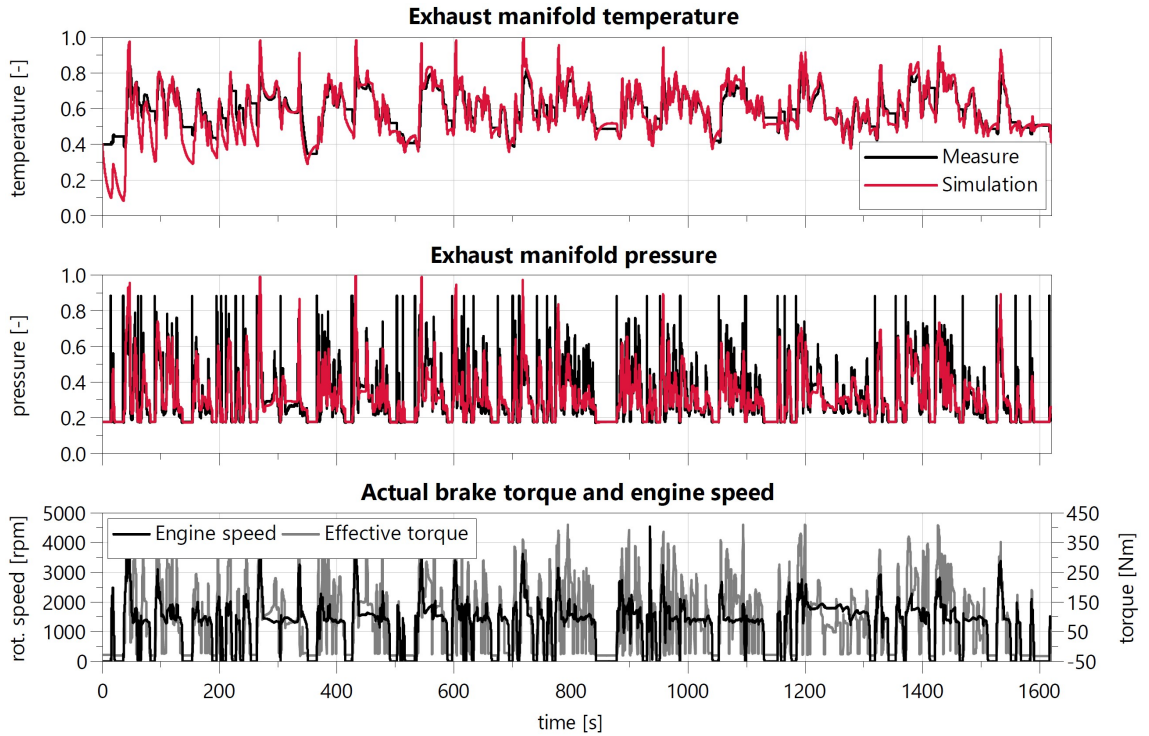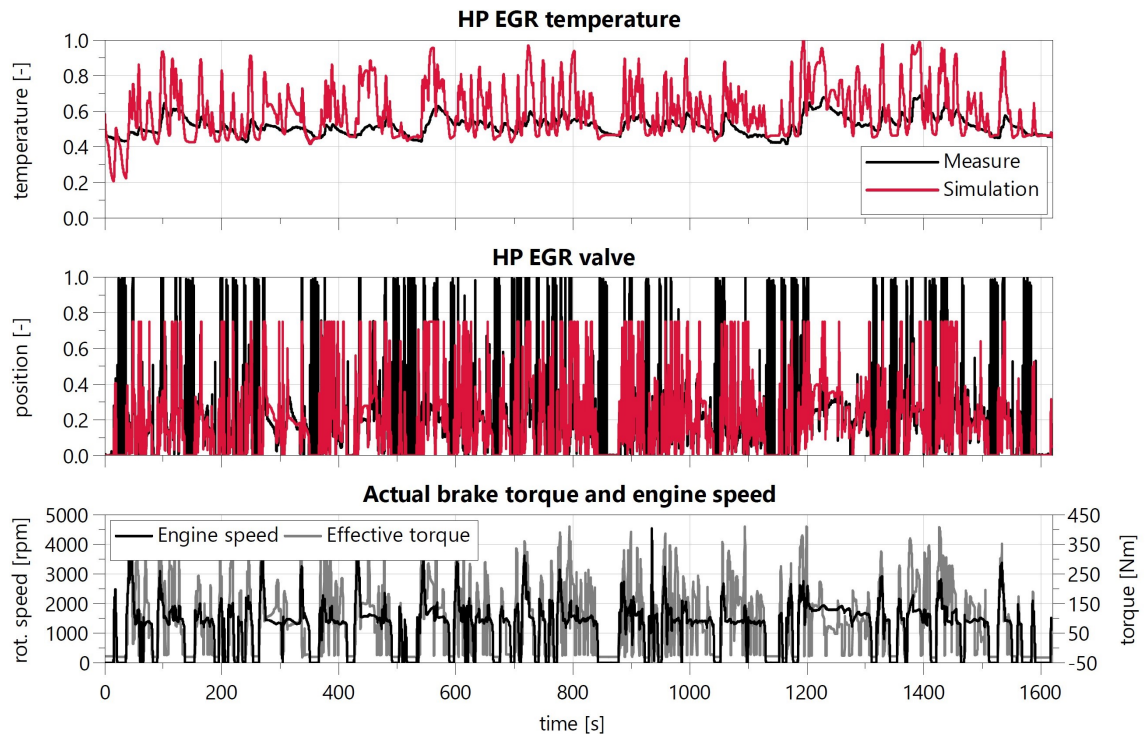**Figure 3.13:** HP EGR flow model validation: normalized experimental (black) and simulated (red) results over a real driving cycle
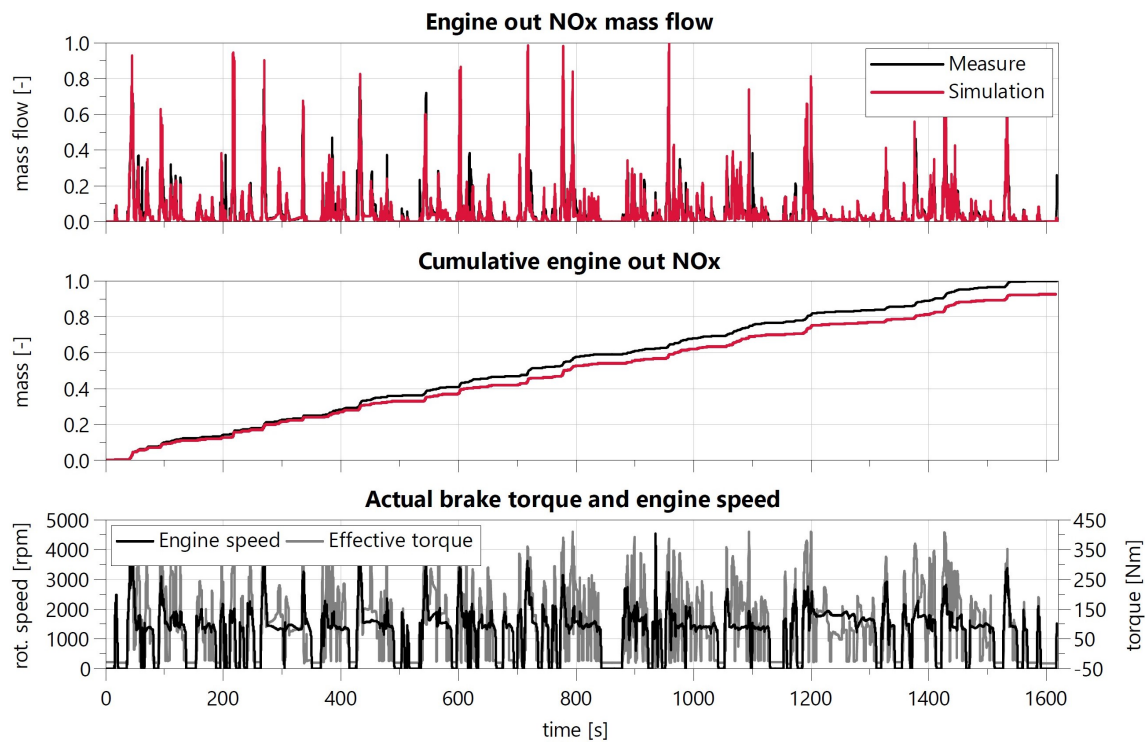
Figure 3.14: NOx model validation: normalized experimental (black) and simulated (red) results over a real driving cycle
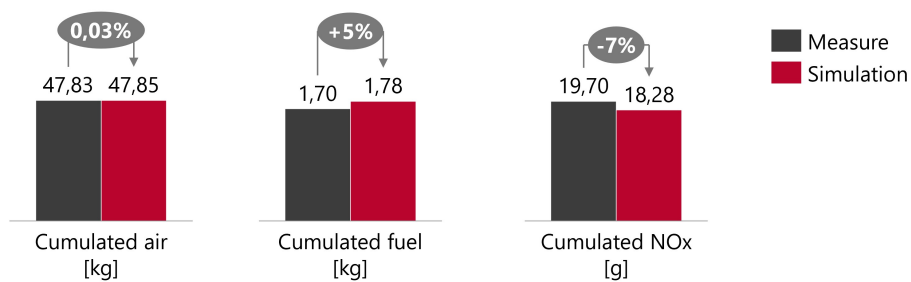
Figure 3.15: Comparison of measured and simulated cumulated air mass, injected fuel mass, and NOx emission on the considered real driving cycle with related percentage deviation

Chapter 4

OBM-oriented data-driven models

In this chapter, an overview of the simulated faults and their effects on pollutant emissions and on other relevant engine operating parameters is given. Then, based on this analysis, the most relevant signals to be considered as input features for the data-driven models have been selected. For both classification and regression models, an introduction to the investigated algorithms is provided, together with the description of the approach that has been adopted to train, optimize and define the most suitable data-driven models for the specific target application.
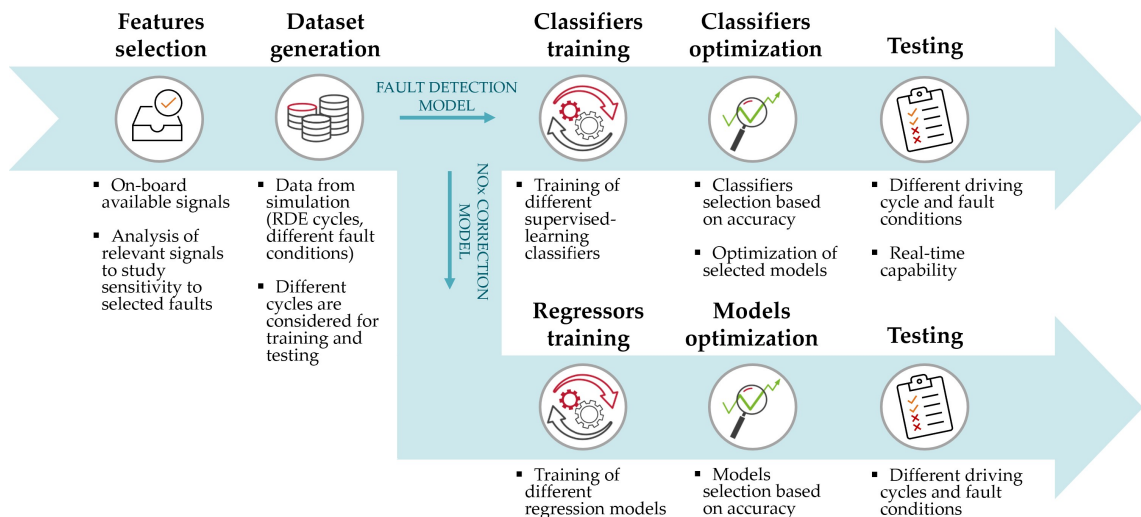
Figure 4.1: Work-flow of the data-driven models for on-board fault detection and NOx emission correction

4.1 Emission-relevant faults investigation

Different types of engine faults have been simulated to evaluate their effect on modelled engine-out pollutant emissions and the capability of the engine controller to react to these faults. A list of the investigated faults is presented below, together with a brief description of each of them and their implementation into the model.

- **MAF sensor low drift**: Mass air flow (MAF) sensors are used to determine the mass flow rate of air that enters the engine, usually based on hot-wire or hot-film technology. The drift error in real sensor readings may occur mainly due to sensor ageing, sensor contamination, or upstream plumbing changes. The result is that the sensor signal received by the ECU is deviating (more specifically it is lower) with respect to the actual value of the measured quantity, leading to poor accuracy on the control of the air-fuel ratio and especially of the EGR rate in Diesel engines, with significant consequences on the emission control. In simulation, it has been implemented by introducing a multiplying factor lower than 1 acting on the air mass flow signal sent from the physical engine model to the controller model;

- **MAF sensor high drift**: the situation is analogous to the previous one, but in this specific case the sensor signal received by the ECU is higher than the actual value of the measured air mass flow, leading to control issues and reduced engine performance. Therefore, this fault is simulated through a multiplying factor higher than 1 acting on the air mass flow signal sent from the physical engine model to the controller model;

- **MAP sensor low drift**: Manifold absoulte pressure (MAP) sensors supply instant manifold absolute pressure information to the ECU, which is needed to calculate air density and, based on that, the actual air mass flow rate entering the cylinder for an accurate control of the air-fuel ratio. In Diesel engines, they are also used to provide a feedback for the boost pressure target and for the control of the EGR rate. As for MAF sensors, the main causes of a degradation in sensor performance are ageing and excessive contamination, with significant consequences on engine performance and emissions. The sensor drift has been implemented into the model by applying a gain lower than 1 on the pressure signal sent from the physical to the controller model.

- **MAP sensor high drift**: this fault is analogous to the previous one, but in this case the sensor signal received by the ECU is higher than the actual pressure level in the intake manifold. Similarly to the previous ones, it is simulated through a multiplying factor, in this case higher than 1, applied to the air intake manifold pressure signal sent from the physical engine model to the controller model;

- **Reduced EGR cooler efficiency**: the function of an EGR cooler is to cool down the exhaust gases that are recirculated back from the exhaust manifold to the intake manifold. It is usually made of a variable number of tubes where the exhaust gases flow through, which

are surrounded by coolant supplied by the engine cooling system. Natural ageing of the system or possible coolant leaks could lead to a reduction of the cooling efficiency compared to the nominal one. This type of malfunction has been implemented into the physical engine model by applying a gain lower than 1 to the nominal EGR cooler efficiency;

- **EGR valve clogging**: the function of this valve is to actuate the control of the amount of exhaust gases recirculated to the intake manifold. Soot and carbon build-up from exhaust gases, as well as oil contamination, can cause clogging and locking of this valve. The fault has been simulated by introducing an upper saturation on the EGR valve opening into the physical engine model.

The mentioned faults are summarized in Tab. 4.1, together with some additional information about model implementation and effects on pollutant emissions. More in detail, starting from the left, the simulated fault is described and a brief explanation of how it has been implemented into the engine physical/controller model is given; then a qualitative summary effect on the main pollutant emissions, based on a deeper analysis involving several simulations in steady-state conditions for the different faults and degrees of severity, is expressed by the signs: plus (+) or minus (–) signs indicate whether an increase or decrease with respect to nominal conditions is observed, respectively; the number of signs, from 1 to 3, is related to the magnitude of the variation, as specified by the legend. On the left side of the table, the effect of each fault on actual pollutant emissions calculated by the physical engine model is presented, quantified as an increase or decrease with respect to nominal emission levels. On the right side of the table, the deviation between the emission level expected by the engine controller and the actual emissions calculated by the physical model in each fault condition is presented, as an index of the controller capability to adapt to non-nominal conditions. The analysis shows that almost all the considered faults have a very significant effect on NOx and soot emissions (up to more than 50% increase compared to nominal in some engine operating points) and that the physical-based models inside the engine controller, calibrated in nominal conditions, are not able to adapt themselves to correctly predict emissions in case of unexpected faults, despite the available virtual measurements coming from the engine physical model.

Moreover, when a fault is introduced, a variation with respect to nominal conditions is observed also for the physical quantities characterizing the intake and exhaust paths, including pressure, temperature, intake air mass flow, as well as oxygen concentration after combustion. In this respect, Fig. 4.2 allows to visualize the sensitivity of selected physical quantities and related controller estimations to each fault condition, based on steady-state simulations of different engine operating points, as for emission-related effects investigation (Tab. 4.1). As for the previous analysis, Fig. 4.2 shows the variation compared to the nominal case in terms of actual emissions calculated by the physical model for each simulated fault (grey bars), together with the deviation between the emission level estimated by the engine controller and the actual emissions calculated by the physical model (red bars). The considered signals have been selected among the on-board

Table 4.1: List of simulated engine faults and their related effects on resulting pollutant emission and controller emissions estimation

Fault	Model implementation	Effect on physical model (fault VS nominal case)			Effect on controller model (controller VS physical model)		
		NOx	HC, CO	soot	NOx	HC, CO	soot
Clogged EGR valve	Upper saturation on EGR valve opening in the physical model	⊕⊕⊕	⊖	⊖⊖	⊖⊖	⊕	⊕⊕⊕
Reduced EGR cooler (EGC) efficiency	Gain <1 on nominal EGC efficiency	⊕⊕⊕	⊖	⊕⊕	⊖⊖	⊕	⊖
Intake mass air flow (MAF) sensor high drift	Gain >1 on air flow signal sent from the physical to the controller model	⊖⊖	⊕	⊕⊕⊕	⊕⊕⊕	⊖	⊖
Intake mass air flow (MAF) sensor low drift	Gain <1 on air flow signal sent from the physical to the controller model	⊕⊕	⊖	⊖⊖	⊖	⊕	⊕⊕
Intake manifold pressure (MAP) sensor high drift	Gain >1 on pressure signal sent from the physical to the controller model	⊖⊖	⊕⊕⊕	⊕⊕⊕	⊕⊕⊕	⊖⊖	⊖⊖
Intake manifold pressure (MAP) sensor low drift	Gain <1 on pressure signal sent from the physical to the controller model	⊕⊕⊕	⊖	⊖⊖	⊖	⊕	⊕⊕⊕
○	Maximum observed variation between 10% and 50% for the considered steady-state engine operating points and fault conditions.						
○○	Maximum observed variation between 50% and 100% for the considered steady-state engine operating points and fault conditions.						
○○○	Maximum observed variation above 100% for the considered steady-state engine operating points and fault conditions.						

measurable signals according to the engine layout already presented in Section 3.2.3 and listed in Tab. 3.2. More in detail, only the signals that are noticeably affected by the introduction of a fault have been considered. Therefore, ambient temperature (T_{amb}), ambient pressure (P_{amb}), pressure upstream low-pressure compressor (P_1) and engine coolant temperature ($T_{coolant}$) have not been included in this analysis. The same applies to temperature downstream DOC (T_5), temperature downstream DPF (T_6), and differential pressure across DPF (ΔP_{DPF}). In fact, these signals characterize mainly the EATS behaviour, which is not the focus of this activity, without providing any significant additional information beyond the ones already provided by other exhaust path-related quantities, including exhaust manifold temperature (T_3), temperature after low pressure turbine (T_4), and exhaust manifold pressure (P_3), when engine-related faults not involving the EATS are considered and engine-out emissions are the focus of the investigation. From this analysis, it can be noticed how each fault acts differently on the considered physical quantities (grey bars): to give some examples, EGR cooler reduced efficiency leads to a significant increase of temperature levels in the exhaust path, as expected, while the opposite effect is observed for EGR valve clogging; opposite deviations can be observed for all the considered signals when comparing sensors drifts with opposite signs (positive or negative); an increase of air mass flow corresponds also to an increase of oxygen concentration; no significant variations of intake manifold temperature (T_{22}) and pressure (P_{22}) are noticed, except for the fault cases involving the related pressure sensor,

mainly due to the fact that the controller is always following a target of pressure and temperature in the intake manifold based on a feedback, eventually acting on EGR rate or turbocharger control. Moving to the controller side (red bars), it can be seen how the models inside the engine controller react to the presence of a fault: if the models are sensitive to the fault, then a limited deviation between the estimated and the actual value is observed, despite a significant variation of the physical signal resulting from the fault introduction. As an example, this can be clearly noticed looking to the exhaust manifold pressure deviation (P_3), except for the case of EGR valve clogging, where the controller deviation tends to be limited even against a significant variation of the physical signal. This is reflected in a deviation between the expected values calculated by the engine controller and the virtual measurements that are sent from the physical model to the controller. The magnitude of this deviation depends both on the type and severity of the investigated faults as well as on the capability of the controller models to adapt the calculations of the considered physical quantities, taking advantage of the other available signals coming from the physical model that are affected by each specific fault. This means that a theoretical zero deviation would be observed whether nominal conditions are considered (provided that the models inside the controller are perfectly calibrated) or the controller is able to completely adapt to the presence of faults, correctly predicting the considered quantity also in non-nominal conditions. Therefore, residual signals can be obtained for all the measured physical quantities, where a residual of a specific quantity X_{Res} is intended as the difference between the expected value calculated by the models within the controller (X_{Ctrl}), and the corresponding actual reference value calculated by the physical engine model (X_{Phys}):

$$X_{Res} = X_{Ctrl} - X_{Phys} \tag{4.1}$$

As an example, Fig. 4.3 shows the trend of the considered residuals in presence of two different intermittent faults, specifically a low drift of the intake manifold absolute pressure (MAP) sensor and a high drift of the air mass flow (MAF) sensor. The residuals have been normalized for a better visualization. Different trends can be observed when a fault is introduced compared to nominal conditions (coloured areas in the plots), which differ also between the two analysed faults. Therefore, the calculated sets of residual signals can be used as input to data-driven models not only to detect an anomaly, namely the presence of a fault, but also to identify it based on the characteristic trends of the considered residuals and their interrelationships. Moreover, a correlation between the same input features and the deviation of NOx emission with respect to nominal conditions can be defined as well, in order to correctly predict the emission trend also in non-nominal conditions. The first task, namely fault identification, will involve data-driven classifiers and will be deeply investigated in Section 4.2; the second task, focused on the development of a regression model for the correction of NOx emission estimation in non-nominal conditions, will be addressed in Section 4.3.
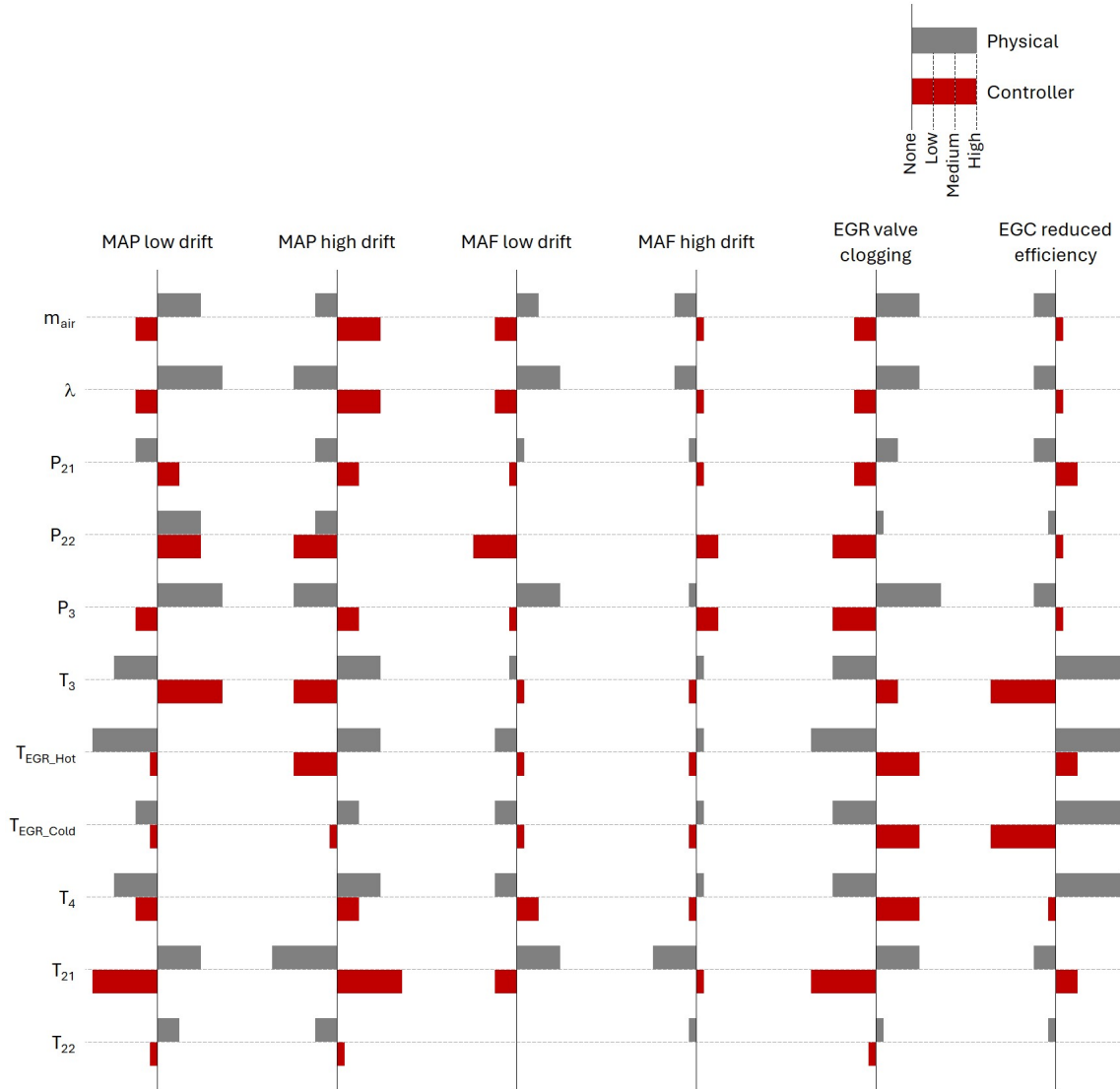
Figure 4.2: Sensitivity of considered physical quantities and corresponding controller predictions to each type of fault

4.2 Fault classification model

In this section, a selection of the most common supervised machine learning classification approaches that could be applied to fault detection and identification are investigated. As a first step, a preliminary performance analysis has been carried out to select the most suitable approach for the specific application, considering model accuracy, training time, and prediction speed. Then, the best performing methods have been optimized to further improve the classification accuracy. Finally the selected models have been implemented into the *Simulink* vehicle model to directly test
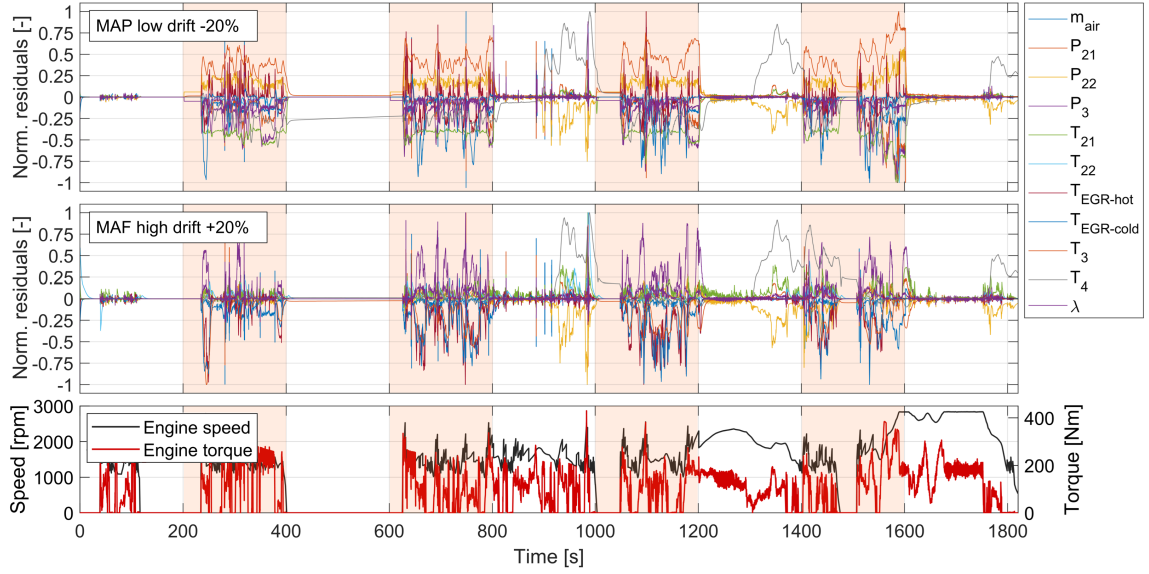
Figure 4.3: Set of residuals for different intermittent faults on WLTC driving cycle: -20% MAP low drift (top), +20% MAF high drift (bottom). The coloured areas identify the intervals in which the fault is present.

their performance on several simulated driving cycles.

4.2.1 Overview of the investigated supervised-learning approaches

The most common supervised learning methods applicable to fault classification are presented in the following paragraphs, with the aim of providing a qualitative description of the underlying algorithm and the main advantages and drawbacks of each group of models [180]. No detailed analytical descriptions have been included in the manuscript since out of the scope of this research activity, which is not dealing with the development of the algorithms itself.

**Decision Tree** DT is a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. It consists of a "rooted tree" including one "root" node, internal or test nodes, and decision nodes called "leaves". According to a certain function of the input attributes values, each test node divides the instance space into two or more subspaces. Each decision node is assigned to one class representing the most relevant target value. It has the great advantage of being easy to understand and interpret and perfect for visual representation. It is a non-parametric model, meaning that no assumption is required about the distribution of the data (differently from Naïve Bayes and Discriminant Analysis, that will be later introduced). Moreover, it requires little data preparation and feature selection happens automatically, so that unimportant features will not influence the result. On the other hand, DT

learners can be unstable due to small variations of the data and also create overfitting, especially for deep structures, which leads to bad generalization capabilities. This issue can be mitigated by using DTs within an ensemble. Coarse Tree (CT), Medium Tree (MT), and Fine Tree (FT), characterized by an increasing number of nodes, have been considered within the performance analysis described in this section.

k-Nearest Neighbours   KNN is a non-parametric classifier, consisting of evaluating the similarity between data points based on a set of numerical features and a selected metric. A target point is assigned to the appropriate class by a majority vote of its neighbours. It is the simplest classification algorithm to implement with just one parameter to be set (i.e., the number of neighbors k). KNN does not explicitly build a model, but it simply tags the new data memorizing the training dataset. Hence, it takes longer time for inference than training and model size grows with the new data incorporated. KNN works well with a small number of input variables, but as the dataset grows efficiency or speed of the algorithm declines very fast. The different types of KNN considered for the preliminary performance analysis include Fine (FKNN), Medium (MKNN), Coarse (CKNN), Cosine (CosKNN), Cubic (CubKNN) and Weighted KNN (WKNN).

Discriminant Analysis   DA consists of finding a linear (or quadratic) combination of features that separates two or more classes of objects, assuming that different classes generate data based on different Gaussian distributions with the same covariance matrices. Furthermore, it is claimed that prior probabilities of class membership are known or can be estimated beforehand. The training phase requires significant computational effort to determine the discriminant functions and their parameters. Once completed, classifying any new data could be achieved simply by solving the corresponding discriminant function for each class and by applying the classification rule, so it is simple, fast, and easy-implementable. Accuracy is limited when its assumptions on predictors distribution are not met. Both Linear (LD) and Quadratic Discriminant (QD) have been considered in the following performance analysis.

Support Vector Machine   SVM classifies data by finding the best hyperplane that separates all data points of one class from those of the other class with the aid of a kernel function. The input data are plotted in a high-dimensional space (with as many dimensions as the number of features), and the SVM algorithm finds the best boundary that separates the classes. This boundary is chosen in such a way that it maximizes the margin, which is the distance between the boundary and the closest data points from each class, also known as "support vectors". SVM can treat non-linear classification problems by applying non-linear kernel functions (Radial Basis Function (RBF), sigmoid, polynomial, etc.). SVM classifiers perform well in high-dimensional space and have excellent accuracy. Moreover, they require limited memory because they only use a portion of the training data. On the other hand, SVM requires a long training period, hence it is not practical for large datasets, and it is not suitable to handle overlapping classes (e.g., due to the presence of noise

in the input signals). Linear (LSVM), Quadratic (QSVM), Cubic (CSVM), Fine Gaussian (FGSVM), Medium Gaussian (MGSVM), and Coarse Gaussian SVM (CGSVM) have been later considered within the performance analysis.

Naïve Bayes   NB is a classification algorithm that applies density estimation to the data. The algorithm employs the principles of Bayes theorem, and "naively" assumes that the presence of one feature in a class does not affect the presence of another one. Another advantage is the ease of implementation, because it requires a small amount of training data, and it is also robust to isolated noise points and to irrelevant features. It is considered as a fast classification method, and it can handle high-dimensional data efficiently. However, it is not easily comprehensible for human readers, and the assumption that all predictors are independent, rarely happens in real life, decreasing the potentially high accuracy of this algorithm. Two types of NB algorithms have been considered within the present investigation, namely Gaussian (GNB) and Kernel Naïve Bayes (KNB).

Ensemble methods   A classification ensemble is a predictive model composed of a weighted combination of multiple classification models. In general, combining multiple classification models increases prediction accuracy. At the same time, a deterioration of interpretability and computational cost is usually observed. Although an unlimited number of ensembles could be developed, the three main classes of ensemble learning methods are: bagging, based on fitting many decision trees on different samples of the same dataset and averaging the predictions; stacking, based on fitting many different model types on the same data and using another model to learn how to best combine the predictions; boosting, based on adding ensemble members sequentially that correct the predictions made by prior models and outputs a weighted average of the predictions. Ensemble methods including Boosted Trees (BoT), Bagged Trees (BagT), Subspace Discriminant (SD), Subspace KNN (SKNN), and RUSBoosted Trees (RUSBT) have been considered in first analysis for this application.

Artificial Neural Network   The standard ANN consists of a set of connected neurons which are organized in an input layer, an output layer, and one or more hidden layers in between. An increase in the number of layers allows for non-linear calculations, improving the weakness of the "perceptron" (namely a single-layer NN) linear nature. In FDI applications, Multi-Layer Perceptron (MLP) networks have shown good accuracy and robustness to noise and errors, as deeply investigated in [181], which is an important feature when dealing with engineering-related systems. However, high-capacity networks with complex structures lead to high computational complexity and thereby to slow convergence and potential overfitting during training. To overcome this issue, a large, diversified dataset is required for training. For the analysis presented in this section, Narrow (NNN), Medium (MNN), Wide NN (WNN), characterized by a single layer with

an increasing number of neurons, have been considered, as well as Bi-layered (BNN) and Tri-layered NN (TNN).

4.2.2 Training dataset and features importance analysis

The dataset available for model training includes 13 signals, namely the 11 residuals previously introduced in addition to engine speed (n_{Eng}) and load (Tq_{Eng}) request. The data have been generated simulating real driving cycles in both nominal and faulty conditions. Specifically, two different real driving cycles in nominal conditions and the same driving cycle with five different degrees of severities for each fault, have been simulated, in order to obtain a well-balanced training dataset. The considered fault classes are listed in Tab. 4.2, together with the related class labels and the degrees of severity that have been simulated for each fault. More in detail, depending on the considered fault, the degree of severity is intended as:

- the percentage error between the measurement and the actual physical quantity for MAP and MAF sensors drifts;

- the imposed maximum EGR valve opening, expressed in percentage with respect to the nominal condition, for EGR valve clogging;

- the actual cooling efficiency of the EGR Cooler (EGC), expressed in percentage compared to the nominal case, for reduced EGC efficiency.

Table 4.2: List of fault classes, related labels, and fault cases included in the training dataset

Class label	Fault	Severity
NOM	None	-
MAP LOW	Intake manifold pressure sensor low drift	-5%, -10%, -15%, -20%, -25%
MAP HIGH	Intake manifold pressure sensor high drift	+5%, +10%, +15%, +20%, +25%
MAF LOW	MAF sensor low drift	-5%, -10%, -15%, -20%, -25%
MAF HIGH	MAF sensor high drift	+5%, +10%, +15%, +20%, +25%
EGR	Clogged EGR valve	40%, 30%, 20%, 10%, 0%
EGC	Reduced EGR cooler efficiency	90%, 70%, 50%, 20%, 0%

The data from the simulations have been sampled with a frequency of 1 Hz, for a total of around 200000 observations (which would correspond to 55 hours of real driving) available for the offline training and validation of the models. Each observation is associated with a label identifying the specific fault class which the observation belongs to. Since supervised learning methods are considered, the known responses to the input data are provided to the model as well during the training phase. The box plots in Fig. 4.4, provide a visual representation of

the statistical distribution of the considered input data, highlighting the behaviour of various calculated residuals in presence of different fault types. More in detail, the horizontal line inside of each box is the median of the considered data, while the top and bottom edges of each coloured box are the upper and lower quartiles, respectively. The distance between the top and bottom edges is known as interquartile range (IQR), namely the width of the range of values that contains the middle half of the observed values. The lines that extend above and below each box are called "whiskers" and connect the upper quartile to the non-outlier maximum and the the lower quartile to the non-outlier minimum. Outliers, defined as the values that are more than $1.5 \cdot IQR$ away from the top or bottom of the box, have been simply discarded for a clearer representation. However, all the observations are included in the training dataset, to further test the robustness of the developed models also in presence of outliers. Focusing on Fig. 4.4, each plot refers to a specific residual among the ones selected as input features for the data-driven models, while the coloured boxes within each plot refer to the different fault classes. It can be observed that residuals calculated for the nominal case (NOM) generally exhibit the lowest dispersion among the considered cases, with a mean value around zero, meaning that a limited deviation between the controller and physical model is observed in nominal conditions. This does not apply to P_{22} and T_4 residuals, due to inaccuracies of the models inside the controller. Moreover, even if it is a summary statistical analysis, it can be clearly noticed that different patterns characterize each fault class, thus allowing to identify a combination of relationships between the trends of all residual signals, that helps distinguishing each case from the others.

As described in Section 4.1, physics-based considerations led to the exclusion of selected signals from the list of input features, including ambient conditions and EATS-related measurements, that could not provide significant information about the presence of engine faults and related effect on engine-out emissions. On the other hand, the study of the behaviour of both physical and controller models reacting to the introduction of a fault, and the described summary statistical analysis of the computed residuals, confirmed the existence of patterns of the considered input features that are characteristic of each fault. In addition to this empirical approaches, different algorithms can be used for feature ranking, in order to select the most significant ones and neglect useless information that would increase model complexity without improving its performance. Two of the most common algorithms, MRMR and ReliefF, have been applied to sort the considered input features based on their importance in predicting the output response, namely the fault class in this specific case. A brief description of the algorithms and the results of their application on the described training dataset and set of input features are provided in the following paragraphs.

MRMR algorithm The Minimum Redundancy Maximum Relevance algorithm ranks model features quantifying their relevance, i.e. the degree to which a feature is useful for predicting the target variable, and redundancy, i.e. the degree to which a feature provides redundant information already captured by other features. Both relevance and redundancy are calculated by quantifying the mutual information I between the involved variables. More in detail, mutual information
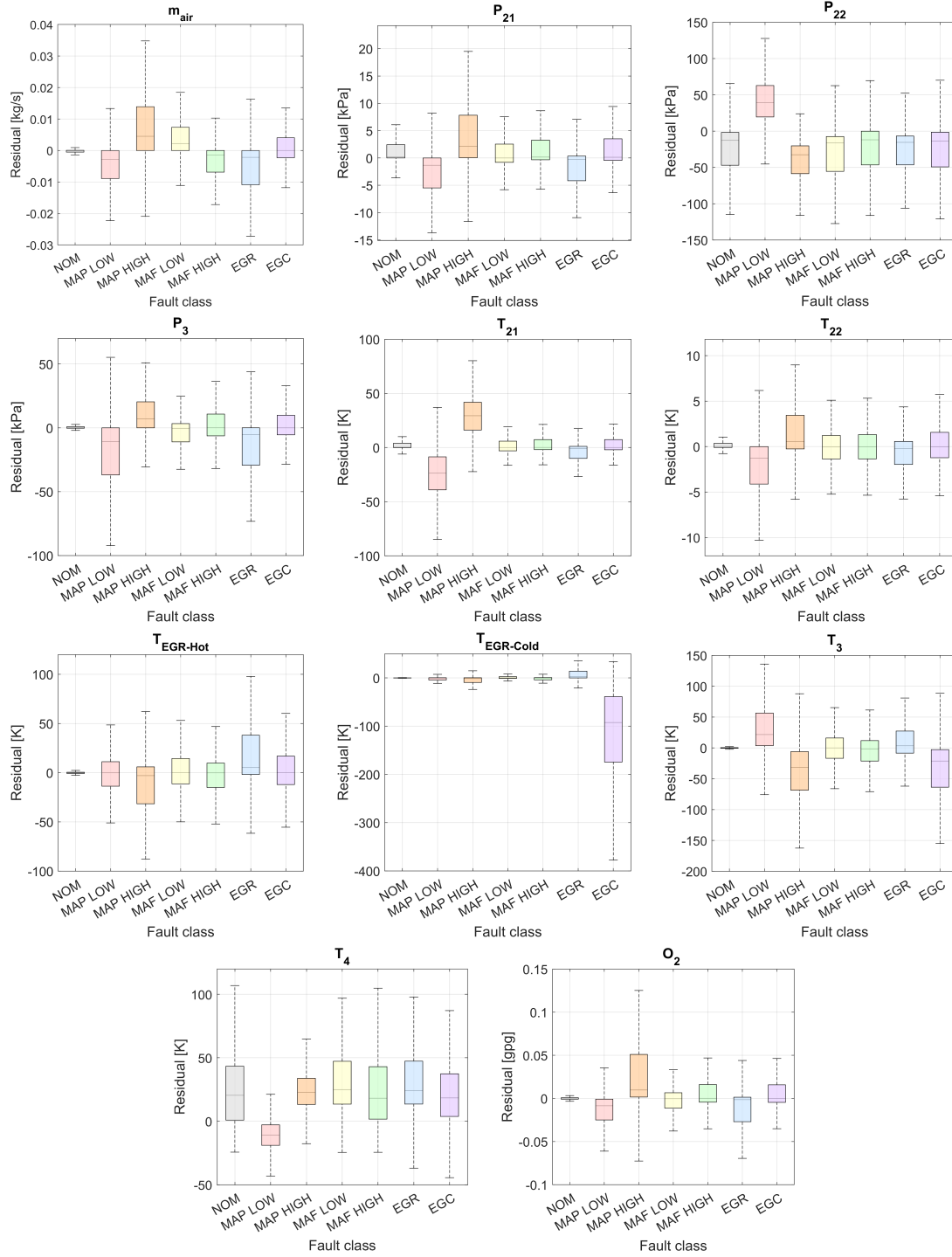
Figure 4.4: Statistical box plots of input features (residuals) for each fault type

quantifies the amount of information obtained about one variable by observing another one. Analytically, the mutual information $I(X;Y)$ between two variables X and Y is expressed as:

$$I(X;Y) = \sum_{i,j} p(X = x_i, Y = y_j) \log \left(\frac{p(X = x_i, Y = y_j)}{p(X = x_i)p(Y = y_j)} \right) \tag{4.2}$$

where $p(X = x_i, Y = y_j)$ is the joint probability of $X = x_i$ and $Y = y_j$ occurring together, while $p(X = x_i)$ and $p(Y = y_j)$ are the marginal probabilities of $X = x_i$ and $Y = y_j$ occurring independently. Therefore, if two variables are independent, their mutual information is zero; if knowing one variable reduces uncertainty about the other, the mutual information is positive. The importance of a feature x is quantified by the MRMR algorithm by means of a score called Mutual Information Quotient (MIQ):

$$MIQ_x = \frac{V_x}{W_x} \tag{4.3}$$

where V_x and W_x are the already introduced relevance and redundancy of a feature x, respectively defined as:

$$V_S = \frac{1}{|S|} \sum_{x \in S} I(x,y) \qquad W_S = \frac{1}{|S|^2} \sum_{x,z \in S} I(x,z) \tag{4.4}$$

where S indicates a specific set of features, $|S|$ is the number of features in in S, x and z are features belonging to S, and y is the response variable. A large score value computed by the algorithm indicates that the corresponding predictor is highly relevant for predicting the target response and that the information provided is slightly or not redundant compared with that provided by other predictors.

ReliefF algorithm The algorithm works by assigning a score to each feature based on how well it distinguishes between instances that are close to each other but belong to different classes. The target is to evaluate how well a feature can differentiate between similar observations (nearest neighbours) belonging to different classes. More in detail, the algorithm iteratively selects a random observation o_r, finds the k-nearest observations to o_r for each class, and calculates for each nearest neighbour o_k all the weights for the predictors X_j, according to the following equations:

$$W_j^i = W_j^{i-1} - \frac{\Delta X_j(o_r, o_k)}{m} \cdot d_{rk} \tag{4.5}$$

for o_r and o_k belonging to the same class,

$$W_j^i = W_j^{i-1} + \frac{p_{y_k}}{1 - p_{y_r}} \cdot \frac{\Delta X_j(o_r, o_k)}{m} \cdot d_{rk} \tag{4.6}$$

for o_r and o_k belonging to different classes, where W_j^i is the weight of the predictor X_j at the ith iteration step; p_{y_r} is the prior probability of the class to which o_r belongs, and p_{y_k} is the prior probability of the class to which o_k belongs; m is the number of specified iterations; $\Delta X_j(o_r,o_k)$ denotes the difference in the value of the predictor X_j between observations o_r and o_k; d_{rk} is the distance between the observations o_r and o_k that can be computed using different suitable distance metric, such as Euclidean distance. Therefore, the algorithm penalizes the predictors that give different values to neighbours of the same class, and rewards with high scores predictors that give different values to neighbours of different classes.

The two feature ranking algorithms have been applied to the considered dataset and input features. Fig. 4.5 shows the corresponding results, with features sorted from the highest to the lowest importance score. A good agreement is observed in the definition of the top-five important features, namely temperature after EGR cooler, temperature after low-pressure turbine, temperature after high-pressure compressor, intake manifold pressure and exhaust manifold temperature. The only features showing a negligible importance score compared to all the others are engine torque and speed, according to the MRMR ranking algorithm, and intake manifold temperature according to the ReliefF algorithm. Based on these results, combined with the previous analysis of residuals patterns, no features have been discarded in the first phase of model development.
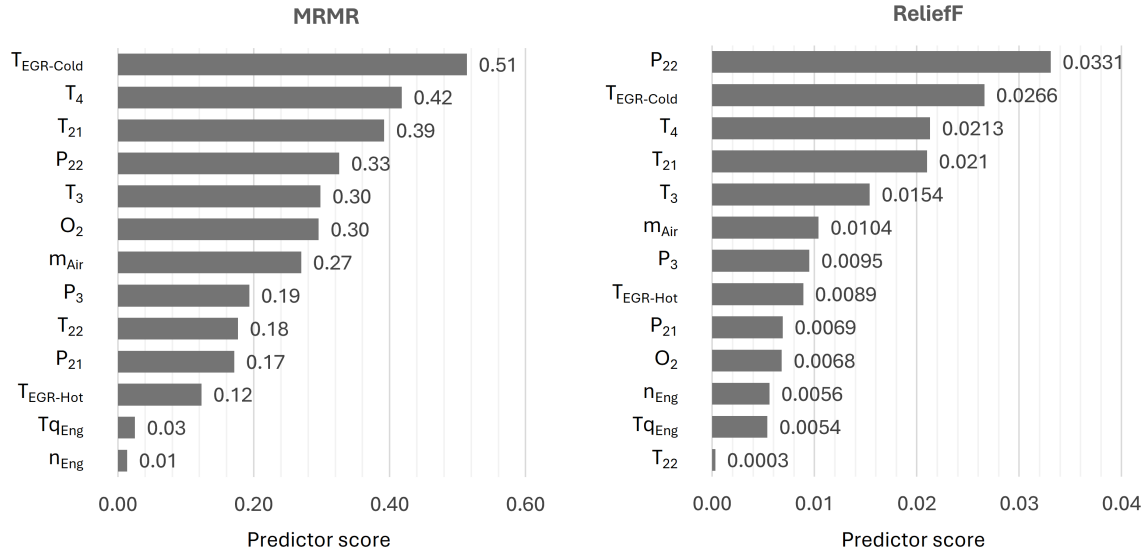

Figure 4.5: Importance scores of selected input features according to MRMR (left) and ReliefF (right) algorithms

4.2.3   Performance assessment of different classifiers

As a preliminary step to choose the best approach for the specific application, the Classification Learner App by *Matlab* has been used to train all the different classification models introduced in Section 4.2.1, investigating their performance in terms of prediction speed, expressed in observations per second (obs/s), training time (s), and accuracy. The latter is a commonly used metric to evaluate the overall effectiveness of a classification model, especially in the case of well-balanced datasets, measuring the percentage of total correctly predicted instances. It is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.7}$$

where TP and TN are the number of true positives and true negatives, respectively (i.e., correct predictions), while FP and FN are the number of false positives and false negatives, respectively (i.e., misclassified observations). Holdout validation, recommended for large datasets, has been applied with 70% of the dataset used for training and 30% held out for validation. The default set of hyperparameters set by the Classification Learner App by *Matlab* [182] has been considered for each model in this preliminary phase. The results of the analysis are graphically represented in Fig. 4.6. It can be observed that Decision Trees and Discriminant Analysis families have the advantage of a very low training time combined with high prediction speed, thanks to the relatively simple model architectures. However, DA methods have been excluded from further investigations due to the limited level of accuracy reached (85% in the best case) compared to other models. The same applies to Naïve Bayes methods. Among the other families, k-Nearest Neighbours achieve good accuracy (up to 96%) but they show relatively high training time and low prediction speed at the same time. Similar considerations are valid also for Support Vector Machines, with the main drawback of a significant required training time. The highest accuracy in the validation phase is reached by the Bagged Trees (BagT) method (98.9%), from the Ensemble family, while the Fine Tree (FT) and the Wide Neural Network (WNN) methods show the highest prediction speed, which makes them interesting in view of real-time application, together with a still high prediction accuracy of 96.8% and 98.2%, respectively. On the other hand, Neural Networks require a long training time, which is significantly reduced for the Ensemble methods and even lower for the Decision Trees. The FT model has the additional advantage of a much lower training time compared to BagT and especially to WNN. Since each of the last mentioned approaches shows different advantages/disadvantages in terms of accuracy, training time, or prediction speed, all three methods, namely Decision Trees, Neural Networks and Ensembles have been considered for further investigations to evaluate their performance on different testing cases.

Even if accuracy is probably the most common and straightforward index to evaluate the performance of classification models, there some additional performance metrics that could be considered to provide additional information:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.8}$$
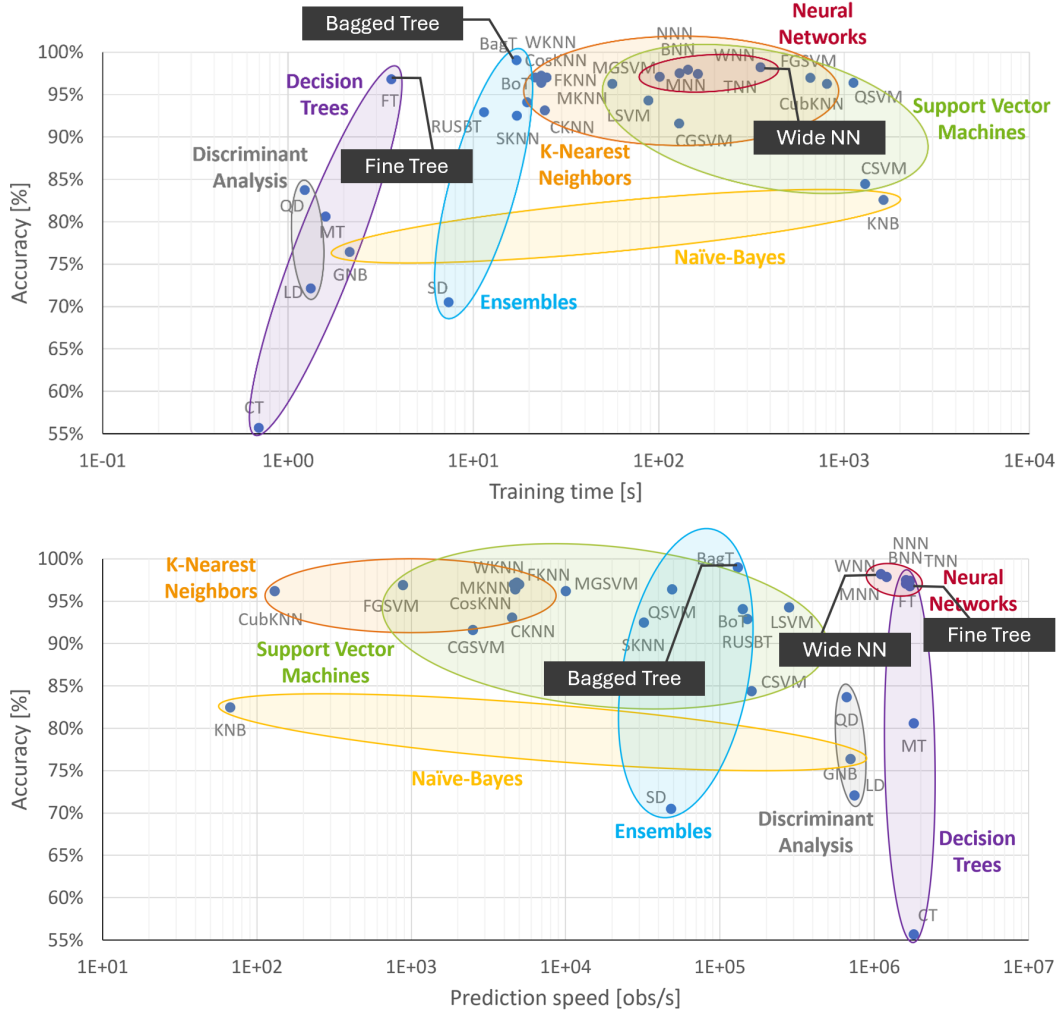
Figure 4.6: Overall accuracy versus training time (top) and prediction speed (bottom) of different families of classification algorithms trained on the considered dataset

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.9}$$

$$\text{F1-score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = 2 \cdot \frac{TP}{2TP + FP + FN} \tag{4.10}$$

Precision is often employed to evaluate the reliability of positive classifications. Higher values of precision, indicating fewer false positives, are desired when the target is reducing incorrect positive classifications. On the other hand, recall index, also known as sensitivity, is preferred when missing a true positive observation is more problematic than getting a false positive. It corresponds to the True Positive Rate (TPR) calculated for the specific fault class, namely the fraction between

related True Positives (TP) and the sum of TP and False Negatives (FN). High values of recall indicate that the model has a low rate of false negatives. However, achieving a high recall often comes at the cost of more false positives, leading to a lower precision. This is why it is important to take into consideration both the indexes. In this regard, F1-score provides a balanced metric between precision and recall. It is particularly useful for unbalanced datasets, where accuracy alone could provide misleading information. For the specific application addressed within this research activity, accuracy can still be considered as a reliable metric to judge the performance of the developed models: the considered training dataset is well-balanced and all misclassification cases (false positives and false negatives) are treated with the same importance. Nonetheless, all the introduced indexes will be considered in the testing phase, described in Chapter 5, and in the experimental application of the developed models, described in Chapter 7 and Chapter 8, for a more comprehensive evaluation of models performance across the different fault cases.

4.2.4 Training, optimization and validation of selected classifiers

To maximize the accuracy of the selected classification models, Bayesian optimization has been used to optimize the main hyperparameters, namely model parameters that specify its architecture and how it learns from training data. Bayesian optimization is a method for hyperparameters tuning that uses a probabilistic model to find the best settings efficiently. The algorithm attempts to minimize a scalar objective function $f(x)$ in a bounded domain. To do that it relies on a model of $f(x)$, predicting how well the set of hyperparametes x will perform. The model is often a Gaussian Process, like the one adopted for this specific application, characterized by a mean function $\mu(x)$ and a covariance function $k(x,x')$. To choose the next set of hyperparameters, different acquisition functions can be used; in this case Expected Improvement (EI) is used, computed as:

$$\text{EI}(x) = \mathbb{E}[\max(f^*(x_{best}) - f(x), 0)] \tag{4.11}$$

where x is the set of hyperparameters being considered, corresponding to the object to be optimized; $f(x)$ is the objective function, i.e. the performance parameter that has been chosen to measure the goodness of the model; $f^*(x_{best})$ is the current best performance result, obtained for the hyperparameter set x_{best}, observed from previous evaluations, which becomes a reference for further improvements. Based on that, $\max(f^*(x_{best}) - f(x), 0)$ represents the improvement over the current best performance, that will be positive if $f(x)$ is lower than $f^*(x_{best})$; otherwise it will be zero. Finally, $\mathbb{E}$ is the expected value of the improvement, computed taking into account the uncertainty in the performance prediction. The algorithm stops after reaching a fixed number of iterations, a fixed time or another stopping criterion selected by the user. This method for hyperparameter tuning is more efficient than grid search, which tests all possible combinations, and random search, which tests random values, as it uses previous results to guide the search, requiring fewer evaluations to find a good set of hyperparameters. Therefore Bayesian optimization is preferred especially when the hyperparameter space is large and high-dimensional, where

the application of the other mentioned methods would be computationally expensive and time-consuming [183].

Focusing on this specific application, the optimizable parameters for each model and their range of optimization are summarized in Tab. 4.3, together with the optimal hyperparameter configurations resulting from the Bayesian optimization. The overall classification error has been chosen as the objective function to be minimized, calculated as the complementary to 1 of model accuracy. Holdout validation has been applied, as for the preliminary analysis described in Section 4.2.3, with 70% of the dataset used for training and 30% kept for the validation phase. A maximum number of 50 iterations has been set as the stopping criterion, which is far above the observed number of iterations after which an asymptote in the evaluated error is observed. Fig. 4.7 shows the trend of the minimum classification error during the optimization phase. More in detail, the estimated minimum classification error (red points) corresponds to an estimate of the minimum classification error computed by the optimization process when considering all the sets of hyperparameter values tried so far; the observed minimum classification error (grey points) corresponds to the minimum classification error computed so far, namely the best performance obtained by the models trained until that moment; the yellow point indicates the iteration that corresponds to the hyperparameters providing the effectively observed minimum classification error, while the blue square indicates the iteration corresponding to the final optimized hyperparameters. Generally these two points coincide, however it can happen that the optimized hyperparameters do not provide the observed minimum classification error, since according to Bayesian optimization, the selected set of hyperparameter values is the one that minimizes an upper confidence interval of the classification error objective model, rather than the one that minimizes the classification error. The overall accuracy reached on the validation dataset by the optimized models is shown in Tab. 4.4, together with the training time (including model optimization), prediction speed and final model size. The last two parameters will gain particular importance in view of real-time on-board application.
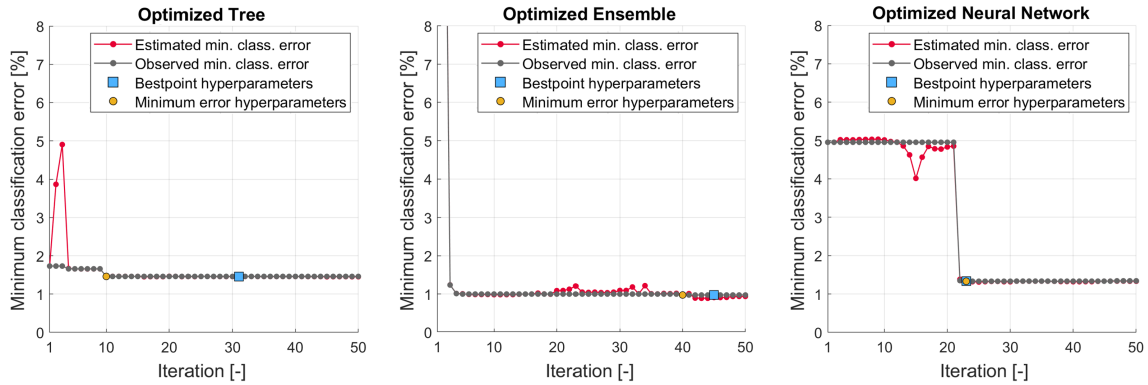

Figure 4.7: Minimum classification error plots resulting from Bayesian optimization of Tree, Ensemble and Neural Network classifiers

Table 4.3: Set of optimizable parameters, optimization range, and final configuration resulting from Bayesian optimization of the classification models

Model	Optimizable hyperparameters	Optimization range	Optimal configuration
Optimizable Tree	Maximum number of splits	$[1, \max(2, n-1)]$	858
	Split criterion	Gini's diversity index, Twoing rule, Maximum deviance reduction	Maximum deviance reduction
Optimizable Ensemble	Ensemble method	AdaBoost, RUSBoost, Bag	AdaBoost
	Maximum number of splits	$[1, \max(2, n-1)]$	655
	Number of learners	[10,500]	455
	Learning rate	[0.001,1]	0.948
	Number of predictors to sample	$[1, \max(2, p)]$	13 (All)
Optimizable Neural Network	Number of fully connected layers	[1,3]	2
	Layers size	[1,300]	43, 173
	Regularization strength	$[0.00001/n, 100000/n]$	$4.8454e-11$
	Activation function	ReLU, Tanh, None, Sigmoid	Tanh

n = number of observations (=217740)
p = number of predictor variables (=13)

Table 4.4: Overall accuracy and training parameters of the optimized classification models

Model	Accuracy [%]	Training time [s]	Prediction speed [obs/s]	Model size [MB]
Ensemble	99.0	1927	6600	128
Neural Network	98.7	17980	490000	0.085
Tree	98.5	105	880000	0.4

Fig. 4.8 shows the confusion matrices evaluated on the validation dataset for the three optimized models. The confusion matrix is a useful tool to understand how the currently selected classifier performs in predicting each class, eventually identifying the areas where the classifier performs poorly. More in detail, the matrix rows show the true class, and the columns show the predicted

class, so that diagonal cells show where the true class and predicted class are matching. On the right of each confusion matrix, the True Positive Rate (TPR) column summarizes the proportion of correctly classified observations per each true class, while the False Negative Rate (FNR) column identifies the proportion of incorrectly classified observations per each true class. TPR and FNR are also known as sensitivity and specificity respectively. The results, expressed in percentage for each fault class, show that the three optimized models can successfully classify the different fault conditions included in the validation dataset. The overall accuracy, calculated as in 4.7, reaches 99% for the Ensemble model, while it is slightly lower for the Tree and Neural Network models, namely 98.5% and 98.7% respectively. More in detail, all the models can perfectly recognize the faults regarding MAP sensor and EGR cooler, with a TPR very close to or equal to 100%. The most significant misclassification error involves the EGR valve clogging, with a maximum FNR above 7% for the Tree model, which is mainly misclassified as a nominal condition. For each model, no significant number of cases in which a fault is incorrectly classified as another fault (excluding nominal condition) is observed, thus minimizing the risk of incorrect fault detection. Given the very good overall performance, all the classification models have been considered for further investigations to assess their accuracy and robustness on different test cases and to evaluate their real-time capability.

4.3　NOx regression model

A parallel approach to the one described in Section 4.2 has been followed also for the development of the regression model aimed at correctly predicting engine-out NOx emission during fault conditions. After a preliminary phase to identify the regression algorithms offering the best trade off in terms of accuracy and computational effort, the selected models have been optimized to further improve their performance. The most significant step of the followed methodology are presented in this section.

4.3.1　Overview and selection of the regression algorithms

In the first stage, several regression models among the most common approaches found in literature have been considered, in order to find the best performing one for this specific application, taking into account model accuracy, training time, and prediction speed. Some of the investigated approaches have been already introduced in Section 4.2.1, since the same algorithms can be applied to both classification and regression tasks. They include:

- Regression Trees

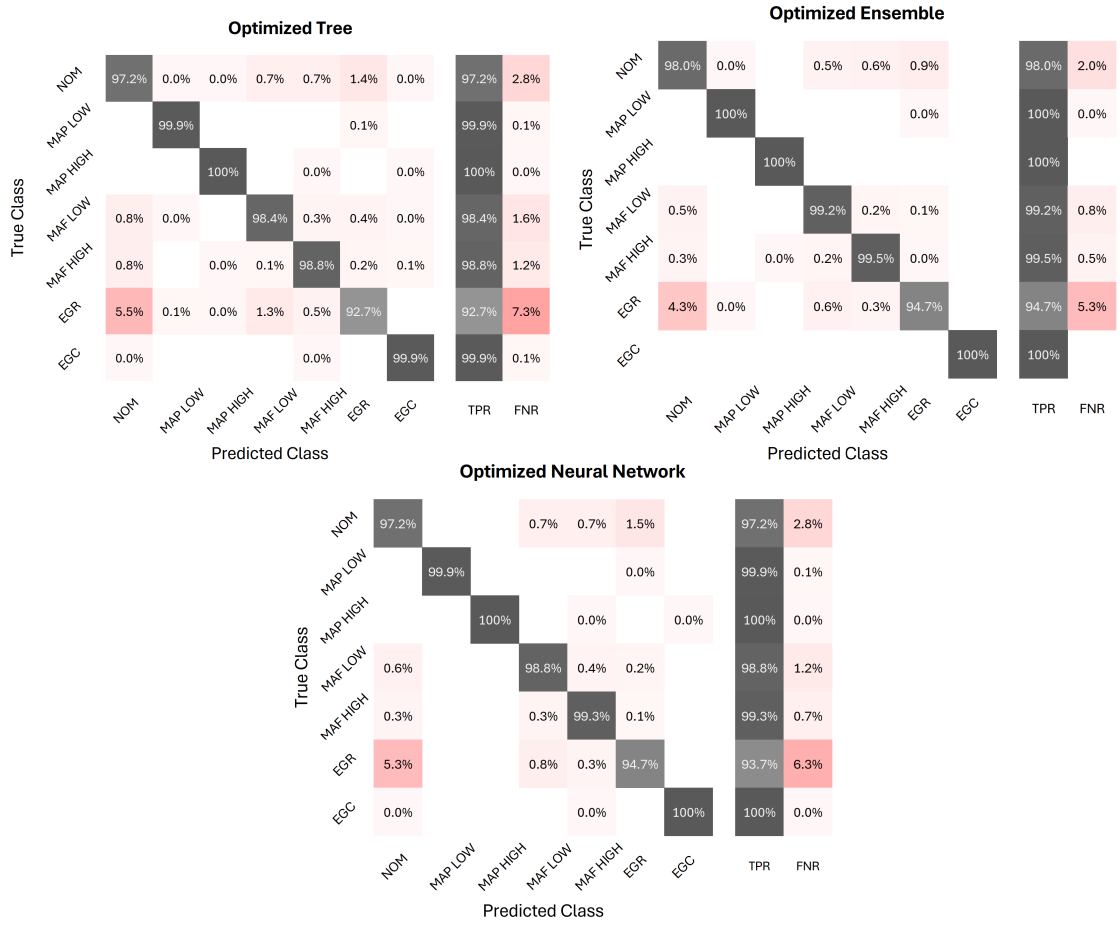- Support Vector Machines

- Ensemble Models

Figure 4.8: Validation results of optimized classification models: related confusion matrices, true positive rates (TPR) and false negative rates (FNR) for each fault class

- Neural Networks

Additional regressors have been considered, which are briefly introduced in the next paragraphs together with related advantages and drawbacks.

Linear Regression   LR algorithms provide a linear relationship between a dependent variable, the model output, and one or more independent variables, the input features. Thanks to this simple structure, these kind of models are easy to interpret and fast for making predictions. However, the highly constrained form of these models often lead to low prediction accuracy, especially if highly non-linear and multidimensional problems are considered. The accuracy can be improved increasing the order of the considered function (e.g. quadratic or cubic), meaning that a curve rather than a straight line is used to fit the data, with a parallel increase of the model complexity.

Gaussian Process Regression GPR models are non-parametric probabilistic models. They are based on the assumption that the function to be learned, that relates the input to the output, is drawn from a Gaussian process, allowing to make predictions with a well-defined uncertainty level. GPR models are often highly accurate, especially when working with small datasets, but can be difficult to interpret and suffer from computational complexity. The latter grows exponentially with the number of training points, affecting both the training time and the prediction speed once the model has been trained.

Kernel Approximation As GPR models, KA is a non-parametric type of regression with the objective to find non-linear relations between data, but unlike GPR it is suitable to handle data with many observations. The algorithm translates input data into a higher-dimensional space, so that a non-linear behaviour in the original input space can be transformed in a linear one in the higher-dimensional space. Since they are working with relatively simple dot products between vectors, KA models tend to train and predict faster than SVM and GPR models, especially for large datasets. Model flexibility depends on the Kernel type, but it is generally lower compared to other regression models like Ensembles and Neural Networks, while model interpretability is comparably hard.

The same input dataset described in Section 4.2.2 has been considered for training and validation, while the response to be learned by the regression models is the correction factor to be applied to the reference NOx emission estimated by the engine controller, in order to match the actual emission calculated by the physical model (corresponding to the NOx effectively emitted by the engine if a real-world case is considered). Since the output provided by the controller model is directly the NOx mass flow, the correction factor is calculated as:

$$f_{NOx} = \frac{\dot{m}_{NOx,act}}{\dot{m}_{NOx,est}} \tag{4.12}$$

where $\dot{m}_{NOx,act}$ and $\dot{m}_{NOx,est}$ are the actual and estimated NOx mass flow, calculated by the engine physical model and the engine controller model respectively. As for the fault classification model, importance scores have been computed for each input feature to quantify their impact in the prediction of the model output, namely the NOx correction factor. The same feature ranking algorithms, MRMR and RReliefF, already described in Section 4.2.2, have been applied also in this case, properly modified to work with regression models [184]. Fig. 4.9 shows the corresponding results, with features sorted from the highest to the lowest importance score. Differently from what has been observed for the classification model in Section 4.2.2, there is no agreement between the two ranking methods in defining the most relevant features. Moreover, it can be noticed that there are no features standing out clearly from the others, except for pressure upstream high-pressure compressor (P_{21}) according to the RReliefF algorithm, which however is not ranked among the features with the highest importance scores according to the MRMR algorithm. At the same time,

102

there are no features to be definitely excluded according to the obtained importance scores; only temperature after EGR cooler ($T_{EGR-Cold}$) is assigned a very low score by RReliefF compared to all the other features, but it is among the most important features according to MRMR. So, taking into account these results in addition to all the previous analysis and considerations on available input features, no features have been further discarded in this phase.
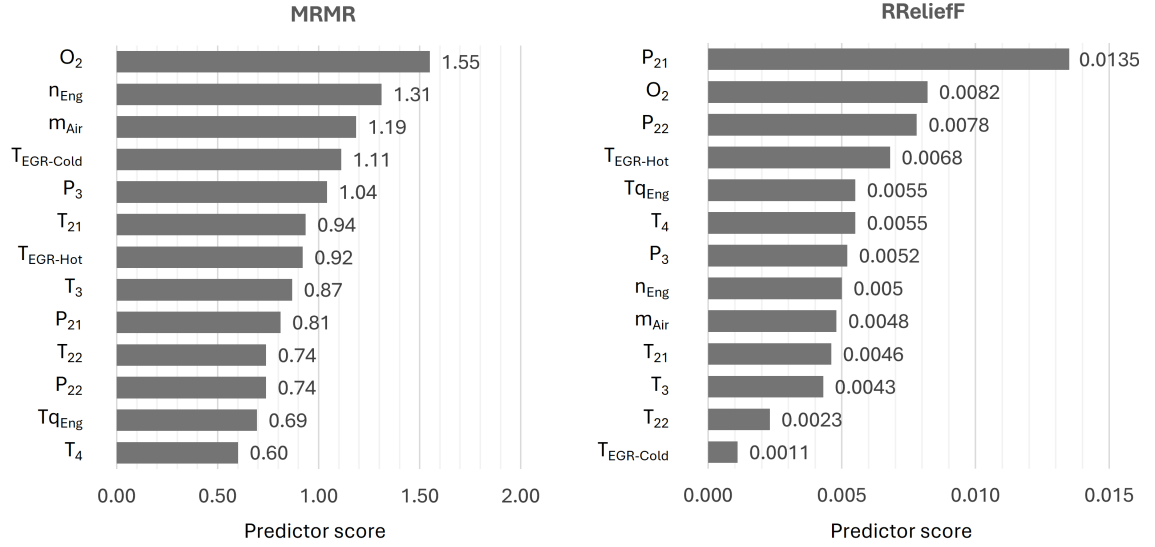

Figure 4.9: Importance scores of selected input features according to MRMR (left) and ReliefF (right) algorithms

The Regression Learner App by *Matlab* has been used to train selected models as representative of the introduced regression algorithms, to evaluate their performance in terms of prediction accuracy, training time and prediction speed. Among the most common performance metrics, Root Mean Squared Error (RMSE), coefficient of determination (R^2), Mean Absolute Error (MAE),and Mean Absolute Percentage Error (MAPE), have been considered to assess and compare the models. The three indexes are calculated as follows:

$$\text{RMSE} = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{4.13}$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{4.14}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{4.15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{4.16}$$

where $\hat{y}_i$ is the predicted value, y_i is the corresponding actual value, $\bar{y}$ is the mean of the actual values, and n is the number of observations. By definition, RMSE is always positive, and its units match the units of the target response, making its interpretation straightforward. MAPE is always non-negative and indicates how the prediction error compares to the response. It is a useful and clear metric to compare models, but it is not appropriate to measure the error when the actual data show zero or very small non-zero values. In those cases, MAE is preferred to MAPE, giving information similar to the RMSE, but being less sensitive to outliers. R^2 expresses the proportion of variance in the response that can be explained by the independent variables. It provides a measure of the goodness of a regression model in fitting the observed data. It is always smaller than 1 and usually larger than 0, and it is particularly useful for linear regression models, while it is less appropriate for highly non-linear models. Therefore, the best performing model will be the one showing smaller values of RMSE and MAPE, and a R^2 values closer to 1. As for the classification models, holdout validation has been applied, with 70% of the dataset used for training and 30% held out for validation. The default set of hyperparameters set by the Regression Learner App by Matlab [185] has been considered for each model in this preliminary phase. The results of the analysis are presented in Tab. 4.5, where the investigated models have been sorted from the best performing one in terms of RMSE to the worst one. The Ensemble regressor outperforms the other algorithms in terms of all considered performance metrics, followed by Tree and Neural Network-based models, which show the second and third best performance in terms of RMSE and R^2. Medium Gaussian SVM shows comparable performance, but it requires a training time that is one order of magnitude higher than the one required by the Neural Network, and two or three order of magnitude higher compared to Tree and Ensemble models respectively. Moreover, SVM prediction speed is significantly lower compared to Ensemble and especially to Tree and Neural Network regressors, while model size is significantly large. Kernel regressors have the advantage of requiring a much lower training time compared to Neural Network and SVM models, but a significant deterioration in RMSE and R^2 values is observed compared to the best performing models. Based on this analysis, Ensemble, Tree, and Neural Network regressors have been selected to be further optimized, as described in the next section. This choice is driven by the fact that Ensemble regressor shows the best overall accuracy, while Tree and Neural Network models are characterized by a much shorter training time and very limited model size respectively, together with higher prediction speed, still achieving very good performance in terms of RMSE and R^2.

4.3.2 Training, optimization and validation of selected regression models

As seen for the classification models, the hyperparameters of the selected regression models have been further optimized to maximize prediction accuracy. Bayesian optimization, already introduced in Section 4.2.2, has been applied. More in detail, the Mean Square Error (MSE) evaluated on the validation dataset has been chosen as the objective function to be minimized during the

Table 4.5: Main performance and training parameters of different regression algorithms trained on the considered dataset

Model	RMSE [-]	MAE [-]	R2 [-]	Training time [s]	Prediction speed [obs/s]	Model size [MB]
Preliminary analysis						
Ensemble (Bagged Trees)	0.069	0.017	0.94	811	29000	33
Fine Tree	0.083	0.021	0.91	48	300000	1
Neural Network (2-layers)	0.094	0.029	0.88	1664	310000	0.009
SVM (Medium Gaussian)	0.105	0.027	0.85	13939	1500	13
SVM (Quadratic)	0.135	0.052	0.76	73096	1600	16
Kernel (Least Squares)	0.156	0.071	0.68	719	63000	0.013
Linear Regression	0.159	0.076	0.67	17	390000	0.016
SVM (Linear)	0.162	0.071	0.65	28866	1900	17
Kernel (SVM)	0.195	0.077	0.50	926	57000	0.013
Optimized models						
Neural Network	0.056	0.019	0.96	79306	530000	0.026
Ensemble	0.065	0.016	0.94	3468	68000	89
Tree	0.082	0.021	0.91	475	1000000	1

optimization; holdout validation has been applied, as for the preliminary analysis, with 70% of the dataset used for training and 30% kept for the validation phase; a maximum number of 50 iterations has been imposed. Similarly to the classification error plots described in 4.2.2, Fig. 4.10 shows the trend of the minimum MSE during the optimization phase: red points correspond the estimated minimum MSE, based on an upper confidence interval of the current MSE objective model; grey points correspond to the minimum MSE obtained with the models trained so far; the yellow point indicates the iteration providing the observed minimum MSE, while the blue square indicates the iteration corresponding to the final configuration of hyperparameters at the end of the optimization process. As already mentioned for the minimum classification error plot, the optimized hyperparameters do not always provide the observed minimum classification error, since according to Bayesian optimization, the chosen set of hyperparameters is the one that minimizes an upper confidence interval of the MSE objective model, rather than the one that minimizes the MSE. The range of optimization of each parameter and the resulting optimal configuration of hyperparameters for each model are summarized in Tab. 4.6. The main performance metrics evaluated on the validation dataset for each optimized model are shown at the bottom of Tab. 4.5, together with the time required for model training and optimization, prediction speed and final model size. According to the validation results, the optimized Neural Network exhibits the lowest RMSE, equal to 0.056, followed by the Ensemble model, with a slightly higher RMSE of 0.065,

and finally the regression Tree, with an RMSE of 0.082. Accordingly, the Neural Network shows the highest R^2, equal to 0.96, followed by the Ensemble and the Tree models, with 0.94 and 0.91 respectively.
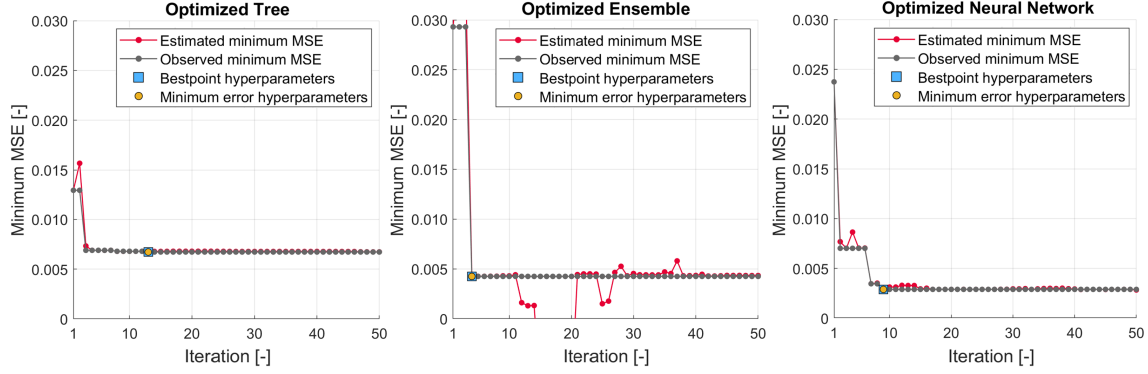

Figure 4.10: Minimum MSE plots resulting from Bayesian optimization of regression Tree, Ensemble and Neural Network models

Table 4.6: Set of optimizable parameters, optimization range, and final configuration resulting from Bayesian optimization of the selected regression models

Model	Optimizable hyperparameters	Optimization range	Optimal configuration
Optimizable Tree	Minimum leaf size	$[1, \max(2, floor(n/2)]$	7
Optimizable Ensemble	Ensemble method	Bag, LSBoost	Bag
	Minimum leaf size	$[1, \max(2, floor(n/2)]$	2
	Number of learners	[10,500]	66
	Learning rate	[0.001,1]	0.241
	Number of predictors to sample	$[1, \max(2,p)]$	13 (All)
Optimizable Neural Network	Number of fully connected layers	[1,3]	2
	Layers size	[1,300]	50, 32
	Regularization strength	[0.00001/n,100000/n]	0.0699
	Activation function	ReLU, Tanh, Sigmoid, None	Relu

n = number of observations (=217740)
p = number of predictor variables (=13)

A visual representation of the performance of the trained models on the validation dataset is provided by Fig. 4.11. The scatter plots on the top show the model response for each observation

related to the corresponding true response; the diagonal line indicates where the predicted response is matching the true response, i.e. a perfect prediction. Therefore, the closer the points are to the diagonal, the better the quality of the prediction. It can be noticed that the Neural Network model presents a lower dispersion of the points, lying mostly very close to the diagonal, especially if compared to the Tree model. This is reflected also by the trend of the differential errors between predicted and true values computed for each observation, showing values that are generally closer to zero for the Neural Network model compared to the other models. These results confirm what already highlighted by the validation performance metrics of the optimized models presented in Tab. 4.5, setting the Neural Network as the best performing model in terms of accuracy, followed by the Ensemble and the Tree models. Based on the overall analysis that has been carried out, the Ensemble model shows the lowest prediction speed, also having a much larger size than the others that would strongly affect the feasibility of implementation on a real-time hardware or standard control unit. However, all three optimized models have been considered for further investigations, in order to assess their accuracy and robustness on different test cases.
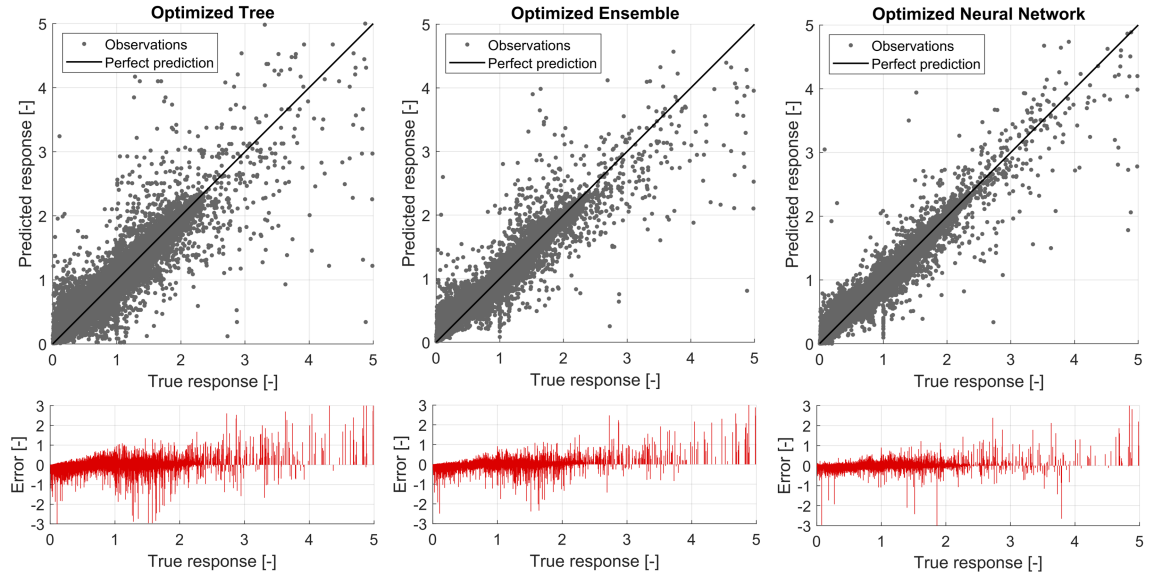

Figure 4.11: Validation results of optimized regression models: model response and related residual error for the considered observations

4.4 Absolute approach against residual approach

The same approach followed to select, optimize and train high-performing classification and regression models, as presented in Section 4.2 and Section 4.3, has been applied to a different input dataset: the same on-board available signals have been considered, but in this case the absolute values measured by the virtual sensors, i.e. the signals coming from the physical engine model,

have been used as input features for the data-driven models instead of the calculated residuals described in Section 4.2.2. Since the same development process already described in detail in this chapter has been followed, for the sake of brevity, only validation results of the best performing optimized models are presented in this section.

Fault detection model More in detail, Tab. 4.7 summarizes the overall accuracy reached on the validation dataset by the optimized Tree, Ensemble and Neural Network models, together with the time required for model optimization and training, prediction speed and final model size. In this case, differently from what has been observed in Section 4.2.4, the accuracy achieved by Neural Network and Ensemble classifiers is significantly higher compared to Tree model, with the Neural Network slightly outperforming the Ensemble model. However, the maximum classification accuracy achieved on the validation dataset is 89%, thus lower than the one observed for the residual approach (99%). Analysing the confusion matrices in Fig. 4.12 evaluated for each model, it can be noticed that the biggest errors are related to the misclassification of EGR and MAF faults, incorrectly identified as nominal conditions. On the other hand, MAP and EGC faults are still characterized by high TPR, especially considering the Neural Network classifier.

Table 4.7: Overall accuracy and training parameters of the optimized classification models trained on absolute dataset

Model	Accuracy [%]	Training time [s]	Prediction speed [obs/s]	Model size [MB]
Neural Network	89.0	21643	470000	0.057
Ensemble	87.7	1968	23000	146
Tree	79.2	85	1000000	2

NOx correction model Moving to the regression model, Tree, Ensemble, and Neural Network are still the best performing regressors compared to other algorithms. The main performance metrics evaluated on the validation dataset for the optimized models are shown in Tab. 4.8, together with training time (including model optimization), prediction speed, and model size. As observed in Section 4.3.2 for the residual approach, the optimized Neural Network exhibits the lowest RMSE, followed by the Ensemble model and finally the regression Tree. However, the best RMSE observed in this case, equal to 0.096, is significantly higher if compared to the best results achieved with the residual approach, where a minimum RMSE of 0.056 was observed. The same applies to the R^2, with values between 0.91 and 0.96 observed for the residual approach, and between 0.79 and 0.88 for the absolute approach introduced in this section. A visual representation of the models response on the validation dataset is given in Fig. 4.13: the scatter plots highlight the lower dispersion of Neural Network and Ensemble responses compared to the Tree model, as confirmed by the corresponding performance metrics.
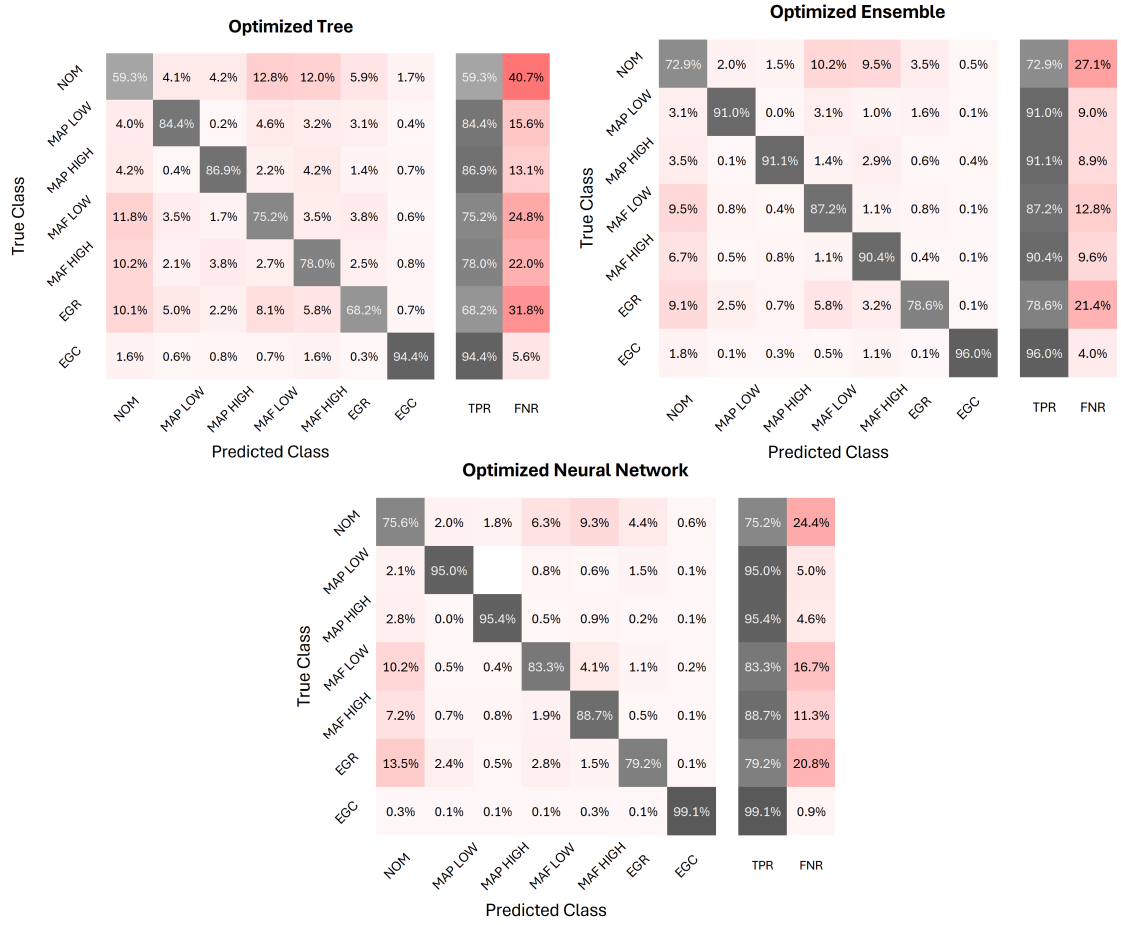
Figure 4.12: Validation results of optimized classification models trained on absolute dataset: related confusion matrices, true positive rates (TPR) and false negative rates (FNR) for each fault class

Table 4.8: Main performance and training parameters of different regression algorithms trained on absolute dataset

Model	RMSE [-]	MAE [-]	R2 [-]	Training time [s]	Prediction speed [obs/s]	Model size [MB]
Neural Network	0.096	0.045	0.88	33405	220000	0.622
Ensemble	0.106	0.035	0.85	1025	200000	42
Tree	0.126	0.046	0.79	312	1200000	1

Based on the validation results, the approach considering residual signals as input features to the data-driven fault detection and NOx correction models proves to be the best. This means that the calculated residuals contain clearer patterns that helps distinguishing between the different fault conditions and identifying non-nominal emission trends. However, in a real-world case, this
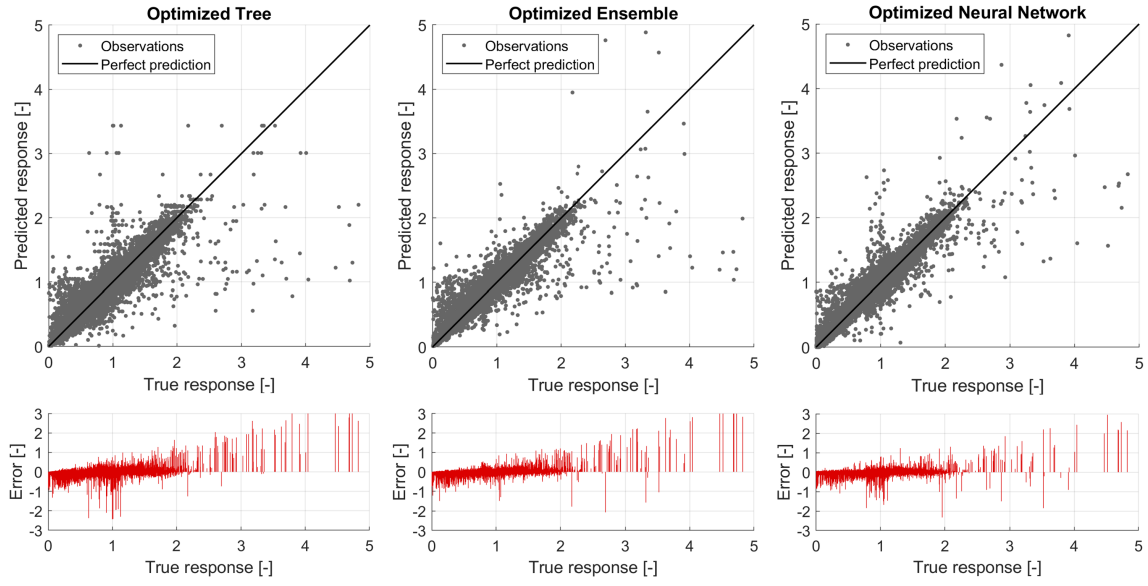
Figure 4.13: Validation results of optimized regression models trained on absolute dataset: model response and related residual error for the considered observations

approach would require the availability (or the additional implementation) of real-time models providing the required estimations to calculate all the residual signals, which is not required if on-board sensors signals are used directly as input to the data-driven models. Moreover, in the residual approach, classification and regression performance strongly depends on the accuracy of controller estimations, since it directly affects residuals calculation, as deeply analysed and discussed in Part II. Therefore, the consistency of ECU models, and consequently of the calculated residuals, must be evaluated to select the most suitable approach. Considering the specific application presented in this chapter, residual approach is preferred to absolute approach. However, for the sake of completeness, further comparisons between the two approaches applied to different test cases will be presented in Chapter 5.

Chapter 5

Application of the developed models

In this chapter, the optimized fault classification and NOx correction models described in Chapter 4 are tested on different simulated driving cycles to evaluate their performance on unseen input data, namely not included in the training dataset. More in detail, the vehicle model introduced in Chapter 3 has been used to simulate different driving cycles and fault conditions, including both single-fault scenarios and combinations of two simultaneous faults. The results obtained by the different models are compared in terms of accuracy, in order to select the best approach for both fault classification and NOx regression models. Finally, the feasibility of real on-board implementation is demonstrated by deploying the developed models on a real-time hardware to evaluate the requested computational load.

5.1   Test case scenarios

The validated classification and regression models have been integrated into the *Simulink* vehicle model by means of dedicated *Statistics and Machine Learning Toolbox*, to be directly tested on different simulated driving cycles and fault conditions. The selected driving cycles are standard homologating driving cycles, including WLTC (Worldwide Harmonized Light Vehicles Test Cycle, [186]), SFTP-US06 (Supplemental Federal Test Procedure, [187]), and FTP-72 (Federal Test Procedure,[188]), which differ from each other in terms of length, maximum speed, and harsh acceleration phases. In addition, an RDE-compliant driving cycle [16], different from the one considered for the training phase, has been included as well to complete the testing framework. The most relevant features of the considered driving cycles are summarized in Tab. 5.1. All the driving cycles have been run considering low battery initial state of charge, in order to force the engine to start within the first seconds of the driving cycle. The same ambient temperature, set to 25°C, has been considered for all the test cases, since the effect of variable environmental conditions on fault detection accuracy has not been investigated in this work.

111

As a first step, the fault detection models have been tested on single-fault scenarios, also including degrees of fault severity that were not included in the training dataset, in order to evaluate the interpolation and extrapolation capabilities of the developed models. Then, combinations of two simultaneous faults have been considered, to assess the robustness of the models in such testing conditions, despite being trained only on single-fault scenarios. In parallel, the NOx correction models have been applied on all mentioned driving cycles and different fault conditions, to evaluate the benefits achieved in terms of accuracy on estimated NOx emission compared to the reference controller model without any applied correction.

Table 5.1: Simulated driving cycles and relevant features. The calculated average speed does not take into account vehicle stops.

		WLTC	US06	FTP-72	RDE
Duration	[s]	1800	596	1369	6483
Distance	[m]	23266	12800	11997	90050
Perc. stop	[%]	13.4	2.2	13.8	7.0
Max. speed	[km/h]	131.4	128.9	91.2	121.2
Avg. speed	[km/h]	53.5	79.6	36.6	53.7
Max. acc.	[m/s^2]	1.61	3.78	1.43	3.16

5.2 Results and discussion

5.2.1 Fault detection and identification

The results in terms of classification accuracy on different driving cycles and fault conditions are summarized in Tab. 5.2 and Tab. 5.3. More in detail, Tab. 5.2 collects several tests where a single fault has been considered for each simulated driving cycle, while combinations of two simultaneous faults are considered for the test cases collected in Tab. 5.3. All the results presented in this section refer to the classifiers trained on residual signals, being the most promising approach according to the results obtained during the validation phase, as comprehensively discussed in Section 4.4.

In view of a real on-board application, the raw prediction given by the classification model has been filtered by means of a moving average. The length of the sliding window has been calibrated to 10 seconds, a trade-off that allows not to lose relevant information, while not being too sensitive to minor model inaccuracies. This means that, for each time instant, the fault detection model outputs the label of the fault class which has been most frequently observed in the past 10 seconds, i.e. the one with the highest probability. The detected classes shown in Tab. 5.2 consider

just the averaged prediction signal, so that negligible, but still unavoidable, misclassification errors are not considered. The related accuracy corresponds to the TPR calculated on the raw prediction signal, which is a useful parameter to evaluate the effective performance of each model for the different test cases, as already discussed in Chapter 4. According to the results, all the models are able to detect the true fault class, with no misclassification errors if the mentioned moving average is applied to the raw prediction signal. The very high accuracy, close to or equal to 100%, on the test cases involving EGR cooler and MAP sensor faults, confirms what already observed in the validation phase (Section 4.2.4). Compared to the Neural Network and Ensemble models, a slightly lower accuracy is observed for the Tree model for both nominal and MAF sensor fault cases, which is still far above 90%, except for the US06 nominal case. A brief consideration must be added with regard to the cases involving EGR valve clogging: this type of fault imposes a limitation on the EGR valve opening which depends on the fault severity, as observed in Fig. 5.1; thus, differently from the other considered faults, which affect all the engine operating points, only in the operating points where the requested EGR valve position is higher than the threshold imposed by the clogging, an effect due to the fault can be observed. For this reason, the fault can be effectively detected only when the requested EGR valve position is above the maximum opening imposed by the fault, which, on the same driving cycle, happens more or less frequently depending on the fault severity. As a result, the classification models correctly recognize both EGR valve fault and nominal conditions when EGR valve clogging is considered, and the percentage of observations classified as EGR valve fault increases if a lower maximum opening of the EGR valve is allowed.

Overall, all the developed models show a robust and reliable behaviour on single fault classification on different unseen driving cycles, including interpolation and extrapolation capabilities on selected fault conditions which are not included in the training dataset (i.e. EGC 40%, EGC 80%, MAP LOW -40%, MAP LOW -30%, MAF HIGH +40%). Moreover, the Ensemble model is confirmed to be the most accurate, as observed in the validation phase, followed by the Neural Network model, which shows similar performance in most of the test cases.

Focusing on Tab. 5.3, the detected fault classes in case of combinations of two simultaneous faults, based on the averaged prediction signal, are shown. Each detected class corresponds to a percentage of observations, calculated on the raw prediction, classified as the considered class, as seen for Tab. 5.2. In the test cases where a combination of MAF and MAP sensors drift with the same absolute value is considered (e.g. MAF LOW -15% + MAP LOW -15%), the models fail to detect both faults, but only MAP sensor fault is detected, for almost 100% of the observations. On the other hand, if a lower severity is considered for the MAP sensor fault with respect to the MAF sensor fault (e.g. MAF LOW -25% + MAP LOW -5%), the models are able to detect also MAF sensor fault. However, only the Neural Network is capable of correctly detecting the true fault combination, while both Ensemble and Tree models show significant false negatives corresponding to EGR valve fault detection. In the remaining test cases, combining MAP or MAF

Table 5.2: Detected classes (after prediction averaging) and related accuracy (calculated on raw prediction) for the three classification models on different single-fault test cases

Test Case			Detected Classes Accuracy [%]					
Driving Cycle	True fault class	Fault severity	Optimized Tree		Optimized Ensemble		Optimized Neural Network	
WLTC	NOM	-	NOM 94.1%		NOM 97.0%		NOM 96.7%	
	MAF HIGH	15%	MAF HIGH 99.2%		MAF HIGH 99.7%		MAF HIGH 99.5%	
	EGR	30%	EGR 50.2%	NOM 47.1%	EGR 49.3%	NOM 49.2%	EGR 49.9%	NOM 48.3%
	EGC	80%	EGC 99.9%		EGC 100%		EGC 100%	
	MAP LOW	-30%	MAP LOW 100%		MAP LOW 100%		MAP LOW 100%	
US06	NOM	-	NOM 87.8%		NOM 93.1%		NOM 92.7%	
	MAF LOW	-10%	MAF LOW 95.1%		MAF LOW 97.7%		MAF LOW 95.2%	
	EGC	70%	EGC 99.8%		EGC 99.9%		EGC 99.9%	
	MAP LOW	-20%	MAP LOW 99.9%		MAP LOW 100%		MAP LOW 99.9%	
FTP-72	MAF HIGH	+10%	MAF HIGH 98.7%		MAF HIGH 99.2%		MAF HIGH 99.0%	
	MAP HIGH	+20%	MAP HIGH 100%		MAP HIGH 100%		MAP HIGH 100%	
	EGR	40%	EGR 13.9%	NOM 79.9%	EGR 15.5%	NOM 81.9%	EGR 15.4%	NOM 81.5%
	EGR	10%	EGR 87.4%	NOM 12.2%	EGR 82.2%	NOM 11.8%	EGR 88.6%	NOM 11.3%
RDE	EGC	40%	EGC 100%		EGC 100%		EGC 100%	
	MAP LOW	-40%	MAP LOW 100%		MAP LOW 100%		MAP LOW 99.5%	
	MAF HIGH	+40%	MAF HIGH 98.7%		MAF HIGH 99.6%		MAF HIGH 99.6%	

sensors faults with EGR cooler or EGR valve faults, all three models show a good behaviour: they are able to correctly detect both fault classes, with no false class detection, except for one case in which also the nominal class is detected, even if in a low percentage, which anyway could be acceptable on a real application since not leading to a wrong fault detection. Based on these results, when a combination of two faults is considered, the Neural Network classification model shows the preferred behaviour: in most of the considered test cases it is able to correctly detect both true faults, and no false fault detections are observed.

In the following paragraphs, three of the considered test cases are deeply analysed to show in detail how the models behave.

Table 5.3: Detected classes (after prediction averaging) and related accuracy (calculated on raw prediction) for the three classification models on different double-fault test cases

Test Case					Detected Classes — Accuracy [%]		
Driving Cycle	True fault classes		Fault severity		Optimized Tree	Optimized Ensemble	Optimized Neural Network
WLTC	MAF LOW	MAP LOW	-15%	-15%	MAP LOW 99.9%	MAP LOW 99.9%	MAP LOW 99.9%
	MAF LOW	MAP LOW	-25%	-5%	MAF LOW 14.2% / MAP LOW 70.4% / EGR 15.3%	NOM 1.8% / MAF LOW 12.1% / MAP LOW 80.1% / EGR 6.0%	MAF LOW 10.4% / MAP LOW 89.6%
	MAP LOW	EGC	-10%	50%	MAP LOW 19.1% / EGC 78.8%	MAP LOW 21.9% / EGC 78.1%	MAP LOW 13.2% / EGC 86.8%
	MAF LOW	EGR	-15%	30%	MAF LOW 86.8% / EGR 11.3%	MAF LOW 87.5% / EGR 12.3%	MAF LOW 97.6% / EGR 2.4%
FTP-72	MAF HIGH	MAP HIGH	+15%	+15%	MAP HIGH 99.9%	MAP HIGH 99.9%	MAP HIGH 100%
	MAF LOW	MAP HIGH	-15	+15%	MAP HIGH 99.9%	MAP HIGH 99.9%	MAP HIGH 99.9%
	MAP HIGH	EGC	+10%	70%	MAP HIGH 21.3% / EGC 78.7%	MAP HIGH 8.0% / EGC 92.0%	MAP HIGH 21.2% / EGC 78.6%
	MAF HIGH	EGR	+15%	20%	NOM 4.7% / MAF HIGH 6.0% / EGR 87.8%	NOM 4.9% / MAF HIGH 5.1% / EGR 90.3%	NOM 8.2% / MAF HIGH 10.7% / EGR 79.1%

Use case 1 A WLTC driving cycle with clogged EGR valve, leading to 30% maximum valve opening, is considered in the use case in Fig. 5.1. Starting from the bottom, engine speed and load are shown; the plot in the centre compares the requested EGR valve position (grey line) and the actual one (black line), which is affected by the induced fault imposing a maximum valve opening (dashed line). The plot on the top shows the instantaneous fault class predicted for each simulation time step, i.e. every 10 ms, by the optimized Tree model (grey line), together with the fault class effectively detected when a moving average with a window length of 10 seconds is applied (red line). The phases in which the engine is switched off have been excluded from the prediction and identified under the label "ICE OFF". The true fault class is represented by the black dashed line. As already explained in the previous paragraphs, the classification model is able to detect the fault when actually present: it can be observed that the model recognizes the presence of the fault each time that the actual EGR valve opening is limited with respect to the requested one, namely when the engine operating condition actually differs from the nominal case.

Use case 2 A US06 driving cycle with a MAF sensor low drift of -10%, is considered in this use case, as shown in Fig. 5.2. The fault has been introduced in the model after 200 s from the beginning of the cycle, to better highlight the reaction of the classification model to the fault introduction. As in the previous case, engine speed and torque are shown in the bottom plot; the instantaneous

fault class predicted by the optimized Ensemble model (grey line), together with the fault classes effectively detected applying the moving average (red line), is shown in the upper plot. The true fault class, corresponding to the black dashed line, highlights the introduction of the fault. As expected, the classification model is able to detect the fault as soon as it is introduced, with a slight delay depending on the moving average, which however proves to be effective in filtering eventual signal spikes due to negligible instantaneous misclassification errors.

Use case 3   The results on FTP-72 driving cycle with a combination of two faults, namely 10% MAP sensor high drift and 70% EGR cooler efficiency, are shown in Fig. 5.3. As in the previous cases, in the upper plot, the instantaneous fault class predicted by the optimized Neural Network model (grey line), together with the fault classes effectively detected applying the moving average (red line), is shown. The true fault classes, corresponding to the black dashed lines, highlight the introduction of the two different faults. Even in the case of fault combination, the classification model is able to correctly detect both faults, even if the EGC fault class is predominant, as confirmed by the results in Tab. 5.3.
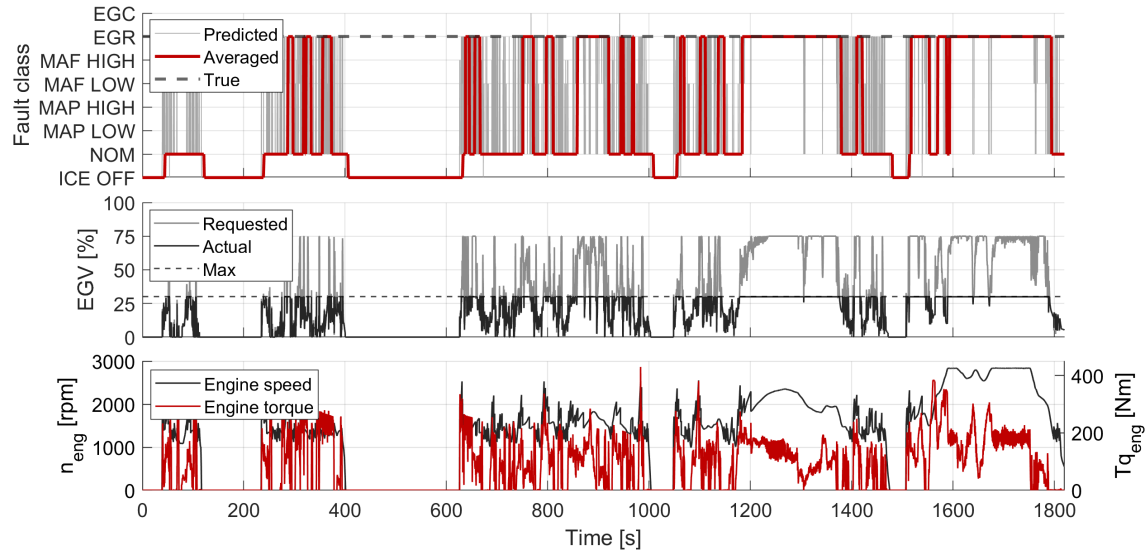
Figure 5.1: Use case 1: test of optimized Tree model on WLTC driving cycle with clogged EGR valve (30% maximum valve opening)

5.2.2   NOx emission estimation

The results in terms of percentage error on the final cumulated NOx emission on different driving cycles and fault conditions are summarized in Fig. 5.4-5.7. These results refer to the models trained on residual signals, being the most promising approach according to the results obtained during the validation phase, as comprehensively discussed in Section 4.4. More in detail, Fig. 5.4
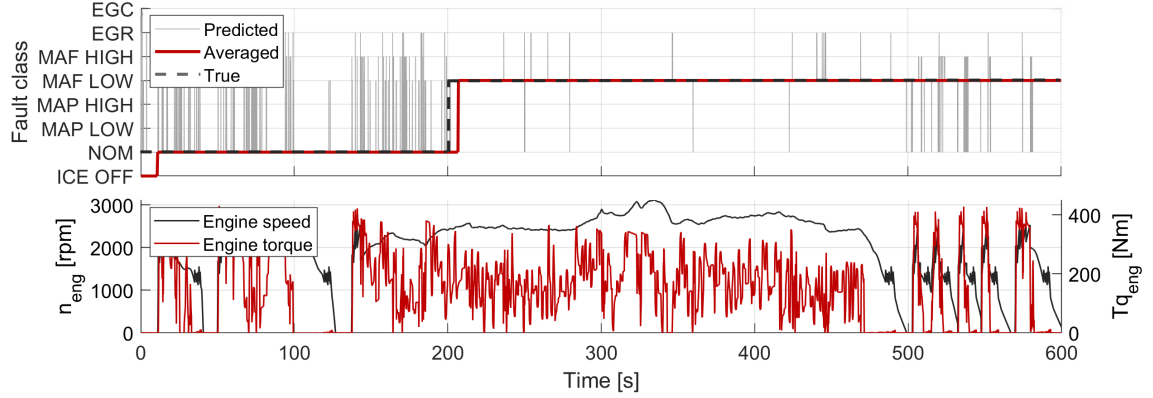
Figure 5.2: Use case 2: test of optimized Ensemble model on US06 driving cycle with 10% MAF sensor low drift starting from 200 s
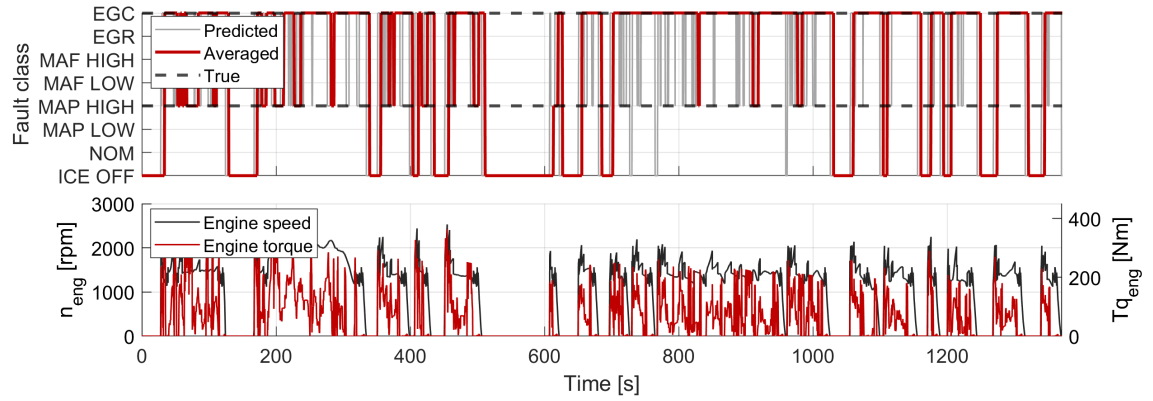

Figure 5.3: Use case 3: test of optimized Neural Network on FTP-72 driving cycle with combined faults, 10% MAP sensor high drift and 70% EGR cooler efficiency

shows the performance of the three optimized regression models on WLTC cycles where different fault conditions have been introduced: the black hatched bars represent the variation of cumulated NOx emission due to the introduction of the fault compared to the nominal condition; the red hatched bars indicate the deviation between the cumulated NOx estimated by the engine controller and the actual NOx emission; analogously, the remaining red, blue and grey bars represent the error between the cumulated NOx estimation corrected by means of the developed models, namely optimized Tree, Ensemble, and Neural Network models respectively, and the actual NOx emission. Based on these results, it can be observed that the applied corrections allows to reduce the error on cumulated NOx emission, compared to the reference ECU estimation, in all the considered fault cases. The three different data-driven models show comparable performance, with the Neural Network achieving the lowest error for MAP low drift, MAP high drift, and MAF high drift test cases, and the Ensemble outperforming the other models in the test cases considering MAF low

drift, EGR valve clogging and EGR cooler reduced efficiency. Similar conclusions can be drawn for the tests performed on FTP-72 cycles, shown in Fig. 5.5, where all the developed regression models allow to effectively correct the ECU NOx estimation, achieving better results than the reference controller model. In this case, the Ensemble model shows the overall best performance, except for the EGR valve fault, even if comparable results are achieved by the Tree and Neural Network models as well. Fig. 5.6 shows the results obtained on the US06 driving cycle: the Ensemble model still outperforms the reference controller model in all the considered test cases, followed by the Neural Network, that shows a worse performance on the MAF high drift case, and finally the Tree model, showing a significant error especially on the MAP high drift case. Fig. 5.7 summarizes the results obtained by testing the same optimized regression models on an RDE driving cycle, different from the one used in the training phase. All the three data-driven approaches allow to reduce the error on the estimation of the final cumulated NOx emission compared to the controller model, showing comparable performance except for the MAP high drift case, where the Neural Network show a significantly worse performance compared to Tree and Ensemble models.
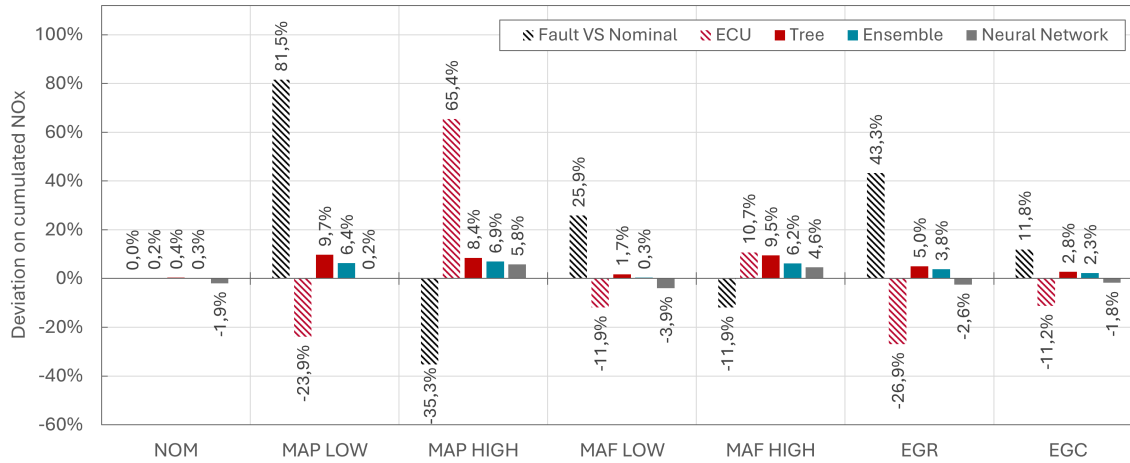
Figure 5.4: Test of optimized regression models with residuals input features on WLTC driving cycle

For the sake of completeness, Fig. 5.8 shows a comparison between the results obtained with Tree and Neural Network models trained on absolute sensor signals (black and green bars respectively) and the ones obtained with Tree and Neural Network models trained on residual signals. In general, excluding the EGR cooler fault case, all the considered regression models allow to improve the NOx prediction performance compared to the reference controller model. However, except for the test cases involving MAP and MAF sensors high drift, where all the models show comparable performance independently from the type of input signals considered, in all the other fault cases leading to an increase of actual NOx emissions, a better performance of the models trained on residual signals is observed. Moreover, the error shown by the models trained on absolute signals is not only larger in absolute value, but also it tends to be a significant underestimation of the actual NOx emission, which is even less accepted than an overestimation. Similar considerations
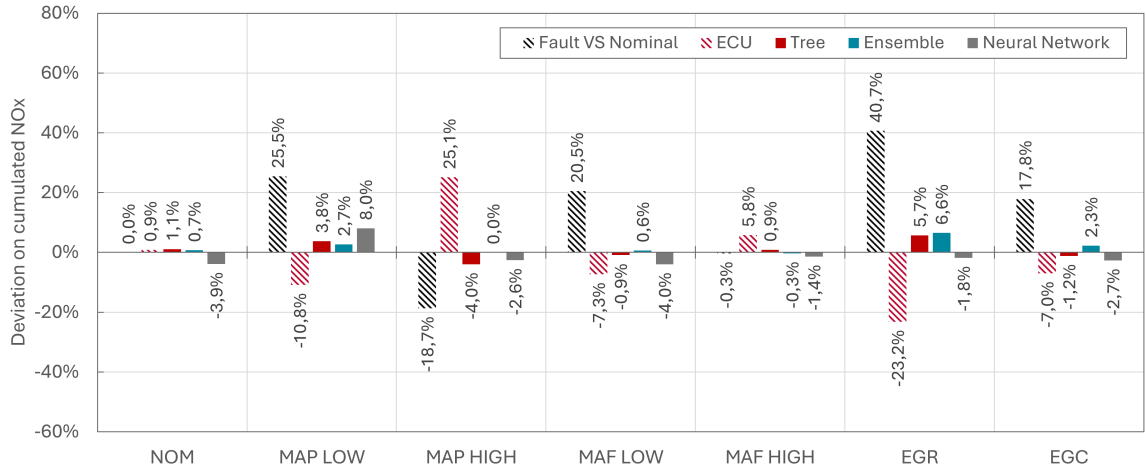
Figure 5.5: Test of optimized regression models with residuals input features on FTP-72 driving cycle
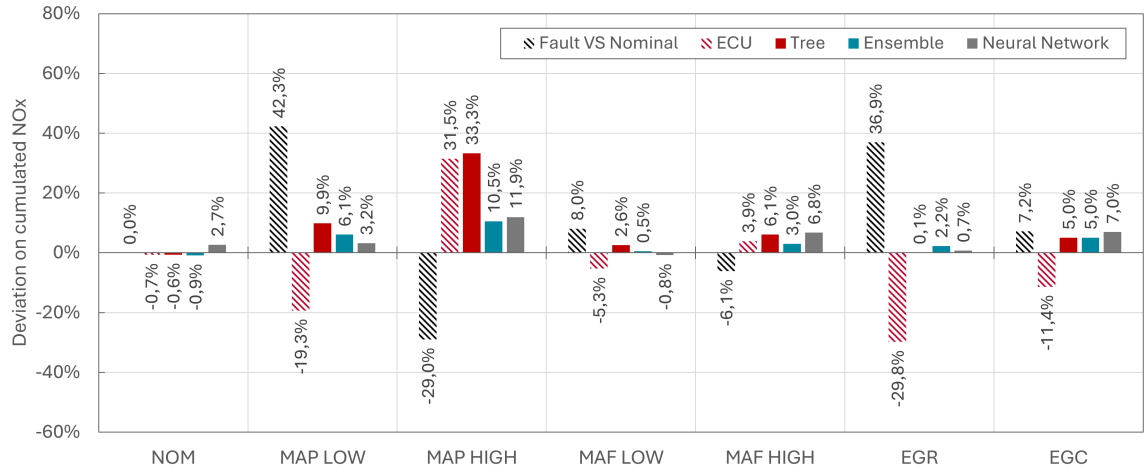

Figure 5.6: Test of optimized regression models with residuals input features on US06 driving cycle

are valid also for additional tests performed on different simulated driving cycles. Therefore, these results confirm what already highlighted in Section 4.4 during the training and validation phase, namely that the residual-based approach is preferable for this specific application.

To summarize, the developed NOx correction models show promising results when tested on different unseen driving cycles, allowing to significantly reduce the gap between actual and estimated NOx emission even if highly emission-relevant faults are considered. Good performance are achieved by all the three tested models, significantly better than the reference controller model in most of the test cases, with none of them clearly outperforming the others. However, the Tree model usually shows the highest errors, followed by the Neural Network and Ensemble models, with the latter achieving the best performance in most of the selected cases, even if the overall
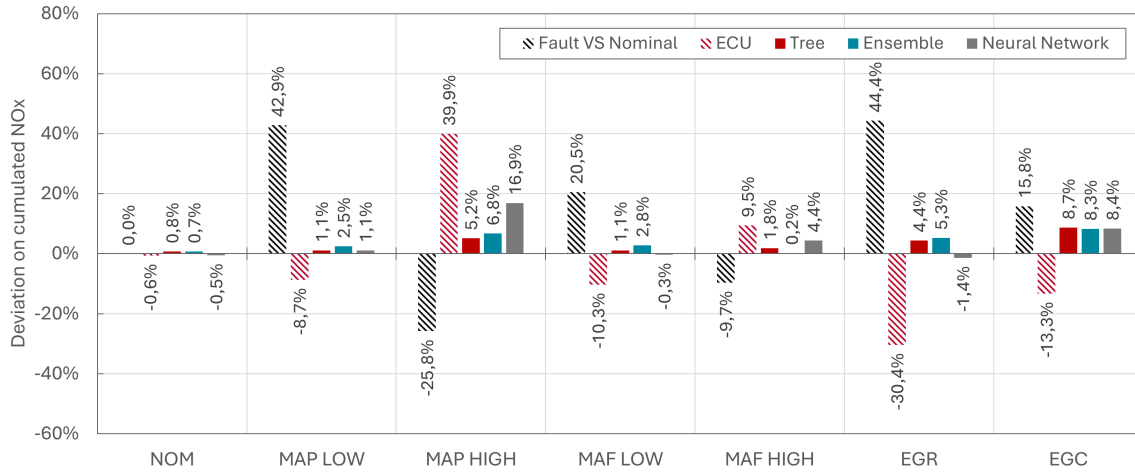
Figure 5.7: Test of optimized regression models with residuals input features on RDE driving cycle
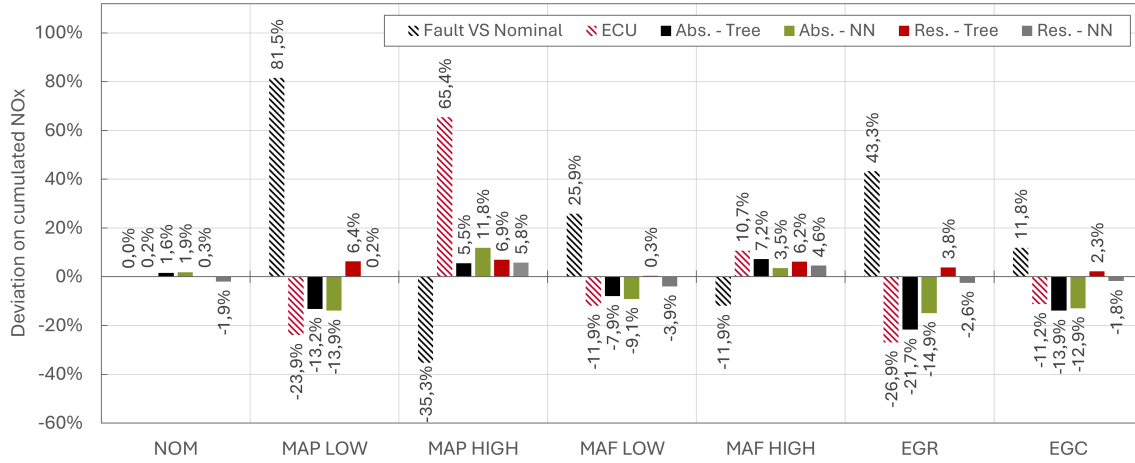

Figure 5.8: Comparison of optimized regression models trained on absolute and residuals input features on WLTC driving cycle

performance is definitely comparable. On the other hand, as later discussed in Section 5.2.3, especially in view of a real-time on-board application, the Neural Network model exhibits significant advantages related to the limited computational load and especially to the extremely small size (in terms of required memory), up to three orders of magnitude less than the Ensemble model. For this reason, for a similar performance in terms of accuracy, the Neural Network is preferred compared to the Ensemble model.

In the following paragraphs, the results of the application of the Neural Network regression model to correct the NOx emission estimated by the engine controller on four different test cases are more deeply analysed, to clarify in detail how the model behaves both in terms of instantaneous and cumulated emission trends. For reasons of confidentiality, in all the presented plots, NOx

emissions have been normalized with respect to the maximum value observed on the same cycle in the nominal case.

Use case 1   A WLTC driving cycle with -20% MAP low drift is considered in the use case represented in the upper plot of Fig. 5.9. The solid lines represent the instantaneous NOx emission, while the dashed lines represent the cumulated emission. As expected, a significant increase of actual NOx emission (black line) is observed compared to the nominal condition (grey line). Even if the ECU model is reacting to the introduction of the fault, as confirmed by a higher NOx emission estimated by the controller (red line) compared to the nominal one, this adaptation is not enough to guarantee an acceptable error on NOx estimation, which still higher than 20%. On the other hand, when the NOx emission correction, calculated by the optimized Neural Network model, is applied to the controller estimation (blue line), the resulting deviation between actual and estimated NOx emission is almost zero, with a residual error below 1%.

Use case 2   A WLTC driving cycle with +20% MAP high drift is considered in this use case, represented in the centre plot of Fig. 5.9. As for the previous case, the solid lines represent the instantaneous NOx emission, while the dashed lines represent the cumulated emission. As expected, a significant decrease of actual NOx emission (black line) is observed compared to the nominal condition (grey line). In this case, the ECU model is not correctly reacting to the introduction of the fault, with a slightly higher NOx emission estimated by the controller (red line) compared to the nominal one, resulting in a significant error between estimated and actual cumulated emission, which is higher than 60%. On the other hand, when the NOx emission correction, calculated by the optimized Neural Network model, is applied to the controller estimation (blue line), the resulting deviation between actual and estimated NOx emission is reduced to less than 6%.

Use case 3   An FTP-72 driving cycle with clogged EGR valve at 20% of the nominal maximum opening, is considered in the test case shown in the upper plot of Fig. 5.10. As for Fig. 5.9, the solid lines represent the instantaneous NOx emission, while the dashed lines represent the cumulated emission. As expected, an increase of actual NOx emission (black line) is observed compared to the nominal condition (grey line). In this case, the ECU model (red line) is not significantly reacting to the introduction of the fault, leading to an error between estimated and actual cumulated emission of more than 20%. On the other hand, when the NOx emission correction, calculated by the optimized Neural Network model, is applied to the controller estimation (blue line), the resulting deviation between actual and estimated NOx emission is reduced to an underestimation of less than 2%.

Use case 4   An FTP-72 driving cycle with 50% EGR cooler efficiency is considered in this use case, represented in the centre plot of Fig. 5.10. As for the previous case, the solid lines represent the instantaneous NOx emission, while the dashed lines represent the cumulated emission. As

expected, an increase of actual NOx emission (black line) is observed compared to the nominal condition (grey line), even if less evident than in the other fault cases. The controller model (red line) is partially adapting its prediction after the introduction of the fault, with a residual error between estimated and actual cumulated emission of 7%. On the other hand, when the NOx emission correction, calculated by the optimized Neural Network model, is applied to the controller estimation (blue line), the resulting deviation between actual and estimated NOx emission is reduced to less than 3%.
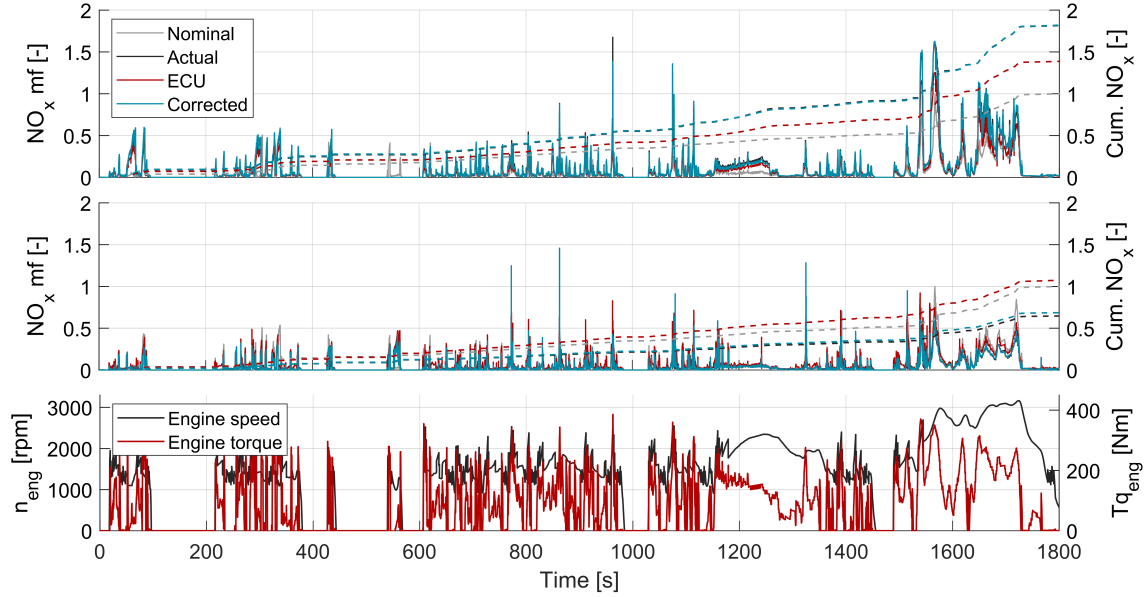
Figure 5.9: Use case 1 and 2: test of optimized Neural Network model with residuals input features on WLTC driving cycle with -20% MAP low drift (top) and +20% MAP high drift (centre). NOx emissions are normalized with respect to the nominal case

5.2.3 Real-time implementation and testing

As a final step, the developed classification models have been implemented on a real-time hardware to evaluate the feasibility of a real on-board implementation in terms of requested computational load. As already introduced in Section 3.4, the hardware selected for this application is a Raspberry Pi 4 computer, coupled with a PiCAN2 board. The developed classification and regression models have been integrated into the *Simulink*-based software deployed on the Raspberry Pi by means of the *Matlab Support Package for Raspberry Pi Hardware* [189]. The Raspberry Pi can successfully communicate with the the real-time pc and the rapid-prototyping control unit, receiving and sending signals via CAN-bus. This way, all signals required as input to the data-driven models are read and processed by the Raspberry Pi, which outputs in real time the computed fault
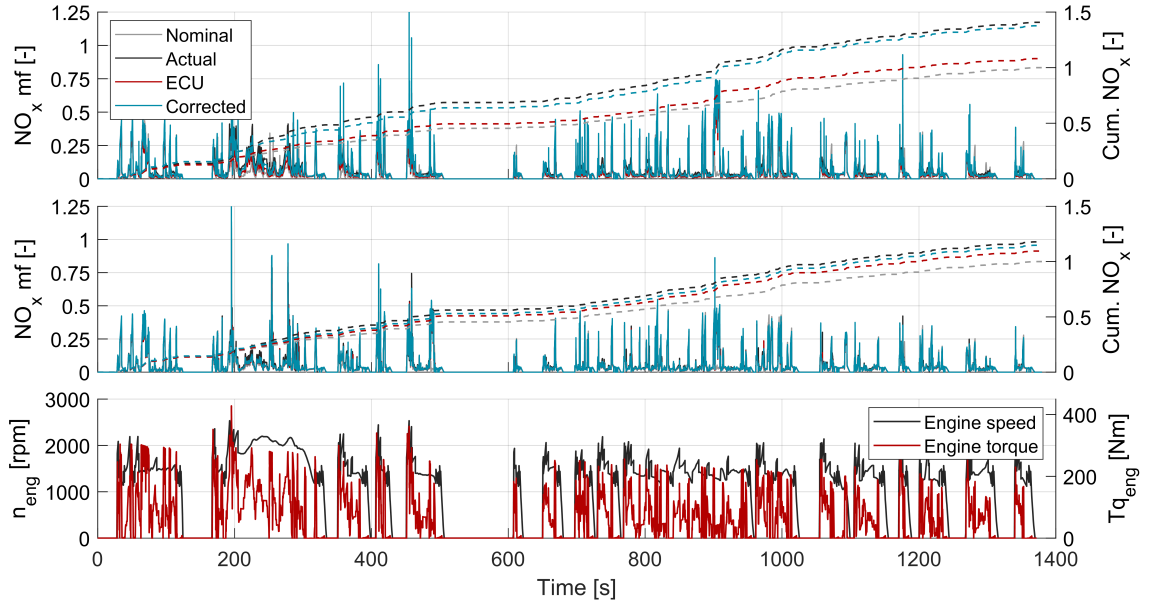
Figure 5.10: Use case 3 and 4: test of optimized Neural Network model with residuals input features on FTP-72 driving cycle with 20% EGR valve clogging (top) and 50% EGR cooler efficiency (centre). NOx emissions are normalized with respect to the nominal case

class and NOx correction factor. The *Code Profile Analyzer* tool by *Matlab* has been used to monitor and analyse the execution time profile of the real-time applications running on the Raspberry Pi. Considering a time-step of 10 ms, which is a commonly set value for standard control units, including the rapid-prototyping control unit used in this case, the results summarized in Tab. 5.4 show that all the tested models are real-time capable, with a computational load that reaches a maximum value of 14% of CPU utilization in the case of the Ensemble classification model and 26% for the Ensemble regression model. As expected from the preliminary performance analysis introduced in Section 4.2.3, the Tree model shows the shortest task execution time among the classifiers, with an average of 0.3 ms and a maximum of 1 ms, followed by the Neural Network classifier, which shows a similar behaviour with 0.5 ms and 1.2 ms of average and maximum task execution time respectively. The Ensemble classifier shows a significantly higher task execution time compared to the other models, with an average value of 1.9 ms and a maximum of 2.8 ms, which however is far within the considered application time-step of 10 ms. Analogously, among the regression models, the Tree and Neural Network regressors show almost the same behaviour, with the shortest average and maximum task execution time, namely 0.2 ms and 0.7 ms respectively. The Ensemble regression model shows a significantly higher task execution time compared to the other models, with an average value of 1.6 ms and a maximum of 2.7 ms, however far below the real-time threshold.

The same results in terms of task execution time can be visualized in Fig. 5.11 and Fig. 5.12

respectively for the tested classification and regression models. Focusing on Fig. 5.11, the plot highlights a wider distribution of execution time for the Ensemble model, while less dispersion is observed for the Neural Network and the Tree model in particular. The maximum execution time for each model, indicated by the dashed lines in the plots, is significantly higher than the average value, but it is only detected within the first iterations, so that the corresponding normalized frequency is almost zero. Similar considerations are valid also for the distribution of the task execution time of the regression models shown in Fig. 5.12.

Table 5.4: Maximum and average execution time, CPU utilization, and required memory for the selected classification and regression models deployed on Raspberry Pi

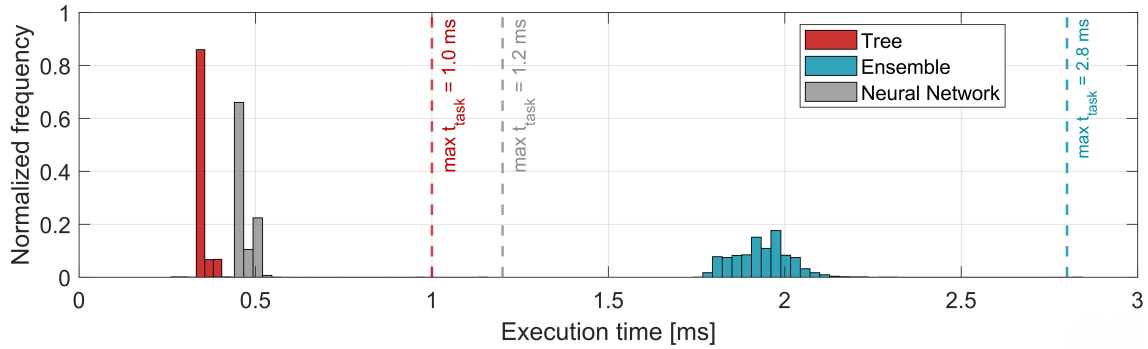| Model	Task execution time [ms]		CPU utilization [%]		Memory [MB]
	Max	Avg	Max	Avg	
Classification models					
Tree	1.0	0.3	4.8	1.7	~0.4
Ensemble	2.8	1.9	14.2	9.7	~128
Neural Network	1.2	0.5	5.9	2.3	<0.1
Regression models					
Tree	0.7	0.2	6.5	2.2	~1
Ensemble	2.7	1.6	26.7	15.7	~89
Neural Network	0.7	0.2	6.7	2.3	<0.1

Figure 5.11: Distribution of task execution time for optimized Tree, Ensemble and Neural Network classification models (real-time application)

5.3 Conclusions of Part I

To summarize, focusing on the fault classification model, the results discussed in Section 5.2.1 highlight that:
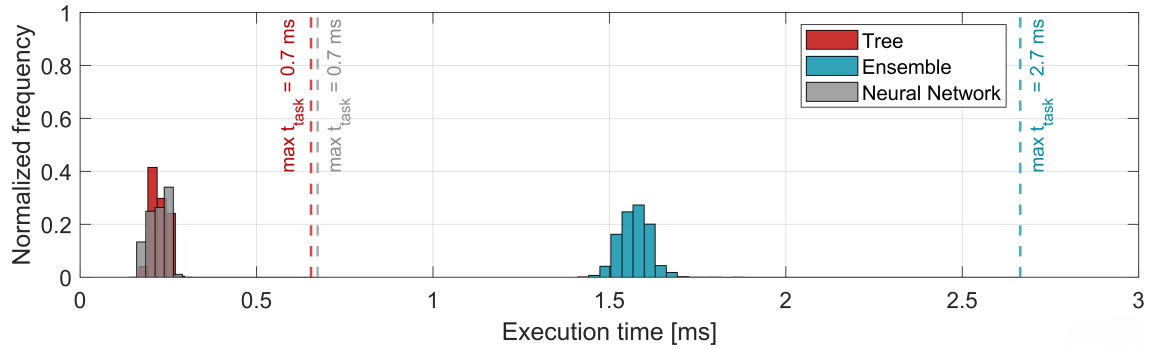
Figure 5.12: Distribution of task execution time for optimized Tree, Ensemble and Neural Network regression models (real-time application)

- all three optimized classification models trained on residual signals are able to correctly detect and identify single faults with high overall accuracy, regardless of the considered driving cycle or fault severity;

- the Ensemble model has proven to be the most accurate, followed by the Neural Network model, which shows similar performance in most of the test cases;

- no false fault detections are observed for the considered test cases, provided that a moving average is applied to the raw prediction signal to neglect instantaneous misclassification errors;

- when a combination of two faults is considered, the Neural Network classification model shows the best behaviour: in most of the considered tests it is able to correctly detect both the induced faults without false fault detections, even if trained just on single-fault conditions.

In parallel, based on the results presented in Section 5.2.1 for the NOx correction model, it can be concluded that:

- all three optimized regression models allows to correct the NOx emission estimated by the reference controller model, significantly reducing the error between the estimated and actual cumulated emission at the end of each simulated driving cycle, for almost all the considered fault cases;

- the models show comparable performance, with the Ensemble model slightly outperforming the others in terms of overall accuracy, being able to reduce the error on the final cumulated emissions in all the considered test cases compared to the reference controller model. The Neural Network model exhibits a similar behaviour, showing even higher accuracy than the Ensemble model in some test cases, but failing to improve the controller performance on one test case, namely the US06 driving cycle with MAF high drift fault, where only the Ensemble model is able to achieve the improvement target;

125

- the models trained on residual signals instead of absolute physical signals achieve the best overall results, confirming the preliminary results obtained during the validation phase.

Finally, the implementation of the developed classification and regression models on a real-time hardware, as described in Section 5.2.3, demonstrates that:

- all the optimized models deployed on the Raspberry Pi are real-time capable, with a maximum task execution time of 2.8 ms for the Ensemble classifier application, and a maximum CPU utilization of 27% observed for the Ensemble regression model;

- thanks to the coupling with a PiCAN board, the Raspberry Pi is able to communicate via CAN-bus with the rapid-prototyping control unit and the virtual sensors providing all the signals required by the data-driven models, as it would happen on a real vehicle;

- despite achieving the highest accuracy, the Ensemble-based classification and regression models are characterized by relatively low prediction speed and a significant model size, much larger than Neural Network and Tree models, that is an important drawback in view of future implementation on standard control units. Therefore, the Neural Network-based models provide the best overall performance if both accuracy and computational load are considered.

Part II

Experimental application

Chapter 6

Experimental data acquisition

I n Part I of this dissertation, a methodology was clearly defined for the development of OBM-oriented data-driven models, taking into account the overall performance in terms of model accuracy, complexity, and computational efficiency in view of on-board application. Given the promising results obtained in simulation for both classification and regression tasks, the developed methodology has been applied to experimental data acquired at the test bench, to definitely prove the effectiveness of the proposed approach for on-board fault detection and NOx emission monitoring, assessing its performance on a real-world application. In this regard, in this chapter, a description of the engine installed at the test bench is provided. Then, an overview of the implemented faults is given, followed by a critical analysis of the effects of each fault on NOx emissions and on the most relevant measured signals. Finally, the accuracy and reliability of built-in ECU models estimating NOx emissions and the most significant engine operating parameters is discussed, considering both nominal and fault conditions.

6.1 Engine architecture and specifications

The engine installed at the test bench is a 2.2-liters 4-cylinder diesel engine for LCV applications, equipped with a single-stage VGT turbocharger and double EGR system, namely High-Pressure (HP) and Low-Pressure (LP) EGR. The injection system is a Bosch common rail type, with a maximum injection pressure at the nozzles of 2000 bar, allowing sequential multiple injections at part load and post injections for DPF control. The shareable engine technical specifications are outlined in Tab. 6.1, while Fig. 6.1 shows the full load torque and power curves. Due to confidentiality constraints, not all engine specifications are made explicit, while torque and power curves have been normalized with reference to their maximum values. All the signals listed in Tab. 6.2 are measured by on-board sensors, according to the sensor layout of the commercial engine, shown in Fig. 6.2. Intake manifold temperature, after HP EGR mixing, is not directly measured by an on-board sensor, however this signal has been acquired at the test bench by an

additional temperature sensor and it has been considered for the sensitivity analysis described in Section 6.2. However, only on-board measured signals have been taken into account for the development of the data-driven models presented in Chapter 7.

Table 6.1: Main engine technical specification

Engine type	Compression Ignition
Number of cylinders	4, in line
Displacement	≈ 2.2 l
Fuel injection system	Central, multi-hole direct injection, 2 high-pressure pumps
Induction system type	Turbocharged, single stage
Compression ratio	> 15 : 1
Rated power	@ 3500 rpm
Rated torque	@ 1500 rpm
Application	LCV
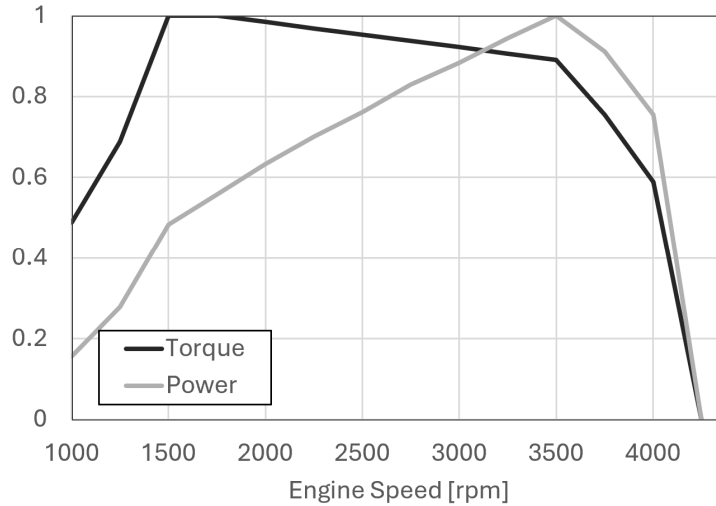
Figure 6.1: Normalized engine full load curve

6.2   Tested faults: overview, implementation, and sensitivity analysis

All the tests have been run following the torque and speed profile corresponding to an FTP-72 (Federal Test Procedure, [188]) driving cycle, shown at the bottom of Fig. 6.3-Fig. 6.5. Different fault conditions have been considered, which were selected based on their potential effect on pollutant emissions (especially NOx emission) and on fault implementation feasibility. In fact,

Table 6.2: List of available on-board sensors signals according to commercial engine layout

Abbreviation	Description
T_{amb}	Ambient temperature
P_{amb}	Ambient pressure
$\dot{m}_{air}$	Intake air mass flow
P_2	Intake manifold pressure
T_{21}	Temperature downstream compressor
T_{22}	Temperature downstream water-charge air cooler
$T_{HPEGR-Cold}$	Temperature downstream HP EGR cooler
$T_{LPEGR-Cold}$	Temperature downstream LP EGR cooler
T_3	Exhaust manifold temperature
T_4	Temperature downstream turbine (upstream DOC)
λ	Oxygen concentration after combustion
$T_{coolant}$	Engine coolant temperature
$y_{NOx_{EO}}$	Engine-out NOx concentration
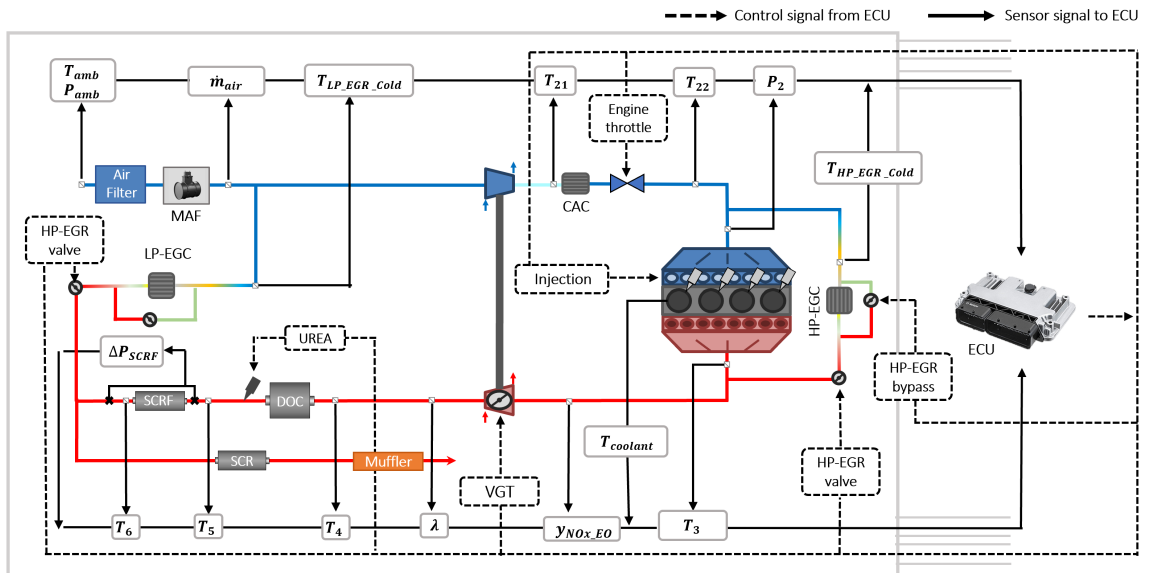
Figure 6.2: Schematic architecture and sensor layout of the commercial engine

ECU diagnosis must not be activated by the injected faults not to compromise the test, but, at the same time, potential engine damages must be prevented as well. Based on these considerations, the performed tests are listed below, together with a brief explanation of the implementation and effects of the considered faults. Each test has been performed twice to compensate possible acquisition errors, while the sampling frequency has been set to 100 ms.

6.2.1 MAF sensor faults

MAF sensor high drift   In general, the drift error in sensor readings may occur due to sensor aging, sensor contamination, or upstream plumbing changes. In this case, it has been implemented at the test bench by changing the sensor characteristic, specifically by introducing a multiplying factor of 1.075 and 1.15 to emulate +7.5% and +15% MAF sensor drifts. This way, the ECU receives a wrong measured signal from the MAF sensor, which is higher than the actual air mass flow. This means that the controller is acting as if the target air mass flow is intaken by the engine, corresponding to the wrong sensor reading (top plot of Fig. 6.3), while the actual air mass flow is lower. To achieve the target intake manifold pressure despite the lower intake air mass, a higher amount of EGR is requested, as it can be observed by the increased HP and LP EGR valves opening in Fig. 6.5, thus resulting in a lower NOx emission, as shown in Fig. 6.6. An increased injected fuel mass is observed compared to the nominal case, as observed in Fig. 6.3, due to the efficiency reduction caused by the increased EGR mass. Accordingly, considering also the lower intake air mass, a lower lambda value is measured compared to the nominal case. A slightly higher HP and LP EGR temperature after cooling can be observed as well compared to the nominal case, due to the increased EGR mass flow that goes through the EGR coolers. Moreover, the higher EGR temperature and increased mass flow result in a higher intake manifold temperature, as confirmed by the temperature trends in Fig. 6.4. A difference in the VGT actuation is observed as well compared to the nominal behaviour, as shown in Fig. 6.4: due to the increased request of HP EGR, the VGT is kept more open, limiting engine performance and affecting engine efficiency if the same torque output is provided.

MAF sensor low drift   Analogously to the low drift case, multiplying factors of 0.925 and 0.85 have been considered to emulate -7.5% and -15% MAF sensor drifts. Thus, contrary to the high drift case, the ECU receives a measured signal from the MAF sensor that is lower than the actual air mass flow. So, the controller acts as if the target air mass flow is intaken by the engine, while the actual air mass flow is higher. Therefore, to achieve the target intake manifold pressure despite the higher intake air mass, a lower amount of EGR is requested, resulting in a strong increase of NOx emission, as shown in Fig. 6.6. As expected, the trends of the main measured quantities observed in Fig. 6.3-Fig. 6.5 are opposite to the high drift case, meaning that a lower injected fuel mass is observed, together with higher lambda values, and lower intake manifold temperature.

The most significant measured quantities for the +15% and -15% MAF drift test cases are compared to the nominal measurements in Figure 3.

6.2.2 EGR systems faults

The same analysis presented in Section 6.2.1 for the MAF fault cases has been performed also for the faults involving the EGR system. However, for the sake of brevity, no figures related to the following test cases have been included in this chapter.
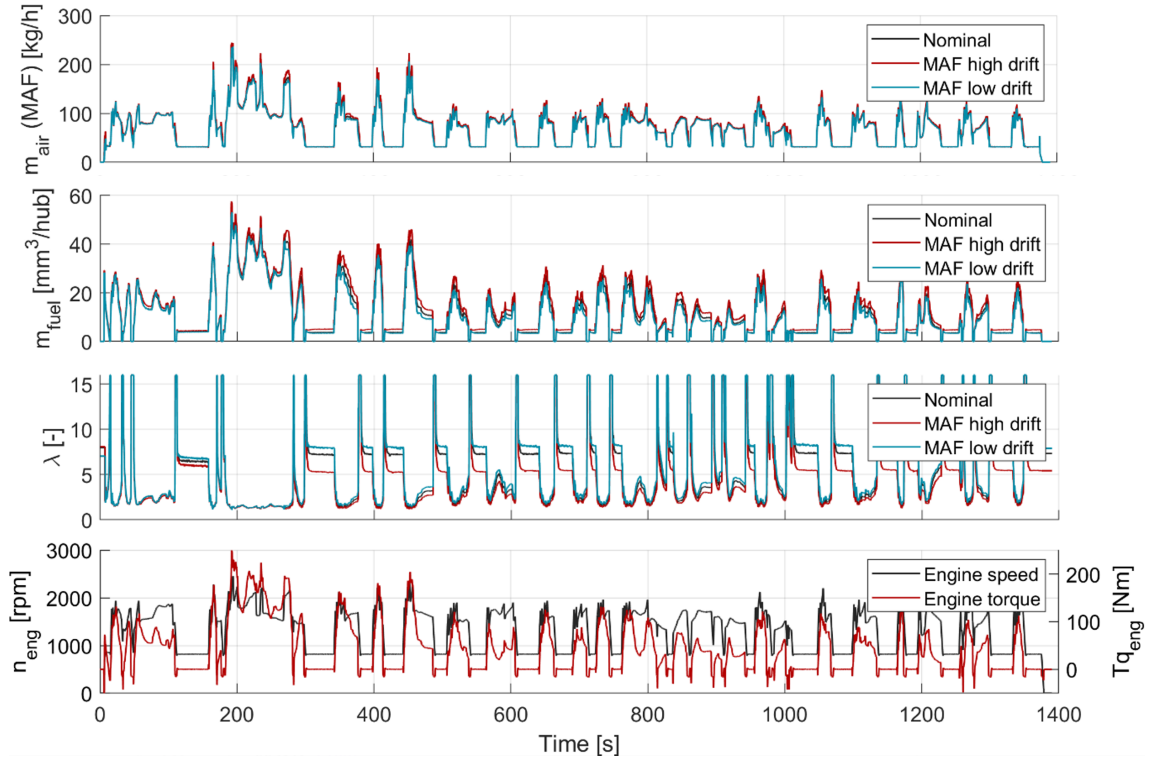
Figure 6.3: Comparison of test bench measurements for +15% MAF high drift, -15% MAF low drift and nominal test cases. From the top to the bottom: air mass flow, injected fuel mass, lambda, engine torque and speed profile

Low-pressure EGR valve clogging   This issue can be caused mainly by soot and carbon buildup from exhaust gases and oil contamination, enhanced by low-quality fuel, poor maintenance and high engine mileage. In the performed tests it has been implemented by limiting the opening profile of the electrically actuated LP EGR valve to a predefined percentage lower than the maximum one, more specifically 20% and 10% of the maximum valve opening. The lower amount of actual LP EGR compared to the request results in the ECU compensating with a higher amount of HP EGR. Accordingly, a higher intake manifold temperature is observed compared to the nominal case, despite the lower LP EGR temperature after cooling. The VGT behaviour is almost the same as in the nominal case, with a slight increase in the VGT opening in the worst case (10% LP EGR), due to the higher HP EGR request to compensate the reduced LP EGR amount, as observed for the MAF high drift case. No significant changes are observed in the injected fuel mass and lambda values, while significantly higher NOx emission is observed, as shown in Fig. 6.6.

High-pressure EGR valve clogging   Similarly to the LP EGR valve clogging, the fault has been implemented by limiting the opening profile of the electrically actuated HP EGR valve. Due to the limited usage of HP EGR in nominal conditions compared to LP EGR, the valve is forced to be
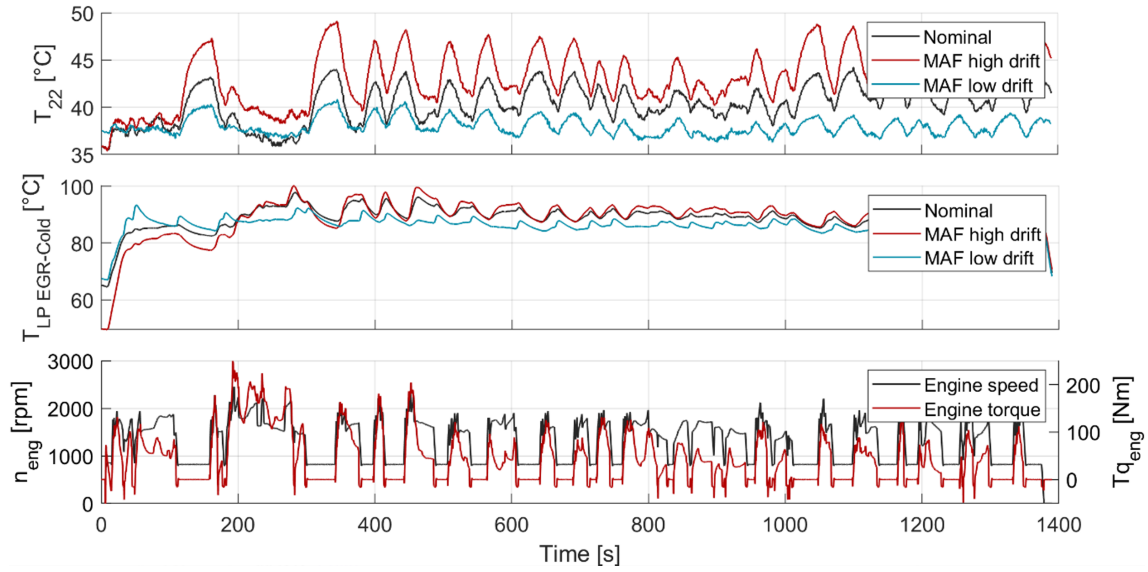
Figure 6.4: Comparison of test bench measurements for +15% MAF high drift, -15% MAF low drift and nominal test cases. From the top to the bottom: intake manifold temperature, temperature downstream LP EGR cooler, engine torque and speed profile

completely closed (0% valve opening) in order to be able to see significant effects of the considered fault. As expected, the absence of HP EGR is fully compensated by LP EGR from an emission point of view, leading to a slight decrease in NOx emissions. VGT is kept more closed than in the nominal case, since no HP EGR is recirculated. Accordingly, intake manifold temperature is significantly lower, while no significant changes are observed for the other measured quantities.

High-pressure EGR cooler bypass clogging Contrary to HP EGR valve, the HP EGR cooler bypass valve is forced fully open, meaning that the HP EGR flow is not passing through the heat exchanger to be cooled. As expected, HP EGR temperature is increasing compared to the nominal case where the bypass is closed, resulting in a higher intake manifold temperature as well. Due to the increased specific volume of the recirculated gases caused by the higher temperature, a lower EGR mass can be stored in the intake manifold (and in the cylinder), thus leading to a slight increase of NOx emissions. No significant changes are observed for the other analysed quantities.

A comparison of instantaneous and cumulated NOx emissions measured at the test bench for different test cases is shown in Fig. 6.6, where the data have been normalized with respect to the nominal case for confidentiality reasons, while the absolute and percentage variation of total cumulated emission for each fault case, compared to the nominal case, is summarized in Tab. 6.3.
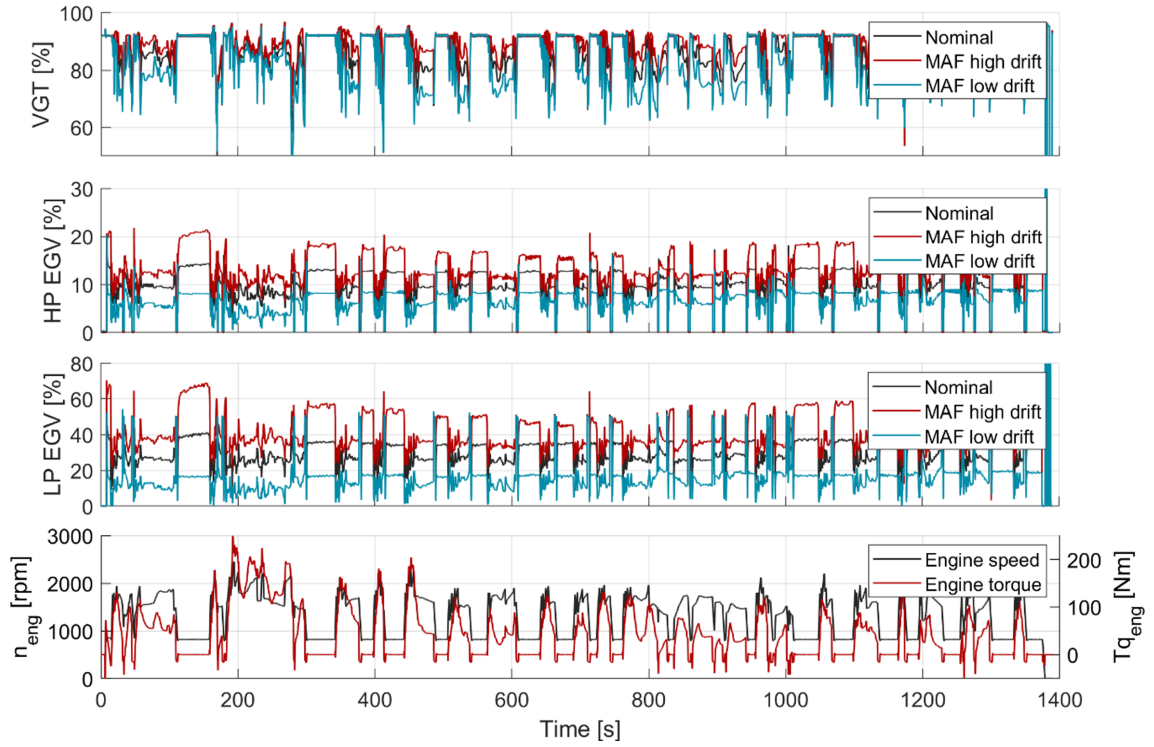
Figure 6.5: Comparison of test bench measurements for +15% MAF high drift, -15% MAF low drift and nominal test cases. From the top to the bottom: VGT position, HP EGR valve position, LP EGR valve position, engine torque and speed

Table 6.3: Absolute and percentage NOx emission variation for each tested fault case compared to nominal

Test case	Abs. variation	Rel. variation	ECU abs. error	ECU rel. error
	[g]	[%]	[g]	[%]
Nominal	-	-	-0.75	-14.0
MAF high drift +7.5%	-1.11	-20.9	+0.43	+10.2
MAF high drift +15%	-2.21	-42.4	+1.79	+60.8
MAF low drift -7.5%	+3.60	+67.7	-4.25	-47.7
MAF low drift -15%	+8.01	+154	-8.67	-64.8
LP EGR valve 20%	+1.05	+20.2	-1.57	-24.8
LP EGR valve 10%	+1.70	+32.6	-2.10	-29.8
HP EGR valve 0%	-0.23	-4.4	-0.83	-16.6
HP EGC bypass 100%	+0.39	+7.5	-0.97	-17.0
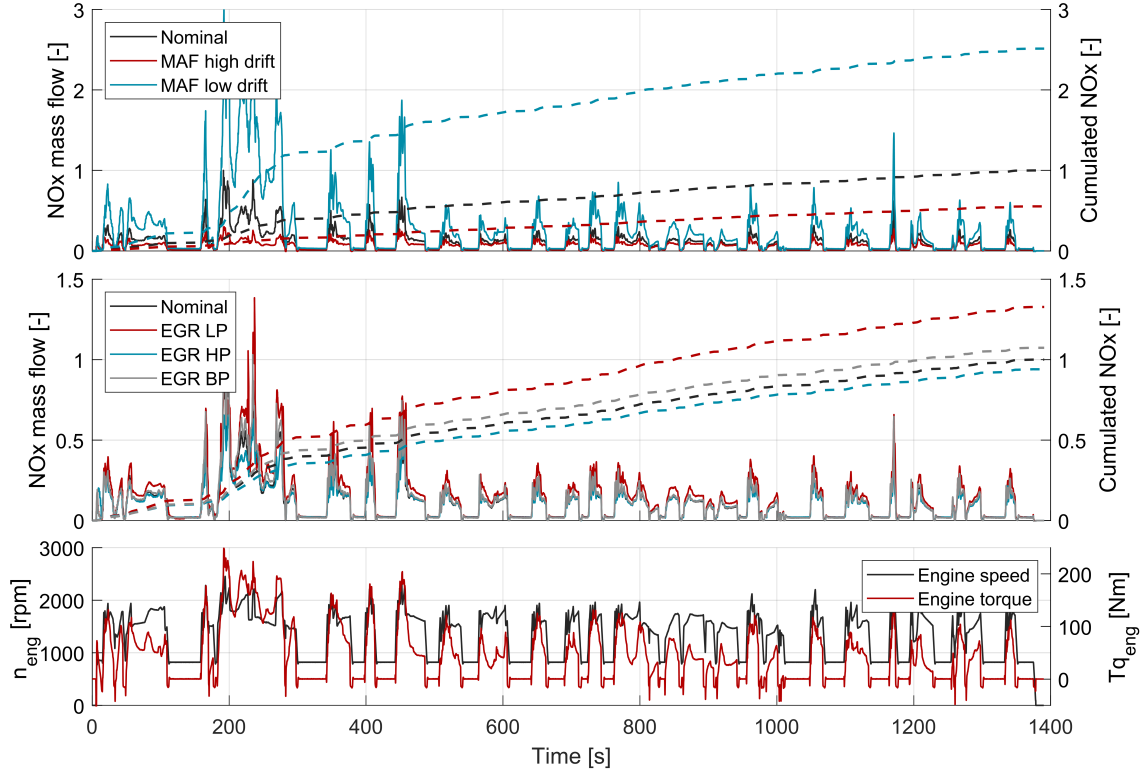
Figure 6.6: Instantaneous and cumulated engine-out NOx emission from test bench measurements for the different fault cases: +15% MAF high drift and -15% MAF low drift on the top, 10% LP EGR valve clogging, HP EGR valve clogging, HP EGC bypass clogging on the bottom. Signals are normalized with respect to the nominal case

6.3 Built-in ECU models

6.3.1 NOx emission model

As extensively introduced in Section 1.3.2, Euro 7 OBM standard requires the evaluation of time-resolved NOx emissions, as well as distance-specific vehicle emission at the end of each trip, which can be provided only by continuous real-time monitoring of the operating vehicle. These data are crucial to implementing the required Excess Emission Driver Warning System (EEDWS) for NOx emissions, and promptly detecting high-emitting vehicles during all-day operation. Due to the limitations of state-of-the-art on-board sensors in terms of accuracy, reliability and availability of the measurement during all operating conditions (e.g. cold starts, DPF regenerative events, stoichiometric or rich combustion etc.), fully sensor-based Euro 7-compliant emission monitoring is currently challenging. Therefore, to face the request of a continuous monitoring of NOx emissions, covering all the possible situations in which direct information from a physical sensor is not available or reliable enough, a model-based approach is needed in parallel. The most critical

136

aspect is that the considered emission models must provide continuous emission monitoring in any possible operating condition, independently from the state of health of system components. As comprehensively discussed in Section 2.2, semi-empirical and map-based emission models commonly implemented into standard on-board control units are not able to properly capture the effects of degraded or faulty components, which is a crucial aspect for OBM.

In this regard, Fig. 6.7 provides a comparison between the actual NOx emission measured at the test bench (black line) and the one estimated by the NOx model running on the commercial ECU of the engine under test (red line), for nominal, 15% MAF sensor high drift, and 15% MAF sensor high drift test cases. Since the ECU model output is NOx concentration, the NOx mass flow has been calculated by multiplying the concentration for the corresponding exhaust gas mass flow, net of EGR flow, calculated as the sum of MAF measured intake air mass flow and injected fuel mass flow (solid line). This way, the cumulated emission over the complete engine cycle can be calculated as well (dashed line). For confidentiality of emission data, all the signals shown in Fig. 6.7 have been normalized with respect to the maximum value observed on the same cycle in the nominal case, in order to retain the information related to the impact of faults on emissions levels. A non-negligible underestimation of NOx emission is noticed still in the nominal case, which strongly increases for the MAF low drift case (up to more than 60%), when a significant increase of actual NOx emission is observed. The opposite happens for the MAF high drift case, leading to lower actual NOx emission compared to the nominal case, for which a significant overestimation of NOx emission is observed. Similar considerations can be drawn for the other fault cases as well, even if, for the sake of brevity, the corresponding NOx estimated trends are not included in this section. More in detail, a significant ECU underestimation is observed for LP EGR valve clogging, while the deviation is closer to the nominal case for HP EGR valve and HP EGC bypass clogging, which have a lighter effect on actual NOx emission. The absolute and relative errors, in terms of total cumulated NOx emission on the FTP-72 cycle, between the ECU estimation and the actual measurement are summarized in Tab. 6.3 for each test case. In general, the NOx model implemented into the considered ECU appears to be insensitive or not sufficiently sensitive to the introduction of faults, so the estimated NOx value does not change significantly when actual emissions increase or decrease considerably compared to the nominal case due to the presence of a fault.

In this context, one of the targets of this research activity is precisely to implement a ML-based correction of the NOx emissions estimated by the reference ECU model, following the same methodology presented in Section 4.3, improving the prediction accuracy for all the considered fault cases.

6.3.2   Additional modelled quantities

As seen for the engine-out NOx emission, additional measured signals are modelled in parallel by the ECU for control and diagnosis purposes. As introduced in Section 4.1, ECU-modelled
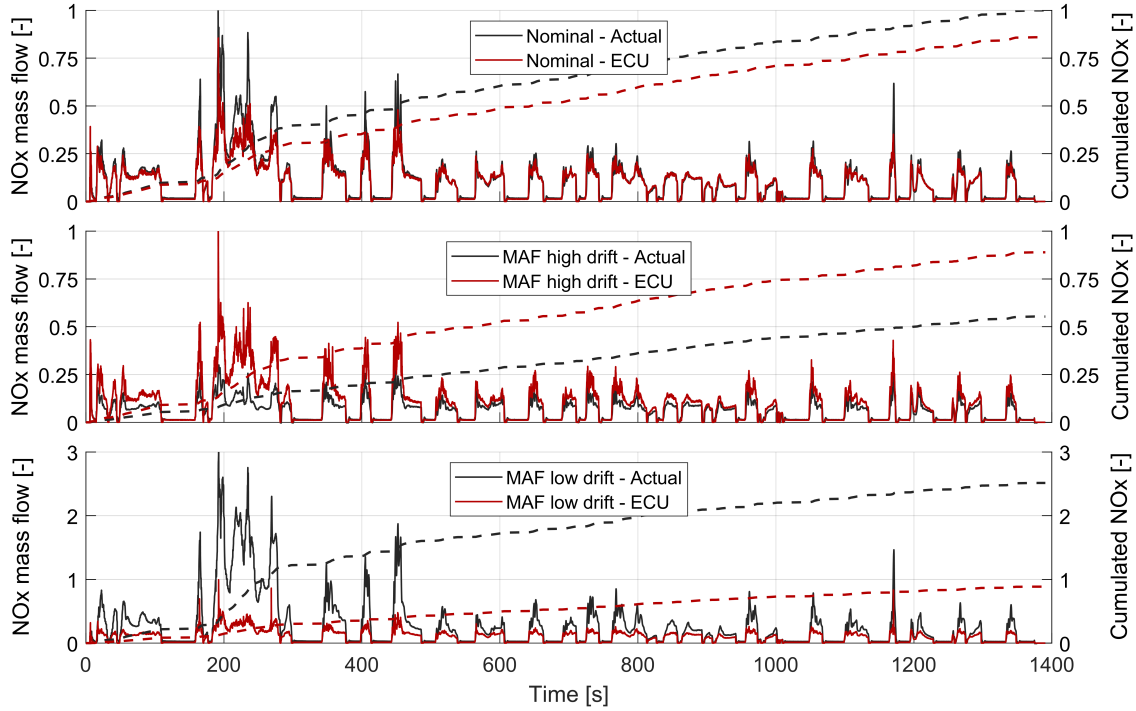
Figure 6.7: Engine-out NOx mass flow: comparison of measured signal and ECU model for nominal, +15% MAF high drift, and -15% MAF low drift. NOx emissions are normalized with respect to the nominal case

signals, together with on-board measurements, can be used to calculate residual signals that can be given as input to data-driven models in order to identify engine faults or predict non-nominal emission trends, as demonstrated by the promising results presented in Chapter 5. However, as already mentioned in Section 4.4, the feasibility and effectiveness of this approach strongly depends on the consistency of ECU models and consequently of the calculated residuals, not only in terms of overall accuracy, but also of reaction to the introduced faults. To give some examples, an overview of ECU estimations against actual on-board measurements for selected quantities, namely intake air mass flow (Fig. 6.8), lambda value (Fig. 6.9), temperature upstream DOC (Fig. 6.10), and temperature downstream HP EGR cooler (Fig. 6.11), is provided for both nominal and selected fault cases.

Focusing on the air mass flow, it can be noticed that there is no significant difference between nominal and MAF fault cases, since the ECU operates to provide the same target air mass flow based on MAF sensor feedback, as already discussed in Section 6.2.1. An almost fixed offset between ECU-estimated (dashed line) and measured signal (solid line) is observed for all the test cases, which is reflected in the trend of the calculated residual signal, namely essentially the same for the different test cases. Moving to Fig. 6.9, when there is a deviation in the actual lambda trend compared to nominal conditions due to a fault, a corresponding change is observed in

138

the estimated lambda trend (matching solid and dashed lines), meaning that the ECU model is sensitive to the considered faults and is consequently adapting its output. The resulting residual signal is oscillating around zero with some positive and negative spikes depending on the considered fault case (mainly positive spikes for the MAF low drift and negative spikes for the MAF high drift). Differently from the air mass flow, this behaviour could be useful to identify patterns characterizing different fault conditions. Focusing on the temperature signal upstream DOC (Fig. 6.10), no significant deviations of actual measured temperature are observed due to the introduction of a fault, and the modelled signal is following quite accurately the actual one both in nominal and fault conditions. Consequently, the calculated residual signal is not highlighting any particular behaviour that could be directly related to different fault conditions (MAF high and low drift in the specific considered case). Similar considerations are valid for other signals like intake manifold pressure, temperature downstream charge air cooler and exhaust manifold temperature, for which a good matching between measured and modelled signals is observed. Moving to Fig. 6.11, a significant variation of the actual temperature downstream the HP EGR cooler is observed, especially for the HP EGR valve fault (grey line). Even if the ECU model is adapting itself to follow the physical sensor trend in all the test cases, a peculiar residual trend can be identified for some of the different considered faults compared to the nominal case.
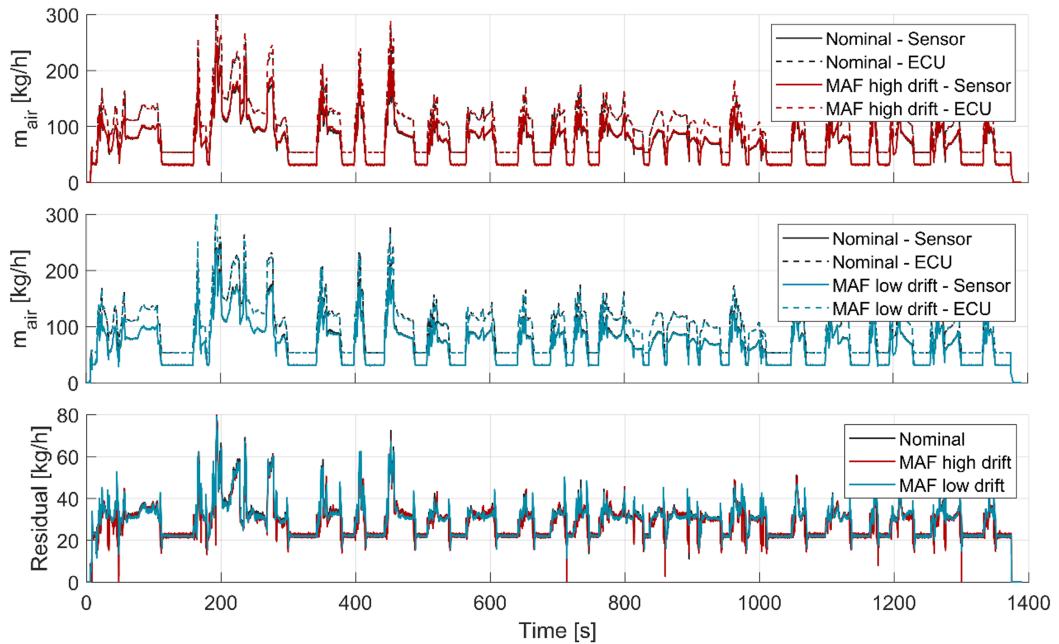

Figure 6.8: Air mass flow: comparison of measured signal (on-board sensor) and ECU model for nominal, +15% MAF high drift, and -15% MAF low drift; the evaluated resulting residual error is shown in the bottom plot
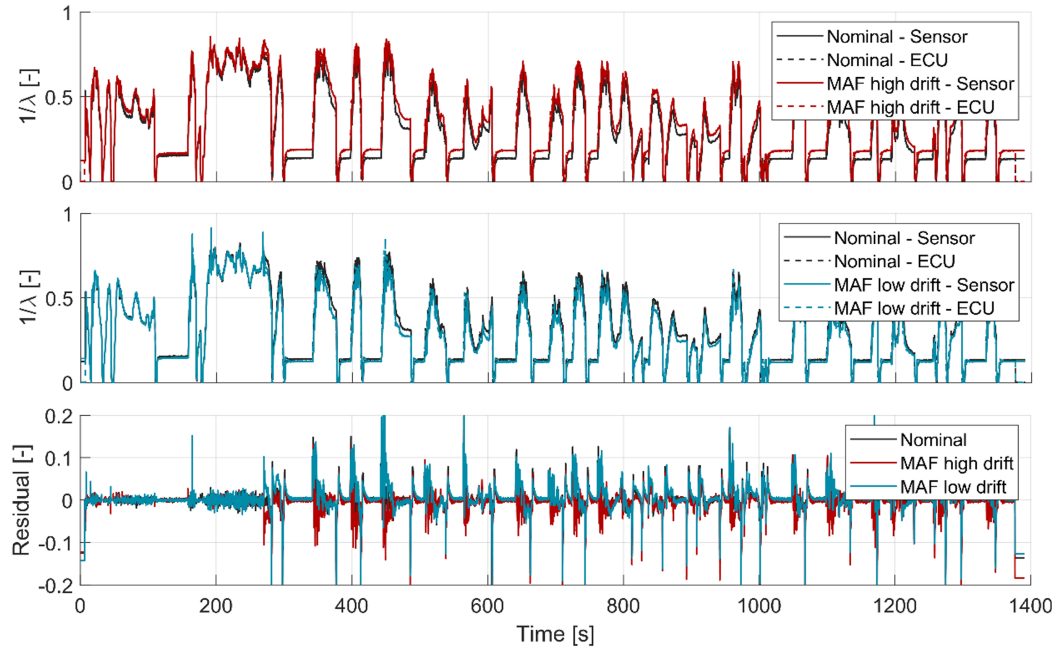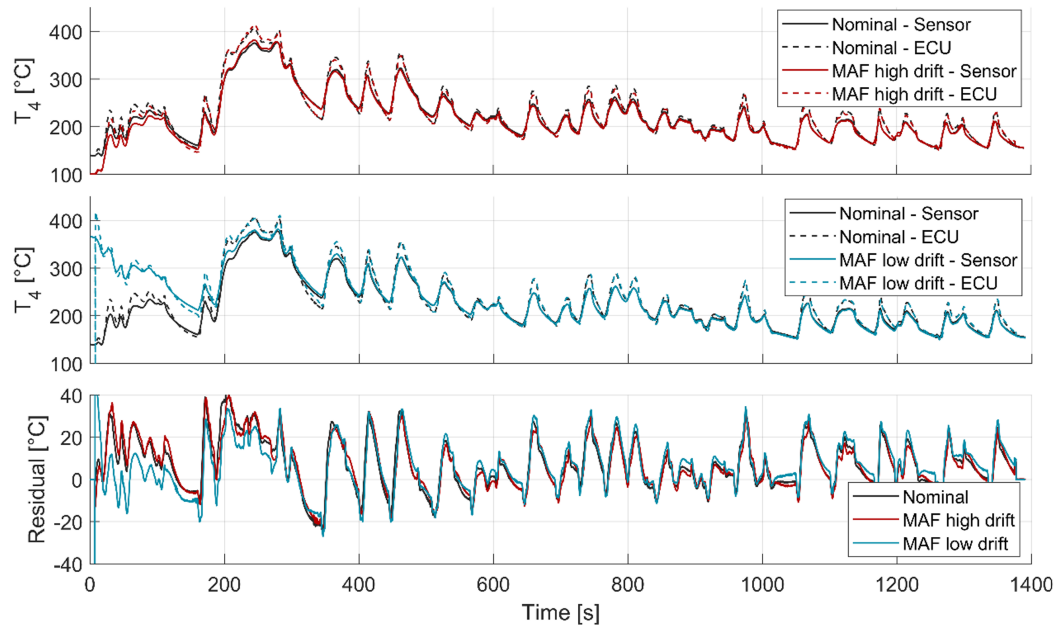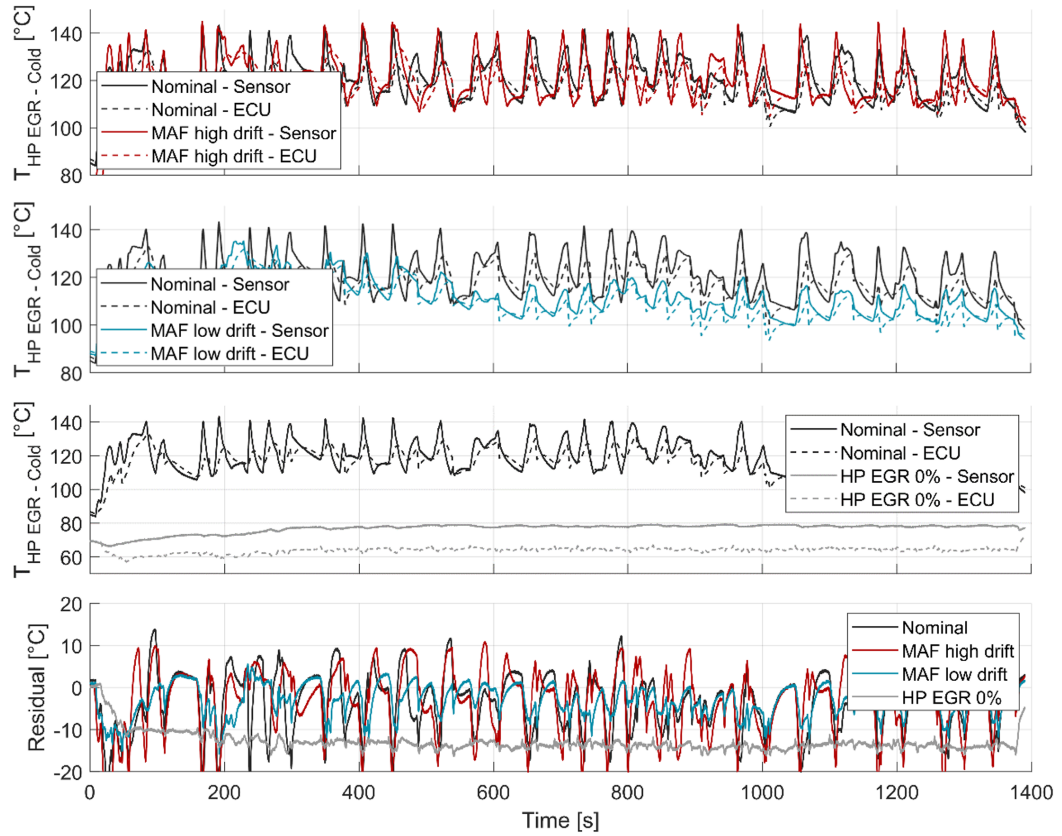
Figure 6.9: Lambda value: comparison of measured signal (on-board sensor) and ECU model for nominal, +15% MAF high drift, and -15% MAF low drift; the evaluated resulting residual error is shown in the bottom plot

Figure 6.10: Temperature upstream DOC (T_4): comparison of measured signal (on-board sensor) and ECU model for nominal, +15% MAF high drift, and -15% MAF low drift; the evaluated resulting residual error is shown in the bottom plot

Figure 6.11: Temperature downstream HP EGR cooler: comparison of measured signal (on-board sensor) and ECU model for nominal, +15% MAF high drift, -15% MAF low drift, and HP EGR valve clogging; the evaluated resulting residual error is shown in the bottom plot

141

Chapter 7

OBM-oriented data-driven models

This chapter is focused on the development of classification and regression models, trained on experimental data, to perform engine fault detection and NOx emission estimation. The same methodology described in detail in Chapter 4 is applied to a real-world dataset, starting with the selection of the training dataset and relevant input features, followed by training and optimization of selected classifiers and regression models.

7.1 Training and testing datasets

The experimental data available for training and testing the data-driven models, according to the real engine and sensor layout presented in Chapter 6, include the signals listed in Tab. 7.1. In addition to engine speed (n_{Eng}) and load (Tq_{Eng}), intake throttle position (r_{Thr}) and injected fuel quantity (m_{Fuel}) have been considered as well in first analysis to investigate if these signals could be clearly related to the presence of a fault. All the other signals are physical quantities measured by available on-board sensors, and for most of them an ECU-estimated signal is available in parallel, as introduced in Section 6.3. This means that residual signals can be calculated by comparing the ECU estimation and the actual measured signals, following the same process adopted in simulation and comprehensively described in Section 4.1. As well as engine speed and torque, residual signals are not calculated for throttle position and injected fuel, since they are control signals sent by the ECU with no corresponding available sensor signals. Moreover, as highlighted in Tab. 7.1, the residual signals can not be calculated even for temperature upstream charge air cooler (CAC) (T_{21}) and temperature downstream LP EGR cooler ($T_{LPEGR-Cold}$), since no corresponding ECU estimation is directly available for these two signals. So finally, 13 available signals have been considered as input features in the dataset including absolute signals from ECU and on-board sensors (absolute approach), while 11 signals are considered in the dataset including available residual signals and ECU control signals (residual approach).

As already introduced in Section 6.2 and summarized in Tab. 6.3, five different types of faults

Table 7.1: List of available absolute and residual input features according to engine on-board sensors and built-in ECU models. A black cross indicates that the input feature is available and included in the training datasets, while a grey cross indicates that the feature is available but discarded according to feature importance analysis

	n_{Eng}		Tq_{Eng}		m_{Air}		P_2		T_{21}		T_{22}		T_3		T_{HP} EGR-Cold		T_4		T_{LP} EGR-Cold		λ		r_{Thr}		m_{Fuel}		Tot.
	Abs	Res	Abs	Res	Abs	Res	Abs	Res	Abs	Res	Abs	Res	Abs	Res	Abs	Res	Abs	Res	Abs	Res	Abs	Res	Abs	Res	Abs	Res	
Absolute approach	✗	(✗)	(✗)		✗		✗		✗		✗		✗		✗		✗		✗		✗		(✗)		✗		11
Residual approach	(✗)		(✗)	(✗)		✗		✗				✗		✗		✗		✗			✗	✗			✗		9

Absolute: physical measurement from on-board sensors
Residual: residual signal calculated as difference between ECU modelled value and sensor measurement (only if ECU modelled value is available)

have been implemented at the test bench , including MAF sensor high and low drifts, LP EGR and HP EGR valves clogging, and EGR cooler bypass clogging. Two degrees of severity have been considered for MAF drifts and LP EGR valve clogging, for a total of 9 different test cases, including the nominal case. Since the same cycle has been considered for all test cases, namely an FTP-72, the test bench-acquired data from each cycle were split into two parts: two different cycles have been generated starting from the FTP-72, rearranging properly selected sections of the cycle, taking into account that the engine operating field should be covered as much as possible in the training cycle. This approach allows to overcome the issue of a single available driving cycle, defining two different cycles to train and test the developed models. The resulting training and driving cycles are shown in Fig. 7.1.
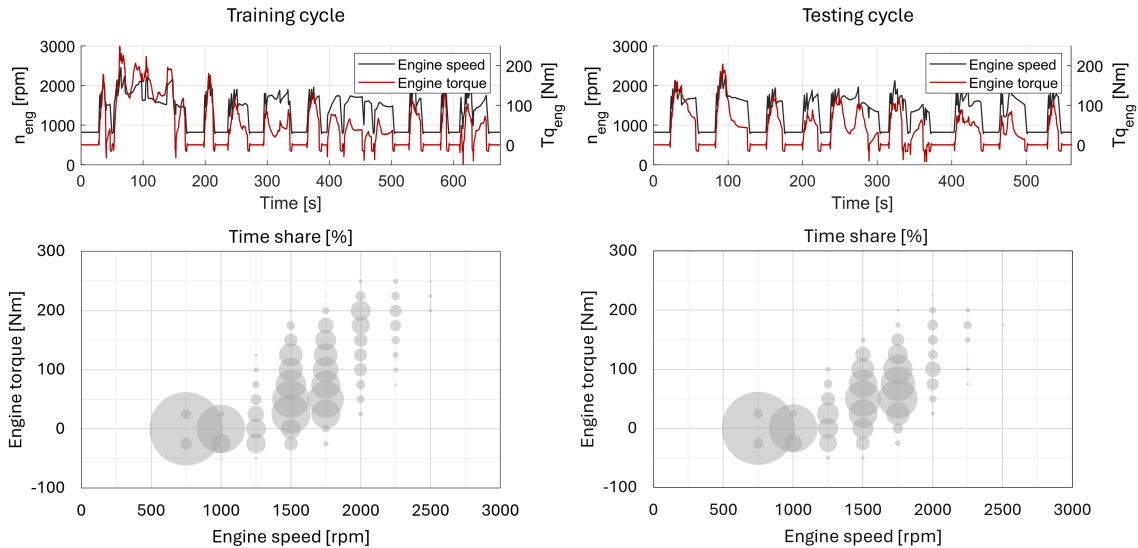

Figure 7.1: Engine speed and torque profiles of training and testing cycles generated from FTP-72

Moreover, as a further step to evaluate the extrapolation capability of the developed models, when two different degrees of fault severity are available for the same type of fault, only one of them has been included in the training dataset, while the other one has been kept for the testing phase. However, it must be considered that, following this approach, a relatively limited amount of training data are available, which could strongly affect model performance. Therefore, four different training datasets have been defined, as summarized in Tab. 7.2:

- Dataset 0, including data from the complete FTP-72 cycle and for all the tested fault cases;

- Dataset 1, including data from the defined custom training cycle and for all the tested fault cases;

- Dataset 2, including data from the defined custom training cycle, but only for selected cases corresponding to the most severe fault conditions;

- Dataset 3, including data from the defined custom training cycle, but only for selected cases corresponding to the least severe fault conditions

Table 7.2: List of fault cases included in each considered training dataset and corresponding total number of available observations for each dataset (sampling frequency is 10 Hz)

	Driving cycle		Fault condition									
	FTP-72	Custom cycle	NOM	MAF HIGH		MAF LOW		EGR HP	EGR BP	EGR LP		Tot.
			-	+7.5%	+15%	-7.5%	-15%	0%	100%	20%	10%	
Dataset 0	✗		✗	✗	✗	✗	✗	✗	✗	✗	✗	125154
Dataset 1		✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	60750
Dataset 2		✗	✗		✗		✗	✗	✗		✗	40500
Dataset 3		✗	✗	✗		✗		✗	✗	✗		40500

Based on its definition, Dataset 0 includes all the available observations, so that no unseen observations to properly test the models are available. For this reason, it has been used for some preliminary evaluations, but the results obtained using this training dataset are not included in this dissertation, being regarded as irrelevant to prove the effectiveness of the proposed data-driven approach. Therefore the analysis presented in this chapter and in Chapter 8 will focus on Dataset 1, 2, and 3.

7.2 Fault classification model

The methodology described in Section 4.2 has been followed for the development of the fault classification model. The experimental data introduced in Section 7.1 have been used to train different classifiers. Feature ranking algorithms, coupled with further considerations on the

statistical distribution of the input data, have been applied to identify the most relevant input features to be included in the training dataset. Then, different classifiers have been trained and optimized on the defined dataset to maximize their performance and select the most suitable approach for the considered application.

7.2.1 Feature importance analysis

As seen in Section 4.2.2 for the simulation dataset, also in this case each experimental observation is associated with a label identifying the specific fault class which the observation belongs to. The box plots in Fig. 7.2, provide a visual representation of the statistical distribution over the FTP-72 cycle of the considered sensor signals to be used as input data for the data-driven models. A detailed explanation on how to read these plots has been given in Section 4.2.2. As for the simulation case, outliers have been discarded for a clearer representation, even if all the observations are included in the training dataset, to further test the robustness of the developed models also in presence of outliers. It can be observed that the distribution of the air mass flow signal from the MAF sensor ($\dot{m}_{air}$) is almost constant for all the considered test cases, in accordance with the ECU control strategy that follows a fixed target of air mass flow depending on the engine operating point, as already explained in Chapter 6. On the other hand, temperature upstream and downstream CAC (T_{21} and T_{22}), as well as temperature after HP and LP EGR coolers ($T_{HPEGR-Cold}$ and $T_{LPEGR-Cold}$), show a distribution that is clearly varying according to the type of fault. As an example, T_{21} distribution is shifted upwards compared to the nominal case for MAF high drift and HP EGR fault classes, while it is shifted downwards for the other fault classes, namely MAF low drift and LP EGR faults. This is completely in agreement with what has already been analysed in detail in Section 6.2. Moving to $T_{HPEGR-Cold}$, it can be noticed that different mean values and distributions characterize the different faults and especially the HP EGR valve clogging, for which a very limited dispersion is observed around a much lower mean value compared to all the other fault cases, as a consequence of the significant decrease in temperature recorded when this defect is introduced (as already analysed in Fig. 6.11).

This summary statistical analysis highlights the different distribution patterns characterizing each fault class, thus allowing to identify a combination of relationships between the trends of the considered signals, that can be used to distinguish each case from the others.

The same feature ranking algorithms already introduced in Section 4.2.2, namely MRMR and ReliefF, have been applied to the considered real-world dataset, more specifically to Dataset 1, and available input features. Fig. 7.3 shows the corresponding results, with features sorted from the highest to the lowest importance score. A perfect agreement is observed in the definition of the top-four important features, namely T_{22}, $T_{HPEGR-Cold}$, $T_{LPEGR-Cold}$, and T_{21}. This result was expected based on the previous statistical analysis. On the other hand, temperature upstream and
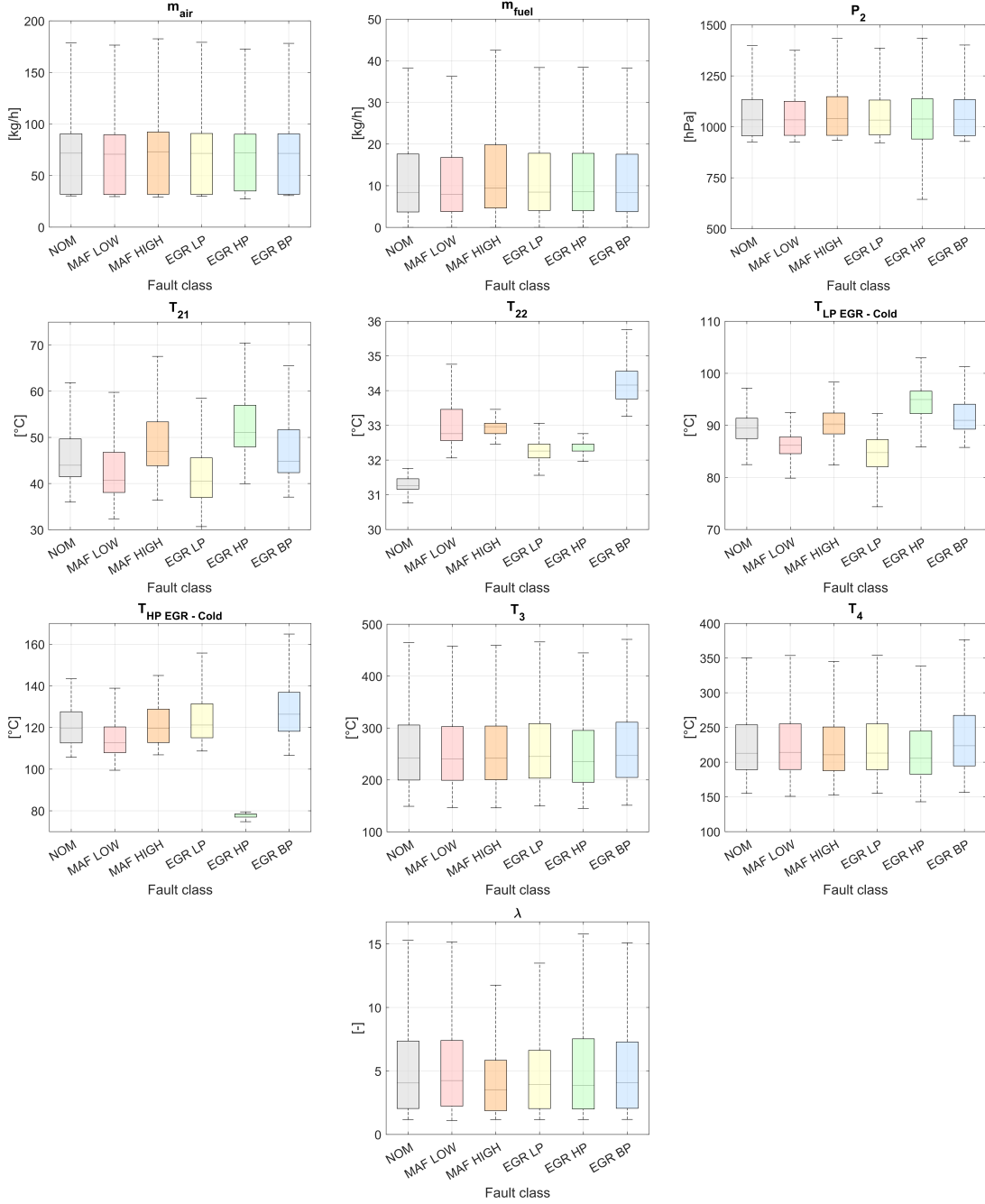
Figure 7.2: Statistical box plots of input features for each fault type

downstream turbine (T_3 and T_4) obtained an almost zero score according to MRMR, while they obtained higher scores according to ReliefF algorithm, so they have been kept among the considered input features. Finally, engine torque (Tq_{Eng}) and throttle position (r_{Thr}), scored zero according

to MRMR algorithm, also occupy the last positions according to ReliefF ranking. Therefore, they have been discarded from the input features for the fault classification model.
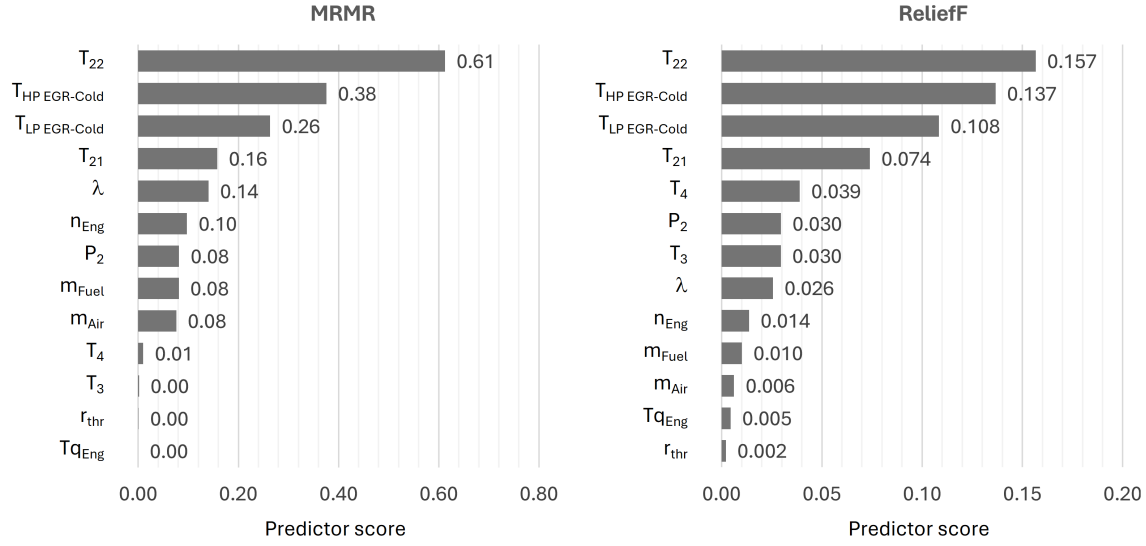
Figure 7.3: Importance scores of selected input features according to MRMR (left) and ReliefF (right) algorithms

7.2.2   Training, optimization and validation of selected classifiers

Based on the performance analysis described in Section 4.2.3 and considering the promising results achieved in simulation in Part I of this dissertation, Tree, Neural Network and Ensemble classifiers have been considered for all the following analysis, to find the best trade-off in terms of accuracy, flexibility, and model complexity. As seen in Section 4.2.4 for the classification models trained on the simulation dataset, Bayesian optimization has been used to optimize the main model hyperparameters and maximize the accuracy of the selected classifiers. The results shown in this section refer to Dataset 1, but the same approach has been applied also to optimize and train models on Dataset 2 and 3. For the sake of brevity, the corresponding validation results are not included in this section, being essentially similar, in term of performance, to those obtained for Dataset 1; on the other hand, a comparison of the performance evaluated on the unseen testing data between the models trained on different datasets will be provided in Chapter 8, pointing out the extrapolation capabilities of the developed models. The optimizable parameters for each model and their range of optimization are summarized in Tab. 7.3, together with the optimal hyperparameter configurations resulting from the Bayesian optimization. The same configuration of optimization parameters selected in Section 4.2.4 has been applied: the overall classification error has been chosen as the objective function to be minimized, and a maximum number of 50 iterations has been set as the stopping criterion. The overall accuracy reached on the validation

dataset by the optimized models is shown in Tab. 7.4, together with the training time (including model optimization), prediction speed and final model size.

Table 7.3: Set of optimizable parameters, optimization range, and final configuration resulting from Bayesian optimization of the classification models trained on Dataset 1 (absolute signals)

Model	Optimizable hyperparameters	Optimization range	Optimal configuration
Optimizable Tree	Maximum number of splits	$[1, \max(2, n-1)]$	378
	Split criterion	Gini's diversity index, Twoing rule, Maximum deviance reduction	Maximum deviance reduction
Optimizable Ensemble	Ensemble method	AdaBoost, RUSBoost, Bag	Bag
	Maximum number of splits	$[1, \max(2, n-1)]$	59169
	Number of learners	$[10, 500]$	11
	Learning rate	$[0.001, 1]$	0.1
	Number of predictors to sample	$[1, \max(2, p)]$	4
Optimizable Neural Network	Number of fully connected layers	$[1, 3]$	1
	Layers size	$[1, 300]$	72
	Regularization strength	$[0.00001/n, 100000/n]$	$2.9355e-09$
	Activation function	ReLU, Tanh, None, Sigmoid	Sigmoid

n = number of observations (=60750)
p = number of predictor variables (=11)

Table 7.4: Overall accuracy and training parameters of the optimized classification models trained on Dataset 1 (absolute signals)

Model	Accuracy [%]	Training time [s]	Prediction speed [obs/s]	Model size [MB]
Neural Network	100.0	11274	160000	0.02
Ensemble	100.0	804	9500	18
Tree	99.8	48	860000	0.4

Fig. 7.4 shows the confusion matrices evaluated on the validation dataset for the three optimized models. The explanation on how to interpret the information provided by the confusion matrices can be found in Section 4.2.4. The results, expressed in percentage for each fault class, confirm what already expressed by the overall model accuracy shown in Tab. 7.4, namely that the three optimized models can almost perfectly classify the different fault conditions included in the validation dataset. More in detail, the overall accuracy reaches 100% for the Ensemble and Neural Network models, while it is slightly lower, namely 99.8%, for the Tree classifier, showing distributed, although not significant, misclassification errors. Given the satisfying overall performance, all the three classification models have been considered for further investigations to assess their accuracy and robustness on the testing dataset, as described in Chapter 8.
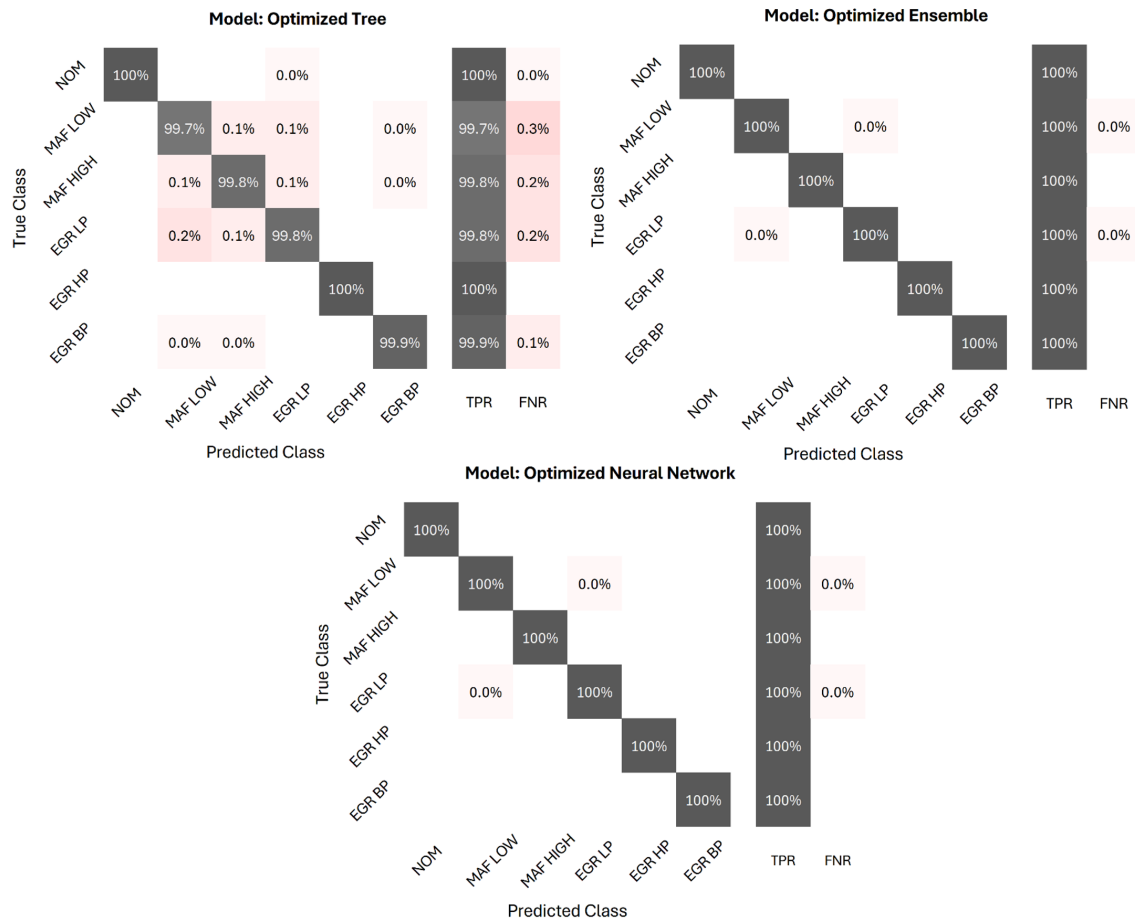
Figure 7.4: Validation results of optimized classification models trained on Dataset 1 (absolute sensor signals): related confusion matrices, true positive rates (TPR) and false negative rates (FNR) for each fault class

7.2.3 Absolute approach against residual approach

Based on the promising results obtained in simulation, presented in Section 5.2.1, the already introduced residual approach has been applied in parallel to the absolute approach to compare the classification performance achieved by the models trained on these different types of input signals. As mentioned in Section 7.1, residuals signals can not be calculated for all the available sensor signals, due to the lack of certain modelled quantities. So a total of 11 input signals have been considered for the feature importance analysis, and the corresponding computed scores are shown in Fig. 7.5. Differently from what has been observed in Fig. 7.3, in this case there is no agreement in the definition of the most important features according to the two ranking algorithms, except for $T_{HPEGR-Cold}$ residual signal, which occupies first and second place in the two rankings. On the other hand, engine torque and speed obtained a zero or nearly zero score according to both MRMR and ReliefF algorithms, so that these signals have been excluded form the input features, as already seen for the absolute signals approach. Therefore, a total of 9 input features have been considered, as shown in Tab. 7.1.

Tree, Neural Network, and Ensemble classifiers have been trained on the custom training cycle defined in Section 7.1, on all the available fault conditions included in Dataset 1. The same optimization process described in Section 7.2.2 has been applied also in this case. The overall accuracy reached on the validation dataset by the optimized models is shown in Tab. 7.5, together with the training time (including model optimization), prediction speed and final model size. As expected, the Ensemble model exhibits the highest overall accuracy, equal to 97.8%, followed by the Neural Network and the Tree models, with 92.9% and 88.5% overall accuracy respectively. More in detail, focusing on Fig. 7.4, showing the confusion matrices evaluated on the validation dataset for the three optimized models, significant misclassification errors resulting in high FNRs can be observed for Tree and Neural Network models in correspondence with Nominal (NOM) and EGR cooler bypass (EGR BP) classes. The Ensemble model shows a better performance, with an FNR below 6%, for both NOM and EGR BP classes. However, the overall performance of the three optimized classifiers trained on residual signals is significantly worse when compared to the ones trained on absolute sensor signals (Fig. 7.4). This will be further confirmed by the results obtained on the testing cycle, discussed in Chapter 8. The explanation for this result lies in the fact that, contrary to what was observed in simulation in Part I, the calculated residual signals do not contain extremely clear patterns that help distinguishing between the different fault classes. This is related to the behaviour of the considered ECU models, as already discussed in Section 6.3, which are not perfectly calibrated to match sensor measurements in nominal conditions, thus resulting in non-zero residuals, and are significantly sensitive to the introduction of faults, which affects the final trend of selected residuals. Therefore, the resulting pattern of residual signals is not uniquely attributable to the presence of a specific fault, but it also depends on the behaviour of the control unit when that fault is introduced, which could make the identification more challenging. Furthermore, it must be emphasized the residual signals $T_{LPEGR-Cold}$ and T_{21}, which were among

151

the most important input features for the models trained on absolute sensor signals, could not be calculated due to the missing reference ECU models, and this may have further affected the classification performance.

To summarize, unlike in the simulation case described in Part I, the approach considering absolute sensor signals rather than computed residual signals proves to be the preferred one for the specific experimental application investigated in Part II of this dissertation. However, since it strongly depends on the considered reference ECU models, this conclusion is not applicable indiscriminately in all cases.
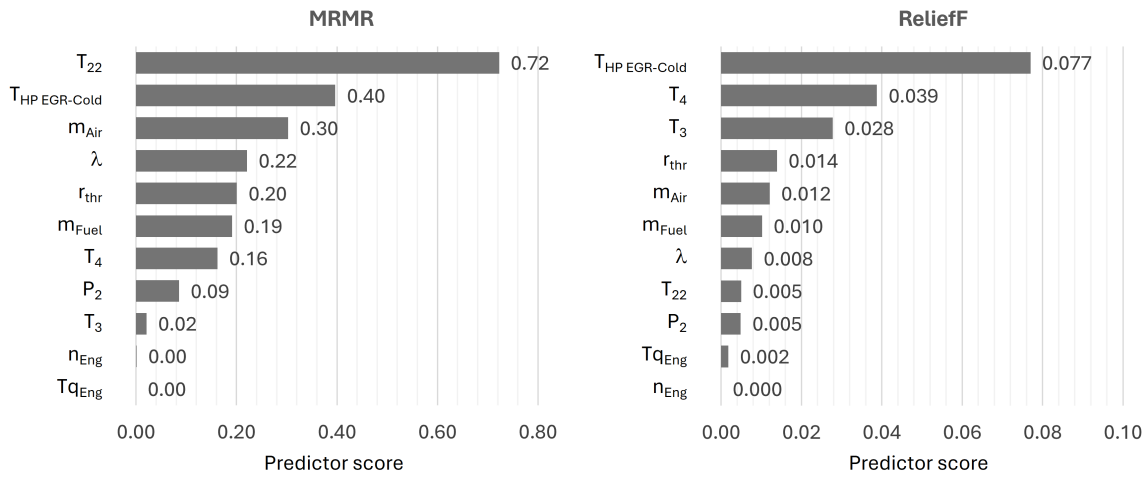
Figure 7.5: Importance scores of selected residual input features according to MRMR (left) and ReliefF (right) algorithms

Table 7.5: Overall accuracy and training parameters of the optimized classification models trained on Dataset 1 (residual signals)

Model	Accuracy [%]	Training time [s]	Prediction speed [obs/s]	Model size [MB]
Ensemble	97.8	2033	5700	79
Neural Network	92.9	21528	440000	0.7
Tree	88.5	115	960000	1

7.3 NOx regression model

The same methodology described in Section 4.3 has been followed for the development of the regression model for NOx emission correction. The same test bench-acquired data used to train the classification model have been considered, focusing on the approach involving absolute sensor

Model: Optimized Tree

Model: Optimized Ensemble

Model: Optimized Neural Network
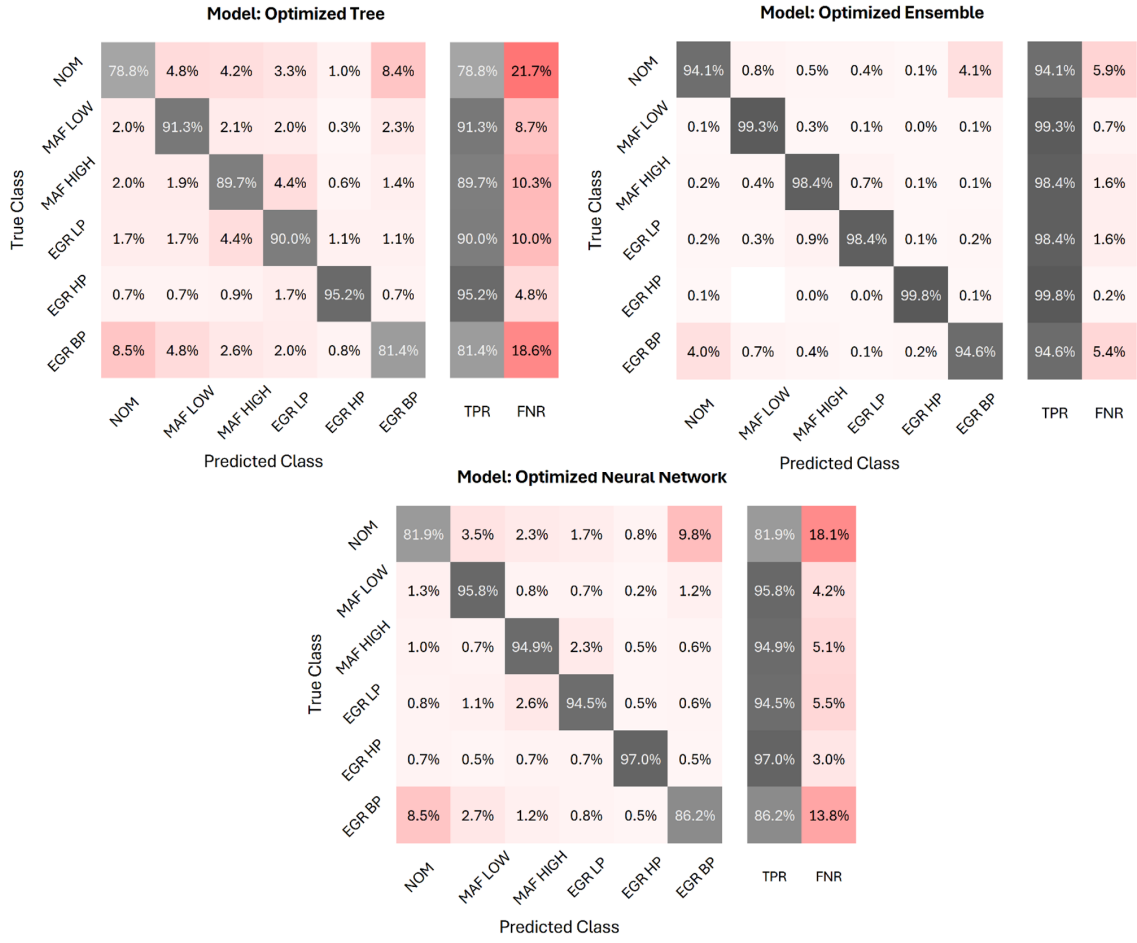
Figure 7.6: Validation results of optimized classification models trained on Dataset 1 (residual signals): related confusion matrices, true positive rates (TPR) and false negative rates (FNR) for each fault class

signals rather than residual signals, based on the results obtained from the classification model (Section 7.2). After a preliminary phase to identify the most relevant input features to be included in the training dataset, different regression models have been trained and optimized, and their performance on the validation dataset are assessed.

7.3.1 Feature importance analysis

As anticipated, the same data used for developing the classification models presented in Section 7.2 have been considered. In this case the target response to be learned by the regression models is the correction factor to be applied to the reference NOx emission estimated by the engine controller, in order to match the actual emission calculated by the physical model. Differently

from the simulation case analysed in Section 4.3, the output provided by the ECU is the NOx concentration, expressed in ppm, and not directly the emission mass flow, so that the corresponding correction factor can be calculated as:

$$f'_{NOx} = \frac{y_{NOx,act}}{y_{NOx,ECU}} \tag{7.1}$$

where $y_{NOx,act}$ and $y_{NOx,ECU}$ are the actual and estimated NOx concentrations, measured by the NOx sensor and calculated by the ECU respectively. Alternatively, following the same approach applied for the simulation case, a correction factor related to the NOx mass flow rather than to the concentration can be calculated as:

$$f_{NOx} = \frac{\dot{m}_{NOx,act}}{\dot{m}_{NOx,est}} = \frac{y_{NOx,act} \cdot \dot{m}_{exh,act}}{y_{NOx,ECU} \cdot \dot{m}_{exh,ECU}} \tag{7.2}$$

where $\dot{m}_{exh,act}$ and $\dot{m}_{exh,ECU}$ are the actual and estimated NOx mass flow, measured at the test bench and calculated by the ECU respectively. This approach, compared to the previous one, has the advantage of taking into account possible effects of the considered faults on exhaust gas mass flow, and not only on NOx concentration. For this reason, the mass flow approach is preferred to the concentration approach and the validation results presented in this section refer to this approach. However, the performance of both approaches on the testing dataset will be discussed and compared more in detail in Chapter 8.

Analogously to the classification case, MRMR and RReliefF ranking algorithms have been applied to investigate the importance of the available input features for predicting the NOx emission correction factor, eventually discarding irrelevant signals. The corresponding results are shown in Fig. 7.7. Differently from what has been observed for the classification model in Fig. 7.3, but similarly to the simulation case presented in Part I of the dissertation, there is no complete agreement between the two ranking methods in defining the most relevant features for the regression models. Moreover, there are no features standing out clearly from the others for their score, except for lambda value (λ) according to the RReliefF algorithm, which is also ranked as the second most important feature according to the MRMR algorithm. At the same time, there are no features to be definitely excluded according to the obtained importance scores. Temperature before and after the turbine (T_3 and T_4) are assigned the last positions in the features ranking according to both algorithms, however the corresponding scores are fully comparable to those assigned to most other signals. So, taking into account these results, differently from the classification case, no features have been further discarded in this phase.

7.3.2 Training, optimization and validation of selected regression models

Based on the performance analysis and the promising results obtained in simulation, as described in Part I of this dissertation, Tree, Neural Network and Ensemble regression models have been selected as the best performing models to be applied for NOx emission correction. As for the
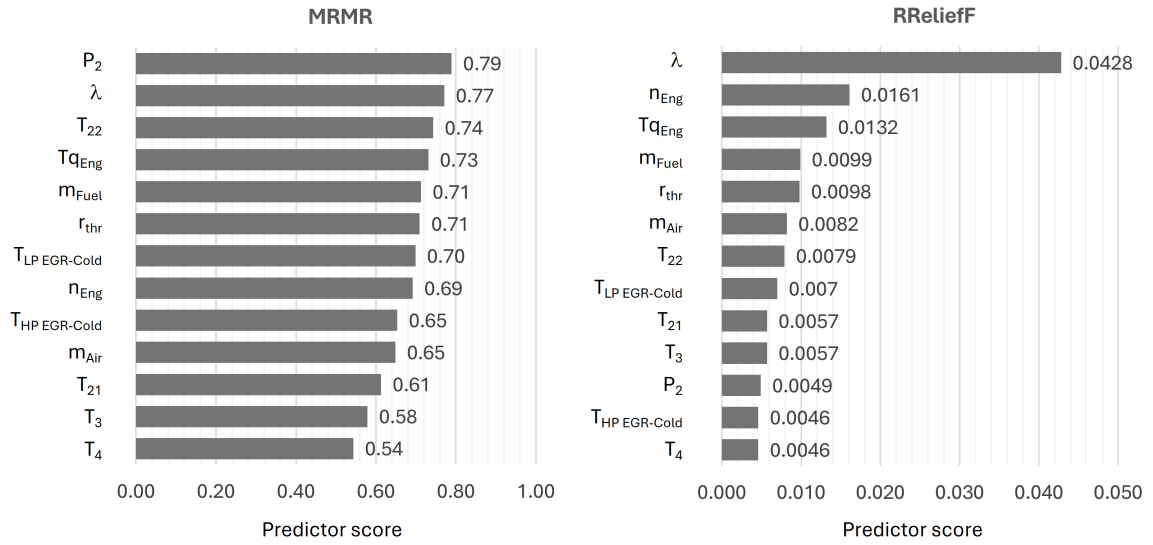
Figure 7.7: Importance scores of selected input features according to MRMR (left) and ReliefF (right) algorithms

regression models trained on simulation data (Section 4.3.2), the hyperparameters of the selected models have been further optimized to maximize prediction accuracy. Bayesian optimization, comprehensively introduced in Section 4.2.2, has been applied also in this case, with MSE chosen as the objective function to be minimized and a maximum limit of 50 iterations imposed for the optimization. The range of optimization of each parameter and the resulting optimal configuration of hyperparameters for each model are summarized in Tab. 7.6.

The main performance metrics evaluated on the validation dataset for each optimized model are summarized in Tab. 7.7, together with the time required for model training and optimization, prediction speed and final model size. According to the validation results, the optimized Ensemble exhibits the lowest RMSE, equal to 0.322, followed by the Neural Network model, with a slightly higher RMSE of 0.363, and finally the regression Tree, with an RMSE of 0.398. Accordingly, the Ensemble regressor shows the highest R^2, equal to 0.89, followed by the Neural Network and the Tree models, with 0.86 and 0.83 respectively. In contrast to the fault classification case, the performance parameters evaluated on the validation phase for the regression models trained on real experimental data turn out to be significantly worse than in the simulation case, even if the level of accuracy is still considered good for the specific application.

The performance of the trained models on the validation dataset is visually represented in Fig. 4.11, showing the model response for each observation on the top, and the differential errors between predicted and true values on the bottom. It can be noticed that for the Ensemble model the points lie closer to the black diagonal line, corresponding to the perfect prediction, with a slightly lower dispersion compared to the other models. This is reflected also in the residual plot, showing values that are generally closer to zero for the Ensemble model compared to the other

regressors. Considering the significant advantages of Neural Network and Tree models in terms of reduced model complexity, and therefore limited prediction speed and model size, compared to the Ensemble model, all the three models have been considered for further investigations, in order to assess their accuracy and robustness on the testing dataset.

Table 7.6: Set of optimizable parameters, optimization range, and final configuration resulting from Bayesian optimization of the selected regression models

Model	Optimizable hyperparameters	Optimization range	Optimal configuration
Optimizable Tree	Minimum leaf size	$[1, \max(2, floor(n/2)]$	5
Optimizable Ensemble	Ensemble method	Bag, LSBoost	Bag
	Minimum leaf size	$[1, \max(2, floor(n/2)]$	1
	Number of learners	[10,500]	94
	Learning rate	[0.001,1]	0.1
	Number of predictors to sample	$[1, \max(2,p)]$	9
Optimizable Neural Network	Number of fully connected layers	[1,3]	2
	Layers size	[1,300]	78, 89
	Regularization strength	$[0.00001/n, 100000/n]$	$3.9966e-10$
	Activation function	ReLU, Tanh, Sigmoid, None	Relu

n = number of observations (=60750)
p = number of predictor variables (=13)

Table 7.7: Main performance and training parameters of different regression algorithms according to NOx concentration/mass flow approaches

Model	RMSE [-]	MAE [-]	R2 [-]	Training time [s]	Prediction speed [obs/s]	Model size [MB]
Ensemble	**0.322** / 0.359	**0.081** / 0.088	**0.89** / 0.86	2174 / 2107	5000 / 10000	119 / 65
Neural Network	**0.363** / 0.380	0.124/**0.099**	**0.86** / 0.85	22315 / 29547	270000 / 340000	~0.2 / 0.6
Tree	**0.398** / 0.425	**0.097**/0.110	**0.83** / 0.81	181 / 195	770000 / 420000	~1

156
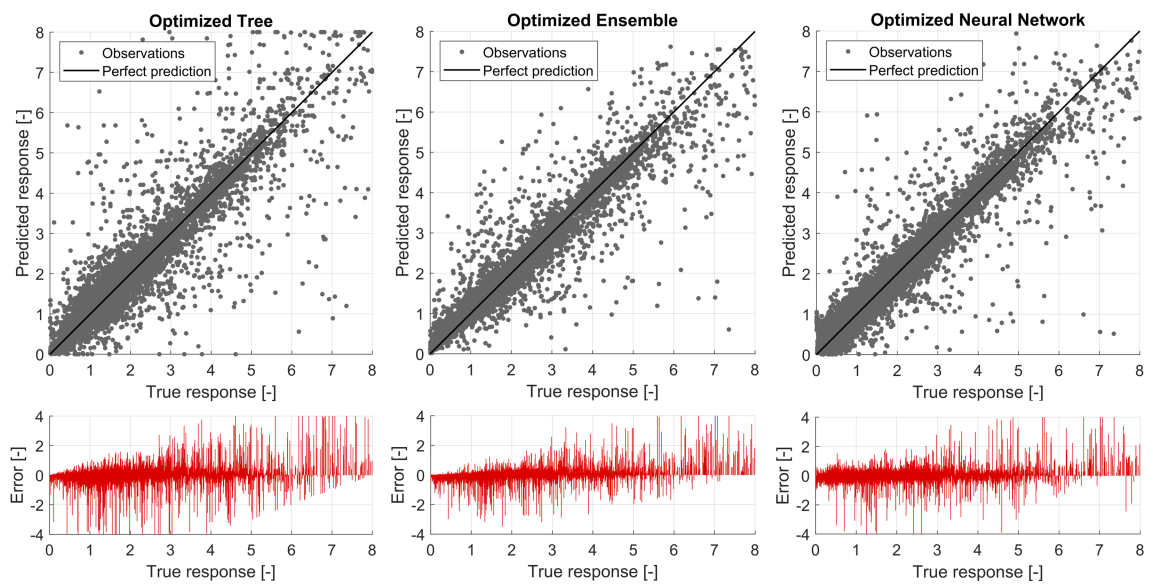
Figure 7.8: Validation results of optimized regression models: model response and related residual error for the considered observations

Chapter 8

Application of the developed models

In this chapter, the results of the application of the models developed in Chapter 7 to an unseen experimental dataset is presented. The effectiveness in detecting and identifying the considered engine faults and the significant improvement achieved in terms of NOx prediction accuracy thanks to the implementation of the developed data-driven models is demonstrated and discussed.

8.1 Test case scenarios

The validated classification and regression models presented in Chapter 7 have been tested on the custom testing cycle shown in Fig. 7.1 in all available fault conditions. The performance of the different classification and regression models trained on the complete set of fault conditions (Dataset 1) will be evaluated to select the best data-driven approach for each application among Tree, Ensemble, and Neural Network models. A comparison with the results obtained with the models trained only on a limited set of fault conditions (Dataset 2 and Dataset 3) will be presented as well, to further assess the extrapolation and interpolation capabilities of the models and the deterioration in terms of prediction accuracy due to the reduction of the fault cases included in the training dataset.

8.2 Results and discussion

8.2.1 Fault detection and identification

The results in terms of overall classification performance parameters evaluated on the testing cycle for the complete set of fault conditions are summarized in Tab. 8.1 and Tab. 8.2. More in detail, Tab. 8.1 collects the results related to the models trained on absolute sensor signals (Section 7.2.2), while Tab. 8.2 the ones related to the models trained on residual signals (Section 7.2.3). The analysed performance parameters include precision, recall, and F1-score, which have been already

introduced in detail in Section 4.2.3. All the results presented in this section refer to the classifiers trained on Dataset 1 (both absolute and residual approaches), thus including all the fault cases, as described in Section 7.1.

Overall, the testing results confirm what observed in the validation phase, namely that the absolute approach is the most suitable one for this application, far outperforming the residual approach for all three models, with an overall accuracy of 93.6%, 96.8%, and 98.9% , against the 56.4%, 68.9%, 61.9% of the residual approach, for Tree, Ensemble, and Neural Network models respectively. Focusing on Tab. 8.1, the Neural Network model shows a perfect performance for NOM, EGR HP, and EGR BP classes, while a slightly worse performance is shown for MAF HIGH, MAF LOW, and EGR LP classes. A similar behaviour is observed also for the Ensemble and Tree models, perfectly recognizing NOM and EGR HP classes, while showing the lowest F1-score for the MAF LOW class. The values of the single precision and recall indexes exhibited by the Neural Network model are not differing significantly from each other for each considered fault class, basically matching the F1-score index. Focusing on the Ensemble model, a lower precision is calculated for the EGR LP class compared to the corresponding recall index: this means that the model provides a relatively low rate of false negatives, meaning that the EGR LP class is mostly identified when occurring, but with more false positives, namely that a non-negligible number of observations is incorrectly classified as EGR LP, while the true fault class is different. Similar considerations are valid also for the Tree model and specifically for the EGR BP fault class. On the other hand, moving to Tab. 8.2, the most critical fault classes to be identified by the models trained on residual signals, even for Ensemble and Neural Network classifiers, are NOM and EGR BP, which on the contrary were perfectly identified with the absolute approach. The recall index exhibited for these classes is always far below 50%, therefore not acceptable for the purposes of this activity. Even focusing on the best performing model, namely the Ensemble one, all the fault classes show an F1-score below 0.75, except for the EGR HP class, with a relatively high recall index, compensated however by a lower precision.

Based on these results, further analysis have been focused on the absolute approach. In this regard, Fig. 8.1 shows the performance trend of model sensitivity to each fault class when different training dataset are considered. As expected, the models trained on the full set of fault conditions, namely Dataset 1, show the best overall performance. On the other hand, a significant performance drop of the models trained on Dataset 2 and Dataset 3 is observed for the fault classes including two degrees of severity, namely MAF HIGH, MAF LOW, and EGR LP: only the most severe or least severe fault case has been included in the training dataset, so that a more significant inter-polation/extrapolation is required for the observations belonging to the unseen fault conditions. However, focusing in particular on the Neural Network classifier, which is the one exhibiting the overall best performance, better results are achieved by the model trained on Dataset 3, especially for the MAF low drift fault class. This means that the model is more likely to predict the fault class when trained on a less severe case and tested on the more severe one than in the opposite situation.

160

Table 8.1: Performance parameters evaluated on the testing cycle for different classifiers trained on absolute sensor signals (Dataset 1)

	NOM	MAF LOW	MAF HIGH	EGR LP	EGR HP	EGR BP
Optimized Tree						
Precision	1.000	0.869	0.993	0.936	1.000	0.948
Recall	0.998	0.889	0.988	0.934	1.000	0.993
F1-score	0.999	0.879	0.991	0.935	1.000	0.970
Optimized Ensemble						
Precision	1.000	0.946	0.966	0.949	1.000	0.994
Recall	1.000	0.938	0.946	0.983	1.000	0.982
F1-score	1.000	0.942	0.956	0.966	1.000	0.988
Optimized Neural Network						
Precision	1.000	0.977	0.982	0.993	1.000	1.000
Recall	1.000	0.984	0.980	0.988	1.000	1.000
F1-score	1.000	0.980	0.981	0.990	1.000	1.000

Table 8.2: Performance parameters evaluated on the testing cycle for different classifiers trained on residual sensor signals (Dataset 1)

	NOM	MAF LOW	MAF HIGH	EGR LP	EGR HP	EGR BP
Optimized Tree						
Precision	0.369	0.610	0.639	0.585	0.696	0.343
Recall	0.326	0.634	0.567	0.619	0.731	0.377
F1-score	0.346	0.622	0.601	0.602	0.714	0.359
Optimized Ensemble						
Precision	0.565	0.669	0.788	0.654	0.789	0.577
Recall	0.354	0.796	0.698	0.758	0.918	0.427
F1-score	0.435	0.727	0.740	0.702	0.849	0.491
Optimized Neural Network						
Precision	0.317	0.719	0.710	0.608	0.824	0.387
Recall	0.308	0.637	0.707	0.652	0.855	0.419
F1-score	0.312	0.675	0.709	0.629	0.839	0.403

In the following paragraphs, two of the considered test cases are deeply analysed to show in detail how the models behave. The raw prediction given by the classification model has been filtered by means of the same moving average introduced in Section 5.2.1. For a clearer representation, only the averaged signal is shown in the following plots.
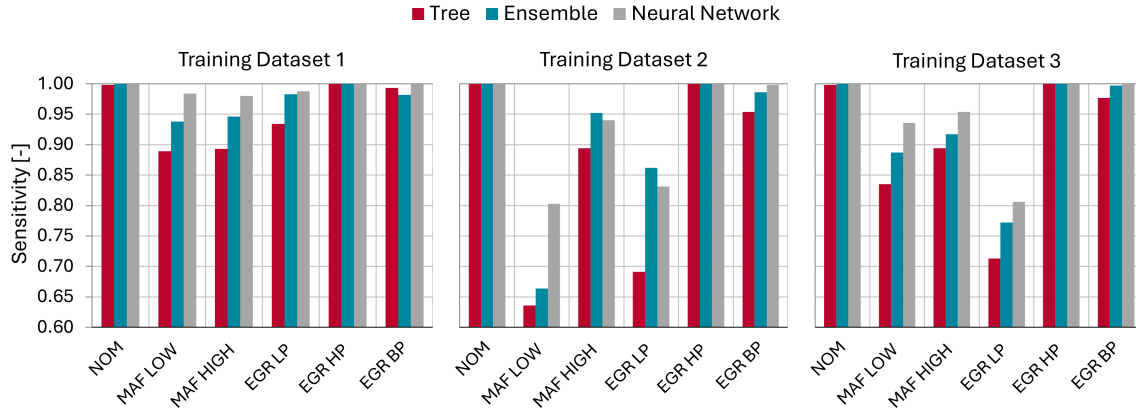
Figure 8.1: Testing results of optimized classification models trained on absolute sensor signals for different training datasets

Use case 1   The fault case considered in Fig. 8.2 is a 7% MAF low drift. In the bottom plot, engine speed and load profiles corresponding to the testing cycle are shown; the upper plot compares the fault class predicted by the different optimized models, namely Tree (red line), Ensemble (blue line), and Neural Network (grey line) classifiers, trained on Dataset 1, which includes all the tested fault conditions. As for the simulation case, the phases in which the engine is switched off have been excluded from the prediction and identified under the label "ICE OFF". The true fault class is represented by the black dashed line. The analysis of this specific test case confirms the general trend already discussed in the previous paragraphs: the Neural Network classifier exhibits the best performance, with an overall accuracy of 100%, followed by the Ensemble and Tree models, with an overall accuracy of 95.9% and 91.7% respectively.

Use case 2   20% LP EGR valve clogging is considered in the test case shown in Fig. 8.3. As in the previous case, engine speed and torque are shown in the bottom plot, while in the top plot, a comparison is provided between the fault class predicted by the optimized Neural Network models trained on the three different datasets, namely Dataset 1 (grey line), Dataset 2 (green line), and Dataset 3 (orange line). As expected, and already evidenced by Fig. 8.1, the models trained on the sets of fault conditions including the considered test case (Dataset 1 and Dataset 3), exhibit the highest accuracy, namely 99.7% and 100% respectively. On the other hand, the model trained on Dataset 2, which includes only the most severe 10% LP EGR valve clogging fault case, is not able to perfectly identify the considered fault case; however the overall accuracy reaches 77.3%, which is still an acceptable result if the fault detection performance is evaluated on the whole cycle. Moreover, looking to the predicted classes, it can be observed that the misclassification errors, corresponding to an overall false negative rate of 22.7%, involve mostly MAF high drift and nominal classes. This means that the actual rate of false fault detection, thus excluding the nominal class detection, is lower, and more specifically 10.8%, which is an even more acceptable

result.
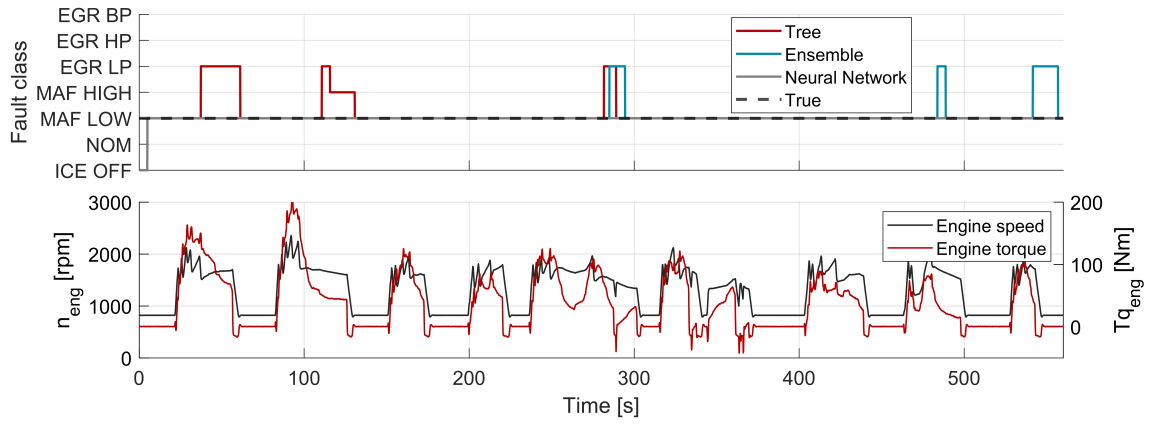
Figure 8.2: Use case 1: test of optimized classification models trained on Dataset 1
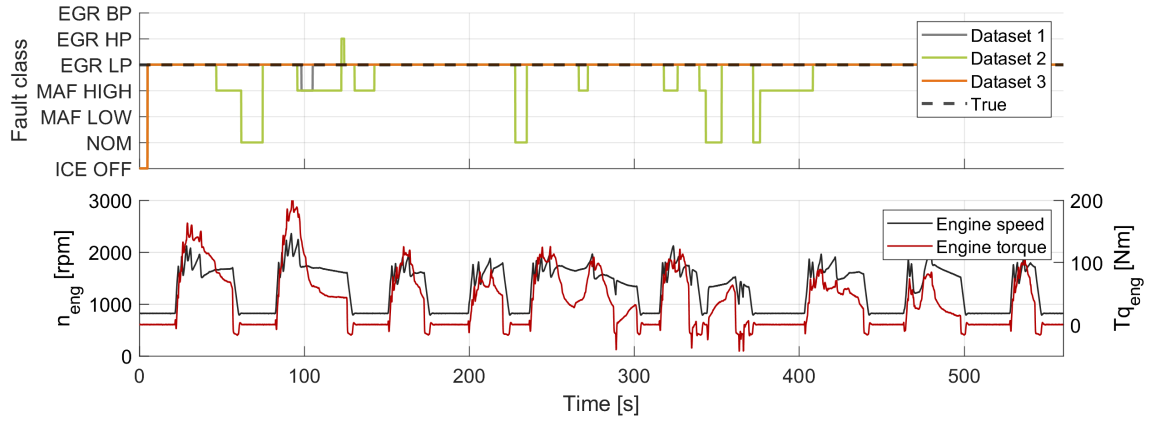
Figure 8.3: Use case 2: test of optimized Neural Network classifiers trained on different datasets

8.2.2   NOx emission estimation

The results in terms of percentage error on the final cumulated NOx emission evaluated on the testing cycle for different fault conditions are summarized in Fig. 8.4-Fig. 8.6. More in detail, Fig. 8.4 shows the performance of the three optimized regression models, trained on Dataset 1 according to the NOx mass flow correction approach, on the same testing cycle where different fault conditions have been considered: as seen in Section 5.2.2, the black hatched bars represent the variation of the cumulated NOx emission, measured at the test bench, due to the introduction of the fault compared to the nominal condition; the red hatched bars indicate the deviation between the cumulated NOx estimated by the ECU and the actual NOx emission; analogously, the remaining red, blue, and grey bars represent the error between the cumulated NOx estimation corrected

by means of the developed models, namely optimized Tree, Ensemble, and Neural Network regression models respectively, and the actual NOx emission. Based on these results, thanks to the applied corrections, a reduction of the error on cumulated NOx emission is observed, compared to the reference ECU estimation, in all the considered fault cases. The three different data-driven models show comparable performance, with the Neural Network achieving the lowest error for most of the test cases. More in detail, a perfect prediction is observed for all MAF low drift and LP EGR valve fault cases. A good performance is exhibited also for the most critical tests, namely +15% MAF high drift and EGR bypass fault cases, where both Ensemble and Tree models show significant errors, up to 40% in the first case. Differently from what has been observed in the validation phase in Section 7.3.2, the Ensemble model is not showing the best overall performance, also being outperformed by the Tree model in some test cases. In parallel, Fig. 8.5, shows the result on the same test cases obtained applying the NOx concentration correction approach. As expected, and already introduced in Section 7.3.1, a deterioration of the final prediction accuracy is observed for the test cases involving MAF sensor drifts, namely when the considered fault leads to a significant variation of the intake air mass flow in addition to the increase or decrease in the emission concentration. These last considerations apply to all three tested models. Finally, even if a reduced error compared to Fig. 8.4 is shown by the Ensemble and Tree models on 20% LP EGR valve clogging and HP EGR bypass fault cases, the Neural Network-based NOx mass flow correction approach is still preferred considering its overall performance, with a maximum overestimation error of 7% corresponding to the +15% MAF high drift case, and only one case, namely the nominal one, in which a slight underestimation of 1% is observed.

Based on these promising results, as a further analysis, Fig. 8.6 presents a comparison between the performance obtained with Neural Networks trained on the three different datasets described in Section 7.1. As expected, the model trained on Dataset 1 (grey bars), including the complete set of fault conditions (grey boxes on the x-axis), shows the best overall performance. Nonetheless, all three models show an improved performance compared to the reference ECU model, in terms of absolute deviation from the actual emission, in most of the test cases. Focusing on MAF sensor fault cases, it can be observed that the Neural Network trained on Dataset 2 (green bars) shows a very good performance on the test cases involving fault conditions included in the training datasets (highlighted by the green boxes on the x-axis), namely -15% MAF low drift and +15% MAF high drift, while a significant deterioration of model performance is observed for the less severe fault cases, not included in the training dataset (orange boxes). Analogously, the Neural Network trained on Dataset 3 (orange bars) shows limited deviations on the tests considering the least severe fault conditions (orange boxes), while significant deviations for the most severe fault cases not included in the training dataset (green boxes). However, focusing on LP EGR valve fault cases, while the Neural Network trained on Dataset 3 exhibits the same behaviour observed for the MAF fault cases, the model trained on Dataset 2 is able to achieve high prediction accuracy even on the test case considering the least severe fault condition not included in the training dataset. This means that for this specific fault case, where a perfect performance of the Network trained

on Dataset 1 was already observed, the model trained only on the most severe degree of severity is able to extend the relations learnt from the training data also to the least severe fault condition.

To summarize, the developed NOx correction models show promising results when tested on an unseen cycle, allowing to significantly reduce the gap between actual and estimated NOx emission even if highly emission-relevant faults are considered. Good performance are achieved by all the investigated models, significantly better than the reference ECU model in most of the test cases, with the Neural Network outperforming Tree and Ensemble models. Moreover, the approach considering the correction of NOx mass flow, rather than NOx concentration, proves to be the best, allowing to achieve the lowest estimation errors especially when MAF-related fault are considered. Moreover, the Neural Network trained in Dataset 2, including only the most severe fault conditions, is able to provide good results in most of the test cases, significantly improving reference ECU performance, except for one single test case, namely +7.5% MAF high drift, where a much more significant error is observed compared to the reference ECU.
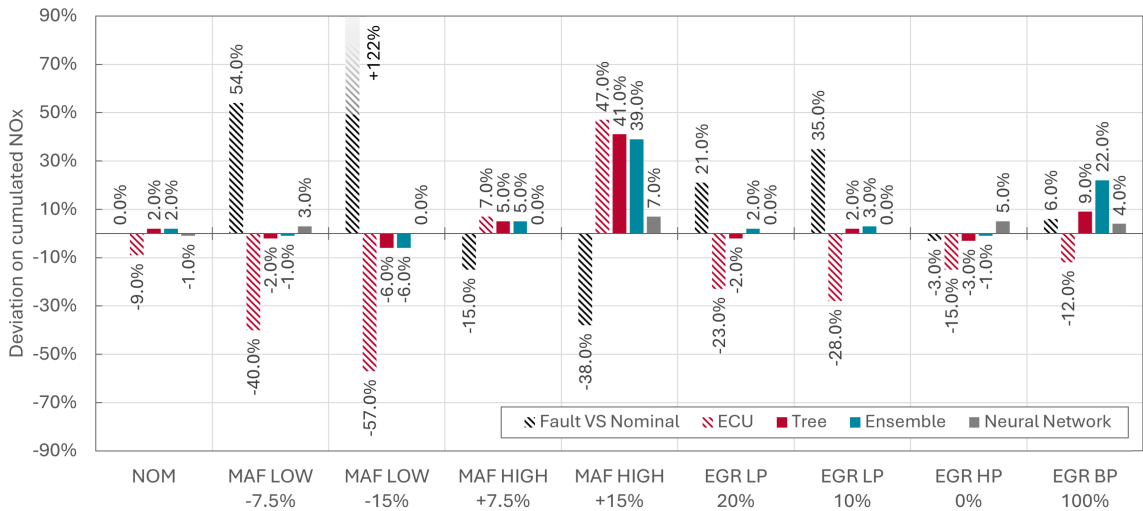
Figure 8.4: Testing results of optimized regression models trained on Dataset 1 according to the NOx mass flow correction approach

In the following paragraphs, the results of the application of Neural Network regression models to correct the NOx emission estimated by the ECU on four different test cases are deeply analysed, to clarify in detail how the models behave both in terms of instantaneous and cumulated emission trends when different approaches and training datasets are considered. For all the presented plots, NOx emissions have been normalized with respect to the maximum value observed on the same cycle in the nominal case.

Use case 1   The test case where +15% MAF high drift is applied is analysed in the upper plot of Fig. 8.7. The solid lines represent the instantaneous NOx emission, while the dashed lines

165
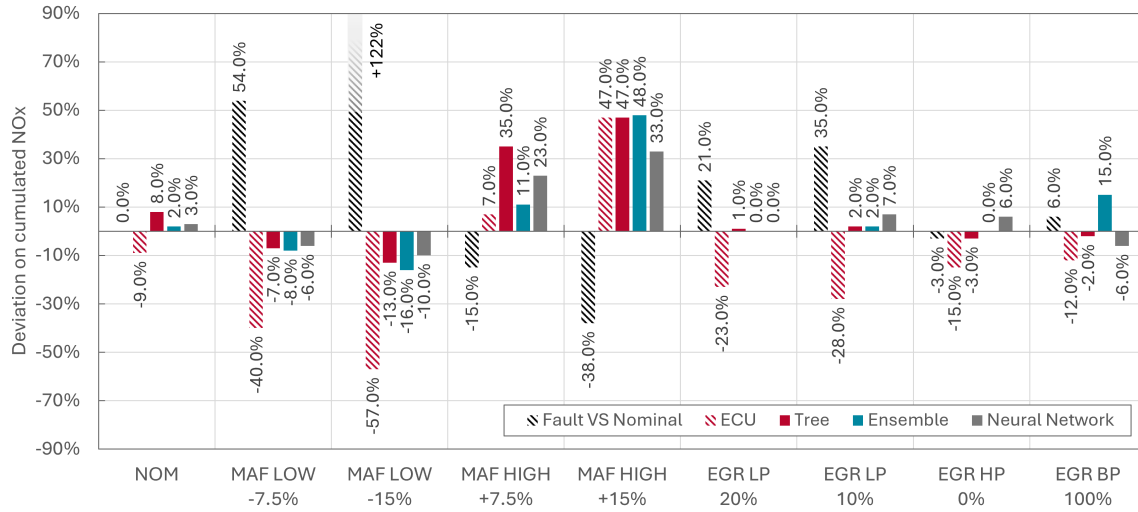
Figure 8.5: Testing results of optimized regression models trained on Dataset 1 according to the NOx concentration correction approach
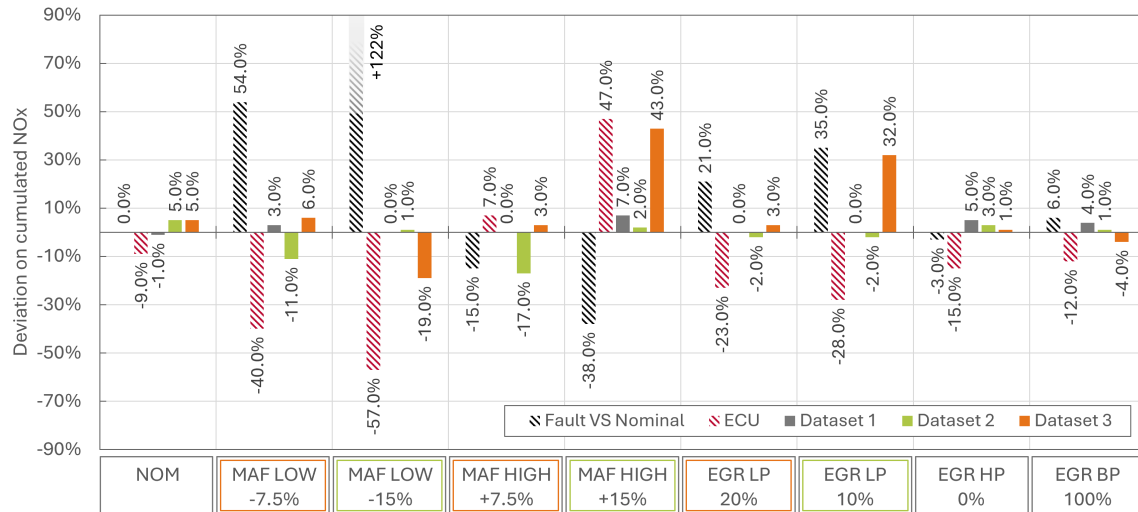
Figure 8.6: Testing results of optimized Neural Network regression model trained on different datasets according to the NOx mass flow correction approach

represent the cumulated emission. As expected for this fault, a significant decrease of actual NOx emission (black line) can be observed compared to the nominal condition (grey line). Even if the ECU model is partially reacting to the introduction of the fault, as confirmed by a lower NOx emission estimated by the controller (red line) compared to the nominal one, this adaptation is not enough to guarantee an acceptable error on NOx estimation, which is still close to 50%. If the NOx emission correction calculated by the optimized Neural Network model, according to the concentration approach, is applied to the controller estimation (green line), the resulting

deviation between actual and estimated NOx emission is reduced, but still leading to significant error on the cumulative emission that is higher than 30%. On the other hand, when the NOx mass flow correction is applied (blue line), the error is reduced to 7%, thus strongly improving ECU estimation accuracy.

Use case 2 In this test case, corresponding to the central plot in Fig. 8.7, -7.5% MAF low drift is considered. Contrary to the previous case, a significant increase of actual NOx emission (black line) is observed compared to the nominal condition (grey line) due to the introduction of the fault. The ECU model is not properly reacting to the introduction of the fault, as confirmed by the slightly lower NOx emission estimated by the ECU (red line) compared to the nominal case, thus resulting in a 40% underestimation. Both the correction approaches allow to strongly improve the estimation accuracy compared to the reference ECU model, however the NOx mass flow correction approach is the one achieving the best performance, resulting in a positive deviation between actual and estimated NOx emission of 3%, against the negative residual error of 6% resulting from the application of the NOx concentration correction approach.

Use case 3 and 4 20% and 10% LP EGR valve clogging are considered in these test cases. The corresponding results, comparing the performance of Neural Networks trained on different datasets according to the NOx mass flow correction approach, are represented in the upper and central plots of Fig. 8.8 respectively. As in the previous case, a significant increase of actual NOx emission (black line) is observed compared to the nominal condition (grey line) for both cases, which is higher for the most severe fault case. The ECU model is essentially not reacting to the introduction of the fault, leading to an underestimation of more than 20% in both cases. As already discussed in this section, the Network trained on Dataset 3 (orange line) exhibits a very good accuracy on the less severe fault condition (use case 3), which is included in the training dataset, while a strong deterioration of model performance is observed for the other test case (use case 4), leading to an overestimation of more than 30%. On the other hand, the correction calculated by the model trained on Dataset 2 (green line) allows to obtain a very accurate NOx emission prediction in both fault cases, even if the fault condition considered in use case 3 is not included in the training dataset. This results in a final deviation of less than 2% from the actual NOx emission in both test cases.

8.3 Conclusions of Part II

To summarize, concerning the fault classification model, the results presented in Section 5.2.1 highlight that:

- good classification accuracy is achieved training the models on a limited portion of the available experimental data, including however all tested fault conditions (Dataset 1), and
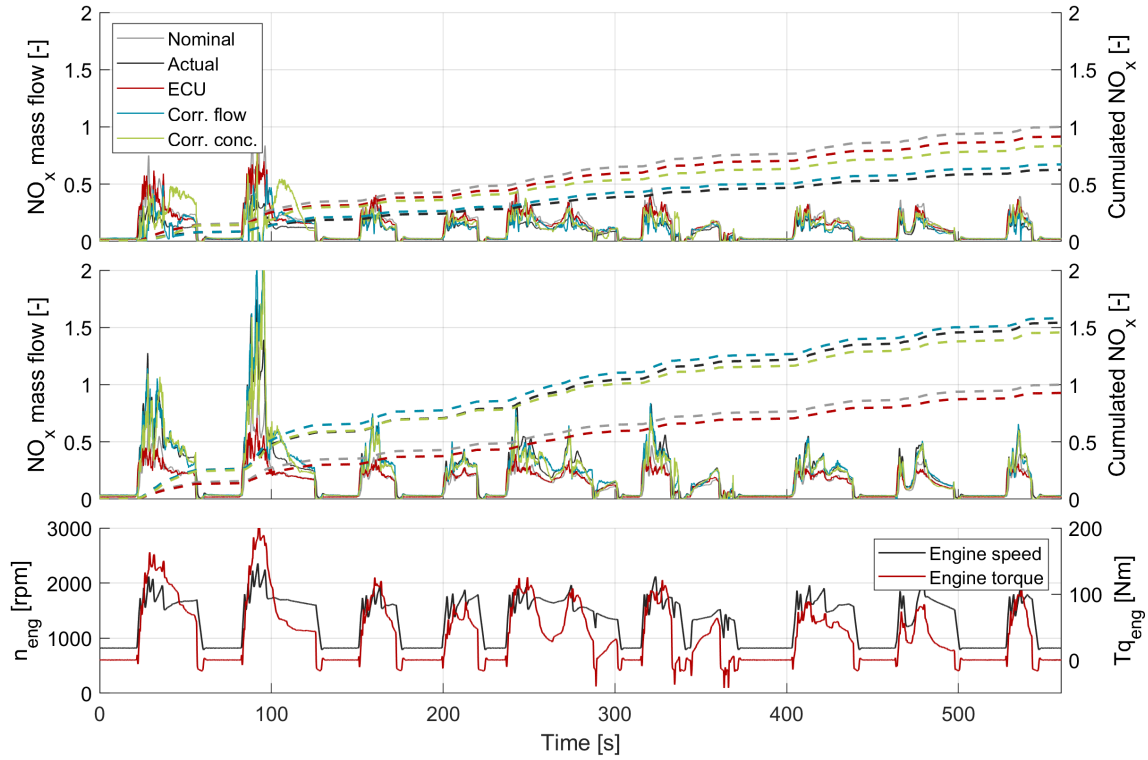
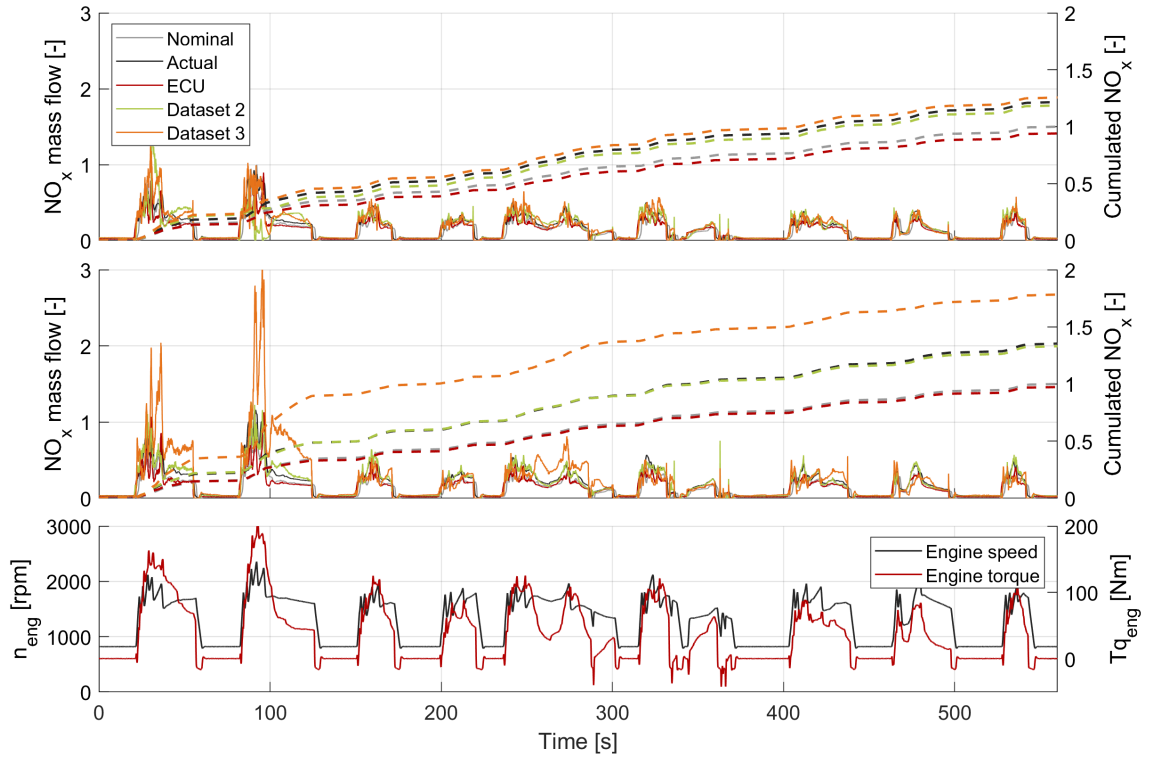**Figure 8.7:** Use cases 1 and 2: test of optimized Neural Network regression models trained on Dataset 1 according to mass flow and concentration correction approaches. NOx emissions are normalized with respect to the nominal case

testing them on the remaining unseen data;

- model performance is strongly affected by the training dataset, with a significant deterioration of classification accuracy when selected fault conditions are not included in the training dataset (Dataset 2 and Dataset 3). However, promising results are achieved also for the models trained on the dataset including only the least severe fault conditions (Dataset 3);

- the approach exploiting physical signals from on-board sensors as input features provides much better overall results compared to the residual approach;

- as for the simulation case described in Part I, the Neural Network is the one providing the best trade-off in terms of overall performance on the considered tests and model complexity, with an accuracy higher than 96% in all fault conditions if trained on Dataset 1.

In parallel, based on the results discussed in Section 5.2.1 related to the NOx correction model, it can be concluded that:

- all three optimized regression models allows to significantly reduce the error between actual

Figure 8.8: Use cases 3 and 4: test of optimized Neural Network regression models trained on Dataset 2 and Dataset 3 according to mass flow approaches. NOx emissions are normalized with respect to the nominal case

and ECU-estimated NOx emission for almost all the considered test cases if trained on the complete set of fault conditions (Dataset 1);

- the models correcting directly the NOx mass flow instead of NOx concentration achieve the best overall results, especially on MAF-related faults;

- model performance is still good when selected fault conditions are not included in the training dataset, and the model is asked to interpolate between known fault conditions, including only the most sever fault cases (Dataset 2);

- the Neural Network model is the one providing the best trade-off in terms of prediction accuracy on different tests and model complexity. If trained on Dataset 1, including all fault conditions, the developed model outperforms the reference ECU model in all test cases, with a maximum error of 7% on the final cumulated NOx emission in the worst case; if trained only on most severe fault cases (Dataset 2), the developed model outperforms the reference ECU model in most of the test cases, except for one isolated fault case (7.5% MAF high drift).

170

Chapter 9

Conclusions and future developments

Over the past decades, the growing level of industrialization and urbanization across several developing countries has led to a marked increase in the atmospheric concentration of pollutants, with all the severe implications involving environmental degradation and human health. The situation is particularly critical in cities, where most of the population lives, and where the transport sector is regularly the main contributor to air pollution [10]. To address these issues, in recent years European regulations on pollutant emissions have become increasingly stringent for nearly all vehicle categories, both in terms of pollutant thresholds and homologating procedures. New challenges are introduced by the Euro 7 regulation that recently came into force, mandating an extended vehicle lifetime and continuous emission On-Board Monitoring (OBM) on all trips [17]. This last point, due to the multitude of factors affecting vehicle emissions as well as the measuring accuracy, represents the dominating challenge to comply with the Euro 7 standards [18].

In this context, the objective of this dissertation is to demonstrate the potential of machine learning approaches in supporting the diagnosis of emission-relevant fault conditions and the implementation of a Euro7-compliant emission on-board monitoring strategy, overcoming the limitations of current on-board sensors and standard control-oriented models. With this aim, a comprehensive and experimentally validated 0-D vehicle model developed in *Matlab/Simulink* has been considered, capable of reliably simulating common engine faults affecting engine-out emissions. The model has been used to investigate the effects of engine-related faults on NOx emissions and on the main on-board measurable signals. Establishing a correlation between the variation of these signals and the presence of different faults allows the latter to be detected and identified, taking advantage of information that is readily available on-board. In this respect,

data-driven approaches offer a viable solution to address this problem, in particular for their relative simplicity and ability to model complex relationships between an input dataset and a target output. Therefore, several driving cycles have been simulated including both nominal and fault conditions, to generate a proper dataset to train and validate different data-driven classification models. To neglect useless information, based on both physical considerations and analytical feature ranking algorithms, only those signals that proved to be relevant for fault identification have been included in the training dataset. Then, a preliminary analysis has been performed to evaluate the performance of different classifiers in terms of accuracy, prediction speed, and requested training time. As a result, three different types of classification models have been selected, namely Decision Tree, Ensemble, and Neural Network-based classifiers, which have been optimized by means of Bayesian optimization to further improve their accuracy. The same methodology has been followed to train and optimize data-driven regression models to correct the NOx emission estimated by the engine controller when a fault is introduced. In fact, commonly adopted map-based emission models, calibrated in nominal conditions, are not able to fully adapt themselves to correctly predict NOx emissions when faults are considered, thus resulting in a significant deviation between actual and estimated emissions. Taking advantage of on-board measurable signals, which contain fault-related information, the developed regression models are able to output a correction factor to be applied to the prediction of the reference model, thereby accounting for malfunction-related emission variations. As a final step, the resulting fault classification and NOx correction models have been integrated into the *Simulink* vehicle model and tested on different driving cycles and fault conditions to evaluate their accuracy, robustness, and interpolation/extrapolation capabilities.

The main results of this first part of the Ph.D. activity, presented and discussed in detail in Chapter 5, demonstrate the great potential of the proposed approach. The developed fault classification models are capable of identifying the considered faults over a wide range of engine operating conditions with negligible misclassification errors, showing promising results if tested on unseen driving cycles and fault severity levels not included in the training dataset. In parallel, satisfactory results were also observed for the NOx correction models on different testing cycles and fault conditions, managing to significantly reduce the error between actual and estimated emissions compared to the reference NOx model. As a fundamental step in view of on-board implementation, the developed models have been deployed on a commercial real-time hardware, namely a Raspberry Pi, and tested at HiL to verify their real-time capability and related computational load. As well as demonstrating the real-time implementability of the models, the proposed hardware architecture offers a ready-to-use and viable solution for vehicle testing: the Raspberry Pi can communicate via CAN-bus with a standard control unit, thus receiving all signals required as input to the data-driven models and sending the computed output signals.

172

After the definition and validation of a clear methodology for the development of OBM-oriented data-driven models, the same approach has been applied to real experimental data acquired at the test bench in both nominal and fault-induced conditions, as presented in Part II of the dissertation. A preliminary analysis of the most significant signals measured at the test bench has been carried out, mainly to study the effects of the induced faults and the corresponding reaction of the standard ECU. The deviation between actual and estimated NOx emissions has been quantified, to assess the capability of the built-in ECU models to adapt to non-nominal conditions. The same machine-learning methods, feature ranking algorithms, and optimization processes applied in simulation have been considered. The results, extensively discussed in Chapter 8, prove the effectiveness of the proposed approach: the fault classification models show excellent accuracy when applied to an unseen testing dataset, with promising results even considering extremely reduced training conditions. The same applies to the NOx correction models, capable of drastically reducing the error on estimated NOx emissions compared to the reference ECU model even for the most critical fault conditions, taking advantage of readily available on-board signals.

The present dissertation can be definitely considered as a valid proof of concept demonstrating the potential and feasibility of the application of data-driven approaches to support standard on-board diagnostics and emission monitoring systems in view of the challenging Euro 7 requirements. The proposed methodology can be used as a starting point for further developments focused on the improvement of the presented approach to overcome its actual limits. These may include: the analysis of additional known engine faults, involving also the EATS system; the development of a parallel approach to handle unknown fault conditions; the investigation of an innovative methodology to distinguish between discrete levels of fault severity and determine the State of Health (SoH) of selected components; the extension of the proposed approach for NOx emission correction to other pollutants. Moreover, thanks to the readily available and flexible hardware architecture employed for the real-time testing of the developed models, vehicle on-board testing will be the natural continuation of this activity to further assess the robustness and reliability of the proposed approach over a wider range of real-world applications.

Bibliography

[1] K. O. Yoro and M. O. Daramola, "CO2 emission sources, greenhouse gases, and the global warming effect," en, *Advances in Carbon Capture*, Elsevier, 2020, pp. 3–28, ISBN 978-0-12-8 19657-1. doi: 10.1016/B978-0-12-819657-1.00001-3. [Online]. Available: https://linkinghu b.elsevier.com/retrieve/pii/B9780128196571000013 (visited on 12/16/2021).

[2] H. Ritchie, P. Rosado, and M. Roser, "Breakdown of carbon dioxide, methane and nitrous oxide emissions by sector," *Our World in Data*, Jan. 2024. [Online]. Available: https://ourw orldindata.org/emissions-by-sector (visited on 07/02/2024).

[3] M. Kampa and E. Castanas, "Human health effects of air pollution," *Environmental Pollution*, Proceedings of the 4th International Workshop on Biomonitoring of Atmospheric Pollution (With Emphasis on Trace Elements), vol. 151, no. 2, pp. 362–367, Jan. 2008, ISSN 0269-7491. doi: 10.1016/j.envpol.2007.06.012. [Online]. Available: https://www.sciencedirect.com/s cience/article/pii/S0269749107002849 (visited on 07/02/2024).

[4] O. US EPA, *Nitrogen Dioxide (NO2) Pollution*, en, Collections and Lists, Mar. 2016. [Online]. Available: https://www.epa.gov/no2-pollution (visited on 07/03/2024).

[5] H. Ritchie, V. Samborska, and M. Roser, "Urbanization," *Our World in Data*, Feb. 2024. [Online]. Available: https://ourworldindata.org/urbanization (visited on 07/02/2024).

[6] *Europe's air quality status 2023*, en, Briefing. [Online]. Available: https://www.eea.europa.e u/publications/europes-air-quality-status-2023/europes-air-quality-status2023 (visited on 07/03/2024).

[7] *Share of people living in urban and rural areas*. [Online]. Available: https://ourworldindata.o rg/grapher/share-urban-and-rural-population (visited on 07/02/2024).

[8] *Air quality statistics*, en, Dashboard (Tableau). [Online]. Available: https://www.eea.euro pa.eu/data-and-maps/dashboards/air-quality-statistics (visited on 07/02/2024).

[9] *Emissions of air pollutants from transport in Europe*, en, Dec. 2023. [Online]. Available: http s://www.eea.europa.eu/en/analysis/indicators/emissions-of-air-pollutants-from (visited on 07/03/2024).

[10] "Commission proposes new euro 7 standards," en, *European Commission - European Commission*, Nov. 2022. [Online]. Available: https://ec.europa.eu/commission/presscorner/d etail/en/ip_22_6495 (visited on 07/09/2024).

[11] *Emissions of the main air pollutants by sector group in the EEA-33*, en, Data Visualization. [Online]. Available: https://www.eea.europa.eu/data-and-maps/daviz/share-of-eea-33 -emissions-5 (visited on 07/02/2024).

[12] *CO2 emission standards for new passenger cars and vans in the European Union*, en-US. [Online]. Available: https://theicct.org/publication/eu-co2-standards-cars-vans-may23/ (visited on 07/03/2024).

[13] *Regulation (EU) 2019/631 of the European Parliament and of the Council of 17 April 2019 setting CO2 emission performance standards for new passenger cars and for new light commercial vehicles, and repealing Regulations (EC) No 443/2009 and (EU) No 510/2011 (recast) (Text with EEA relevance)Text with EEA relevance*, en, Legislative Body: OP_DATPRO, Jan. 2024. [Online]. Available: http://data.europa.eu/eli/reg/2019/631/2024-01-01/eng (visited on 07/03/2024).

[14] *Commission Regulation (EU) 2017/1151 of 1 June 2017 supplementing Regulation (EC) No 715/2007 of the European Parliament and of the Council on type-approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (Euro 5 and Euro 6) and on access to vehicle repair and maintenance information, amending Directive 2007/46/EC of the European Parliament and of the Council, Commission Regulation (EC) No 692/2008 and Commission Regulation (EU) No 1230/2012 and repealing Commission Regulation (EC) No 692/2008 (Text with EEA relevance)*, en, Legislative Body: GROW, COM, Jul. 2017. [Online]. Available: http://data.europa.eu/eli/reg/2017/1151/oj/eng (visited on 06/16/2024).

[15] L. Ntziachristos, G. Mellios, D. Tsokolis, M. Keller, *et al.*, "In-use vs. type-approval fuel consumption of current passenger cars in Europe," *Energy Policy*, vol. 67, pp. 403–411, Apr. 2014, ISSN 0301-4215. doi: 10.1016/j.enpol.2013.12.013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421513012573 (visited on 07/03/2024).

[16] *Commission Regulation (EU) 2018/1832 of 5 November 2018 amending Directive 2007/46/EC of the European Parliament and of the Council, Commission Regulation (EC) No 692/2008 and Commission Regulation (EU) 2017/1151 for the purpose of improving the emission type approval tests and procedures for light passenger and commercial vehicles, including those for in-service conformity and real-driving emissions and introducing devices for monitoring the consumption of fuel and electric energy (Text with EEA relevance.)* en, Legislative Body: COM, GROW, Nov. 2018. [Online]. Available: http://data.europa.eu/eli/reg/2018/1832/oj/eng (visited on 06/16/2024).

[17] *Regulation (EU) 2024/1257 of the European Parliament and of the Council of 24 April 2024 on type-approval of motor vehicles and engines and of systems, components and separate technical units intended for such vehicles, with respect to their emissions and battery durability (Euro 7), amending Regulation (EU) 2018/858 and repealing Regulations (EC) No 715/2007 and (EC) No 595/2009 of the European Parliament and of the Council, Commission Regulation (EU) No 582/2011, Commission Regulation (EU) 2017/1151, Commission Regulation (EU) 2017/2400 and Commission Implementing Regulation (EU) 2022/1362*, en, Legislative Body: OP_DATPRO, Apr. 2024. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401257 (visited on 07/03/2024).

[18] V. Müller, H. Pieta, J. Schaub, M. Ehrly, *et al.*, "On-Board Monitoring to meet upcoming EU-7 emission standards – Squaring the circle between effectiveness and robust realization," en, *Transportation Engineering*, vol. 10, p. 100 138, Dec. 2022, ISSN 2666-691X. doi: 10.1016/j.treng.2022.100138. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666691X22000367 (visited on 10/25/2022).

[19]   Z. C. Samaras, A. Kontses, A. Dimaratos, D. Kontses, *et al.*, "A European Regulatory Per-
       spective towards a Euro 7 Proposal," en, *SAE International Journal of Advances and Current
       Practices in Mobility*, vol. 5, no. 3, pp. 998–1011, Jun. 2022, ISSN 2641-9645. doi: 10.4271/20
       22-37-0032. [Online]. Available: https://www.sae.org/content/2022-37-0032 (visited on
       09/04/2023).

[20]   C. Morgan and J. Goodwin, "Impact of the Proposed Euro 7 Regulations on Exhaust
       Aftertreatment System Design: New Euro standards of global importance to the automotive
       industry," en, *Johnson Matthey Technology Review*, vol. 67, no. 2, pp. 239–245, Apr. 2023,
       Publisher: Johnson Matthey, ISSN 2056-5135. doi: 10.1595/205651323X16805977899699.
       [Online]. Available: https://technology.matthey.com/content/journals/10.1595/2056513
       23X16805977899699 (visited on 07/04/2024).

[21]   L. Kang, D. Lou, Y. Zhang, L. Fang, *et al.*, "Research on cross sensitivity of NOx sensor
       and Adblue injection volume in accordance with the actual situation based on cubature
       Kalman filter," *Energy*, vol. 284, p. 128 666, Dec. 2023, ISSN 0360-5442. doi: 10.1016/j.ener
       gy.2023.128666. [Online]. Available: https://www.sciencedirect.com/science/article/pii
       /S0360544223020601 (visited on 09/28/2023).

[22]   Y. Tan, P. Henderick, S. Yoon, J. Herner, *et al.*, "On-Board Sensor-Based NOx Emissions from
       Heavy-Duty Diesel Vehicles," *Environmental Science & Technology*, vol. 53, no. 9, pp. 5504–
       5511, May 2019, Publisher: American Chemical Society, ISSN 0013-936X. doi: 10.1021/ac
       s.est.8b07048. [Online]. Available: https://doi.org/10.1021/acs.est.8b07048 (visited on
       09/28/2023).

[23]   A. Dimaratos, D. Kontses, S. Doulgeris, N. Zacharof, *et al.*, "The Potential of On-Board Data
       Monitoring for the Characterization of Real-World Vehicle Fuel and Energy Consumption
       and Emissions," en, Capri, Italy, Aug. 2023, pp. 2023–24–0113. doi: 10.4271/2023-24-0113.
       [Online]. Available: https://www.sae.org/content/2023-24-0113 (visited on 09/28/2023).

[24]   B. Du, L. Zhang, Y. Geng, Y. Zhang, *et al.*, "Testing and evaluation of cold-start emissions in
       a real driving emissions test," en, *Transportation Research Part D: Transport and Environment*,
       vol. 86, p. 102 447, Sep. 2020, ISSN 1361-9209. doi: 10.1016/j.trd.2020.102447. [Online].
       Available: https://www.sciencedirect.com/science/article/pii/S1361920920306349
       (visited on 09/28/2022).

[25]   C. Quérel, O. Grondin, and C. Letellier, "Semi-physical mean-value NOx model for diesel
       engine control," *Control Engineering Practice*, vol. 40, pp. 27–44, Jul. 2015, ISSN 0967-0661.
       doi: 10.1016/j.conengprac.2015.02.005. [Online]. Available: https://www.sciencedirect.co
       m/science/article/pii/S0967066115000386 (visited on 10/02/2023).

[26]   C. Guardiola, J. Martín, B. Pla, and P. Bares, "Cycle by cycle NOx model for diesel engine
       control," *Applied Thermal Engineering*, vol. 110, pp. 1011–1020, Jan. 2017, ISSN 1359-4311.
       doi: 10.1016/j.applthermaleng.2016.08.170. [Online]. Available: https://www.sciencedire
       ct.com/science/article/pii/S1359431116315307 (visited on 07/04/2024).

[27]   S.-Y. Lee, J. Andert, S. Pischinger, M. Ehrly, *et al.*, "Scalable Mean Value Modeling for Real-
       Time Engine Simulations with Improved Consistency and Adaptability," English, ISSN:
       0148-7191, 2688-3627, Warrendale, PA: SAE International, Apr. 2019. doi: 10.4271/2019-01
       -0195. [Online]. Available: https://www.sae.org/publications/technical-papers/content
       /2019-01-0195/ (visited on 10/04/2023).

[28] *Emissions of the main air pollutants in Europe*, en, Sep. 2023. [Online]. Available: https://www.eea.europa.eu/en/analysis/indicators/emissions-of-the-main-air (visited on 07/04/2024).

[29] J. Heywood, *Internal Combustion Engine Fundamentals.pdf*, English. McGraw-Hill Education, 2018.

[30] J. Kasab and A. Strzelec, *Automotive Emissions Regulations and Exhaust Aftertreatment Systems*, en. SAE International, Aug. 2020, ISBN 978-0-7680-9955-3. doi: 10.4271/9780768099560. [Online]. Available: https://saemobilus.sae.org/content/R-477/ (visited on 07/04/2024).

[31] K. Senecal, F. Leach, and Associate Professor of Engineering Science at the University of Oxford, *Racing Toward Zero: The Untold Story of Driving Green*, en. SAE International, Jun. 2021, ISBN 978-1-4686-0147-3. doi: 10.4271/9781468601473. [Online]. Available: https://saemobilus.sae.org/content/R-501 (visited on 07/04/2024).

[32] *Council Directive 70/220/EEC of 20 March 1970 on the approximation of the laws of the Member States relating to measures to be taken against air pollution by gases from positive-ignition engines of motor vehicles*, en, Mar. 1970. [Online]. Available: http://data.europa.eu/eli/dir/1970/220/oj/eng (visited on 07/05/2024).

[33] *Regulation (EC) No 715/2007 of the European Parliament and of the Council of 20 June 2007 on type approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (Euro 5 and Euro 6) and on access to vehicle repair and maintenance information (Text with EEA relevance)*, en, Legislative Body: EP, CONSIL, Jun. 2007. [Online]. Available: http://data.europa.eu/eli/reg/2007/715/oj/eng (visited on 07/05/2024).

[34] *Emission standards in europe for cars and light trucks: Rde testing*. [Online]. Available: https://dieselnet.com/standards/eu/ld_rde.php (visited on 07/08/2024).

[35] "Commission proposes new Euro 7 standards," en, *European Commission - European Commission*, Nov. 2022. [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_6495 (visited on 07/09/2024).

[36] Z. Samaras, S. Hausberger, and D. G. Mellios, *Preliminary findings on possible Euro 7 emission limits for LD and HD vehicles*, en, Oct. 2020. [Online]. Available: https://circabc.europa.eu/.

[37] ACEA, "ACEA Proposal for Euro7," Tech. Rep., Jun. 2021. [Online]. Available: https://www.acea.auto/files/ACEA_proposal_for_Euro7.pdf (visited on 02/28/2022).

[38] J. Dornoff and F. Rodríguez, "Euro 7: The new emission standard for light- and heavy-duty vehicles in the European Union," en, Mar. 2024.

[39] *Directive 98/69/EC of the European Parliament and of the Council of 13 October 1998 relating to measures to be taken against air pollution by emissions from motor vehicles and amending Council Directive 70/220/EEC*, en, Oct. 1998. [Online]. Available: http://data.europa.eu/eli/dir/1998/69/oj/eng (visited on 07/10/2024).

[40] *Commission Directive 2003/76/EC of 11 August 2003 amending Council Directive 70/220/EEC relating to measures to be taken against air pollution by emissions from motor vehicles (Text with EEA relevance)*, en, Aug. 2003. [Online]. Available: http://data.europa.eu/eli/dir/2003/76/oj/eng (visited on 07/10/2024).

[41] *2019/2020 Worldwide Emissions Standards: Passenger Cars and Light Duty Vehicles*, en. Delphi, 2016, Google-Books-ID: Dx8JyAEACAAJ.

[42] *ISO 15031-3: Road vehicles — Communication between vehicle and external equipment for emissions-related diagnostics — Part 3: Diagnostic connector and related electrical circuits: Specification and use*, en, Feb. 2023. [Online]. Available: https://www.iso.org/standard/84210.html (visited on 07/11/2024).

[43] *Road vehicles — Controller area network (CAN)*, en, Dec. 2003. [Online]. Available: https://www.iso.org/standard/33422.html (visited on 07/17/2024).

[44] *OBD2 Explained - A Simple Intro*, en, 2023. [Online]. Available: https://www.csselectronics.com/pages/obd2-explained-simple-intro (visited on 07/17/2024).

[45] A. Barbier, J. M. Salavert, C. E. Palau, and C. Guardiola, "Analysis of the Euro 7 on-board emissions monitoring concept with real-driving data," *Transportation Research Part D: Transport and Environment*, vol. 127, p. 104 062, Feb. 2024, ISSN 1361-9209. doi: 10.1016/j.trd.2024.104062. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361920924000191 (visited on 02/28/2024).

[46] E. Directorate-General for Internal Market Industry and SMEs, *Technical studies for the development of Euro 7: testing, pollutants and emission limits*, eng. Publications Office of the European Union, 2022, ISBN 978-92-76-56406-5. [Online]. Available: https://data.europa.eu/doi/10.2873/97170 (visited on 07/16/2024).

[47] P. Zhao, X. Wu, S. Zhang, L. He, *et al.*, "Regulatory Insights for On-Board Monitoring of Vehicular NOx Emission Compliance," *Environmental Science & Technology*, vol. 58, no. 18, pp. 7968–7976, May 2024, Publisher: American Chemical Society, ISSN 0013-936X. doi: 10.1021/acs.est.4c00079. [Online]. Available: https://doi.org/10.1021/acs.est.4c00079 (visited on 05/20/2024).

[48] V. Franco, P. Dilara, N. Hennig, D. Manara, *et al.*, "On-Board Monitoring of Emissions in the Future Euro 7 Standard," en, Capri, Italy, Aug. 2023, pp. 2023–24–0111. doi: 10.4271/2023-24-0111. [Online]. Available: https://www.sae.org/content/2023-24-0111 (visited on 10/31/2023).

[49] *Commission Regulation (EU) 2023/443 of 8 February 2023 amending Regulation (EU) 2017/1151 as regards the emission type approval procedures for light passenger and commercial vehicles (Text with EEA relevance)*, en, Legislative Body: GROW, COM, Feb. 2023. [Online]. Available: http://data.europa.eu/eli/reg/2023/443/oj/eng (visited on 07/17/2024).

[50] S. Funk, J. Potter, E. Pruski, S. Funk, *et al.*, "Technical Challenges with on Board Monitoring," English, ISSN: 0148-7191, 2688-3627, 0148-7191, 2688-3627, SAE International, Apr. 2024. doi: 10.4271/2024-01-2597. [Online]. Available: https://saemobilus-sae-org.ezproxy.unibo.it/papers/technical-challenges-board-monitoring-2024-01-2597 (visited on 07/09/2024).

[51] T. Selleri, C. Ferrarese, J. Franzetti, R. Suarez-Bertoa, *et al.*, "Real-Time Measurement of NOx Emissions from Modern Diesel Vehicles Using On-Board Sensors," en, *Energies*, vol. 15, no. 22, p. 8766, Jan. 2022, Number: 22 Publisher: Multidisciplinary Digital Publishing Institute, ISSN 1996-1073. doi: 10.3390/en15228766. [Online]. Available: https://www.mdpi.com/1996-1073/15/22/8766 (visited on 01/02/2023).

[52] Y. Todo, H. Ichikawa, H. Yotou, K. Aoki, *et al.*, "Development of High Accuracy and Quick Light-off NOx Sensor," English, ISSN: 0148-7191, 2688-3627, 0148-7191, 2688-3627, SAE International, Apr. 2018. doi: 10.4271/2018-01-0334. [Online]. Available: https://saemobil us-sae-org.ezproxy.unibo.it/papers/development-high-accuracy-quick-light-off-nox-se nsor-2018-01-0334 (visited on 07/20/2024).

[53] P. Thompson, "Low-cost On-board ePM for Continuous Compliance Monitoring and Geofencing," en, Mar. 2022. [Online]. Available: https://www.cert.ucr.edu/sites/default /files/2022-04/2.4%20Low-Cost%20On-Board%20Electrostatic%20PM%20Sensing%20 for%20Continous%20Compliance%20Monitoring%20and%20Geofencing.pdf (visited on 07/22/2024).

[54] S. X. Ding, *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms and Tools* (Advances in Industrial Control), en. London: Springer, 2013, ISBN 978-1-4471-4798-5978-1-4471-4799-2. doi: 10.1007/978-1-4471-4799-2. [Online]. Available: https://link.springer .com/10.1007/978-1-4471-4799-2 (visited on 10/08/2023).

[55] K. Tidriri, N. Chatti, S. Verron, and T. Tiplica, "Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges," *Annual Reviews in Control*, vol. 42, pp. 63–81, Jan. 2016, ISSN 1367-5788. doi: 10.1016/j.arcontrol.2016.09.008. [Online]. Available: https://www.sciencedirect.com /science/article/pii/S1367578816300669 (visited on 07/23/2024).

[56] C. Sankavaram, A. Kodali, K. R. Pattipati, and S. Singh, "Incremental Classifiers for Data-Driven Fault Diagnosis Applied to Automotive Systems," *IEEE Access*, vol. 3, pp. 407–419, 2015, Conference Name: IEEE Access, ISSN 2169-3536. doi: 10.1109/ACCESS.2015.24 22833. [Online]. Available: https://ieeexplore.ieee.org/document/7089165 (visited on 10/05/2023).

[57] C. Keliris, M. M. Polycarpou, and T. Parisini, "A robust nonlinear observer-based approach for distributed fault detection of input–output interconnected systems," *Automatica*, vol. 53, pp. 408–415, Mar. 2015, ISSN 0005-1098. doi: 10.1016/j.automatica.2015.01.042. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109815000527 (visited on 08/13/2024).

[58] H. M. Odendaal and T. Jones, "Actuator fault detection and isolation: An optimised parity space approach," *Control Engineering Practice*, vol. 26, pp. 222–232, May 2014, ISSN 0967-0 661. doi: 10.1016/j.conengprac.2014.01.013. [Online]. Available: https://www.sciencedire ct.com/science/article/pii/S0967066114000434 (visited on 08/13/2024).

[59] R. Doraiswami, C. P. Diduch, and J. Tang, "A Diagnostic Model For Identifying Parametric Faults," *IFAC Proceedings Volumes*, 17th IFAC World Congress, vol. 41, no. 2, pp. 7369–7374, Jan. 2008, ISSN 1474-6670. doi: 10.3182/20080706-5-KR-1001.01246. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474667016401291 (visited on 08/13/2024).

[60] N. Chatti, B. Ould-Bouamama, A.-L. Gehin, and R. Merzouki, "Signed Bond Graph for multiple faults diagnosis," *Engineering Applications of Artificial Intelligence*, vol. 36, pp. 134–147, Nov. 2014, ISSN 0952-1976. doi: 10.1016/j.engappai.2014.07.018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197614001869 (visited on 08/13/2024).

[61] W. Li, H. Li, S. Gu, and T. Chen, "Process fault diagnosis with model- and knowledge-based approaches: Advances and opportunities," *Control Engineering Practice*, vol. 105, p. 104 637, Dec. 2020, ISSN 0967-0661. doi: 10.1016/j.conengprac.2020.104637. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967066120302070 (visited on 08/13/2024).

[62] Y. Qian, L. Xu, X. Li, L. Lin, *et al.*, "LUBRES: An expert system development and implementation for real-time fault diagnosis of a lubricating oil refining process," *Expert Systems with Applications*, vol. 35, no. 3, pp. 1252–1266, Oct. 2008, ISSN 0957-4174. doi: 10.1016/j.eswa.2007.07.061. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417407004186 (visited on 08/13/2024).

[63] B. He, X. Yang, T. Chen, and J. Zhang, "Reconstruction-based multivariate contribution analysis for fault isolation: A branch and bound approach," *Journal of Process Control*, vol. 22, no. 7, pp. 1228–1236, Aug. 2012, ISSN 0959-1524. doi: 10.1016/j.jprocont.2012.05.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959152412001254 (visited on 08/13/2024).

[64] M. T. Amin, F. Khan, and S. Imtiaz, "Fault detection and pathway analysis using a dynamic Bayesian network," *Chemical Engineering Science*, vol. 195, pp. 777–790, Feb. 2019, ISSN 0009-2509. doi: 10.1016/j.ces.2018.10.024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0009250918307371 (visited on 08/13/2024).

[65] M. Sayed-Mouchaweh, Ed., *Fault Diagnosis of Hybrid Dynamic and Complex Systems*, en. Cham: Springer International Publishing, 2018, ISBN 978-3-319-74013-3978-3-319-74014-0. doi: 10.1007/978-3-319-74014-0. [Online]. Available: http://link.springer.com/10.1007/978-3-319-74014-0 (visited on 10/06/2023).

[66] K. T. P. Nguyen, K. Medjaher, and D. T. Tran, "A review of artificial intelligence methods for engineering prognostics and health management with implementation guidelines," en, *Artificial Intelligence Review*, vol. 56, no. 4, pp. 3659–3709, Apr. 2023, ISSN 1573-7462. doi: 10.1007/s10462-022-10260-y. [Online]. Available: https://doi.org/10.1007/s10462-022-10260-y (visited on 10/06/2023).

[67] F. Harrou, M. N. Nounou, H. N. Nounou, and M. Madakyaru, "Statistical fault detection using PCA-based GLR hypothesis testing," *Journal of Loss Prevention in the Process Industries*, vol. 26, no. 1, pp. 129–139, Jan. 2013, ISSN 0950-4230. doi: 10.1016/j.jlp.2012.10.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950423012001581 (visited on 08/13/2024).

[68] A. Abid, M. T. Khan, and J. Iqbal, "A review on fault detection and diagnosis techniques: Basics and beyond," en, *Artificial Intelligence Review*, vol. 54, no. 5, pp. 3639–3664, Jun. 2021, ISSN 1573-7462. doi: 10.1007/s10462-020-09934-2. [Online]. Available: https://doi.org/10.1007/s10462-020-09934-2 (visited on 10/06/2023).

[69] R. Kumari, Sheetanshu, M. K. Singh, R. Jha, *et al.*, "Anomaly detection in network traffic using K-mean clustering," *2016 3rd International Conference on Recent Advances in Information Technology (RAIT)*, Mar. 2016, pp. 387–393. doi: 10.1109/RAIT.2016.7507933. [Online]. Available: https://ieeexplore.ieee.org/document/7507933 (visited on 08/13/2024).

[70] A. R. Sahu, S. K. Palei, and A. Mishra, "Data-driven fault diagnosis approaches for industrial equipment: A review," en, *Expert Systems*, vol. n/a, no. n/a, e13360, May 2023, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/exsy.13360, ISSN 1468-0394. doi: 10.1111/exsy.13360. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.13360 (visited on 09/26/2023).

[71] S. Ren, F. Si, and Y. Cao, "Development of Input Training Neural Networks for Multiple Sensor Fault Isolation," *IEEE Sensors Journal*, vol. 22, no. 15, pp. 14 997–15 009, Aug. 2022, Conference Name: IEEE Sensors Journal, ISSN 1558-1748. doi: 10.1109/JSEN.2022.3184078. [Online]. Available: https://ieeexplore-ieee-org.ezproxy.unibo.it/document/9805461 (visited on 07/26/2024).

[72] M. Hajji, M.-F. Harkat, A. Kouadri, K. Abodayeh, *et al.*, "Multivariate feature extraction based supervised machine learning for fault detection and diagnosis in photovoltaic systems," *European Journal of Control*, vol. 59, pp. 313–321, May 2021, ISSN 0947-3580. doi: 10.1016/j.ejcon.2020.03.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0947358019304054 (visited on 09/26/2023).

[73] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, Aug. 2018, ISSN 0888-3270. doi: 10.1016/j.ymssp.2018.02.016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327018300748 (visited on 10/06/2023).

[74] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Challenges and Opportunities of Deep Learning Models for Machinery Fault Detection and Diagnosis: A Review," *IEEE Access*, vol. 7, pp. 122 644–122 662, 2019, Conference Name: IEEE Access, ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2938227. [Online]. Available: https://ieeexplore.ieee.org/document/8819956 (visited on 07/27/2024).

[75] C. Hu, J. Wu, C. Sun, R. Yan, *et al.*, "Interinstance and Intratemporal Self-Supervised Learning With Few Labeled Data for Fault Diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 6502–6512, May 2023, Conference Name: IEEE Transactions on Industrial Informatics, ISSN 1941-0050. doi: 10.1109/TII.2022.3183601. [Online]. Available: https://ieeexplore.ieee.org/document/9797863 (visited on 10/09/2023).

[76] Y. Ji, S. Jeong, Y. Cho, H. Seo, *et al.*, "A Study on the Anomaly Detection of Engine Clutch Engagement/Disengagement Using Machine Learning for Transmission Mounted Electric Drive Type Hybrid Electric Vehicles," en, *Applied Sciences*, vol. 11, no. 21, p. 10 187, Jan. 2021, Number: 21 Publisher: Multidisciplinary Digital Publishing Institute, ISSN 2076-3417. doi: 10.3390/app112110187. [Online]. Available: https://www.mdpi.com/2076-3417/11/21/10187 (visited on 10/06/2023).

[77] J. A. Carino, M. Delgado-Prieto, J. A. Iglesias, A. Sanchis, *et al.*, "Fault Detection and Identification Methodology Under an Incremental Learning Framework Applied to Industrial Machinery," *IEEE Access*, vol. 6, pp. 49 755–49 766, 2018, Conference Name: IEEE Access, ISSN 2169-3536. doi: 10.1109/ACCESS.2018.2868430. [Online]. Available: https://ieeexplore.ieee.org/document/8454453 (visited on 10/05/2023).

[78] M. Ul Islam Khan, M. I. H. Pathan, M. Mominur Rahman, M. M. Islam, *et al.*, "Securing Electric Vehicle Performance: Machine Learning-Driven Fault Detection and Classification," *IEEE Access*, vol. 12, pp. 71 566–71 584, 2024, Conference Name: IEEE Access, ISSN 2169-3536. doi: 10.1109/ACCESS.2024.3400913. [Online]. Available: https://ieeexplore.ieee.org/document/10530324 (visited on 06/28/2024).

[79] M. Aishwarya and R. M. Brisilla, "Design and Fault Diagnosis of Induction Motor Using ML-Based Algorithms for EV Application," *IEEE Access*, vol. 11, pp. 34 186–34 197, 2023, Conference Name: IEEE Access, ISSN 2169‑3536. doi: 10.1109 / ACCESS.2023.3263588. [Online]. Available: https://ieeexplore-ieee-org.ezproxy.unibo.it/document/10089411 (visited on 07/26/2024).

[80] Z. Sun, Z. Wang, P. Liu, Z. Qin, *et al.*, "An Online Data-Driven Fault Diagnosis and Thermal Runaway Early Warning for Electric Vehicle Batteries," *IEEE Transactions on Power Electronics*, vol. 37, no. 10, pp. 12 636–12 646, Oct. 2022, Conference Name: IEEE Transactions on Power Electronics, ISSN 1941‑0107. doi: 10.1109/TPEL.2022.3173038. [Online]. Available: https://ieeexplore-ieee-org.ezproxy.unibo.it/document/9770404 (visited on 07/26/2024).

[81] X. Zhu, Y. Li, G. Yin, and R. J. Patton, "Interval Observer-Based Fault Detection and Isolation for Quadrotor UAV With Cable-Suspended Load," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–13, 2024, Conference Name: IEEE Transactions on Systems, Man, and Cybernetics: Systems, ISSN 2168‑2232. doi: 10.1109/TSMC.2024.3411315. [Online]. Available: https://ieeexplore-ieee-org.ezproxy.unibo.it/document/10580949 (visited on 07/26/2024).

[82] A. Naimi, J. Deng, P. Doney, A. Sheikh-Akbari, *et al.*, "Machine Learning-Based Fault Diagnosis for a PWR Nuclear Power Plant," *IEEE Access*, vol. 10, pp. 126 001–126 010, 2022, Conference Name: IEEE Access, ISSN 2169‑3536. doi: 10.1109 / ACCESS.2022.3225966. [Online]. Available: https://ieeexplore-ieee-org.ezproxy.unibo.it/document/9967986 (visited on 07/26/2024).

[83] M. Kordestani, M. E. Orchard, K. Khorasani, and M. Saif, "An Overview of the State of the Art in Aircraft Prognostic and Health Management Strategies," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–15, 2023, Conference Name: IEEE Transactions on Instrumentation and Measurement, ISSN 1557‑9662. doi: 10.1109/TIM.2023.3236342. [Online]. Available: https://ieeexplore-ieee-org.ezproxy.unibo.it/document/10015835 (visited on 07/26/2024).

[84] M. Cheliotis, I. Lazakis, and G. Theotokatos, "Machine learning and data-driven fault detection for ship systems operations," *Ocean Engineering*, vol. 216, p. 107 968, Nov. 2020, ISSN 0029-8018. doi: 10.1016/j.oceaneng.2020.107968. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801820309203 (visited on 10/06/2023).

[85] Y. S. Wang, N. N. Liu, H. Guo, and X. L. Wang, "An engine-fault-diagnosis system based on sound intensity analysis and wavelet packet pre-processing neural network," *Engineering Applications of Artificial Intelligence*, vol. 94, p. 103 765, Sep. 2020, ISSN 0952‑1976. doi: 10.1016/j.engappai.2020.103765. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197620301688 (visited on 10/06/2023).

[86] R. Ahmed, M. El Sayed, S. A. Gadsden, J. Tjong, *et al.*, "Automotive Internal-Combustion-Engine Fault Detection and Classification Using Artificial Neural Network Techniques," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 1, pp. 21–33, Jan. 2015, Conference Name: IEEE Transactions on Vehicular Technology, ISSN 1939-9359. doi: 10.1109/TVT.2014.2317736. [Online]. Available: https://ieeexplore.ieee.org/document/6803044 (visited on 10/05/2023).

[87] M. Khajavi, S. Nasiri, and A. Eslami, "Combined fault detection and classification of internal combustion engine using neural network," *Journal of Vibroengineering*, vol. 16, pp. 3912–3921, Dec. 2014.

[88] L. T. Francis, V. E. Pierozan, G. Gracioli, and G. M. de Araujo, "Data-driven Anomaly Detection of Engine Knock based on Automotive ECU," *2022 XII Brazilian Symposium on Computing Systems Engineering (SBESC)*, ISSN: 2324-7894, Nov. 2022, pp. 1–8. doi: 10.1109/SBESC56799.2022.9965059. [Online]. Available: https://ieeexplore.ieee.org/document/9965059 (visited on 10/05/2023).

[89] P. Wolf, A. Mrowca, T. T. Nguyen, B. Bäker, *et al.*, "Pre-ignition Detection Using Deep Neural Networks: A Step Towards Data-driven Automotive Diagnostics," *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, ISSN: 2153-0017, Nov. 2018, pp. 176–183. doi: 10.1109/ITSC.2018.8569908. [Online]. Available: https://ieeexplore.ieee.org/document/8569908 (visited on 10/05/2023).

[90] M. A. Maruf and A. Azim, "Anomaly Detection and Functional Testing for Automotive CAN Communication," *2024 IEEE International Systems Conference (SysCon)*, ISSN: 2472-9647, Apr. 2024, pp. 1–8. doi: 10.1109/SysCon61195.2024.10553463. [Online]. Available: https://ieeexplore.ieee.org/document/10553463?dld=dW5pYm8uaXQ%3D&source=SEARCHALERT (visited on 06/28/2024).

[91] S. Zhu, M. K. Tan, R. K. Y. Chin, B. L. Chua, *et al.*, "Engine Fault Diagnosis using Probabilistic Neural Network," *2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Sep. 2021, pp. 1–6. doi: 10.1109/IICAIET51634.2021.9573654. [Online]. Available: https://ieeexplore.ieee.org/document/9573654 (visited on 10/06/2023).

[92] P. K. Wong, J. Zhong, Z. Yang, and C. M. Vong, "Sparse Bayesian extreme learning committee machine for engine simultaneous fault diagnosis," *Neurocomputing*, vol. 174, pp. 331–343, Jan. 2016, ISSN 0925-2312. doi: 10.1016/j.neucom.2015.02.097. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231215011765 (visited on 10/05/2023).

[93] M. E. Mumcuoglu, S. M. Farea, M. Unel, S. Mise, *et al.*, "Fuel consumption classification for heavy-duty vehicles: A novel approach to identifying driver behavior and system anomalies," *2023 AEIT International Conference on Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, Jul. 2023, pp. 1–6. doi: 10.23919/AEITAUTOMOTIVE58986.2023.10217234. [Online]. Available: https://ieeexplore.ieee.org/document/10217234 (visited on 10/05/2023).

[94] A. Theissler, "Detecting known and unknown faults in automotive systems using ensemble-based anomaly detection," *Knowledge-Based Systems*, vol. 123, pp. 163–173, May 2017, ISSN 0950-7051. doi: 10.1016/j.knosys.2017.02.023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705117301077 (visited on 10/05/2023).

[95] M. Youssef and H. Ibrahim, "Automotive engine fault detection and isolation using LSTM for model-based residual sequence classification," *2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, ISSN: 2642-3766, Nov. 2021, pp. 1–6. doi: 10.1109/CCE53527.2021.9633053. [Online]. Available: https://ieeexplore.ieee.org/document/9633053 (visited on 10/05/2023).

[96] D. Jung and C. Sundström, "A Combined Data-Driven and Model-Based Residual Selection Algorithm for Fault Detection and Isolation," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 2, pp. 616–630, Mar. 2019, Conference Name: IEEE Transactions on Control Systems Technology, ISSN 1558-0865. doi: 10.1109/TCST.2017.2773514. [Online]. Available: https://ieeexplore.ieee.org/document/8122029 (visited on 10/05/2023).

[97] D. Jung, "Data-Driven Open-Set Fault Classification of Residual Data Using Bayesian Filtering," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 5, pp. 2045–2052, Sep. 2020, Conference Name: IEEE Transactions on Control Systems Technology, ISSN 1558-0865. doi: 10.1109/TCST.2020.2997648. [Online]. Available: https://ieeexplore.ieee.org/document/9106315 (visited on 07/01/2024).

[98] M. Cervantes-Bobadilla, J. García-Morales, Y. I. Saavedra-Benítez, J. A. Hernández-Pérez, *et al.*, "Multiple fault detection and isolation using artificial neural networks in sensors of an internal combustion engine," *Engineering Applications of Artificial Intelligence*, vol. 117, p. 105 524, Jan. 2023, ISSN 0952-1976. doi: 10.1016/j.engappai.2022.105524. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197622005140 (visited on 10/06/2023).

[99] M. Abboush, C. Knieke, and A. Rausch, "Intelligent Identification of Simultaneous Faults of Automotive Software Systems under Noisy and Imbalanced Data based on Ensemble LSTM and Random Forest," *IEEE Access*, pp. 1–1, 2023, Conference Name: IEEE Access, ISSN 2169-3536. doi: 10.1109/ACCESS.2023.3340865. [Online]. Available: https://ieeexplore.ieee.org/document/10348584?dld=dW5pYm8uaXQ%3D&source=SEARCHALERT (visited on 12/15/2023).

[100] A. Namigtle-Jiménez, R. Escobar-Jiménez, J. Gómez-Aguilar, C. García-Beltrán, *et al.*, "On-line ANN-based fault diagnosis implementation using an FPGA: Application in the EFI system of a vehicle," en, *ISA Transactions*, vol. 100, pp. 358–372, May 2020, ISSN 00190578. doi: 10.1016/j.isatra.2019.11.003. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0019057819304835 (visited on 07/01/2024).

[101] M. Vergara, L. Ramos, N. D. Rivera-Campoverde, and F. Rivas-Echeverría, "EngineFaultDB: A Novel Dataset for Automotive Engine Fault Classification and Baseline Results," *IEEE Access*, vol. 11, pp. 126 155–126 171, 2023, Conference Name: IEEE Access, ISSN 2169-3536. doi: 10.1109/ACCESS.2023.3331316. [Online]. Available: https://ieeexplore.ieee.org/document/10311597 (visited on 12/18/2023).

[102] M. Aliramezani, C. R. Koch, and M. Shahbakhti, "Modeling, diagnostics, optimization, and control of internal combustion engines via modern machine learning techniques: A review and future directions," *Progress in Energy and Combustion Science*, vol. 88, p. 100 967, Jan. 2022, ISSN 0360-1285. doi: 10.1016/j.pecs.2021.100967. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360128521000654 (visited on 07/29/2024).

[103] S. Estrada-Flores, I. Merts, B. De Ketelaere, and J. Lammertyn, "Development and validation of "grey-box" models for refrigeration applications: A review of key concepts," *International Journal of Refrigeration*, Issue with Special Emphasis on Data and Models on Food Refrigeration, vol. 29, no. 6, pp. 931–946, Sep. 2006, ISSN 0140-7007. doi: 10.1016/j.ijrefrig.2006.03.018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140700706000983 (visited on 08/05/2024).

[104] F. Baldi, G. Theotokatos, and K. Andersson, "Development of a combined mean value–zero dimensional model and application for a large marine four-stroke Diesel engine simulation," *Applied Energy*, vol. 154, pp. 402–415, Sep. 2015, ISSN 0306-2619. doi: 10.1016/j.apenergy.2015.05.024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261915006340 (visited on 07/30/2024).

[105] J. Tang, G. G. Zhu, and Y. Men, "Review of engine control-oriented combustion models," en, *International Journal of Engine Research*, vol. 23, no. 3, pp. 347–368, Mar. 2022, Publisher: SAGE Publications, ISSN 1468-0874. doi: 10.1177/1468087421992955. [Online]. Available: https://doi.org/10.1177/1468087421992955 (visited on 07/31/2024).

[106] "Mean Value Modeling of SI and Diesel Engines," en, *Vehicle Dynamics and Control*, R. Rajamani, Ed., Boston, MA: Springer US, 2006, pp. 257–285, ISBN 978-0-387-28823-9. doi: 10.1007/0-387-28823-6_9. [Online]. Available: https://doi.org/10.1007/0-387-28823-6_9 (visited on 07/30/2024).

[107] L. Guzzella and C. H. Onder, *Introduction to Modeling and Control of Internal Combustion Engine Systems*, en. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, ISBN 978-3-642-10774-0978-3-642-10775-7. doi: 10.1007/978-3-642-10775-7. [Online]. Available: http://link.springer.com/10.1007/978-3-642-10775-7 (visited on 10/11/2023).

[108] M. Tamborski, G. D'Errico, T. Lucchini, and A. Onorati, "Detailed prediction of HRR and NOx emissions in CI engines via a novel thermodynamic model with constant equivalence ratio zones," en, *International Journal of Engine Research*, vol. 24, no. 6, pp. 2315–2337, Jun. 2023, Publisher: SAGE Publications, ISSN 1468-0874. doi: 10.1177/14680874221128645. [Online]. Available: https://doi.org/10.1177/14680874221128645 (visited on 07/31/2024).

[109] X. Wu, K. Liu, Q. Liu, J. Fu, *et al.*, "Effects of direct water injection timings on knock suppression, combustion, and emission performance of high compression ratio hydrogen-enriched natural gas engine," *Energy Conversion and Management*, vol. 250, p. 114 887, Dec. 2021, ISSN 0196-8904. doi: 10.1016/j.enconman.2021.114887. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0196890421010633 (visited on 07/31/2024).

[110] A. H. Farajollahi, R. Firuzi, M. Rostami, and F. Bagherpor, "Consideration of the effects of increasing spray cone angle and turbulence intensity on heavy-duty diesel engine pollution and specific outputs using CFD," en, *International Journal of Engine Research*, vol. 24, no. 2, pp. 373–392, Feb. 2023, Publisher: SAGE Publications, ISSN 1468-0874. doi: 10.1177/14680874211052735. [Online]. Available: https://doi.org/10.1177/14680874211052735 (visited on 07/31/2024).

[111] Y. Ma, D. Yang, D. Xie, and J. E, "Investigating the effect of fuel properties and environmental parameters on low-octane gasoline-like fuel spray combustion and emissions using machine learning-global sensitivity analysis method," *Energy*, vol. 306, p. 132 551, Oct. 2024, ISSN 0360-5442. doi: 10.1016/j.energy.2024.132551. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544224023259 (visited on 08/03/2024).

[112] M. Mohammadi and E. Neshat, "Kinetic modeling of ion chemistry in diesel engines using a novel reduced ionic chemical mechanism," *Journal of the Taiwan Institute of Chemical Engineers*, vol. 152, p. 105 172, Nov. 2023, ISSN 1876-1070. doi: 10.1016/j.jtice.2023.105172. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1876107023005011 (visited on 07/31/2024).

[113] M. Mohammadi and E. Neshat, "Accurate prediction of NOx emissions from diesel engines considering in-cylinder ion current," *Environmental Pollution*, vol. 266, p. 115 347, Nov. 2020, ISSN 0269-7491. doi: 10.1016/j.envpol.2020.115347. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0269749120360358 (visited on 07/31/2024).

[114] F. Berni, V. Pessina, L. Teodosio, A. d'Adamo, *et al.*, "An integrated 0D/1D/3D numerical framework to predict performance, emissions, knock and heat transfer in ICEs fueled with NH3–H2 mixtures: The conversion of a marine Diesel engine as case study," *International Journal of Hydrogen Energy*, vol. 50, pp. 908–938, Jan. 2024, ISSN 0360-3199. doi: 10.1016/j.ijhydene.2023.09.158. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360319923047730 (visited on 07/31/2024).

[115] S. Hu, S. d'Ambrosio, R. Finesso, A. Manelli, *et al.*, "Comparison of Physics-Based, Semi-Empirical and Neural Network-Based Models for Model-Based Combustion Control in a 3.0 L Diesel Engine," en, *Energies*, vol. 12, no. 18, p. 3423, Jan. 2019, Number: 18 Publisher: Multidisciplinary Digital Publishing Institute, ISSN 1996-1073. doi: 10.3390/en12183423. [Online]. Available: https://www.mdpi.com/1996-1073/12/18/3423 (visited on 08/05/2024).

[116] S. d'Ambrosio, R. Finesso, G. Hardy, A. Manelli, *et al.*, "Model-Based Control of Torque and Nitrogen Oxide Emissions in a Euro VI 3.0 L Diesel Engine through Rapid Prototyping," en, *Energies*, vol. 14, no. 4, p. 1107, Jan. 2021, Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, ISSN 1996-1073. doi: 10.3390/en14041107. [Online]. Available: https://www.mdpi.com/1996-1073/14/4/1107 (visited on 08/05/2024).

[117] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY: Springer, 2000, ISBN 978-1-4419-3160-3978-1-4757-3264-1. doi: 10.1007/978-1-4757-3264-1. [Online]. Available: http://link.springer.com/10.1007/978-1-4757-3264-1 (visited on 08/05/2024).

[118] C. D. Jayaweera and N. Aziz, "Reliability of Principal Component Analysis and Pearson Correlation Coefficient, for Application in Artificial Neural Network Model Development, for Water Treatment Plants," en, *IOP Conference Series: Materials Science and Engineering*, vol. 458, no. 1, p. 012 076, Dec. 2018, Publisher: IOP Publishing, ISSN 1757-899X. doi: 10.1088/1757-899X/458/1/012076. [Online]. Available: https://dx.doi.org/10.1088/1757-899X/458/1/012076 (visited on 08/05/2024).

[119] Z. Yuan, X. Shi, D. Jiang, Y. Liang, *et al.*, "Data-Based Engine Torque and NOx Raw Emission Prediction," en, *Energies*, vol. 15, no. 12, p. 4346, Jan. 2022, Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, ISSN 1996-1073. doi: 10.3390/en15124346. [Online]. Available: https://www.mdpi.com/1996-1073/15/12/4346 (visited on 08/05/2024).

[120] N. Wei, Q. Zhang, Y. Zhang, J. Jin, *et al.*, "Super-learner model realizes the transient prediction of CO2 and NOx of diesel trucks: Model development, evaluation and interpretation," *Environment International*, vol. 158, p. 106 977, Jan. 2022, ISSN 0160-4120. doi: 10.1016/j.envint.2021.106977. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160412021006024 (visited on 07/29/2024).

[121] J. Liao, J. Hu, F. Yan, P. Chen, *et al.*, "A comparative investigation of advanced machine learning methods for predicting transient emission characteristic of diesel engine," *Fuel*, vol. 350, p. 128 767, Oct. 2023, ISSN 0016-2361. doi: 10.1016/j.fuel.2023.128767. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0016236123013807 (visited on 09/04/2023).

[122] C. Chandrashekar, P. Chatterjee, and D. S. Pawar, "Modeling real-world diesel car tailpipe emissions using regression-based approaches," *Transportation Research Part D: Transport and Environment*, vol. 128, p. 104 092, Mar. 2024, ISSN 1361-9209. doi: 10.1016/j.trd.2024.104092. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S136192092400 049X (visited on 08/02/2024).

[123] M. I. Ozmen, A. Yilmaz, C. Baykara, and O. A. Ozsoysal, "Modelling Fuel Consumption and NOx Emission of a Medium Duty Truck Diesel Engine With Comparative Time-Series Methods," *IEEE Access*, vol. 9, pp. 81 202–81 209, 2021, Conference Name: IEEE Access, ISSN 2169-3536. doi: 10.1109/ACCESS.2021.3082030. [Online]. Available: https://ieeexpl ore-ieee-org.ezproxy.unibo.it/document/9435333 (visited on 08/01/2024).

[124] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," en, *Future Internet*, vol. 15, no. 8, p. 255, Aug. 2023, Number: 8 Publisher: Multidisciplinary Digital Publishing Institute, ISSN 1999-5903. doi: 10.3390/fi1508025 5. [Online]. Available: https://www.mdpi.com/1999-5903/15/8/255 (visited on 08/05/2024).

[125] *Types of Machine Learning Models Explained*, en. [Online]. Available: https://it.mathworks.c om/discovery/machine-learning-models.html (visited on 08/05/2024).

[126] S. Khurana, S. Saxena, S. Jain, and A. Dixit, "Predictive modeling of engine emissions using machine learning: A review," en, *Materials Today: Proceedings*, 2nd International Conference on Future Learning Aspects of Mechanical Engineering, vol. 38, pp. 280–284, Jan. 2021, ISSN 2214-7853. doi: 10.1016/j.matpr.2020.07.204. [Online]. Available: https://www.sciencedir ect.com/science/article/pii/S2214785320352974 (visited on 11/18/2022).

[127] Y. Ge, P. Hou, T. Lyu, Y. Lai, *et al.*, "Machine Learning-Aided Remote Monitoring of NOx Emissions from Heavy-Duty Diesel Vehicles Based on OBD Data Streams," en, *Atmosphere*, vol. 14, no. 4, p. 651, Apr. 2023, Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, ISSN 2073-4433. doi: 10.3390/atmos14040651. [Online]. Available: https://www .mdpi.com/2073-4433/14/4/651 (visited on 05/24/2023).

[128] N. D. Rivera-Campoverde, J. L. Muñoz-Sanz, and B. d. V. Arenas-Ramirez, "Estimation of Pollutant Emissions in Real Driving Conditions Based on Data from OBD and Machine Learning," en, *Sensors*, vol. 21, no. 19, p. 6344, Jan. 2021, Number: 19 Publisher: Multidis-ciplinary Digital Publishing Institute, ISSN 1424-8220. doi: 10.3390/s21196344. [Online]. Available: https://www.mdpi.com/1424-8220/21/19/6344 (visited on 10/25/2022).

[129] J. Liao, J. Hu, P. Chen, L. Zhu, *et al.*, "Prediction of the transient emission characteristics from diesel engine using temporal convolutional networks," *Engineering Applications of Artificial Intelligence*, vol. 127, p. 107 227, Jan. 2024, ISSN 0952-1976. doi: 10.1016/j.engapp ai.2023.107227. [Online]. Available: https://www.sciencedirect.com/science/article/pii /S0952197623014112 (visited on 08/01/2024).

[130] Ü. Ağbulut, A. E. Gürel, and S. Sarıdemir, "Experimental investigation and prediction of performance and emission responses of a CI engine fuelled with different metal-oxide based nanoparticles–diesel blends using different machine learning algorithms," *Energy*, vol. 215, p. 119 076, Jan. 2021, ISSN 0360-5442. doi: 10.1016/j.energy.2020.119076. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544220321836 (visited on 07/30/2024).

[131] A. Falai and D. Misul, "Data-Driven Model for Real-Time Estimation of NOx in a Heavy-Duty Diesel Engine," English, *Energies*, vol. 16, no. 5, 2023, ISSN 1996-1073. doi: 10.3390/en16052125.

[132] N. Gruber and A. Jockisch, "Are GRU Cells More Specific and LSTM Cells More Sensitive in Motive Classification of Text?" English, *Frontiers in Artificial Intelligence*, vol. 3, Jun. 2020, Publisher: Frontiers, ISSN 2624-8212. doi: 10.3389/frai.2020.00040. [Online]. Available: https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2020.00040/full (visited on 08/06/2024).

[133] S. Liu, J. McGree, Z. Ge, and Y. Xie, "2 - Classification methods," *Computational and Statistical Methods for Analysing Big Data with Applications*, S. Liu, J. McGree, Z. Ge, and Y. Xie, Eds., San Diego: Academic Press, Jan. 2016, pp. 7–28, ISBN 978-0-12-803732-4. doi: 10.1016/B978-0-12-803732-4.00002-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128037324000027 (visited on 08/06/2024).

[134] M. Jean, P. Granier, and T. Leroy, "Experimental Evaluation of a Bayesian Estimation and Control of Engine Knocking Level," *IEEE Transactions on Control Systems Technology*, vol. 31, no. 4, pp. 1934–1940, Jul. 2023, Conference Name: IEEE Transactions on Control Systems Technology, ISSN 1558-0865. doi: 10.1109/TCST.2023.3247194. [Online]. Available: https://ieeexplore-ieee-org.ezproxy.unibo.it/document/10061704 (visited on 08/06/2024).

[135] B.-J. Yoon, "Hidden Markov Models and their Applications in Biological Sequence Analysis," *Current Genomics*, vol. 10, no. 6, pp. 402–415, Sep. 2009, ISSN 1389-2029. doi: 10.2174/138920209789177575. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2766791/ (visited on 08/06/2024).

[136] M. A. Rizvi, A. I. Bhatti, and Q. R. Butt, "Hybrid Model of the Gasoline Engine for Misfire Detection," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 8, pp. 3680–3692, Aug. 2011, Conference Name: IEEE Transactions on Industrial Electronics, ISSN 1557-9948. doi: 10.1109/TIE.2010.2090834. [Online]. Available: https://ieeexplore-ieee-org.ezproxy.unibo.it/document/5620975 (visited on 08/06/2024).

[137] H. Huang, Y. Song, X. Peng, S. X. Ding, *et al.*, "A Sparse Nonstationary Trigonometric Gaussian Process Regression and Its Application on Nitrogen Oxide Prediction of the Diesel Engine," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8367–8377, Dec. 2021, Conference Name: IEEE Transactions on Industrial Informatics, ISSN 1941-0050. doi: 10.1109/TII.2021.3068288. [Online]. Available: https://ieeexplore-ieee-org.ezproxy.unibo.it/document/9384285 (visited on 08/06/2024).

[138] C. Joerg, S.-Y. Lee, C. Reuber, J. Schaub, *et al.*, "Accurate Mean Value Process Models for Model-Based Engine Control Concepts by Means of Hybrid Modeling," Apr. 2019, pp. 2019–01–1178. doi: 10.4271/2019-01-1178. [Online]. Available: https://www.sae.org/content/2019-01-1178/ (visited on 10/14/2022).

[139] Z. Yang, D. Eddy, S. Krishnamurty, I. Grosse, *et al.*, "Investigating Grey-Box Modeling for Predictive Analytics in Smart Manufacturing," en, American Society of Mechanical Engineers Digital Collection, Nov. 2017. doi: 10.1115/DETC2017-67794. [Online]. Available: https://dx-doi-org.ezproxy.unibo.it/10.1115/DETC2017-67794 (visited on 08/07/2024).

[140] H. P. Selvam, S. Shekhar, and W. F. Northrop, "Prediction of nox Emissions from Compression Ignition Engines Using Ensemble Learning-Based Models with Physical Interpretability," English, ISSN: 0148-7191, 2688-3627, 0148-7191, 2688-3627, SAE International, Sep. 2021. doi: 10.4271/2021-24-0082. [Online]. Available: https://saemobilus-sae-org.ezproxy.unibo.it/papers/prediction-no-x-emissions-compression-ignition-engines-using-ensemble-learning-based-models-physical-interpretability-2021-24-0082 (visited on 08/07/2024).

[141] B. Jayaprakash, B. Wilmer, and W. F. Northrop, "Initial Development of a Physics-Aware Machine Learning Framework for Soot Mass Prediction in Gasoline Direct Injection Engines," en, Capri, Italy, Aug. 2023, pp. 2023–24–0174. doi: 10.4271/2023-24-0174. [Online]. Available: https://www.sae.org/content/2023-24-0174 (visited on 10/31/2023).

[142] A. Bajwa, G. Zou, F. Zhong, X. Fang, *et al.*, "Development of a semi-empirical physical model for transient NOx emissions prediction from a high-speed diesel engine," en, *International Journal of Engine Research*, p. 14 680 874 241 255 165, Jun. 2024, Publisher: SAGE Publications, ISSN 1468-0874. doi: 10.1177/14680874241255165. [Online]. Available: https://doi.org/10.1177/14680874241255165 (visited on 08/07/2024).

[143] M. Bidarvatan, V. Thakkar, M. Shahbakhti, B. Bahri, *et al.*, "Grey-box modeling of HCCI engines," *Applied Thermal Engineering*, vol. 70, no. 1, pp. 397–409, Sep. 2014, ISSN 1359-4311. doi: 10.1016/j.applthermaleng.2014.05.031. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1359431114003925 (visited on 08/07/2024).

[144] A. E. Campos-Ferreira, J. d. J. Lozoya-Santos, J. C. Tudon-Martinez, R. A. R. Mendoza, *et al.*, "Vehicle and Driver Monitoring System Using On-Board and Remote Sensors," en, *Sensors*, vol. 23, no. 2, p. 814, Jan. 2023, Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, ISSN 1424-8220. doi: 10.3390/s23020814. [Online]. Available: https://www.mdpi.com/1424-8220/23/2/814 (visited on 02/08/2023).

[145] Z. Lv, Y. Zhang, Z. Ji, F. Deng, *et al.*, "A real-time NOx emission inventory from heavy-duty vehicles based on on-board diagnostics big data with acceptable quality in China," *Journal of Cleaner Production*, vol. 422, p. 138 592, Oct. 2023, ISSN 0959-6526. doi: 10.1016/j.jclepro.2023.138592. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652623027506 (visited on 10/05/2023).

[146] R. Finesso, O. Marello, E. Spessa, Y. Yang, *et al.*, "Model-Based Control of BMEP and NOx Emissions in a Euro VI 3.0L Diesel Engine," English, *SAE International Journal of Engines*, vol. 10, no. 5, pp. 2288–2304, Sep. 2017, Publisher: SAE International, ISSN 1946-3936,1946-3944,1946-3936,1946-3944. doi: 10.4271/2017-24-0057. [Online]. Available: https://saemobilus-sae-org.ezproxy.unibo.it/articles/model-based-control-bmep-nox-emissions-a-euro-vi-30l-diesel-engine-2017-24-0057 (visited on 08/05/2024).

[147] R. Finesso, O. Marello, D. Misul, E. Spessa, *et al.*, "Development and Assessment of Pressure-Based and Model-Based Techniques for the MFB50 Control of a Euro VI 3.0L Diesel Engine," English, *SAE International Journal of Engines*, vol. 10, no. 4, pp. 1538–1555, Mar. 2017, Publisher: SAE International, ISSN 1946-3936,1946-3944,1946-3936,1946-3944. doi: 10.4271/2017-01-0794. [Online]. Available: https://saemobilus-sae-org.ezproxy.unibo.it/articles/development-assessment-pressure-based-model-based-techniques-mfb50-control-a-euro-vi-30l-diesel-engine-2017-01-0794 (visited on 08/08/2024).

[148] R. Finesso, E. Spessa, Y. Yang, R. Finesso, *et al.*, "Development and Validation of a Real-Time Model for the Simulation of the Heat Release Rate, In-Cylinder Pressure and Pollutant Emissions in Diesel Engines," English, *SAE International Journal of Engines*, vol. 9, no. 1, pp. 322–341, Jan. 2016, Publisher: SAE International, ISSN 1946-3936,1946-3944,1946-3936 ,1946-3944. doi: 10.4271/2015-01-9044. [Online]. Available: https://saemobilus.sae.org/articles/development-validation-a-real-time-model-simulation-heat-release-rate-cylinder-pressure-pollutant-emissions-diesel-engines-2015-01-9044 (visited on 08/08/2024).

[149] S. d'Ambrosio, R. Finesso, L. Fu, A. Mittica, *et al.*, "A control-oriented real-time semi-empirical model for the prediction of NOx emissions in diesel engines," *Applied Energy*, vol. 130, pp. 265–279, Oct. 2014, ISSN 0306-2619. doi: 10.1016/j.apenergy.2014.05.046. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261914005418 (visited on 08/03/2024).

[150] S. Duraiarasan, R. Salehi, A. Stefanopoulou, S. Mahesh, *et al.*, "Control-Oriented Physics-Based NOx Emission Model for a Diesel Engine With Exhaust Gas Recirculation," *ASME Letters in Dynamic Systems and Control*, vol. 1, no. 011008, Mar. 2020, ISSN 2689-6117. doi: 10.1115/1.4046450. [Online]. Available: https://doi.org/10.1115/1.4046450 (visited on 08/01/2024).

[151] C. Guardiola, B. Pla, D. Blanco-Rodriguez, and P. O. Calendini, "ECU-oriented models for NOx prediction. Part 1: A mean value engine model for NOx prediction," en, *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 229, no. 8, pp. 992–1015, Jul. 2015, Publisher: IMECHE, ISSN 0954-4070. doi: 10.1177/0954407014550191. [Online]. Available: https://doi.org/10.1177/0954407014550191 (visited on 08/05/2024).

[152] C. Guardiola, H. Climent, B. Pla, and D. Blanco-Rodriguez, "ECU-oriented models for NOx prediction. Part 2: Adaptive estimation by using an NOx sensor," en, *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 229, no. 10, pp. 1345–1360, Sep. 2015, Publisher: IMECHE, ISSN 0954-4070. doi: 10.1177/0954407014561278. [Online]. Available: https://doi.org/10.1177/0954407014561278 (visited on 08/05/2024).

[153] H. Zhong, K. Chen, C. Liu, M. Zhu, *et al.*, "Models for predicting vehicle emissions: A comprehensive review," *Science of The Total Environment*, vol. 923, p. 171324, May 2024, ISSN 0048-9697. doi: 10.1016/j.scitotenv.2024.171324. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969724014633 (visited on 08/03/2024).

[154] S. Makridakis, "The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms," *Futures*, vol. 90, pp. 46–60, Jun. 2017, ISSN 0016-3287. doi: 10.1016/j.futures.2017.03.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0016328717300046 (visited on 08/09/2024).

[155] A. M. Howlader, D. Patel, and R. Gammariello, "Data-driven approach for instantaneous vehicle emission predicting using integrated deep neural network," en, *Transportation Research Part D: Transport and Environment*, vol. 116, p. 103654, Mar. 2023, ISSN 1361-9209. doi: 10.1016/j.trd.2023.103654. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361920923000512 (visited on 04/04/2023).

[156] R. Liu, H.-d. He, Z. Zhang, C.-l. Wu, *et al.*, "Integrated MOVES model and machine learning method for prediction of CO2 and NO from light-duty gasoline vehicle," *Journal of Cleaner Production*, vol. 422, p. 138 612, Oct. 2023, ISSN 0959-6526. doi: 10.1016/j.jclepro.2023.1386 12. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095965262 3027701 (visited on 10/05/2023).

[157] J. Seo and S. Park, "Optimizing model parameters of artificial neural networks to predict vehicle emissions," en, *Atmospheric Environment*, vol. 294, p. 119 508, Feb. 2023, ISSN 1352-2310. doi: 10.1016/j.atmosenv.2022.119508. [Online]. Available: https://www.sciencedire ct.com/science/article/pii/S1352231022005738 (visited on 12/07/2022).

[158] L. Hu, C. Wang, Z. Ye, and S. Wang, "Estimating gaseous pollutants from bus emissions: A hybrid model based on GRU and XGBoost," *Science of The Total Environment*, vol. 783, p. 146 870, Aug. 2021, ISSN 0048-9697. doi: 10.1016/j.scitotenv.2021.146870. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969721019409 (visited on 07/29/2024).

[159] L. Yang, Y. Ge, L. Lyu, J. Tan, *et al.*, "Enhancing vehicular emissions monitoring: A GA-GRU-based soft sensors approach for HDDVs," *Environmental Research*, vol. 247, p. 118 190, Apr. 2024, ISSN 0013-9351. doi: 10.1016/j.envres.2024.118190. [Online]. Available: https://ww w.sciencedirect.com/science/article/pii/S001393512400094X (visited on 07/29/2024).

[160] Y. Yu, Y. Wang, J. Li, M. Fu, *et al.*, "A Novel Deep Learning Approach to Predict the Instantaneous NOx Emissions From Diesel Engine," *IEEE Access*, vol. 9, pp. 11 002–11 013, 2021, Conference Name: IEEE Access, ISSN 2169-3536. doi: 10.1109/ACCESS.2021.30 50165. [Online]. Available: https://ieeexplore.ieee.org/document/9317799 (visited on 07/29/2024).

[161] G. Sundaram, T. Gehra, J. Ulmen, M. Heubaum, *et al.*, "Modeling of Transient Gasoline Engine Emissions using Data-Driven Modeling Techniques," en, Detroit, Michigan, United States, Apr. 2023, pp. 2023–01–0374. doi: 10.4271/2023-01-0374. [Online]. Available: https://www.sae.org/content/2023-01-0374 (visited on 09/04/2023).

[162] K. Nikzadfar and A. H. Shamekhi, "An extended mean value model (EMVM) for control-oriented modeling of diesel engines transient performance and emissions," *Fuel*, vol. 154, pp. 275–292, Aug. 2015, ISSN 0016-2361. doi: 10.1016/j.fuel.2015.03.070. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001623611500366X (visited on 08/01/2024).

[163] M. A. Koç and R. Şener, "Prediction of emission and performance characteristics of reactivity-controlled compression ignition engine with the intelligent software based on adaptive neural-fuzzy and neural-network," *Journal of Cleaner Production*, vol. 318, p. 128 642, Oct. 2021, ISSN 0959-6526. doi: 10.1016/j.jclepro.2021.128642. [Online]. Available: h ttps://www.sciencedirect.com/science/article/pii/S0959652621028456 (visited on 08/01/2024).

[164] P. Sun, J. Zhang, W. Dong, D. Li, *et al.*, "Prediction of oxygen-enriched combustion and emission performance on a spark ignition engine using artificial neural networks," *Applied Energy*, vol. 348, p. 121 466, Oct. 2023, ISSN 0306-2619. doi: 10.1016/j.apenergy.2023.121466. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S030626192300 8309 (visited on 08/02/2024).

192

[165] X. Chen, S. Bai, G. Li, G. Wang, *et al.*, "A new transient NOx prediction model for diesel engine based on neural network model with Pelican Optimization Algorithm," en, *International Journal of Engine Research*, vol. 25, no. 1, pp. 231–239, Jan. 2024, Publisher: SAGE Publications, ISSN 1468-0874. doi: 10.1177/14680874231196050. [Online]. Available: https://doi.org/10.1177/14680874231196050 (visited on 08/02/2024).

[166] H.-T. Wen, J.-H. Lu, and D.-S. Jhang, "Features Importance Analysis of Diesel Vehicles' NOx and CO2 Emission Predictions in Real Road Driving Based on Gradient Boosting Regression Model," en, *International Journal of Environmental Research and Public Health*, vol. 18, no. 24, p. 13 044, Jan. 2021, Number: 24 Publisher: Multidisciplinary Digital Publishing Institute, ISSN 1660-4601. doi: 10.3390/ijerph182413044. [Online]. Available: https://www.mdpi.com/1660-4601/18/24/13044 (visited on 08/02/2024).

[167] C. Wu, Y. Pei, C. Liu, X. Bai, *et al.*, "Insights into the Fusion Correction Algorithm for On-Board NOx Sensor Measurement Results from Heavy-Duty Diesel Vehicles," en, *Energies*, vol. 16, no. 16, p. 6082, Jan. 2023, Number: 16 Publisher: Multidisciplinary Digital Publishing Institute, ISSN 1996-1073. doi: 10.3390/en16166082. [Online]. Available: https://www.mdpi.com/1996-1073/16/16/6082 (visited on 08/02/2024).

[168] W. Zhang and J. Li, "Research on fault diagnosis and signal reconstruction technology of diesel engine NOx sensor based on deep learning algorithm," en, *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 238, no. 5, pp. 1057–1068, Nov. 2022, Publisher: IMECHE, ISSN 0954-4070. doi: 10.1177/09544070221144658. [Online]. Available: https://doi.org/10.1177/09544070221144658 (visited on 08/02/2024).

[169] R. Pillai, V. Triantopoulos, A. S. Berahas, M. Brusstar, *et al.*, "Modeling and Predicting Heavy-Duty Vehicle Engine-Out and Tailpipe Nitrogen Oxide (NOx) Emissions Using Deep Learning," English, *Frontiers in Mechanical Engineering*, vol. 8, Mar. 2022, Publisher: Frontiers, ISSN 2297-3079. doi: 10.3389/fmech.2022.840310. [Online]. Available: https://www.frontiersin.org/journals/mechanical-engineering/articles/10.3389/fmech.2022.840310/full (visited on 08/02/2024).

[170] A. < Capancioni, "Development of predictive energy management strategies for hybrid electric vehicles supported by connectivity," en, Doctoral Thesis, Alma Mater Studiorum - Università di Bologna, Mar. 2022. doi: 10.48676/unibo/amsdottorato/10044. [Online]. Available: https://amsdottorato.unibo.it/10044/ (visited on 09/07/2024).

[171] L. < Brunelli, "Performance assessment of vehicular connectivity in hybrid electric vehicles for fuel and emissions reduction," en, Doctoral Thesis, Alma Mater Studiorum - Università di Bologna, Mar. 2023. doi: 10.48676/unibo/amsdottorato/10735. [Online]. Available: https://amsdottorato.unibo.it/10735/ (visited on 09/01/2024).

[172] S.-Y. Lee, J. Andert, C. Quérel, J. Schaub, *et al.*, "X-in-the-Loop-basierte Kalibrierung: HiL Simulation eines virtuellen Dieselantriebsstrangs," de, *Simulation und Test 2017*, J. Liebl and C. Beidl, Eds., ser. Proceedings, Wiesbaden: Springer Fachmedien, 2018, pp. 53–79, ISBN 978-3-658-20828-8. doi: 10.1007/978-3-658-20828-8_4.

[173] D. Blanco-Rodriguez, G. Vagnoni, S. Aktas, and J. Schaub, "Model-based Tool for the Efficient Calibration of Modern Diesel Powertrains," de, *MTZ worldwide*, vol. 77, no. 10, pp. 54–59, Oct. 2016, ISSN 2192-9114. doi: 10.1007/s38313-016-0103-5. [Online]. Available: https://doi.org/10.1007/s38313-016-0103-5 (visited on 10/04/2023).

[174] S. Canè, L. Brunelli, S. Gallian, A. Perazzo, *et al.*, "Performance assessment of a predictive pre-heating strategy for a hybrid electric vehicle equipped with an electrically heated catalyst," en, *Applied Thermal Engineering*, vol. 219, p. 119 341, Jan. 2023, ISSN 1359-4311. doi: 10.1016/j.applthermaleng.2022.119341. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1359431122012716 (visited on 04/06/2023).

[175] *SCALEXIO*, en. [Online]. Available: https://www.dspace.com/en/inc/home/products/hw/simulator_hardware/scalexio.cfm (visited on 09/07/2024).

[176] *MicroAutoBox II*, en. [Online]. Available: https://www.dspace.com/en/inc/home/products/hw/micautob/microautobox2.cfm (visited on 09/07/2024).

[177] R. P. Ltd, *Buy a Raspberry Pi 4 Model B*, en-GB. [Online]. Available: https://www.raspberrypi.com/products/raspberry-pi-4-model-b/ (visited on 07/17/2024).

[178] *PiCAN2 CAN-Bus Board for Raspberry Pi 2-3*, en. [Online]. Available: https://www.skpang.co.uk/products/pican2-can-bus-board-for-raspberry-pi-2-3 (visited on 07/17/2024).

[179] L. Brunelli, A. Capancioni, P. Gonnella, R. Casadio, *et al.*, "A Hybrid Vehicle Hardware-in-the-Loop System With Integrated Connectivity for Ehorizon Functions Validation," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 4340–4352, May 2021, Conference Name: IEEE Transactions on Vehicular Technology, ISSN 1939-9359. doi: 10.1109/TVT.2021.3073807.

[180] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer Series in Statistics). New York, NY: Springer, 2009, ISBN 978-0-387-84857-0978-0-387-84858-7. doi: 10.1007/978-0-387-84858-7. [Online]. Available: http://link.springer.com/10.1007/978-0-387-84858-7 (visited on 10/12/2023).

[181] A. A. A. Mohd Amiruddin, H. Zabiri, S. A. A. Taqvi, and L. D. Tufa, "Neural network applications in fault diagnosis and detection: An overview of implementations in engineering-related systems," en, *Neural Computing and Applications*, vol. 32, no. 2, pp. 447–472, Jan. 2020, ISSN 1433-3058. doi: 10.1007/s00521-018-3911-5. [Online]. Available: https://doi.org/10.1007/s00521-018-3911-5 (visited on 10/23/2023).

[182] *Classification Learner App - MATLAB & Simulink - MathWorks Italia*. [Online]. Available: https://it.mathworks.com/help/stats/classification-learner-app.html?s_tid=CRUX_lftnav (visited on 10/13/2023).

[183] Y. Rimal, N. Sharma, and A. Alsadoon, "The accuracy of machine learning models relies on hyperparameter tuning: Student result classification using random forest, randomized search, grid search, bayesian, genetic, and optuna algorithms," en, *Multimedia Tools and Applications*, vol. 83, no. 30, pp. 74 349–74 364, Sep. 2024, ISSN 1573-7721. doi: 10.1007/s11042-024-18426-2. [Online]. Available: https://doi.org/10.1007/s11042-024-18426-2 (visited on 09/14/2024).

[184] *Feature Selection and Feature Transformation Using Regression Learner App - MATLAB & Simulink - MathWorks Italia*. [Online]. Available: https://it.mathworks.com/help/stats/feature-selection-and-feature-transformation-using-regression-learner-app.html (visited on 09/19/2024).

[185] *Regression Learner App - MATLAB & Simulink - MathWorks Italia*. [Online]. Available: https://it.mathworks.com/help/stats/regression-learner-app.html (visited on 09/19/2024).

[186] *Emission Test Cycles: WLTC*. [Online]. Available: https://dieselnet.com/standards/cycles/wltp.php (visited on 10/16/2023).

[187] *Emission Test Cycles: SFTP-US06*. [Online]. Available: https://dieselnet.com/standards/c ycles/ftp_us06.php (visited on 10/16/2023).

[188] *Emission Test Cycles: FTP-72 (UDDS)*. [Online]. Available: https://dieselnet.com/standard s/cycles/ftp72.php (visited on 10/16/2023).

[189] *MATLAB Support Package for Raspberry Pi Hardware Documentation - MathWorks Italia*. [On-line]. Available: https://it.mathworks.com/help/supportpkg/raspberrypiio/ (visited on 07/17/2024).