A novel prior-informed machine learning model for diesel engine emission estimation

Fuel 407 (2026) 137435

Contents lists available at ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/jfue

Full length article

A novel prior-informed machine learning model for diesel engine emission estimation

Filippo Aglietti [a, b, *], Andrea Piano [a], Francesco Della Santa [c, d], Andrea Capra [b], Maria Pia Centini [b], Marcello Rimondi [b], Federico Millo [a]

[a] Energy Department, Politecnico di Torino, Corso Duca degli Abruzzi 24, Turin, 10129, Turin, Italy
[b] Dumarey Automotive Italia S.p.A., Corso Castelfidardo 36, Turin, 10129, Italy
[c] Department of Mathematical Sciences, Politecnico di Torino, Corso Duca degli Abruzzi 24, Turin, 10129, Turin, Italy
[d] Gruppo Nazionale per il Calcolo Scientifico, Istituto Nazionale di Alta Matematica, Piazzale Aldo Moro 5, Rome, 00185, Rome, Italy

HIGHLIGHTS

- Gradient-Informed NNs (GradINNs) applied to diesel emission modeling.
- GradINNs use prior beliefs on gradient smoothness to improve learning.
- Models trained and validated using experimental engine test bench data.
- GradINN outperforms MLP and GPR on NOx, CO, THC, and PM predictions.
- GradINN reaches benchmark accuracy with 25 % fewer experimental data.

ARTICLE INFO

Keywords:
Diesel emission modeling
Deep learning
Gradient prior beliefs

ABSTRACT

The increasing complexity of internal combustion engines, coupled with stringent emission regulations, has made virtualization essential for efficient and effective exploration of engine design spaces. This evolution demands predictive models that balance accuracy with computational efficiency, particularly when applied to emission estimation tasks. Data-driven approaches have been widely adopted in this field due to their flexibility and strong predictive capabilities. However, achieving high accuracy with these methods typically requires large training datasets.

To overcome this limitation, this study explores the application of Gradient-Informed Neural Networks (GradINNs) for diesel-engine emission prediction. GradINN combines a primary neural network, responsible for the emission estimates, with an auxiliary network that encodes prior beliefs about the gradients of the output with respect to the model's input parameters. A specialized loss function enforces consistency between the predicted gradients and these prior beliefs. The proposed model is benchmarked against traditional data-driven approaches, specifically Neural Networks (NNs) and Gaussian Process Regression (GPR), using data from both a Design of Experiments (DoE) campaign and an engine map.

Results demonstrate that GradINN consistently outperforms both benchmark methods across key emission targets, including nitrogen oxides, particulate matter, unburned hydrocarbons, and carbon monoxide. The proposed approach achieves lower prediction errors and improved generalization, notably maintaining comparable accuracy with up to 25 % fewer training samples compared to the best-performing benchmark, highlighting its potential to reduce experimental effort without compromising accuracy.

1. Introduction

The tightening of emission regulations has significantly increased the complexity of modern diesel engine systems, requiring calibration and optimization of numerous control parameters to comply with emission standards [1–3]. This complexity primarily arises from the integration of advanced after-treatment technologies and sophisticated combustion

* Corresponding author at: Energy Department, Politecnico di Torino, Corso Duca degli Abruzzi 24, Turin, 10129, Turin, Italy.
Email address: filippo.aglietti@dumarey.com (F. Aglietti).

https://doi.org/10.1016/j.fuel.2025.137435
Received 29 May 2025; Received in revised form 11 September 2025; Accepted 7 November 2025
Available online 18 November 2025

control strategies that are introduced to meet these stricter regulatory limits, adding additional degrees of freedom to the engine design space. As the number of parameters increases, the resulting large and multidimensional design space makes the identification of optimal configurations increasingly challenging. Traditional calibration approaches, heavily reliant on extensive experimental testing, are becoming less efficient and economically unsustainable due to high resource demands and long development cycles. Therefore, robust and accurate modeling techniques are essential to enable virtual calibration processes, significantly reducing the reliance on physical testing and consequently shortening development time and lowering associated costs [4–6]. Traditional modeling approaches have been extensively used to predict in-cylinder spray and combustion, which significantly influence engine fuel consumption and pollutant emissions [7–11]. These models are typically classified into three categories, listed by increasing complexity and computational cost: zero-dimensional thermodynamic, quasi-dimensional (phenomenological), and multidimensional models [12]. Although all approaches rely on experimental data, they differ considerably in the level of detail provided, making model selection dependent on analysis objectives. For example, multidimensional simulations provide detailed spatial and temporal resolution but their high computational demands limit their practicality in rapid design iterations and real-time applications [13].

To address these limitations, recent research has increasingly adopted data-driven modeling strategies. By leveraging large-scale datasets, these methods enable the construction of accurate predictive models of engine behavior and emissions [14] being particularly beneficial when conventional physics-based models struggle to capture complex phenomena, such as soot formation and oxidation [15], while also offering lower computational costs. These techniques enable efficient modeling of nonlinear relationships between engine parameters and emissions without detailed knowledge of the underlying physics [16–21]. Within this framework, Neural Networks (NNs) [22–29] and Gaussian Process Regression (GPR) [29–35] have demonstrated strong potential.

However, despite their computational efficiency during prediction, data-driven models generally require extensive datasets to achieve high accuracy. Collecting these datasets often demands significant experimental effort, which can partially offset the advantages of virtualization. A promising strategy to mitigate this dependency is integrating physical knowledge into the model training process. One established approach is prior-knowledge-based feature selection, where domain expertise guides the identification of the most relevant input features, reducing dimensionality and enhancing model generalization [36,37]. Physics-Informed Neural Networks (PINNs) provide another effective method, directly embedding physical equations into the training process [38]. PINNs enforce compliance with underlying physical laws alongside data fitting, improving accuracy and generalization, particularly when available data is limited. Such physics-informed approaches are particularly advantageous in scenarios with costly or limited high-resolution data acquisition [39–41]. Recent studies have successfully applied PINNs to modeling diesel-engine dynamics [42,43]. An alternative but complementary strategy involves leveraging information about gradient magnitude: models are trained not only using function values but also gradient information, thus improving accuracy and generalization capabilities with fewer training points [44–46].

However, in many engineering applications, prior information is either unavailable or cannot be easily expressed in a precise mathematical form, such as in diesel engine emission modeling, where the underlying processes are often too complex to formalize accurately. Although several trade-offs between engine control variables and emissions are well known [47,48], these relationships are generally qualitative rather than quantitative, as their exact impact varies considerably depending on specific engine architectures, combustion geometries, operating conditions, and application contexts. For example, increasing Exhaust Gas Recirculation is known to generally reduce nitrogen oxide emissions, but the precise quantitative effect depends heavily on combustion parameters, air management, and engine load conditions. Similarly, higher rail pressure typically reduces soot emissions by improving fuel atomization, yet the quantitative relationship is significantly influenced by factors such as injection timing, injection duration, and combustion chamber geometry. These complex, nonlinear dependencies make it challenging to directly incorporate precise physical constraints into machine learning models.

In such cases, model development relies on empirical data and general domain knowledge, such as the expectation that emission gradients with respect to input parameters exhibit smooth, non-oscillatory behavior without abrupt discontinuities. In this context, Gradient-Informed Neural Networks (GradINN) have been introduced as a new deep learning model capable of integrating such general prior hypotheses, i.e., prior beliefs, into the training of a NN, without requiring explicit mathematical formulations of physical relationships [49]. GradINN utilizes an auxiliary NN to impose soft constraints on the gradients of a primary NN, guiding the training process toward solutions consistent with expected physical behavior. By incorporating prior beliefs, GradINN aims to enhance the generalization of NN models and reduce their dependence on large experimental datasets. This increased data efficiency is particularly beneficial in engineering applications where collecting extensive experimental data is expensive and time-consuming, enabling effective model development even with limited available data.

Within this framework, this study compares standard neural networks and Gradient-Informed Neural Networks (GradINNs), which in this context are employed to enforce smoothness in the model gradients, for the prediction of diesel engine emissions, focusing on NO_x, PM, CO, and THC. In particular, the investigation examines how gradient-based regularization, introduced through smooth prior beliefs, affects model accuracy and data efficiency across these different emission types. Furthermore, the analysis includes a comparison with Gaussian Process Regression (GPR), another data-driven approach that incorporates prior beliefs via kernel function selection.

The rest of the paper is structured as follows: Section 2 describes the test case and data acquisition process. Section 3 presents the modeling approaches, including baseline Neural Networks, GradINN and Gaussian Process Regressions. Section 4 outlines the evaluation procedure adopted to compare model performance. Section 5 discusses the experimental results and model comparisons. Finally, Section 6 reports the main findings of the study.

2. Test case

The considered test case (Table 1) is a 4-cylinder, 2-L diesel engine designed for passenger vehicles, featuring a compression ratio of 15.5 (Fig. 1). The engine is equipped with a Variable Geometry Turbine (VGT) turbocharger and a Water Charge Air Cooler (WCAC). The fuel delivery system is based on a common rail architecture capable of reaching injection pressures up to 2000 bar. Additionally, the engine employs a dual-loop Exhaust Gas Recirculation (EGR) system, composed of high-pressure (HP) and low-pressure (LP) circuits. The HP EGR loop (uncooled) is primarily active during low-load conditions and engine warm-up phases, whereas the LP EGR loop is utilized under normal driving conditions.

Table 1
Engine technical specifications.

Parameter	Specification
Engine Type	4-cylinder diesel
Displacement	2.0 L
Compression Ratio	15.5
Turbocharger Type	VGT
Air Cooling System	WCAC
Fuel Injection System	Common rail
EGR System	Dual-loop (HP and LP)

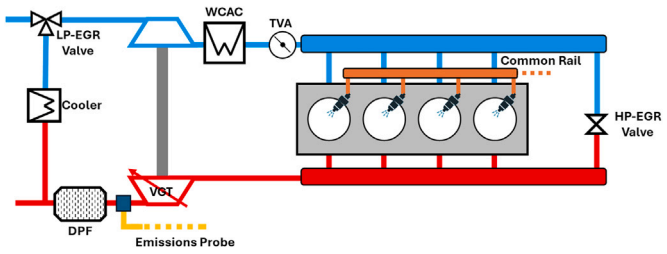

Fig. 1. Schematic of engine setup.

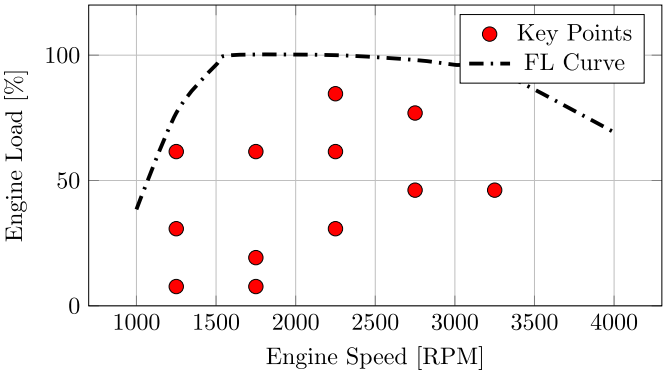

Fig. 2. DoE key-points distribution over engine map.

2.1. Test matrix

To train and evaluate the data-driven models, experimental datasets obtained from an experimental campaign were employed. This campaign consisted of 12 Design of Experiments (DoEs), each conducted at specific engine speed and load conditions to characterize engine behavior under warm operating scenarios. By combining the results of all 12 DoEs, a dataset of 1116 experimental measurements was obtained. Within each local DoE, additional control parameters were systematically varied to comprehensively explore the operating range.

The specific engine operating points for each local DoE were selected based on an operational time analysis of various representative driving cycles (i.e., NEDC and WLTC). A clustering approach was then applied to group similar operating conditions and identify key representative points [50]. Fig. 2 illustrates the distribution of selected key points across the engine map.

Table 2 summarizes the control parameters varied in each local DoE, which, along with engine speed and load, were used as input features for the data-driven models. The injection pattern employed in the tests included two pilot injections, one main injection, and one after-injection (denoted respectively as P_2, P_1, M, A). All injections were controlled both in terms of quantities and timings while the main injection quantity was adjusted to maintain the target engine load after setting the pilot and after-injection quantities. Injection timings were controlled directly for the main injection by setting the Start of Injection (SOI), whereas dwell times (DT), defined as the intervals between consecutive injections, were adjusted for pilot and after-injections. Additionally, the EGR system was regulated by defining the intake airflow setpoint and controlling the split ratio between High Pressure (HP) and Low Pressure (LP) EGR loops.

The range of control parameters for each local DoE was defined based on the characteristics of the corresponding operating point. An orthogonal sampling strategy was then applied within these ranges to generate the experimental grid. The distribution of the main control variables across the full experimental dataset is shown in Fig. 3. For the emissions targeted by this study, their distributions are shown in Fig. 4. These include NO_x, PM, THC, and CO. PM is measured in terms of Filter Smoke Number (FSN) and is referred to as Smoke throughout the paper. Both input and output variables are scaled by their respective maximum values and expressed as percentages (with the exception of the SOI of the main injection, which is reported as an absolute value).

Table 2
Parameters used as input features for the data-driven models. Engine speed and main injection quantity identify the key points, while the remaining parameters are control variables varied within each local DoE.

Key Point Parameters	
Eng Speed [RPM]	Engine speed
Q_M [mm^3]	M injection quantity (controlled to maintain load)
Control Parameters	
Boost [kPa]	Intake manifold pressure
M_{Air} [mg]	Air mass flow
P_{Rail} [MPa]	Fuel rail injection pressure
EGR Split [%]	Ratio between LP and HP EGR loops
SOI_M [°BTDC]	SOI of M
Q_A [mm^3]	A injection quantity
Q_{P1} [mm^3]	P_1 injection quantity
Q_{P2} [mm^3]	P_2 injection quantity
DT_A [ms]	DT between M and A
DT_{P1} [ms]	DT between P_1 and M injection
DT_{P2} [ms]	DT between P_2 and P_1 injection

3. Modeling approach

The objective of this study is to compare the predictive performance of the Gradient-Informed Neural Network (GradINN) against benchmark models commonly used for emission estimation. Specifically, the comparison includes: Multi-Layer Perceptron (MLP), i.e., a feed-forward fully-connected Neural Network, and Gaussian Process Regression (GPR). The following sections provide a description of the models considered in this study.

3.1. Neural networks and multi-layer perceptrons

Neural Networks (NNs) are widely employed for regression tasks due to their ability to approximate complex nonlinear functions [51]. In a regression task, given a target function $F : \mathbb{R}^d \to \mathbb{R}^m$, a NN trained for learning F is a parametric function $\widehat{F}(\cdot;\boldsymbol{\theta}) : \mathbb{R}^d \to \mathbb{R}^m$, where $\boldsymbol{\theta}$ is the vector that collects all the trainable parameters of the NN model. Given a training set $\{(\mathbf{x}_i, F(\mathbf{x}_i))\}_{i=1}^N$ with N samples, the NN learns to approximate F by optimizing $\boldsymbol{\theta}$ through the minimization of an appropriate loss function via Stochastic Gradient Descent [51]. For regression tasks, the Mean Squared Error (MSE) loss is typically considered (Eq. 1):

$$\mathcal{L}_{\text{data}}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{F}(\mathbf{x}_i;\boldsymbol{\theta}) - F(\mathbf{x}_i)\right)^2. \tag{1}$$

Among the various architectures proposed in the literature, this study focuses on Multi-Layer Perceptrons (MLPs) as commonlly used architecture for emission regression problems [22–28]. MLPs is a fully-connected feedforward neural networks consisting of an input layer followed by $H+1$ fully-connected (FC) layers, where the last one serves as the output layer. The input layer simply receives the input vector $\boldsymbol{x} \in \mathbb{R}^d$, while each FC layer performs a transformation defined as:

$$\text{FC}_{h+1}(\boldsymbol{x}^{(h)}) = \boldsymbol{\sigma}^{(h+1)}\left(W^{(h+1)}\boldsymbol{x}^{(h)} + \boldsymbol{b}^{(h+1)}\right) \tag{2}$$

with $h = 0, \dots, H$, where:

- $\boldsymbol{x}^{(h)} \in \mathbb{R}^{n_h}$ is the input to the layer, with $\boldsymbol{x}^{(0)} = \boldsymbol{x} \in \mathbb{R}^d$ (i.e., inputs layer).
- $W^{(h+1)} \in \mathbb{R}^{n_{h+1}\times n_h}$ and $\boldsymbol{b}^{(h+1)} \in \mathbb{R}^{n_{h+1}}$ are the trainable weights and biases of the layer (note that the collection of all such parameters across layers defines the global parameter vector $\boldsymbol{\theta}$).

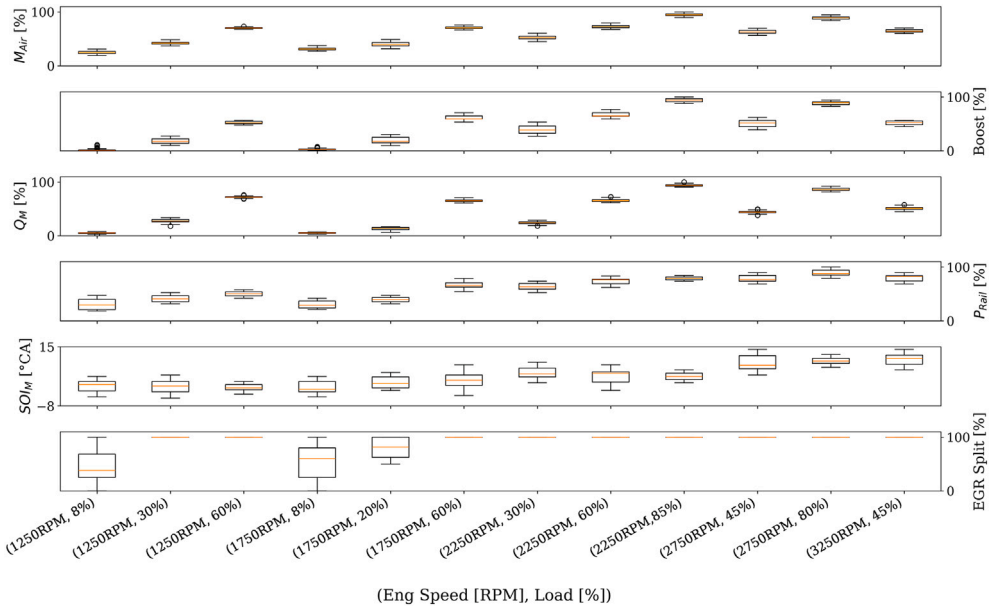

Fig. 3. Distribution of the main control variables across DoEs.

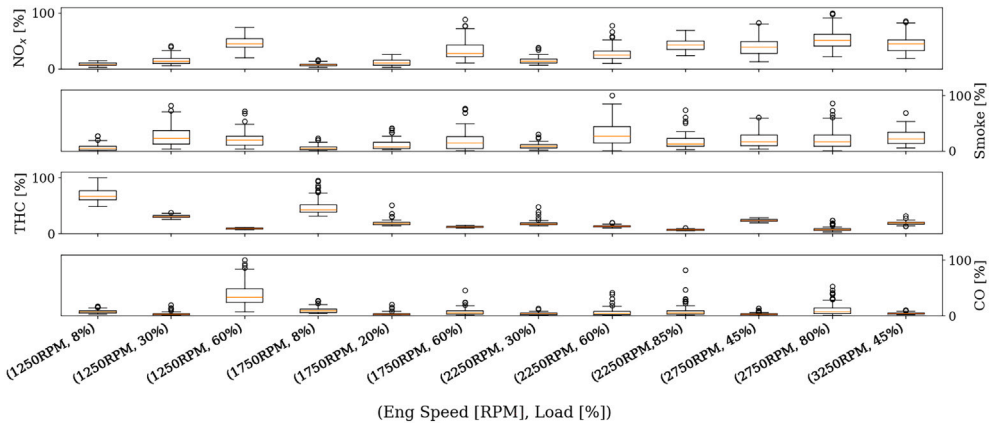

Fig. 4. Distribution of target emissions across DoEs.

- $\boldsymbol{\sigma}^{(h+1)}$ is the element-wise application of a nonlinear activation function (replaced in this study by a linear activation in the output layer).

Note that $h = 0$ corresponds to the first hidden layer and $h = H$ to the output layer.

A representation of a generic MLP architecture with H hidden layers, each containing n neurons, is provided in Fig. 5, where H denotes the network depth (hereafter referred to as D), and n the layer width (denoted as W).

3.2. Gaussian processes

Gaussian Process Regressors (GPRs) are a class of nonparametric probabilistic models widely used for regression tasks [52]. They define a prior over functions, assuming that any finite set of function values follows a multivariate Gaussian distribution. Specifically, we assume that the target function $F(\mathbf{x})$ follows a Gaussian Process prior:

$$F(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right), \tag{3}$$

where $m(\mathbf{x})$ is the mean function, representing the prior expectation of the function values (typically set to 0), and $k(\mathbf{x}, \mathbf{x}')$ is the covariance function, i.e., kernel, which defines the similarity between different inputs and encodes assumptions about smoothness and correlation in the data.

Given a training set $\{(\mathbf{x}_i, F(\mathbf{x}_i))\}_{i=1}^N$, the model learns a posterior distribution over functions conditioned on the observed data. The predictive distribution at a new input $\mathbf{x}^*$ is Gaussian with mean μ and variance σ^2 given by:

$$\mu(\mathbf{x}^*) = m(\mathbf{x}^*) + K_*^\top K^{-1}(\mathbf{F} - m(\mathbf{X})), \tag{4}$$

$$\sigma^2(\mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{x}^*) - K_*^\top K^{-1} K_*, \tag{5}$$

where $\mathbf{X} = \{(\mathbf{x}_i)\}_{i=1}^N$ and $\mathbf{F} = \{(F(\mathbf{x}_i))\}_{i=1}^N$ are respectively the inputs and outputs at the training points, the term K denotes the covariance matrix computed over $\mathbf{X}$, while K_* is the covariance vector between the new input $\mathbf{x}^*$ and the training data $\mathbf{X}$.

3.3. Gradient-informed neural networks

Gradient-Informed Neural Network (GradINN) is a novel deep learning approach that leverages prior beliefs about the target function's gradients to guide the training process [49]. Such beliefs can range

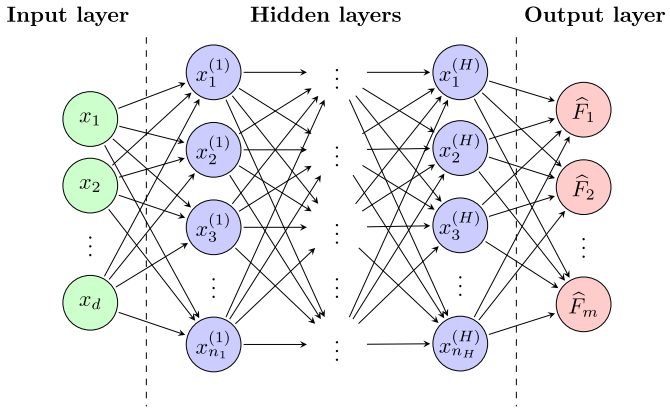

Fig. 5. Schematic representation of a generic MLP with d input features, m output units, and H hidden layers (i.e., depth $D = H$), each composed of n neurons (i.e., width $W = n$).

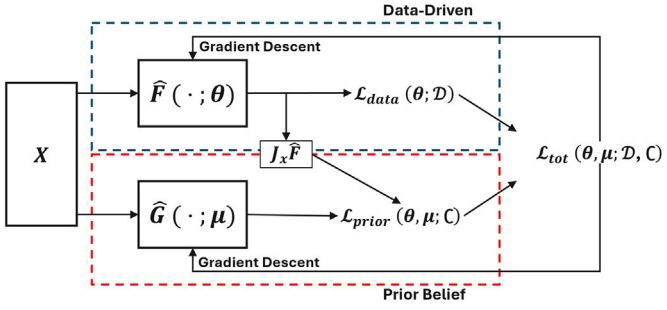

Fig. 6. Gradient-informed neural network and loss functions.

from smoothness assumptions to monotonicity or oscillatory behavior, thereby reflecting domain-specific knowledge of how the target function F should behave.

A GradINN model comprises two neural networks (Fig. 6):

- **Primary Network** $\widehat{F}(\cdot,\theta) : \mathbb{R}^n \to \mathbb{R}^m$ is responsible for approximating the target function $F(\mathbf{x})$.
- **Auxiliary Network** $\widehat{G}(\cdot,\mu) : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ encodes prior beliefs regarding the gradients of F with respect to $\mathbf{x}$, i.e., $\mathbf{J}_{\mathbf{x}}F$ in the case of a generic vector-valued output, while it reduces to $\nabla_{\mathbf{x}}F$ for a scalar output.

The prior belief is encoded into $\widehat{G}$ primarily through the choice of activation functions and neural network weights initialization. For instance, using a smooth activation function such as sigmoid or ELU, combined with a weight initialization strategy sampling from the Glorot uniform distribution, which produces an approximately constant output at initialization, incorporates a prior belief of smoothness in the gradients. This setup tends to penalize gradient spikes, discontinuities, and oscillations, guiding the primary network towards learning more stable and well-behaved gradient representations.

The networks are coupled via a *customized loss function* that enforces consistency between the gradients of the primary network and the expected gradient behavior embedded in the auxiliary network. In particular, the training process of a GradINN minimizes a total loss function defined as:

$$\mathcal{L}_{\text{tot}}(\theta,\mu;\mathcal{D},C) = \mathcal{L}_{\text{data}}(\theta;\mathcal{D}) + \mathcal{L}_{\text{prior}}(\theta,\mu;C), \tag{6}$$

where $\mathcal{L}_{\text{data}}$ is the standard supervised loss given in Eq. (1), evaluated on a training dataset $\mathcal{D} = \{(\bar{\mathbf{x}}_k, F(\bar{\mathbf{x}}_k))\}_{k=1}^{N}$, which consists of labeled data used to fit $\widehat{F}$ to the ground truth values. The term $\mathcal{L}_{\text{prior}}$ incorporates prior beliefs about the gradients by minimizing the discrepancy between the Jacobian of the primary network and the output of the auxiliary network at the collocation points $C = \{\bar{\mathbf{x}}_k\}_{k=1}^{N_C}$, i.e., unlabeled data where gradient constraints are imposed. Since these points do not require corresponding ground truth labels, they can be selected without the need for dedicated experiments.

$$\mathcal{L}_{\text{prior}}(\theta,\mu;C) = \frac{1}{N_C}\sum_{k=1}^{N_C}\left\|\mathbf{J}_{\mathbf{x}}\widehat{F}(\bar{\mathbf{x}}_k,\theta) - \widehat{G}(\bar{\mathbf{x}}_k,\mu)\right\|_F^2, \tag{7}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

During training, the parameters θ of the primary network $\widehat{F}$ are optimized to fit both the labeled data D and the gradient constraints imposed in C. Simultaneously, the auxiliary network $\widehat{G}$, governed by parameters μ, refines the prior gradient information to ensure consistency with the observed data (i.e., D).

By incorporating gradient priors into the learning process, GradINNs enhance predictive accuracy, improve generalization, and reduce reliance on large training datasets, making them particularly valuable in data-scarce scenarios.

4. Models evaluation procedure

The models evaluation process follows three main steps. First, a hyperparameter optimization is conducted for each model to determine the configuration that achieves the best performance within its respective class. This ensures a fair comparison by allowing each approach to reach its optimal predictive capabilities. Next, a comparison is performed using an additional engine map dataset, covering an engine operating range from 1250 to 2750 RPM. This evaluation provides insights into the generalization ability of each modeling approach under real operating conditions. Finally, the optimal model for each class is trained on datasets of varying sizes to assess the impact of data availability on the performance of these data-driven models.

4.1. Dataset

The data generated from the experimental activity described in Section 2.1 was used to construct the training and test datasets for the data-driven models. Specifically, the full dataset was first randomly split into two equal halves. One half was used for training and validation, while the other half was reserved for testing (see Fig. 7(a)). This choice ensures a robust performance evaluation, as a large testing dataset provides a comprehensive assessment of model performance across diverse operating conditions, reducing the risk of biased or over-optimistic results.

To evaluate how the performance of the models varies with the size of the training dataset, the training set was further partitioned using a random split strategy. Specifically, training subsets corresponding to 75 % and 50 % of the training set were considered (respectively, Fig. 7(b) and (c)), while the remaining data was discarded. For what concerns the testing dataset, it was kept fixed at 50 % without any modification

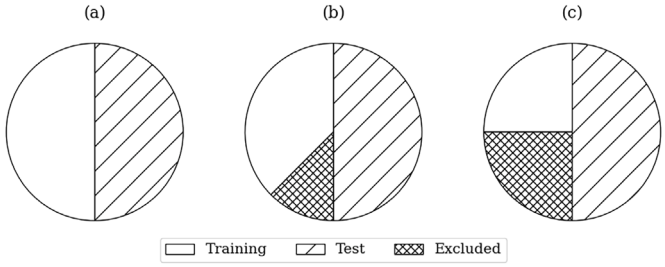

Fig. 7. Dataset splitting strategy for model evaluation: (a) Initial division into training and test sets, (b) and (c) Progressive reduction of the training set to 75 % and 50 % of its original size to assess data efficiency.

Table 3
DoE for NN hyperparameter optimization.

	Level 1	Level 2	Level 3
W	16	32	64
D	2	4	6
g	Sigmoid	ELU	–
bs	128	64	32
LR_0	0.001	0.01	–

Table 4
DoE for GradINN hyperparameter optimization.

	Level 1	Level 2	Level 3
$W_{\hat{F}}$	16	32	64
$D_{\hat{F}}$	2	4	6
$g_{\hat{F}}$	Sigmoid	ELU	–
$W_{\hat{G}}$	16	32	64
$D_{\hat{G}}$	2	4	–
$g_{\hat{G}}$	Sigmoid	ELU	–
bs	128	64	32
LR_0	0.001	0.01	–

Table 5
DoE for GPR hyperparameter optimization.

	Level 1	Level 2	Level 3	Level 4
Kernel	RBF	Matern ($\nu = 0.5$)	Matern ($\nu = 1.5$)	Matern ($\nu = 2.5$)
Length Scale	0.1	0.5	1.0	2.0
Alpha	10^{-6}	10^{-3}	10^{-2}	10^{-1}
Noise Level	10^{-3}	10^{-2}	10^{-1}	10^{0}

(Fig. 7). This consistency ensures a fair performance comparison across models trained with different dataset sizes, isolating the effect of training data availability on predictive accuracy.

The engine map dataset, used to test the models under real operating conditions, was obtained from a dedicated experimental campaign in which the engine calibration maps for the control parameters were fixed. A total of 127 operating points were tested over a uniform grid spanning 0 %–75 % of engine load and 1250–2750 rpm of engine speed.

For the GradINNs, the input variables of all points in the DoE were used as the collocation dataset. This choice follows from the nature of the experimental grid, which was defined a priori based on the number of control factors for each design of experiment, resulting in a standardized distribution of inputs across the operating space.

4.2. Hyperparameters optimization

The hyperparameter search was designed following a full-factorial approach, considering a set of predefined levels for each parameter.

The configurations explored for the NN and the GradINN are reported in Tables 3 and 4. For the network architecture, we explored the width (W) and depth (D) of the network, as well as the activation function (g). Regarding the training process, different mini-batch sizes (bs) were considered and a Reduce-on-Plateau schedule was adopted for the learning rate update. Specifically, the training starts with an initial learning rate (LR_0), which is reduced by half whenever the model performance does not improve for 100 consecutive epochs. To monitor training progress and avoid overfitting, 20 % of the training dataset was used as a validation set. This validation set was employed both for learning rate scheduling and for implementing early stopping, ensuring that training stops when performance no longer improves. The hyperparameter combinations tested for the standard NN model are the same as those considered for the primary network $\hat{F}$ in GradINN. In addition, for GradINN, different architectures of the auxiliary network $\hat{G}$ were explored, including its width ($W_{\hat{G}}$), depth ($D_{\hat{G}}$), and activation function ($g_{\hat{G}}$). Since the goal is to impose a prior belief of gradient smoothness in the emission estimation task, $\hat{G}$ employs smooth activation functions, and its weights are initialized using a Glorot scheme [49].

In total, the hyperparameter optimization involved 108 experiments for NN and 1296 experiments for GradINN, reflecting the additional complexity introduced by the auxiliary network.

For the GP models, the hyperparameter search focused on optimizing the kernel function and its associated parameters. The tested configurations are reported in Table 5. The choice of kernel plays a crucial role in determining the flexibility of the model, influencing both the smoothness of predictions and the ability to capture complex dependencies.

The GP hyperparameters include: the kernel type, where RBF (Radial Basis Function) and Matern kernels with different smoothness parameters (ν, with smaller values corresponding to less smooth functions) were tested; the length scale, which controls the spatial correlation between points; Alpha, representing the regularization parameter that helps prevent overfitting; and Noise Level, which accounts for the observation noise in the training data. In total, the GP hyperparameter optimization involved 256 experiments.

Unlike neural networks, Gaussian Process models do not require a validation dataset during training. Model parameters are optimized by maximizing the marginal likelihood, allowing the entire training set to be used without the need for early stopping or performance monitoring.

5. Results

The performance of the modeling approaches described in Section 3 was first assessed through a hyperparameter optimization study. Fig. 8 shows box plots of the root mean squared error (RMSE) values obtained from different hyperparameter configurations tested for GPR, MLP, and GradINN across all considered emission targets.

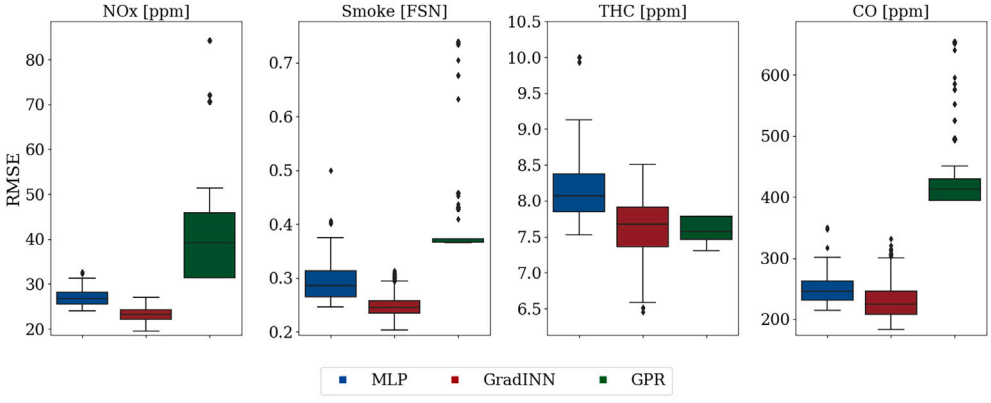

Fig. 8. Hyperparamenters optimization: RMSE box plot for GP, MLP and GradINN.

Table 6
Performance of the models in terms of RMSE and MRE for training and testing datasets. Best testing performance in bold.

		RMSE			MRE		
		MLP	GradINN	GPR	MLP	GradINN	GPR
Training	NO_x	13.4 ppm	10.7 ppm	21.8 ppm	4.7 %	4.0 %	7.9 %
	Smoke	0.12 FSN	0.05 FSN	0.27 FSN	20.8 %	7.8 %	52.6 %
	THC	6.3 ppm	1.5 ppm	5.2 ppm	13.9 %	3.5 %	10.6 %
	CO	95.9 ppm	94.2 ppm	348.3 ppm	18.6 %	16.1 %	48.7 %
Testing	NO_x	24.1 ppm	**19.6** ppm	31.4 ppm	7.1 %	**5.7** %	11.5 %
	Smoke	0.25 FSN	**0.20** FSN	0.37 FSN	33.9 %	**21.5** %	60.4 %
	THC	7.5 ppm	**6.4** ppm	7.3 ppm	15.2 %	**10.8** %	14.1 %
	CO	214.3 ppm	**188.9** ppm	394.7 ppm	38.2 %	**25.9** %	66.5 %

When comparing the best-performing model for each approach, it can be observed that for NOx, Smoke, and CO emissions, the optimal MLP models achieve lower RMSE values than the optimal GPR. Conversely, for THC emissions, the best GPR model shows slightly lower RMSE values than the best MLP model. This indicates that the relative performance of these approaches varies depending on the specific emission type being modeled.

In contrast, the optimal GradINN models, through the introduction of a smoothness prior belief, demonstrate significant improvement in predictive accuracy resulting in lower RMSE values than both the best GPR and MLP models across all emission types.

Table 6 presents a quantitative comparison of the errors achieved by the best-performing model in each class, reported in terms of RMSE and Mean Relative Error (MRE) for both training and testing datasets. The GradINN approach yields the lowest RMSE and MRE values across all emission types. Compared to MLP, GradINN achieves RMSE reductions of 18.7 % for NO_x, 11.9 % for CO, 20.0 % for Smoke and 14.7 % for THC. When compared to GPR models, the improvements are even more substantial with reductions of 37.6 % for NO_x, 52.1 % for CO, and 45.9 % for Smoke, while showing similar improvements of 12.3 % for THC.

An additional analysis was conducted on the error distribution across the test dataset for the best-performing models from the three model classes. As shown in Fig. 9, the test dataset was divided into three ranges based on the full range of each normalized emission, i.e., low (0 %–33 %), mid (33 %–66 %), high (66 %–100 %). Absolute prediction errors were then evaluated within each subrange. The analysis confirms the superior performance of the best GradINN configuration compared to the best MLP and GPR models, showing consistent improvements in error distribution across all three ranges and for all emissions considered. This global improvement indicates that the reduction in overall RMSE is not limited to specific emission level but reflects a broader enhancement in predictive accuracy.

To further assess the performance and evaluate their ability to predict emissions under real operating conditions, the three models analyzed in Fig. 9 were tested on the engine maps dataset. Fig. 10 shows a comparison of emission predictions where the first row displays the normalized ground truth (GT) emission distributions for NO_x, Smoke, THC, and CO. Note that the emission values of the engine map dataset are normalized with respect to the maximum of the corresponding emissions observed across the entire DoE campaign of Section 2.1. The subsequent rows illustrate the absolute prediction errors (computed as the difference between GT values and Model predictions) for the MLP, GradINN, and GPR models, respectively. For each emission type and model, the RMSE and MRE values are reported above the corresponding plots.

The results show that GradINN consistently provides the lowest prediction errors across the engine map for all emissions, while the GPR model exhibits substantially higher prediction errors compared to both MLP and GradINN. Despite the overall trend, all three models exhibit lower performance in specific operating regions, particularly in the mid-to-high RPM and low-load zones. In particular, the error increase is most pronounced for THC emission, where the emission gradients become significantly steeper and more difficult to be extrapolated. These challenging areas were underrepresented in the original DoE experimental campaign, leading to limited training information. This pattern illustrates that while GradINN successfully enhances generalization within well-represented domains through its smooth prior belief implementation, its effectiveness diminishes in regions where the underlying emission behavior lacks sufficient characterization due to insufficient training data.

As a final step, the data efficiency was investigated. This analysis was conducted by progressively reducing the training dataset size from 100 % to 50 % (Fig. 7). The hyperparameters of each model class corresponded to the best-performing configurations identified during the hyperparameter optimization process. Fig. 11 illustrates the RMSE and MRE over the DoE test dataset as a function of training dataset size for MLP, GradINN, and GPR models. The results show that GradINN maintains its superior performance across all tested training dataset sizes. From a practical implementation perspective, the achievable dataset reduction is limited by the emission type with the highest data requirements. When comparing GradINN to MLP it is possible to see how, for NO_x, THC, and CO emissions, GradINN maintains performance equivalent to full-dataset MLP models even when trained on less than 75 % of the data. However, Smoke emissions represent the constraining factor, limiting the achievable reduction to approximately 25 % of the original dataset. This increase in data efficiency demonstrated by GradINN offers significant flexibility in experimental testing strategies. In particular, this gain can be leveraged in two main ways: either as a direct reduction in experimental testing time and associated resources, or alternatively, as an opportunity to redistribute the testing points toward

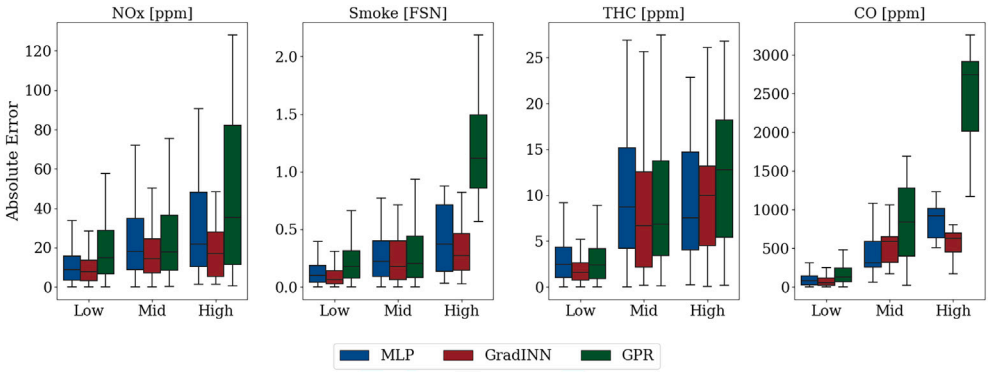

Fig. 9. Absolute error distribution across three emission intensity ranges (low: 0 %–33 %, mid: 33 %–66 %, high: 66 %–100 %) for the best-performing models.

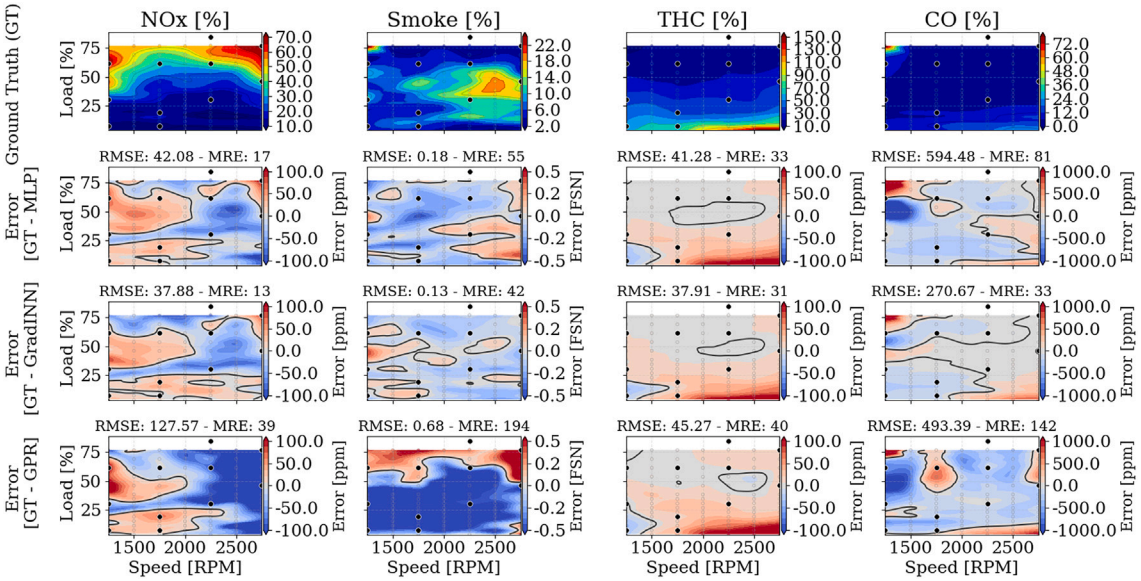

Fig. 10. Comparison of model prediction performance across the engine map. The top row shows normalized Ground Truth (GT) plotted against engine speed [RPM] and Load [%]. The subsequent rows display the absolute prediction errors for MLP, GradINN, and GPR models with their respective RMSE and MRE values. In all the plots, the black dots represents the DoE key-points.

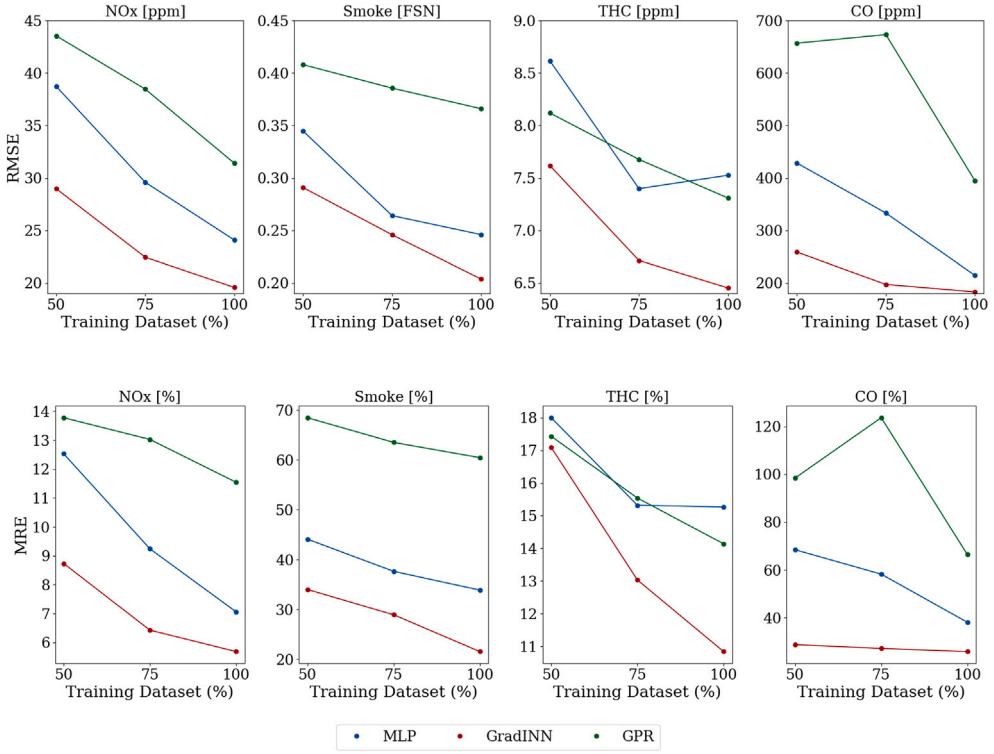

Fig. 11. Prediction error (RMSE and MRE) as a function of training dataset size for MLP, GradINN, and GPR.

less explored regions of the operating domain. The latter approach could further enhance model performance in challenging regions such as the mid-to-high RPM and low-load zones identified in the engine map validation, potentially addressing the limitations observed in these areas while maintaining the same overall experimental effort.

6. Conclusion

This work focused on comparing benchmark machine learning models, i.e., Multi-Layer Perceptrons (MLPs) and Gaussian Process Regression (GPR), with a novel Gradient-Informed Neural Network (GradINN) model for predicting key pollutants emitted by a diesel

engine, including Nitrogen Oxides, Particulate Matter, Hydrocarbons, and Carbon Monoxide. GradINN employs two neural networks: a primary network responsible for emission prediction, and an auxiliary network that encodes smoothness prior beliefs about the gradient of the target function. This auxiliary network guides training by encouraging consistency between prior belief and gradient of the primary network, thus driving the model toward more physically meaningful solutions.

To evaluate the best achievable performance, a dedicated hyperparameter optimization was conducted for each model using a dataset derived from a Design of Experiments testing campaign, which systematically explored the engine's operating space. Under these conditions, the best-performing configurations of GradINN, MLP, and GPR were compared. GradINN consistently outperformed the other models across all emission targets, achieving lower prediction errors.

To further validate the models under real-world operating conditions, the best-performing configurations were evaluated on a complete engine map dataset covering a wide range of speed and load scenarios. GradINN again demonstrated superior predictive capabilities, outperforming the other models across all emissions. However, performance degradation was observed in specific regions of the engine map, particularly at mid-to-high RPM and low-load conditions. These areas were underrepresented in the training data, confirming that while GradINN benefits from smooth gradient priors, it still requires sufficient input coverage to maintain accuracy across the entire domain.

Since GradINN consistently demonstrated superior predictive performance, the study further investigated its data efficiency. Specifically, GradINN maintained high accuracy even when trained on reduced datasets, achieving performance comparable to MLPs with up to 25 % fewer training data points. This increased data efficiency provides a significant practical advantage, as it allows experimental effort to be reduced or redirected towards underrepresented regions of the operating space.

CRediT authorship contribution statement

Filippo Aglietti: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andrea Piano:** Writing – review & editing, Validation, Supervision, Investigation. **Francesco Della Santa:** Writing – review & editing, Supervision. **Andrea Capra:** Writing – review & editing, Visualization, Supervision. **Maria Pia Centini:** Writing – review & editing, Visualization, Investigation. **Marcello Rimondi:** Writing – review & editing, Supervision, Resources, Project administration. **Federico Millo:** Writing – review & editing, Validation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

F.D. acknowledges that this study was carried out within the FAIR-Future Artificial Intelligence Research and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)–MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3—D.D. 1555 11/10/2022, PE00000013). F.D. acknowledges support from Italian MIUR PRIN project 20227K44ME, Full and Reduced Order Modeling of Coupled Systems: focus on Non-Matching Methods and Automatic Learning (FaReX).

Data availability

The authors do not have permission to share data.

References

[1] Caroca JC, Millo F, Vezza D, Vlachos T, Filippo A, Bensaid S, Russo N, Fino D. Detailed investigation on soot particle size distribution during DPF regeneration, using standard and bio-diesel fuels. Am Chem Soc 2010;50(5):2650–8. https://doi.org/10.1021/ie1006799

[2] Wang X, Zhang S, Bechkoum K. Model-based multi-critical optimisation of combustion engine fuel consumption and emissions. IOP Publishing 2019;517:12005. https://doi.org/10.1088/1757-899x/517/1/012005.

[3] Johnson T. Vehicular emissions in review. SAE Int J Engines Apr 2016;9. https://doi.org/10.4271/2016-01-0919

[4] Millo F, Piano A, Zanelli A, Boccardo G, Rimondi M, Fuso R. Development of a fully physical vehicle model for off-line powertrain optimization: a virtual approach to engine calibration. SAE Tech Pap Ser 2021. https://api.semanticscholar.org/CorpusID:239637485.

[5] Finesso R, Marello O, Spessa E, Yang Y, Hardy G. Model-based control of bmep and NOx emissions in a EURO VI 3.0l diesel engine. SAE Int J Engines 2017;10(5):2288–304. https://www.jstor.org/stable/26422614.

[6] Pal A, Zhu L, Wang Y, Zhu G. Data-driven model-based calibration for optimizing electrically boosted diesel engine performance. Int J Engine Res 2023;24(4):1515–29. https://doi.org/10.1177/14680874221090307

[7] Piano A, Scalambro A, Millo F, Catapano F, Sementa P, Di Iorio S, Bianco A. CFD-based methodology for the characterization of the combustion process of a passive pre-chamber gasoline engine. Transp Eng 2023;13:100200. https://doi.org/10.1016/j.treng.2023.100200

[8] D'Errico G, Lucchini T, Onorati A, Hardy G. Computational fluid dynamics modeling of combustion in heavy-duty diesel engines. Int J Engine Res 2015;16 (1) (112–24. https://doi.org/10.1177/1468087414561276

[9] Posch S, Gößnitzer C, Lang M, Novella R, Steiner H, Wimmer A. Turbulent combustion modeling for internal combustion engine CFD: a review. Prog Energy Combust Sci 2025;106:101200. https://doi.org/10.1016/j.pecs.2024.101200

[10] Barth M, An F, Norbeck JM, Ross M. Modal emissions modeling: a physical approach. SAGE Publishing 1996;1520(1):81–8. https://doi.org/10.1177/0361198196152000110

[11] Dempsey AB, Seiler PJ, Svensson KI, Qi Y. A comprehensive evaluation of diesel engine CFD modeling predictions using a semi-empirical soot model over a broad range of combustion systems. SAE Int J Engines 2018;11(6):1399–420. https://www.jstor.org/stable/26649166.

[12] Stiesch G. Modeling engine spray and combustion processes; 2003. https://api.semanticscholar.org/CorpusID:92902423

[13] Richard S, Bougrine S, Font G, Francois L, Berr F. On the reduction of a 3d CFD combustion model to build a physical 0d model for simulating heat release, knock and pollutants in SI engines. Oil Gas Sci Technol Rev IFP 2009;64:223–42. https://doi.org/10.2516/ogst/2008055

[14] Sediako A, Andrić J, Sjöblom J, Faghani E. Heavy duty diesel engine modeling with layered artificial neural network structures; Apr 2018. https://doi.org/10.4271/2018-01-0870

[15] Zhao F, Yang W, Yu W. A progress review of practical soot modelling development in diesel engine combustion. J Traffic Transp Eng (English Ed) 2020;7(3):269–81, Special Issue: Clean Alternative Fuels for Transport Vehicles. doi:https://doi.org/10.1016/j.jtte.2020.04.002. https://www.sciencedirect.com/science/article/pii/S2095756420300611.

[16] Rahimi M, Moosavi SM, Smit B, Hatton TA. Toward smart carbon capture with machine learning. Elsevier BV 2021;2(4):100396. https://doi.org/10.1016/j.xcrp.2021.100396

[17] Canal R, Riffel F, Gracioli G. Machine learning for real-time fuel consumption prediction and driving profile classification based on ecu data. IEEE Access 2024;PP:1. https://doi.org/10.1109/ACCESS.2024.3400933

[18] Imdadul Alam GM, Tanim S, Sarker S, Watanobe Y, Islam R, Mridha M, Nur K. Deep learning model based prediction of vehicle CO2 emissions with explainable AI integration for sustainable environment. Sci Rep Jan 2025;15. https://doi.org/10.1038/s41598-025-87233-y

[19] Liao J, Hu J, Yan F, Chen P, Zhu L, Zhou Q, Xu H, Li J. A comparative investigation of advanced machine learning methods for predicting transient emission characteristic of diesel engine. Fuel 2023;350:128767. https://doi.org/10.1016/j.fuel.2023.128767. https://www.sciencedirect.com/science/article/pii/S0016236123013807.

[20] Khurana S, Saxena S, Jain S, Dixit A. Predictive modeling of engine emissions using machine learning: a review. Mater Today Proc 2021;38:280–4, 2nd International Conference on Future Learning Aspects of Mechanical Engineering. doi:https://doi.org/10.1016/j.matpr.2020.07.204.

[21] Zhong H, Chen K, Liu C, Zhu M, Ke R. Models for predicting vehicle emissions: a comprehensive review. Sci Total Environ 2024;923:171324. https://doi.org/10.1016/j.scitotenv.2024.171324

[22] Thompson G, Atkinson CM, Clark N, Long TW, Hanzevack E. Technical note: neural network modelling of the emissions and performance of a heavy-duty diesel engine. SAGE Publishing 2000;214(2):111–26. https://doi.org/10.1177/095440700021400201

[23] Li H, Butts K, Zaseck K, Liao-McPherson D, Kolmanovsky I. Emissions modeling of a light-duty diesel engine for model-based control design using multi-layer perceptron neural networks. IFAC-Papersonline Mar 2017. https://doi.org/10.4271/2017-01-0601

[24] Wang MY, Zhang SJ, Zhang X. Prediction of pahs emitted from marine diesel engine using artificial neural networks combining genetic algorithms. Trans Tech Publications 2014:1233–6. www.scientific.net/amm.599-601.1233.

[25] Pennycott A, Zhang Q, Brace C, Burke R, Akehurst S. An empirical model for the carbon dioxide emissions of a diesel engine. SAGE Publishing 2012;226(11):1507–13. https://doi.org/10.1177/0954407012444524
[26] Zhang J, Li X, Amini MR, Kolmanovsky I, Tsutsumi M, Nakada H. Modeling and control of diesel engine emissions using multi-layer neural networks and economic model predictive control. Elsevier BV 2023;56(2):10696–702. https://doi.org/10.1016/j.ifacol.2023.10.724
[27] Odufuwa OY, Tartibu LK, Kusakana K. Artificial neural network modelling for predicting efficiency and emissions in mini-diesel engines: key performance indicators and environmental impact analysis. Fuel 2025;387:134294. https://doi.org/10.1016/j.fuel.2025.134294
[28] Samuel PM, Gnanamoorthi V, Purushothaman P, Gurusamy A, Gobalakichenin D. Prediction efficiency of artificial neural network for CRDI engine output parameters. Transp Eng 2021;3:100041. https://doi.org/10.1016/j.treng.2020.100041
[29] Yuan H, Goyal H, Islam R, Giles K, Howson S, Lewis A, Parsons D, Esposito S, Akehurst S, Jones P, McAllister M, Littlefair B, Lu Z, Zhu S. Thermodynamics-based data-driven combustion modelling for modern spark-ignition engines. Energy Dec. 2024;313. https://doi.org/10.1016/j.energy.2024.134074
[30] Shi X, Jiang D, Qian W, Liang Y. Application of the Gaussian process regression method based on a combined kernel function in engine performance prediction. ACS Omega Nov 2022;7. https://doi.org/10.1021/acsomega.2c05952
[31] Pal A, Wang Y, Zhu L, Zhu G. Engine calibration optimization based on ITS surrogate models, In: Dynamic Systems and Control Conference; 2019. https://doi.org/10.1115/DSCC2019-8984
[32] Zinage S, Bilionis I, Meckl P. A causal graph-enhanced Gaussian process regression for modeling engine-out NOx. 2024; arXiv:2410.18424.
[33] Shi X, Jiang D, Qian W, Liang Y. Application of the Gaussian process regression method based on a combined kernel function in engine performance prediction. ACS Omega 2022;7(45):41732–43. https://doi.org/10.1021/acsomega.2c05952
[34] Pan T, Cai Y, Chen S. Development of an engine calibration model using Gaussian process regression. Int J Automot Technol 2021;22(2):327–34. https://doi.org/10.1007/s12239-021-0031-5
[35] Castric S. Modeling pollutant emissions of diesel engine based on kriging models: a comparison between geostatistic and Gaussian process approach. IFAC Proc Volumes 2012;45:1708–15. https://doi.org/10.3182/20120523-3-RO-2023.00038
[36] Zhang L, Wen J. A systematic feature selection procedure for short-term data-driven building energy forecasting model development. Elsevier BV 2018;183:428–42. https://doi.org/10.1016/j.enbuild.2018.11.010
[37] Boccardo G, Piano A, Zanelli A, Babbi M, Cambriglia L, Mosca S, Millo F. Development of a virtual methodology based on physical and data-driven models to optimize engine calibration. Transp Eng 2022.
[38] Diab W, Chaabi O, Alkobaisi S, Awotunde AA, Kobaisi MA. Learning generic solutions for multiphase transport in porous media via the flux functions operator. Cornell University; Jan 2023. https://doi.org/10.48550/arxiv.2307.01354.
[39] Mattey R, Ghosh S. A Physics informed neural network for time-dependent nonlinear and higher order partial differential equations. Cornell University; Jun 2021. http://export.arxiv.org/pdf/2106.07606.
[40] Liu W, Pyrcz MJ. Physics-informed graph neural network for spatial-temporal production forecasting. Cornell University; Jan 2022. https://doi.org/10.48550/arxiv.2209.11885.
[41] Chen W, Wang Q, Hesthaven JS, Zhang C. Physics-informed machine learning for reduced-order modeling of nonlinear problems. J Comput Phys 2021;446:110666. https://doi.org/10.1016/j.jcp.2021.110666
[42] Nath K, Kumar V, Smith D, Karniadakis G. A digital twin for diesel engines: operator-infused pinns with transfer learning for engine health monitoring. Dec 2024; arXiv:2412.11967.
[43] Nath K, Meng X, Smith DJ, Karniadakis GE. Physics-informed neural networks for predicting gas flow dynamics and unknown parameters in diesel engines. Sci Rep 2023;13(1):13683. https://doi.org/10.1038/s41598-023-39989-4.
[44] McDonagh C, Chen X. Leveraging intrinsic gradient information for further training of differentiable machine learning models. Cornell University; Jan 2021. https://doi.org/10.48550/arxiv.2112.00094.
[45] McDonagh C, Chen X. Leveraging intrinsic gradient information for machine learning model training. Cornell University; Nov 2021. https://arxiv.org/pdf/2112.00094.
[46] Son H, Jang JW, Han WJ, Hwang HJ. Sobolev training for physics informed neural networks. 2021; arXiv preprint arXiv:2101.08932.
[47] Jacob A, Ashok B, Rajasekar V, Balusamy S, Alagumalai A. NOx and PM trade-off in IC engines. NOx Emiss Control Technol Station Automot Intern Combust Engines 2021:69–91. https://doi.org/10.1016/B978-0-12-823955-1.00003-6
[48] Tamilvanan A, Ashok B, Mohanraj T, Jayalakshmi P, Dhamodharan P, Sakthivel R. Chapter 5 - effect of engine operating parameters in NOx reduction. In Ashok B, editor. NOx emission control technologies in stationary and automotive internal combustion engines. Elsevier; 2022. pp. 125–53. https://doi.org/10.1016/B978-0-12-823955-1.00005-X
[49] Aglietti F, Santa FD, Piano A, Aglietti V. GradINN: gradient informed neural network. 2024; arXiv:2409.01914.
[50] Castagné M, Bentolila Y, Chaudoye F, Hallé A, Nicolas F, Sinoquet D. Comparison of engine calibration methods based on design of experiments (DOE). Oil Gas Sci Technol Rev IFP 2008;63(4):563–82. https://doi.org/10.2516/ogst:2008029.
[51] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016. http://www.deeplearningbook.org.
[52] Rasmussen C, Bousquet O, Luxburg U, Rätsch G. Gaussian processes in machine learning. Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures 2004;3176:63–71.