algorithms

IMPACT FACTOR 2.1

CITESCORE 4.5

Article

A Hybrid Physics-Informed ML Framework for Emission and Energy Flow Prediction in a Retrofitted Heavy-Duty Vehicle

Talha Mujahid, Teresa Donateo and Pietropaolo Morrone

Special Issue

Computational Modeling and Intelligent Simulation of Next-Generation Energy Systems

Edited by

Dr. Arya Abdollahi and Prof. Dr. Frank Werner

MDPI

https://doi.org/10.3390/a19040317

algorithms

MDPI

Article

A Hybrid Physics-Informed ML Framework for Emission and Energy Flow Prediction in a Retrofitted Heavy-Duty Vehicle

Talha Mujahid [1], Teresa Donateo [1,*] and Pietropaolo Morrone [2]

[1] Department of Engineering for Innovation, University of Salento, via Monteroni, 73100 Lecce, Italy
[2] Department of Mechanical, Energy and Management Engineering (DIMEG), University of Calabria, via P. Bucci, 87036 Rende, Italy; pietropaolo.morrone@unical.it
* Correspondence: teresa.donateo@unisalento.it

Abstract

This study introduces a physics-informed machine learning framework for predicting transient emissions and energy variables in a retrofitted heavy-duty diesel vehicle. It merges data-driven modeling with physically derived features for reliable real-world analysis. A Random Forest regressor is trained on a public dataset (26 trips from one instrumented vehicle) to predict CO_2 and NO_x mass rates, exhaust temperature, exhaust mass flow rate, and fuel flow rate from synchronized multi-sensor inputs using past-only, time-lagged features. On held-out trips, exhaust temperature prediction achieves R^2 = 0.9997 and RMSE = 0.53 g/s; for CO_2, with R^2 = 0.9985 and RMSE= 0.38 g/s, comparable performance is reported for NO_x, exhaust flow, and fuel rate. The trained model is integrated into a simulation framework to enable the evaluation of alternative operating conditions and powertrain configurations. First, the impact of cold-start versus hot-start operation is assessed, showing cumulative emission penalties of up to +28% for CO_2 and +30% for NO_x. Second, the effect of hybridization is investigated by comparing the baseline thermal configuration with a hybrid electric architecture, resulting in estimated reductions of -12.2% in CO_2 and -10.5% in NO_x emissions. This tool excels in high-fidelity emission prediction and system-level energy analysis, aiding advanced powertrain assessments under realistic driving conditions.

Keywords: cold-start; hard-to-electrify transportation; heavy-duty vehicles; Random Forest; temporal modeling; predictive analytics

1. Introduction

Heavy-duty diesel engines (HDDEs) play a central role in freight transport because of their high thermal efficiency, durability, and sustained-load capability [1,2]. At the same time, heavy-duty diesel vehicles (HDDVs) are a major source of CO_2, NO_x, and other pollutants, particularly under real driving conditions and in high-density traffic environments [3,4]. Freight trucks and buses account for a substantial share of road-transport emissions, and their activity has returned to near pre-pandemic levels in recent years, reinforcing the urgency of effective mitigation strategies [5,6]. This need is further amplified by increasingly stringent regulatory frameworks, including the European Union Regulation (EU) 2019/1242 [6] and forthcoming standards such as Euro VII, which explicitly target real-world emissions and cold-start performance [7,8]. The estimate targets in CO_2 emission for HDVs are shown in Figure 1.

Academic Editors: Arya Abdollahi and Frank Werner

Received: 24 March 2026
Revised: 13 April 2026
Accepted: 14 April 2026
Published: 17 April 2026

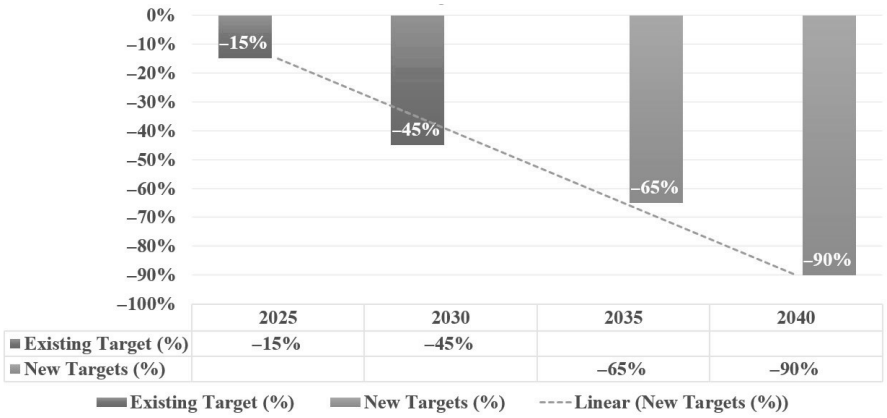

Figure 1. European Union CO emission reduction targets for heavy-duty vehicles (HDVs) under Regulation (EU) 2019/1242, showing progressive limits from 2025–2040 [6].

Among the various operating conditions, cold-start and warm-up phases contribute disproportionately to the total emissions [9]. During this period, the engine and aftertreatment systems operate far from optimal thermal conditions, leading to incomplete combustion, longer ignition delays, and reduced catalyst efficiency [10–12]. In particular, low exhaust temperatures limit NO_x conversion efficiency in selective catalytic reduction (SCR) systems, while transient combustion behavior increases variability in pollutant formation [13]. These effects are especially relevant in urban driving, where stop–start operation prolongs the warm-up phase.

Traditional methodologies for the estimation of energy flows and emissions are predominantly based on mean-value engine models and steady-state engine maps, in which fuel consumption and emissions are expressed as functions of engine speed and load. These approaches remain widely adopted due to their physical interpretability and relatively low computational burden, as discussed in classical works such as those by Guzzella et al. [14] and Rakopoulos and Giakoumis [15]. Several studies have demonstrated that under such conditions, steady-state approaches fail to capture the complex interactions governing emissions, particularly for pollutants such as NO_x, whose formation is highly sensitive to transient thermal and mixing phenomena [16–18].

To overcome these limitations, more advanced physics-based approaches have been developed, including multi-zone combustion models and detailed dynamic simulations capable of resolving turbocharger behavior, air–fuel mixing, and heat transfer processes. While these models can achieve high predictive fidelity, they typically require extensive calibration efforts and detailed knowledge of engine parameters, which are not always accessible. Furthermore, their computational cost often makes them unsuitable for system-level analyses or real-time applications, thereby limiting their use in vehicle simulation environments aimed at evaluating alternative powertrain configurations [19,20]. As a result, data-driven approaches have gained increasing attention as complementary tools when synchronized onboard diagnostics (OBD) and portable emission measurement system (PEMS) data are available.

A variety of models, including artificial neural networks, support vector machines, and ensemble methods such as Random Forests, have demonstrated strong predictive capabilities when trained on experimental datasets obtained from chassis dynamometer or portable emissions measurement system campaigns. These approaches are particularly effective in capturing non-linear relationships between operating variables and emissions, without requiring the explicit modeling of the underlying physical processes achieving high predictive accuracy [21]. These models can capture non-linear interactions and temporal

dependencies. A detailed analysis of the literature on the ML model applied to engine output prediction is reported in Appendix A.

Despite these advantages, most existing machine learning applications in this domain focus primarily on predictive accuracy, often treating the problem as a purely data-driven regression task. More importantly, these models are rarely integrated into simulation frameworks, which restricts their applicability to design-oriented analyses. In many cases, they function as standalone predictors, providing estimates of emissions for the given input conditions but lacking the capability to support the evaluation of alternative powertrain configurations or operational strategies. Recent developments in explainable artificial intelligence have also highlighted the need to improve the interpretability of machine learning models in emission prediction tasks, further emphasizing the limitations of purely black-box approaches [22].

Two key limitations remain insufficiently addressed in the literature. First, many high-performing models rely on high-dimensional or non-deployable input variables, such as in-cylinder pressure, injection parameters, or internal aftertreatment states [23,24]. While these inputs enhance predictive performance, they limit applicability in real-world onboard implementations, where only a restricted set of signals is typically available. Second, methodological issues related to time-series modeling are not always rigorously addressed, particularly with respect to causality and data leakage. The use of future information—either explicitly or implicitly through feature construction or data splitting—can lead to overly optimistic performance estimates that are not reproducible in real-time deployment scenarios.

A further gap concerns the integration of emission prediction models into system-level analyses, such as hybridization and waste heat recovery (WHR). While several studies focus on improving predictive accuracy, relatively few explicitly address how data-driven models can be used as engineering tools for evaluating the effect of hybridization and waste heat recovery on pollutant emissions.

Physically meaningful features derived from vehicle dynamics or simplified engine models that can be used as inputs to machine learning algorithms, thereby enhancing model robustness and interpretability, have been proposed [25]. While these approaches represent a significant step forward, they often remain focused on component-level prediction and do not fully exploit the potential of integrating data-driven models into comprehensive vehicle simulation environments. As a result, their use in assessing system-level modifications, such as powertrain hybridization, remains limited.

In light of the above considerations, a clear research gap emerges in the need for modeling frameworks that can simultaneously ensure high predictive accuracy under transient real-world conditions, maintain consistency with physical principles, and be readily integrated into system-level simulation tools for design and analysis purposes.

1.1. Goals and Original Contributions of the Present Work

This work was conducted within the IRIDESCENT research project (“biodIesel hybRID Electric buS with waste heat reCovEry aNd sTorage”), which seeks to leverage synergies between hybridization, waste heat recovery, and biodiesel to attain carbon neutrality and reduce cold-start emissions in heavy-duty vehicles. A map-based empirical approach was applied to related objectives in a prior work [26], where fuel consumption could be estimated over complete trips, but transient tailpipe pollutant mass rates and exhaust-temperature dynamics were not reproduced with sufficient fidelity for warm-up analysis. The present work therefore develops a data-driven estimator trained directly on synchronized OBD and PEMS measurements for the reference vehicle of the IRIDESCENT project.

The proposed methodology was developed and validated using a real-world dataset comprising 26 trips from a single instrumented heavy-duty vehicle operating under vary-

ing thermal and loading conditions [27]. While this dataset does not support fleet-wide generalization, it provides a controlled and reproducible testbed for evaluating the modeling framework and its application to scenario-based analyses. Compared with previous studies, the proposed approach prioritizes robustness, interpretability, and deployability over model complexity. While deep learning architectures such as CNN–LSTM and Transformer-based models have shown strong performance in controlled settings [18], they often rely on large datasets and extended sensing capabilities. In contrast, the present framework demonstrates that high predictive accuracy can be achieved using a limited and physically meaningful set of inputs, provided that temporal dependencies and validation procedures are handled rigorously.

The present work proposes a hybrid modeling framework that combines physics-informed feature engineering with data-driven emission prediction and system-level simulation. In the proposed approach, the input variables to the machine learning model are not limited to directly measured signals but include quantities derived from a simplified longitudinal vehicle model, such as traction power demand and resistive forces. This choice enables the incorporation of fundamental physical relationships into the modeling process, thereby improving generalization and interpretability. Moreover, the results of the ML model are meant to be used for the sizing, modeling, and optimization of a complex powertrain combining WHR and hybridization to optimize engine operating conditions and avoid any cold-start emissions penalty. From this perspective, the novelty of the present work does not reside in the use of a specific machine learning algorithm, but rather in the integration of data-driven emission modeling within a physically grounded and simulation-oriented framework. This integration enables the transition from purely predictive models to tools that can effectively support system-level design and assessment, thereby contributing to the development of more efficient and environmentally sustainable heavy-duty vehicle technologies. Furthermore, by explicitly linking emission prediction to engineering applications such as hybridization and thermal management, the proposed methodology moves beyond purely data-driven regression tasks and contributes toward decision-support tools for the design of advanced heavy-duty propulsion systems.

Other elements of novelty can be found in the following aspects of the proposed approach:

- Deployment-feasible modeling: The estimator is developed using a compact set of predictors available from standard OBD signals (engine speed, coolant temperature, and ambient temperature), ensuring direct applicability without additional instrumentation.
- Strictly causal temporal feature construction: Time-dependent behavior is captured through past-only (causal) features, including lags, derivatives, and rolling statistics, ensuring consistency with real-time implementation and avoiding look-ahead bias.
- Leakage-safe validation protocol: Model performance is evaluated using trip-level data partitioning, preventing overlap between training and testing samples and providing a realistic estimate of generalization to unseen driving conditions.
- Multi-output prediction including thermal variables: The model simultaneously predicts pollutant mass rates (CO_2, NO_x, CO) and exhaust temperature and mass flow rates, which are critical for analyzing thermal management and WHR potential [28].
- Integration into system-level scenario analysis: The trained model is used as a surrogate tool to evaluate the impact of cold-start versus hot-start operation and the effect of hybrid-electric powertrain configurations with waste heat recovery, extending previous quasi-static analyses [29,30].

1.2. Organization of the Paper

The remainder of the manuscript is organized as follows. Section 2 describes the dataset and the estimation of engine working points. Section 3 presents the estimator,

feature construction, and evaluation protocol. Section 4 reports predictive performance on held-out trips. Section 5 presents the warm-up scenario analysis and the hybridized trajectory demonstration, respectively, with limitations stated in Section 6.

2. Description and Processing of the Dataset

The dataset employed in this work originates from an experimental campaign reported in Ref. [27], which investigated the behavior of a heavy-duty diesel vehicle under real-world driving conditions, encompassing a wide range of operating states, cargo loads, and environmental conditions relevant to cold-start behavior. The trips are performed with a single instrumented truck on the same route with the same driver, so the database cannot be used to obtain a fleet-representative estimator or to estimate the effect of driving style, etc. In particular, the test platform was an Isuzu FTR850 AMT, classified as a Class 6 heavy-duty truck equipped with a six-cylinder, turbocharged, intercooled, common-rail diesel engine. This configuration is representative of engines widely used in medium- to heavy-duty fleets and provides a suitable baseline for the IRIDESCENT project.

Field tests were conducted within the City of Tshwane Metropolitan Municipality (Pretoria, South Africa) along a 61.7 km route designed to capture representative driving conditions. Each experimental run lasted approximately two hours, subject to traffic variability.

Real-time exhaust emissions were measured using a SEMTECH DS+ Portable Emissions Measurement System (PEMS) connected directly to the exhaust line. The system continuously recorded the concentrations of CO, CO_2, NO, and NO_2, together with exhaust temperature and flow rate. Unfortunately, the PEMS used in Ref. [27] does not comprise the soot measurement unit; therefore, it was not possible to model such an important diesel engine pollutant.

Parallel engine diagnostics were collected via OBD-II (engine speed, fuel flow, coolant temperature) and supplemented with ambient measurements (temperature, pressure, humidity). All channels were synchronized to a common timestamp, yielding a multimodal dataset suitable for transient modeling.

2.1. Classification of the Trips

The dataset made available by Ref. [27] enabled a detailed characterization of transient engine behavior under variable driving conditions. To develop the model, the trips were grouped by payload and initial engine temperature. In terms of payload, the dataset includes eight unloaded (baseline), eight with approximately 3000 kg (maximum load), and eight with 1500 kg (medium load). The classification of cold-start versus hot-start events was based on the initial coolant temperature and vehicle mass. Trips beginning with coolant temperatures above 79 °C were labeled as hot-start, while those starting below 30 °C were labeled as cold-start.

$$S_{start} = \begin{cases} Cold-start, \;\; T_{cool}\,(t_o) < 30\,^{\circ}C \\ Mid-start, \;\; 30\,^{\circ}\text{C} \le T_{cool}\,(t_o) < 79\,^{\circ}C \\ Hot-start, \;\; T_{cool}\,(t_o) \ge 79\,^{\circ}C \end{cases} \quad (1)$$

Trips 1 and 11 fell between these limits and are labeled as mid-starts, providing a more nuanced differentiation of thermal regimes than a single threshold for interpreting emission variability.

An overview of the 26 trips and their classification is shown in Figure 2.

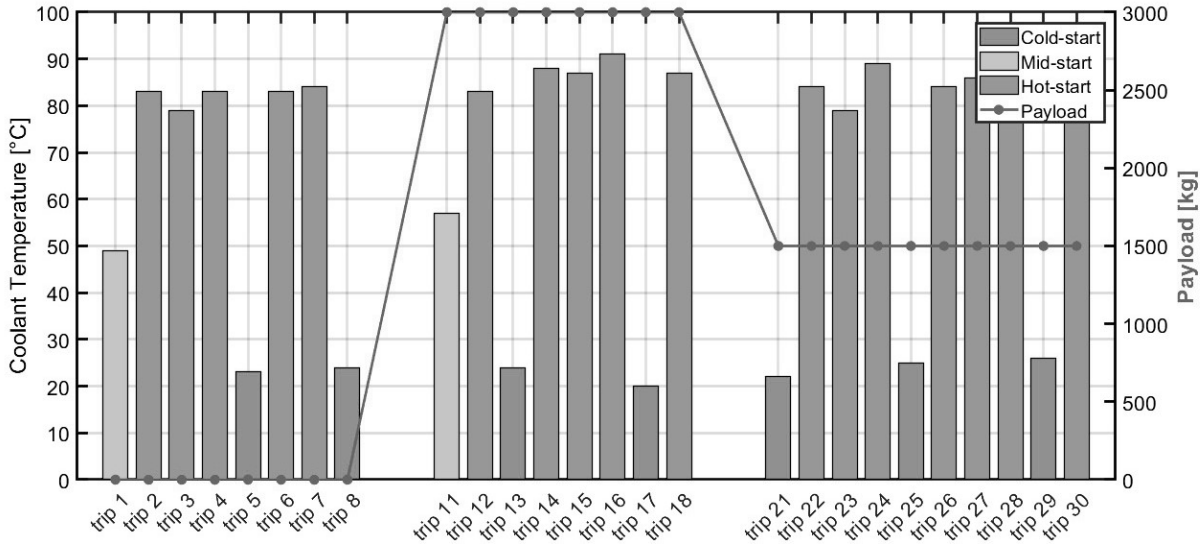

Figure 2. Overview of the 26 experimental trips, classified by payload (unloaded, 1500 kg, 3000 kg) and thermal start condition (cold: T_{cool} < 30 °C; mid: 30–79 °C; hot: T_{cool} > 79 °C).

As an example of the relevant data, Figure 3 shows the cumulative values of NO_x emissions during the first ten minutes of Trips 1–8. Despite being performed with the same driver, on the same route, and without a trailer, there are relevant differences between the final values of emissions due to the effect of different ambient conditions and traffic levels. In particular, it is possible to notice how the dotted lines, which refer to cold-start trips, tendentially give higher values of NO_x emissions.

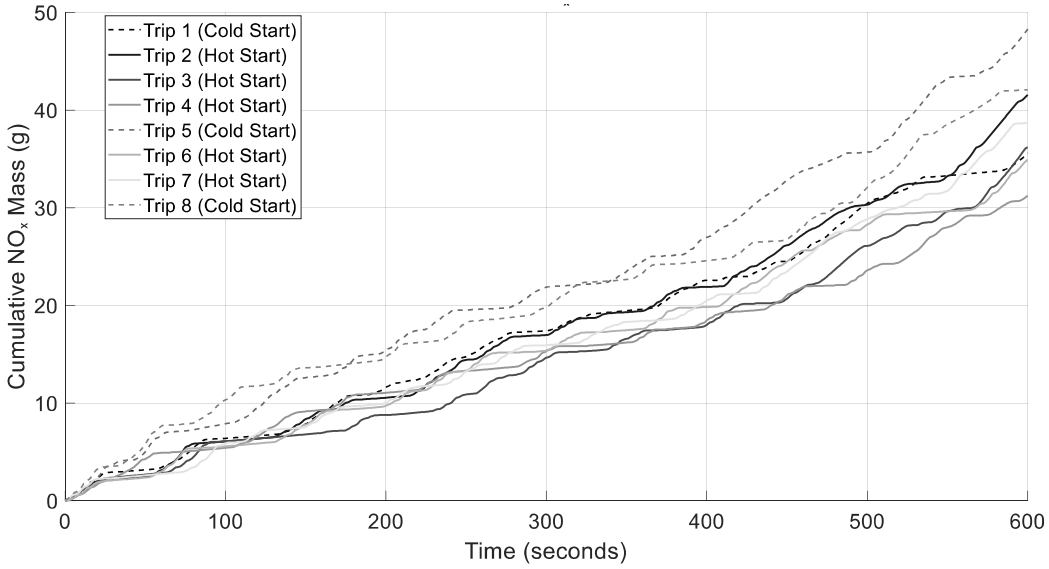

Figure 3. Cumulative NO_x emissions in the first 10 min of Trips 1–8 (unloaded), comparing cold-start (dotted) vs. hot-start (solid) conditions.

Figure 4 highlights the difference in the coolant temperature time history for the first eight trips, all corresponding to zero trailing load. Note the time to reach the desired coolant temperature of 80 °C ranges between 350 to 450 s, being affected by the initial coolant temperature.

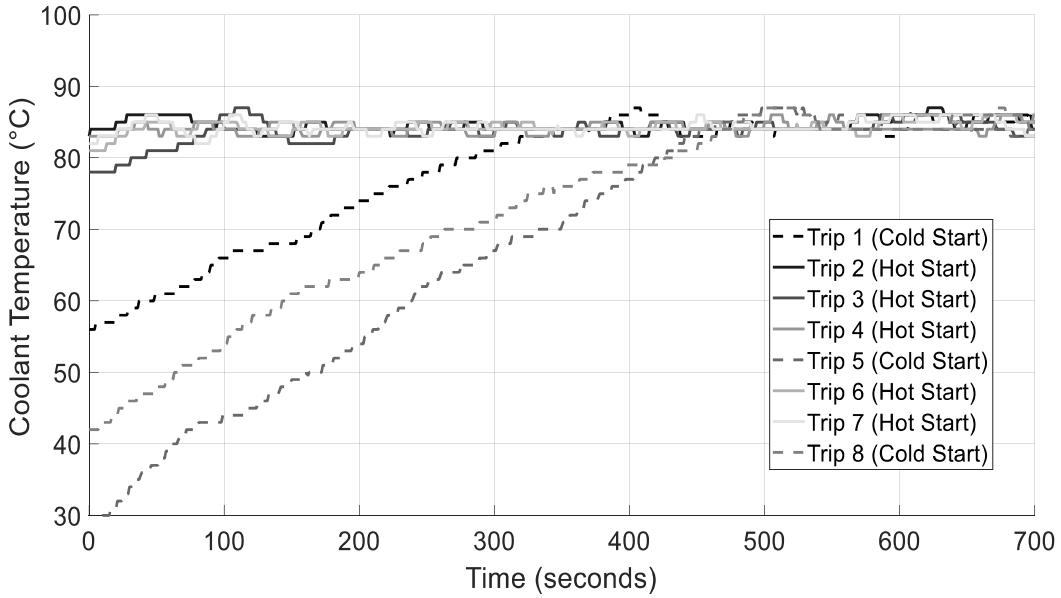

Figure 4. Coolant temperature profiles for Trips 1–8 (no payload), showing warm-up times of 350–450 s to reach 80 °C, influenced by initial conditions.

2.2. Engine Torque Estimation

A signal related to engine torque, i.e., "engine load" is available in the dataset; however, it was found to be inconsistent with the other measured variables and therefore not sufficiently reliable for direct use in the analysis. As a result, engine torque was not directly measured but instead estimated with a physical-based approach described here.

The required power at the wheel for each sample was calculated by considering factors such as vehicle speed, road grade, rolling resistance, and drag, using the following relation:

$$P_{\mathrm{w}}(t) = \frac{1}{2} C_d \cdot A_f \cdot \rho_a \cdot V(t)^3 + M \cdot g \cdot C_r \cdot V(t) \cdot \cos\alpha(t) + M \cdot g \cdot \sin\alpha(t) \cdot V(t) + M_i \cdot a(t) \cdot V(t) \quad (2)$$

where $P_{\mathrm{w}}(t)$ represents the instantaneous power demand, C_d is the drag coefficient, A_f is the frontal area, ρ_a is the air density, $V(t)$ is the vehicle speed, M is the total vehicle mass, $\alpha(t)$ is the road grade, and $a(t)$ is the acceleration.

The engine speed $\omega_e(i)$ from OBD was used to estimate the engine torque τ:

$$\tau(t) = \frac{P_w(t) + P_{aux}(t)}{\eta_{gb}\omega_e(t)} \quad (3)$$

where η_{gb} is the overall efficiency of the transmission that accounts for all mechanical losses in the driveline connecting the engine to the wheels. $P_{aux}(i)$ is the parasitic power of the vehicle auxiliaries. The details of the model and its validation can be found in a previous work of the authors [26].

3. Development of the Model

The goal of this investigation was to implement and validate a multi-output Random Forest model for predicting the engine behavior under dynamic conditions based on the input measured on the whole vehicle. The inputs and outputs of the model are described in Table 1. Inputs include ambient temperature, coolant temperature, engine speed, and engine torque, selected for real-time feasibility and strong correlation with thermal and emission dynamics. Let t be the index the synchronized timeline; the predictor and target vectors are:

$$\mathrm{x}(t) = [T_{\mathrm{amb}}(t), T_{\mathrm{cool}}(t), n(t), \tau(t)],\ \mathrm{y}(t) = \left[\dot{m}_{fuel}(t),\ \dot{m}_{\mathrm{CO_2}}(t), \dot{m}_{\mathrm{CO}}(t), \dot{m}_{\mathrm{NO}_x}(t), T_{\mathrm{exh}}(t), \dot{m}_{\mathrm{exh}}(t)\right] \quad (4)$$

Table 1. Input and output variables for the Random Forest model, with units and training/validation trip splits.

	Variable	Symbol	Unit
Inputs	Engine torque	τ	N.m
	Engine speed	$n(t)$	Rpm
	Coolant temperature	T_{cool}	°C
	Ambient temperature	$T_{\mathrm{amb}}(t)$	°C
Outputs	Exhaust mass flow rate	$\dot{m}_{\mathrm{exh}}$	kg/s
	Exhaust temperature	T_{exh}	°C
	CO_2 emissions	$\dot{m}_{\mathrm{CO_2}}$	g/s
	NO_x emissions	$\dot{m}_{\mathrm{NO}_x}$	g/s
	CO emissions	$\dot{m}_{\mathrm{CO}}$	g/s
Training trips (80.77%)	3, 29, 18, 11, 4, 8, 23, 14, 1, 24, 30, 7, 13, 5, 26, 6, 22, 17, 21, 27, 12		
Trips for validation	28, 2, 16, 25, 15		

To preserve causality and deployment feasibility, contemporaneous target values were excluded from inputs; only past and current predictor states were used.

Note that the inputs include not only raw signals but also a variable (torque) derived from the simplified powertrain balance. The quasi-static formulation used in Equation (2) disregards rotational inertia effects, which may lead to inaccuracies during high-frequency transients (e.g., rapid accelerations). Nevertheless, the ML model is trained on experimental emission data that inherently capture these dynamic phenomena; as a result, some of the discrepancies introduced at the input stage are mitigated through the data-driven mapping. In this context, the torque input should be interpreted as the engine set point rather than the actual delivered torque. Viewed this way, the approach is well suited for evaluating the hybrid configuration, where engine set points are determined by energy management strategies within the simulation framework, while the actual engine torque remains influenced by the aforementioned dynamic effects.

The targets include regulated species in mass-rate form (CO_2, CO, NO_x and exhaust operating conditions (temperature and mass flow) that co-evolve during warm-up and are central to cold-start assessment. To prevent leakage and preserve real-time feasibility, contemporaneous target quantities were not used as inputs at the same timestamp; only $x(t)$ was provided to the model at time t. This strict partition retains sufficient observability of the process while ensuring that subsequent model training remains causal and deployment-ready.

Random Forest was selected for its proven robustness and predictive stability on noisy small-to-medium-sized datasets characterized by strong non-linearities and mixed-feature interactions. From the review of the state-of-the-art presented in the Introduction, ensemble tree-based models were identified as being particularly effective for emission and combustion prediction tasks, where input and output relationships are highly non-linear and the data volume is limited by experimental cost. Several alternative regression algorithms, such as Support Vector Machines (SVMs), Decision Trees, k-Nearest Neighbors (k-NN), and Multilayer Perceptrons (MLPs), were evaluated using identical feature sets and trip-level cross-validation. However, these models either exhibited lower R^2 values, higher sensitivity to noise, or degraded generalization when trained on the relatively small dataset available. In contrast, the Random Forest consistently achieved the highest predictive accuracy and R^2 across all output variables (CO_2, NO_x, exhaust temperature, and

exhaust mass flow) while maintaining low variance and minimal overfitting. Its inherent ensemble averaging and bootstrapping mechanisms reduce model variance and improve resilience against outliers and data imbalance conditions typical of real-world emission datasets. Consequently, the Random Forest was selected as the optimal predictor for this study, balancing model complexity, interpretability, and performance robustness under constrained data availability.

Note that mass flow rate and CO_2 emissions are not independent outputs since the ratio depends on the carbon content of the fuel. A preliminary analysis of the dataset confirmed the correlations between the total fuel mass measured by the OBD-II and the total emissions of CO_2 obtained from the Portable Emissions Measurement System. Figure 5 shows a strong correlation ($R^2 \approx 0.9917$) between the total fuel mass measured by the OBD and the CO_2 emissions recorded by the PEMS, validating the carbon-mass balance approach adopted later to estimate fuel consumption from CO_2.

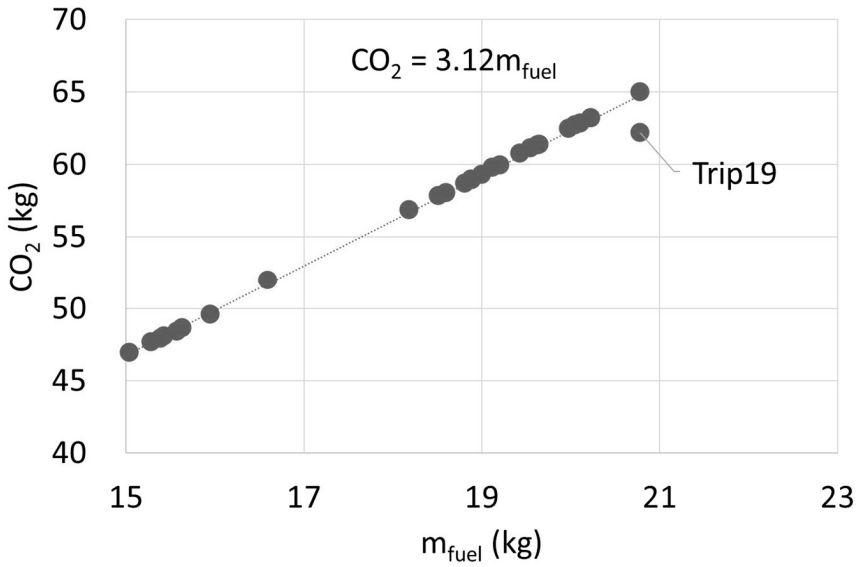

Figure 5. Correlation between total fuel mass from OBD-II (x-axis) and CO_2 emissions from PEMS (y-axis) across all trips (R^2 = 0.9917), validating carbon mass balance with the exception of Trip 19The 26 retained trips were split at the trip level: twenty-one trips (80.77%) for training and five trips (19.23%) held out for final testing (Table 1). No time samples from a given trip appear in both sets.

Splits were by trip so that no time samples from a given trip appeared in both sets. Metric computations respected the native sampling rates via sample-time weighting.

Model effectiveness was assessed on the held-out validation/test data using regression metrics computed for each target (mass rate of CO_2, CO, and NO; exhaust temperature; exhaust mass flow) and then averaged for a single summary score. Let $\{(y_a, \hat{y}_a)\}_{a=1}^{n}$ denote the ground-truth and predicted values for a given target, $\overline{y}$ the mean of the ground-truth values, and $r_a = y_a - \hat{y}_a$ the residual. The coefficient of determination is:

$$R^2 = 1 - \frac{\sum_{a=1}^{n}(y_a - \hat{y}_a)^2}{\sum_{a=1}^{n}(y_a - \overline{y})^2} \tag{5}$$

The root–mean–squared error and mean–squared error are:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{a=1}^{n}(y_a - \hat{y}_a)^2}, \ \text{MSE} = \frac{1}{n}\sum_{a=1}^{n}(y_a - \hat{y}_a)^2 \tag{6}$$

The mean–absolute error and maximum absolute error are:

$$MAE = \frac{1}{n}\sum_{a=1}^{n}|y_a - \hat{y}_a|,\ Max\ Error = \max_{1\le a\le n}|y_a - \hat{y}_a| \tag{7}$$

These metrics jointly characterize average accuracy (MAE/MSE/RMSE), explained variance (R^2), and worst-case deviation (*Max Error*). Final models were refit on the full training set using the selected hyperparameters and then evaluated on held-out trips using the same metrics to provide an unbiased estimate of performance on unseen data.

3.1. Temporal Feature Engineering

Temporal feature engineering is essential for modeling cold-start behavior because exhaust-gas dynamics are transient and history-dependent. As noted in our previous publication [26], map-based approaches are inadequate under cold-start conditions since they do not account for transport and thermal inertia in the exhaust system. To represent these time dependencies explicitly, we constructed features that capture short-horizon fluctuations and longer-horizon memory using strictly causal operators, so that inputs at time t depend only on data available up to t.

For each primary channel $x(t)$ (engine speed, vehicle speed, fuel flow, coolant temperature, exhaust temperature and flow, manifold pressure and temperature, ambient variables), causal lags $\{x\ (t - k\ \Delta t)\}$ with $k = 1,\ 2,\ 3, \ldots, n$ were formed to encode short-term memory and autoregressive structure. Rapid changes were represented by discrete derivatives $\Delta x(t) = x(t) - x(t-1)$ and the corresponding causal slope $\frac{\Delta x(t)}{\Delta t}$; where useful to capture curvature, the second backward difference $\Delta^2 x(t) = x(t) - 2x\ (t-1) + x(t-2)$ was also included. Local volatility and extremes were quantified with trailing (past-only) rolling statistics over a window length W, so that each feature at time t depends only on samples at times $\leq$ t. This avoids look-ahead leakage that can occur with centered windows in time-series evaluation.

$$\mu_{\text{rolling}} = \frac{1}{2w+1}\sum_{i=t-w}^{t+w} x_i \tag{8}$$

$$\sigma_{\text{rolling}} = \sqrt{\frac{1}{2w+1}\sum_{i=t-w}^{t+w}\left(x_i - \mu_{\text{rolling}}\right)^2} \tag{9}$$

$$\begin{aligned} x_{min}^{(w)} &= min(x_{t-w}, \ldots, x_{t+w}) \\ x_{max}^{(w)} &= max(x_{t-w}, \ldots, x_{t+w}) \end{aligned} \tag{10}$$

These statistics highlight short-term emission bursts and transient engine behavior during cold starts. Longer memory with bounded latency was captured using exponentially weighted moving averages (EWMA) for selected channels,

$$\text{EWMA}_\alpha(t) = \alpha\ x(t) + (1-\alpha)\ \text{EWMA}_\alpha(t-1) \tag{11}$$

with α selected on the training set. Cumulative exposure variables tracked thermal progress using causal integrals (left-Riemann sums on cleaned/masked streams), including cumulative fuel $\int_0^t \dot{m}_f(\tau)d\tau$ and a proxy for exhaust thermal energy $\int_0^t \dot{m}_{exh}(\tau)c_p[T_{\text{exh}}(\tau) - T_{\text{amb}}(\tau)]d\tau$; elapsed time since trip start and since thermal thresholds, e.g., $(T_{\text{cool}}\ \geq 30\ ^\circ\text{C}, T_{\text{exh}}\ \geq T_{\text{amb}}\ + \delta)$, was also included. Lagged target emissions are not used as predictors in the deployment-feasible estimator because they would require pollutant sensing that is not available from standard OBD streams. The feature set is restricted to past-only transformations of the predictor channels and elapsed-time variables. $\{y(t-1), y(t-2), \ldots\}$ for each target y using only past values; any

timestamp masked for a target remained masked in all derived features to avoid leakage from missing-data handling.

The final feature set used for model training comprised the original variables (e.g., ambient temperature, engine speed, brake power, coolant and exhaust temperatures/flows, manifold conditions), their causal lags, the centered rolling statistics above (with trailing/causal counterparts used in deployment to avoid look-ahead), gradient and derivative features, the exponentially weighted averages, and the cumulative exposures. Redundant features were pruned on the training split by variance-inflation and correlation screening, and all continuous features were standardized with training-set statistics using $z = (x - \mu)/\sigma$ (robust scaling for heavy-tailed channels such as fuel and exhaust flows).

The details of data-cleaning process and the sensitivity analysis performed to select the trips for model development and validation are reported in Appendix B.

3.2. Model Training

Hyperparameters controlling model capacity and regularization were tuned by a grid search with trip-grouped cross-validation on the training set only. The search varied the number of trees ($n_{\text{estimators}}$), maximum tree depth (max_depth), minimum samples required to split an internal node ($min\_samples\_split$), minimum samples per leaf ($min\_samples\_leaf$), the number of features considered at each split ($max\_features$), and bootstrap sampling. Increasing $n_{\text{estimators}}$ improves estimator stability at additional computational cost. The bias–variance trade-off was controlled primarily by max_depth, $min\_samples\_split$, and $min\_samples\_leaf$. Each candidate configuration was evaluated using the same trip-grouped protocol described below, and the setting with the best validation performance was selected. Final Random Forest hyperparameters (selected on training trips only) are as follows:

$$\begin{aligned} n_{\text{estimators}} &= [\text{FILL}], max\_depth = [\text{FILL OR None}], min\_samples\_split = [\text{FILL}], min\_samples\_leaf \\ &= [\text{FILL}], max\_features = [\text{FILL}], bootstrap = [\text{FILL True/False}], random\ seed \\ &= [\text{FILL}]. \end{aligned} \tag{12}$$

Cross-validation was used to estimate generalization performance and reduce overfitting during model selection. Trip-grouped K-fold cross-validation (K = [FILL]) was applied to the training trips: in each round, $K - 1$ folds (complete trips) were used for training and the remaining fold (complete trips) was used for validation. Splits were constructed so that samples from the same trip never appeared in both training and validation within the same round. Within each trip, observations preserved their natural time order (no shuffling).

3.3. Feature Importance and Ablation

Feature attribution was assessed with permutation importance on the validation folds to avoid model-internal biases and to respect the multi-output setting. For each target $y \in \left[\dot{m}_{\text{CO}_2}(t), \dot{m}_{\text{CO}}(t), \dot{m}_{\text{NO}_x}(t), T_{\text{exh}}(t), \dot{m}_{\text{exh}}(t)\right]$, a baseline score (RMSE and R^2) was recorded. Each input in $\mathrm{x}(t) = [T_{\text{amb}}(t), T_{\text{cool}}(t), n(t), \tau(t)]$ was then permuted within the validation fold (preserving trip grouping and time order), and the drop in performance $\Delta S = S_{perm} - S_0$ was measured. Importance was reported as the mean ΔS across folds with 95% confidence intervals from fold-to-fold variability. A conditional variant that shuffles within narrow strata of the remaining features (e.g., speed–torque bins) was also computed to mitigate interaction effects.

Ablation analysis quantified marginal contribution under the same preprocessing, masking, and grouped cross-validation used for training. Starting from the full model, inputs were removed one at a time (i) without T_{cool} (thermal state), (ii) without n (kinematics), (iii) without τ (load/enthalpy proxy), and (iv) without T_{amb}(boundary condition), and the model was retrained and re-evaluated on the same held-out trips. Effect sizes were

expressed as the relative RMSE increase $\frac{RMSE_{abl}-RMSE_{full}}{RMSE_{full}} \times 100\%$ and ΔR^2 per target, then averaged across targets for a summary measure. Repeats across folds (and two random seeds for tree bootstrapping) confirmed stability; dispersion is shown as standard error. This protocol establishes which of the four production-feasible inputs most strongly drive cold-start predictions, provides a robustness check against spurious correlations, and keeps all analyses consistent with the causal data handling of Section 3.3.

4. Validation of the ML Model

The trips excluded from the training were used to validate the model in a qualitative and quantitative way.

The qualitative comparison between experimental and predicted time histories of the output variables CO_2 and NO_x mass rates, exhaust temperature, and exhaust mass flow, is shown in Figure 6 for Trip 15. Figure 6 confirms accurate phase alignment across all targets, with minor smoothing of NO_x peaks and slight over-prediction of exhaust flow during early transients. Moreover, late-trip ramps in CO_2 and exhaust temperature are reproduced cleanly, indicating that the chosen inputs capture thermal and flow dynamics effectively.

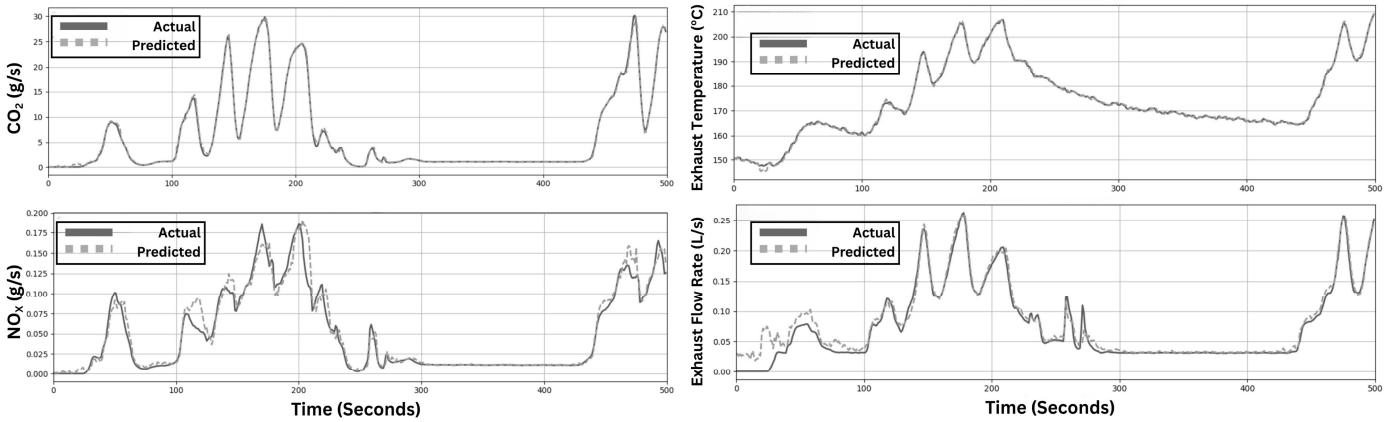

Figure 6. Measured (blue solid) vs. predicted (orange dashed) time series for CO_2, NO_x mass rates, exhaust temperature, and exhaust flow on held-out Trip 15.

Note that the estimator reproduces the experimental data with strong agreement, particularly CO_2 mass rate and exhaust temperature. Small, localized deviations during sharp transients can be observed for NO_x mass rate and exhaust mass flow rate.

Quantitatively, CO_2 achieves R^2= 0.9985 and $RMSE = 0.3763$ g/s, indicating that load-driven variations in fuel use are captured with near-deterministic fidelity. NO_x shows similarly strong alignment ($R^2 = 0.9677$, $RMSE = 0.0138$), consistent with tracking engine-out variability and the first-order effect of after-treatment readiness early in operation. Exhaust temperature is reproduced with near-perfect coherence ($R^2 = 0.9997$, $RMSE = 0.5275\,^{\circ}$C), and exhaust flow is accurately tracked ($R^2 = 0.9612$, $RMSE = 0.0136$), confirming that the chosen inputs encode the dominant thermal and flow dynamics.

Maximum absolute errors on Trip 15 are 5.79 g/s (CO_2), 0.13 (NO_x), 8.85 °C (exhaust temperature), 0.107 (exhaust flow), and 0.41(CO); these excursions are confined to brief transients and do not alter the overall fit trend.

CO was more challenging due to low variance and short spikes that disproportionately influence R^2, but overall accuracy remained acceptable, with $RMSE = 0.0252$ and $R^2 = 0.8006$.

Table 2 and Figures 7 and 8 summarize performance across the full test, demonstrating robust generalization under diverse operating conditions and corroborating the behavior observed in Figure 6.

Table 2. Results of the model (testing trips).

Metric	Variable	TRIP 2	TRIP 15	TRIP 16	TRIP 22	TRIP 28
R^2	CO_2 (kg)	0.9144	0.9985	0.9984	0.9144	0.9980
RMSE		2.4413	0.3763	0.3812	2.4413	0.4127
MSE		5.9601	0.1416	0.1453	5.9601	0.1703
MAE		1.4602	0.2169	0.2151	1.4602	0.2214
Max Error		17.420	5.7910	7.6706	17.420	34.269
R^2_	CO (g)	0.3318	0.8006	0.7839	0.3318	0.8338
RMSE		0.0272	0.0252	0.0286	0.0272	0.0221
MSE		0.0007	0.0006	0.0008	0.0007	0.0005
MAE		0.0145	0.0098	0.0111	0.0145	0.0102
Max Error		0.2919	0.4126	0.4519	0.2919	0.2864
R^2_NO_x	NO_x (g)	0.9068	0.9677	0.9685	0.9068	0.9632
RMSE_NO_x		0.0206	0.0138	0.0133	0.0206	0.0140
MSE_NO_x		0.0004	0.0002	0.0002	0.0004	0.0002
MAE		0.0140	0.0083	0.0080	0.0140	0.0087
Max Error		0.1048	0.1297	0.1085	0.1048	0.2615
R^2_	T_{exh} (°C)	0.9962	0.9997	0.9998	0.9962	0.9998
RMSE		2.0123	0.5275	0.4742	2.0123	0.5192
MSE		4.0492	0.2783	0.2248	4.0492	0.2696
MAE		1.2557	0.3726	0.3538	1.2557	0.3512
Max Error		12.850	8.8490	14.311	12.850	35.904
R^2	$\dot{m}_{exh}$ (g/s)	0.8985	0.9612	0.9617	0.8985	0.9553
RMSE		0.0185	0.0136	0.0132	0.0185	0.0137
MSE		0.0003	0.0002	0.0002	0.0003	0.0002
MAE		0.0132	0.0086	0.0084	0.0132	0.0087
Max Error		0.1061	0.1070	0.1000	0.1061	0.1793

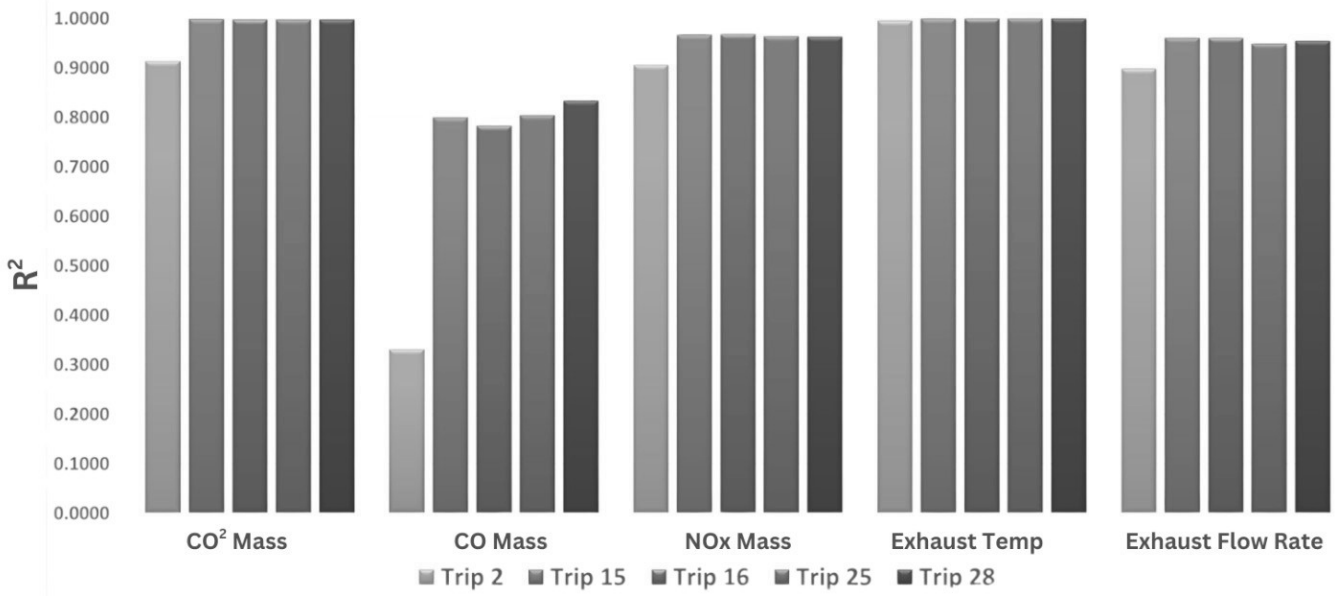

Figure 7. Model performance metrics (R^2) across validation Trips 2, 15, 16, 22, and 28, showing near-perfect prediction for CO_2 ($R^2 > 0.99$) and exhaust temperature.

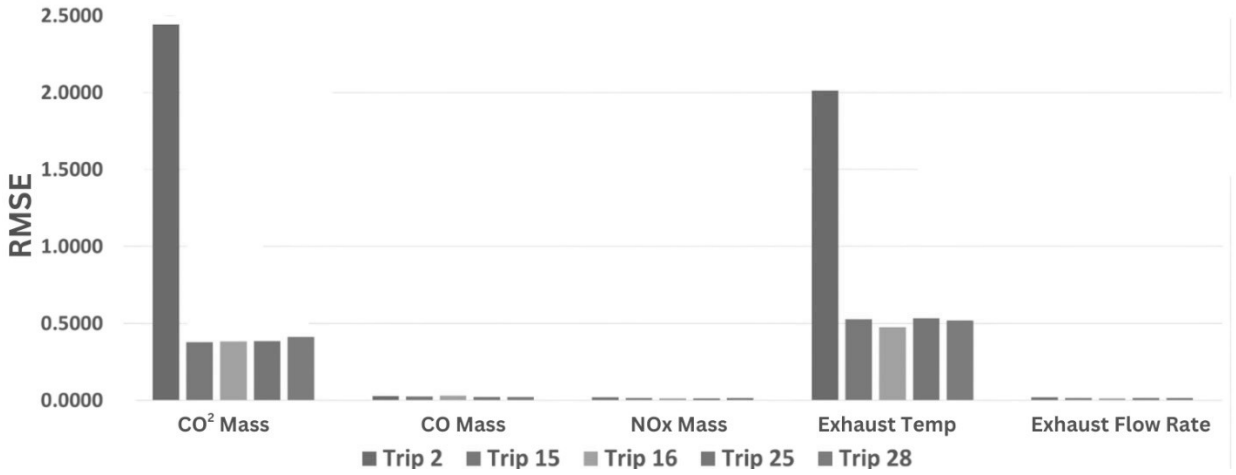

Figure 8. RMSE values for all targets across validation trips, highlighting lowest errors for exhaust temperature (2.01 °C) and CO_2 (2.44 g/s) in Trip 2.

Overall, the model achieves very high accuracy for CO_2, NO_x, and exhaust temperature, with R^2 values typically above 0.90 and often close to 0.99, while exhaust flow remains slightly lower but still strong. The model's high accuracy for CO2 predictions arises from its direct correlation with engine operating conditions (power and speed). Thus, CO_2 is the easiest output to model. Notably, a map-based approach from prior work already achieved strong results [26].

On the other hand, CO confirms its largest variability due to the very low amplitude of the signal and only short-duration spike. This makes R^2 for CO sensitive to minor deviations, even though absolute errors remain small.

The fidelity of the model is notably lower for Trip 2 compared with the other validation trips, particularly in the prediction of CO mass, which exhibits the lowest R^2 among all targets. Trip 2 represents an unloaded (no-payload) cold-start event at lower ambient temperature and shorter duration, producing lighter engine load and slower thermal rise compared to the payload trips used for training. Consequently, key input features like engine torque and coolant temperature gradient partially fell outside the training data's statistical domain. This distributional shift impaired Random Forest generalization, particularly for CO—whose low emission rates in diesel engines already challenge signal-noise separation.

The low CO signal amplitude inherent to light-load operation further amplifies the relative effect of measurement noise on the R^2 metric, magnifying apparent error. Overall, the poorer performance observed for Trip 2 is attributable to the combined effects of no-payload operation, colder start temperature, limited dynamic range, and feature distribution mismatch, whereas Trips 15, 16, 25, and 28, covering broader load and temperature conditions, align closely with the training distribution and therefore exhibit consistently higher prediction fidelity across all output variables.

In contrast, the predictions for CO_2, NO_x, and exhaust flow rate remain stable, each achieving approximately $R^2 = 0.90$, indicating that the model can still accurately reproduce the dominant combustion and flow trends even under unloaded conditions.

To sum up, the Random Forest model showed strong predictive performance on unseen trips. Predictions closely followed measured trends for CO_2, NO_x, exhaust temperature, and exhaust flow, with only minor deviations during sharp transients, while CO retains the expected sensitivity to narrow, high-amplitude excursions. Phase alignment was tight, confirming that temporal features capture the relevant dynamics, within the bounds of the reference vehicle dataset.

These results justify the use of the model to address scenario-based warm-up comparisons and hybridization studies. RMSE values are generally small, confirming that prediction errors are well bounded.

Comparison with the Previous Model (Map-Based)

Compared with the previous quasi-static approach [26], the ML model demonstrates a clear improvement in accuracy. Figure 9 shows that the proposed model reduces the relative error on exhaust temperature from 21% to 10% and guarantees a very accurate estimation of the average exhaust flow rate (+1%), compared with the quasi-static model (−36%).

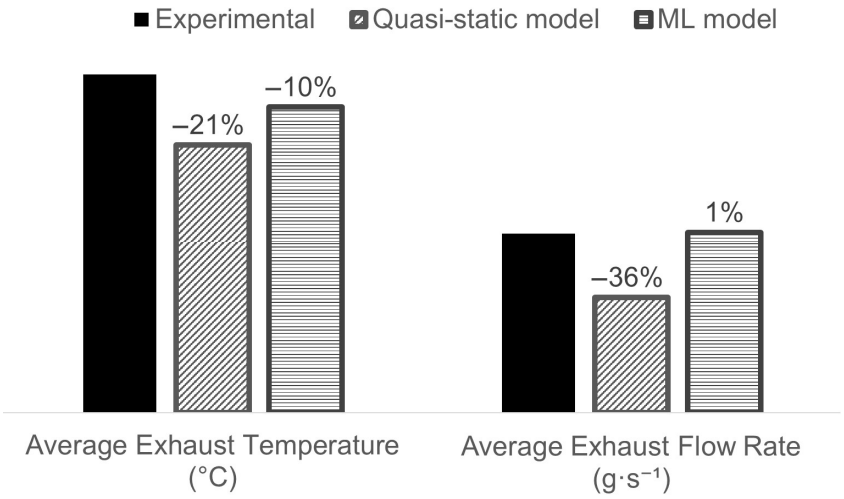

Figure 9. Relative accuracy comparison of quasi-static (map-based) vs. ML model for average exhaust temperature and mass flow rate across validation trips.

5. Results

The model developed and validated in the previous section was applied to estimate the additional fuel consumption and emissions resulting from cold-start operation and to predict the emissions of a virtual hybrid electric configuration of the same vehicle.

5.1. Cold Start vs. Hot Start

The model developed and validated in the previous section was applied to estimate the additional fuel consumption and emissions resulting from cold-start operation. To this end, we introduced the Matching Coolant Temperature Time (MCTT), defined as the time at which a cold-start or mid-start trip reaches a coolant temperature of 80 °C, as determined from the analysis of the experimental data.

The computed MCTT for cold-start and mid-start trips in the dataset is shown in Figure 10 as a function of the initial coolant temperature. Trips are classified according to payload, with the size of the bubbles representing the payload value.

MCTT ranges from a minimum of 206 s for mid-temperature Trip 11 to a maximum of 646 s for Trip 29. For trips with the highest payload (3000 kg), MCTT decreases almost linearly with increasing initial coolant temperature. The minimum of 206 s is reached by mid-temperature Trip 11, thanks to the favorable initial condition of the engine.

However, this trend is not consistent across other payloads. For example, Trip 1 shows one of the highest MCTT values despite a high initial coolant temperature, and its MCTT exceeds that of Trip 5, even though both trips have the same payload and Trip 1 has a higher initial temperature. The two trips with a 1500 kg payload are too similar in initial coolant temperature to reveal a clear trend.

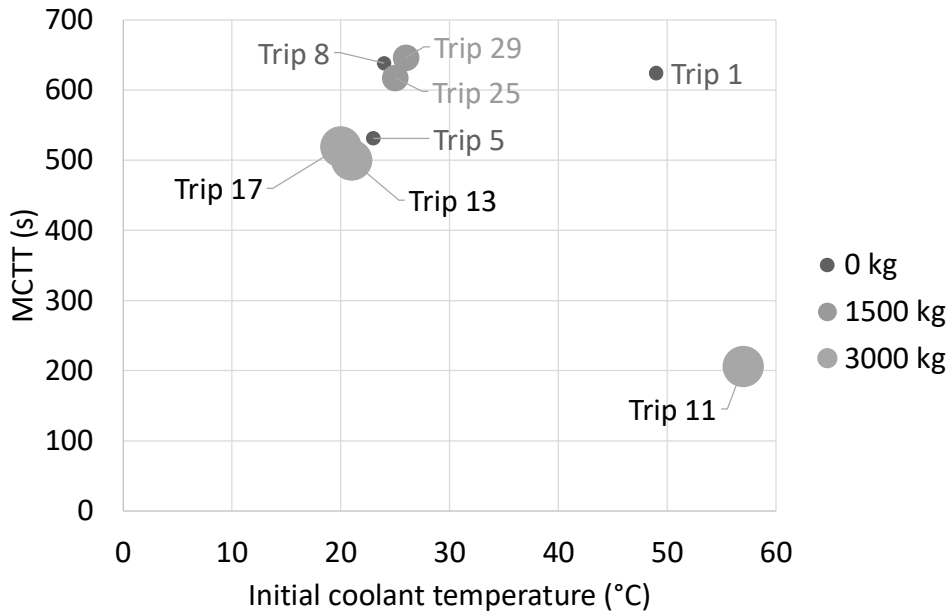

Figure 10. Matching Coolant Temperature Time (MCTT) to 80 °C for cold/mid-start trips vs. initial coolant temperature; bubble size indicates payload (0–3000 kg).

Lower payloads would be expected to impose less load on the engine, resulting in slower warming. However, this trend is not clearly observed in the experimental data. For instance, Trip 8 exhibits nearly the same MCTT as Trips 29 and 25.

In summary, the experimental data shows that MCTT is not directly correlated with either coolant temperature or payload and is influenced also by the specific differences in terms of speed pattern that can derive from other boundary conditions, like the level of traffic. This observation justifies the use of the ML model, which takes vehicle speed and acceleration as inputs, to analyze the effect of cold-start operation on engine outputs. As pointed out in the analysis of the scientific literature.

Cold-start operation has a significant impact on engine fuel consumption and emissions. In particular, it generally leads to an increase in fuel consumption and CO_2 emissions due to the engine operating at suboptimal temperatures and reduced combustion efficiency. The effect on NO_x emissions, however, is more complex and depends critically on the coolant temperature.

Based on the thermal mechanism of NO_x formation, NO_x formation is strongly influenced by the local gas temperature within the cylinder, as higher temperatures promote the formation of nitrogen oxides [20]. During a cold start, the lower initial coolant temperature produces a cooler overall mixture, which tends to suppress NO_x formation. However, the cold engine also experiences longer ignition delays, which lead to a more rapid and intense rate of heat release once combustion occurs. This localized temperature peak can offset the initial cooling effect, resulting in higher NO_x emissions despite the lower overall engine temperature.

Therefore, the net effect of cold-start conditions on NO_x emissions is the result of two competing mechanisms: the cooling of the mixture, which tends to reduce NO_x, and the delayed, more intense combustion, which tends to increase it. Understanding this interplay is critical for accurately predicting emissions under cold-start conditions and for designing strategies to mitigate their impact, such as optimized ignition timing, pre-heating of the intake charge, or improved thermal management of the engine.

Note that the reference vehicle does not include an aftertreatment device that could potentially be affected by the lower temperature of the engine exhaust.

To quantify these effects, the model was applied to each cold-start trip in two scenarios: first, using the measured coolant temperature, and then using a constant coolant temperature of 80 °C to simulate the same trip under hot-start conditions. Trips 5, 8, 29, and 17 were selected for detailed analysis, with attention focused on the first few minutes of each trip. This is because the impact of cold-start conditions diminishes as the engine reaches normal operating temperature, making the effect negligible over the entire trip, as noted in Ref. [26]. The results are reported in the following plots.

Figure 11 compares cold-start and hot-start operation under identical driving conditions for Trip 5, highlighting differences in exhaust temperature, CO_2 and NO_x rates, and cumulative emissions. Hot-start begins with a much higher exhaust temperature, whereas cold-start remains 30–60 °C lower for extended periods. This thermal advantage translates directly into the emission-rate channels. The NO_x mass trace under a hot start shows visibly attenuated peaks and faster decay following each acceleration event, resulting in a cumulative NO_x value consistently below that of the cold-start run. Although instantaneous CO_2 mass occasionally rises during short load pulses in the hot-start scenario, the overall cumulative CO_2 grows more slowly. By the end of the 700 s window, cumulative CO_2 under hot start is lower than for cold start. Exhaust-flow traces largely follow the speed/torque profile in both cases; minor increases associated with the higher hot-start temperature do not offset the reduction in integrated emissions.

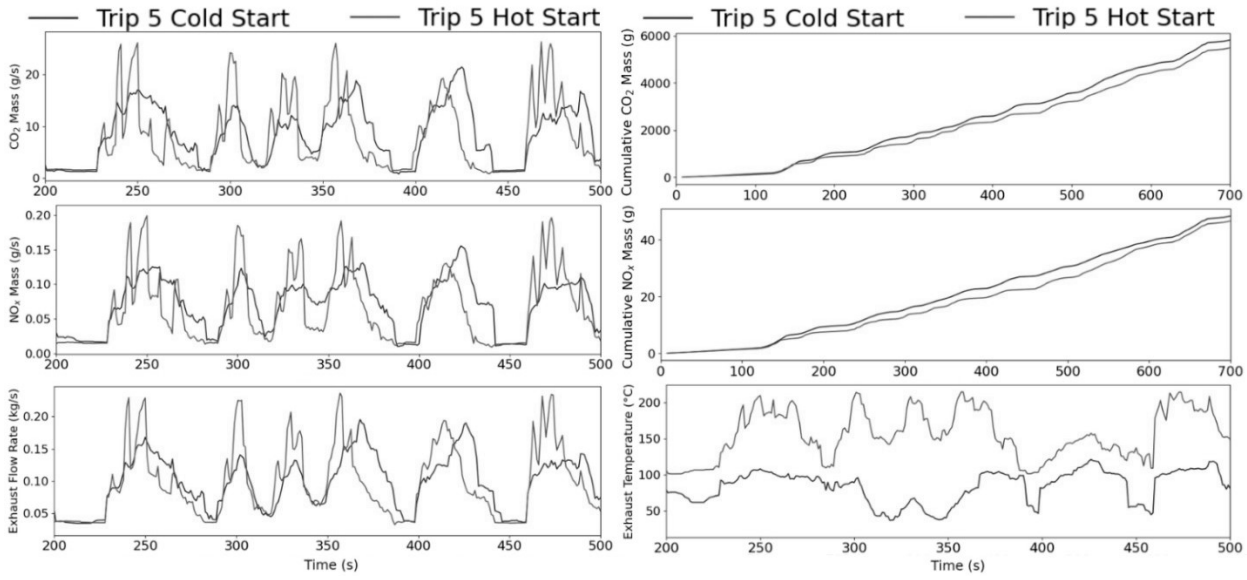

Figure 11. Cold start (measured coolant) vs. hot start (constant 80 °C) under identical kinematics (Trip 5)—CO_2/NO_x rates, exhaust temperature/flow, cumulatives.

Figure 12 shows Trip 8, which exhibits the same qualitative behavior with even clearer penalization in the total emissions of NO_x and CO_2. The hot-start exhaust temperature rapidly reaches and maintains a higher plateau, whereas the cold-start temperature remains lower and recovers more slowly after decelerations. NO_x peaks are consistently suppressed in the hot-start run, and the inter-peak baseline is lower, yielding a cumulative NO_x curve that remains below the cold-start counterpart throughout. CO_2 mass rates again show localized instances where hot-start temporarily matches or slightly exceeds cold-start during brief load changes; however, the cumulative CO_2 remains meaningfully lower under hot-start across the analyzed period. Exhaust-flow series are closely aligned as expected, with minor variations secondary to the thermal effect on conversion efficiency captured by the model.

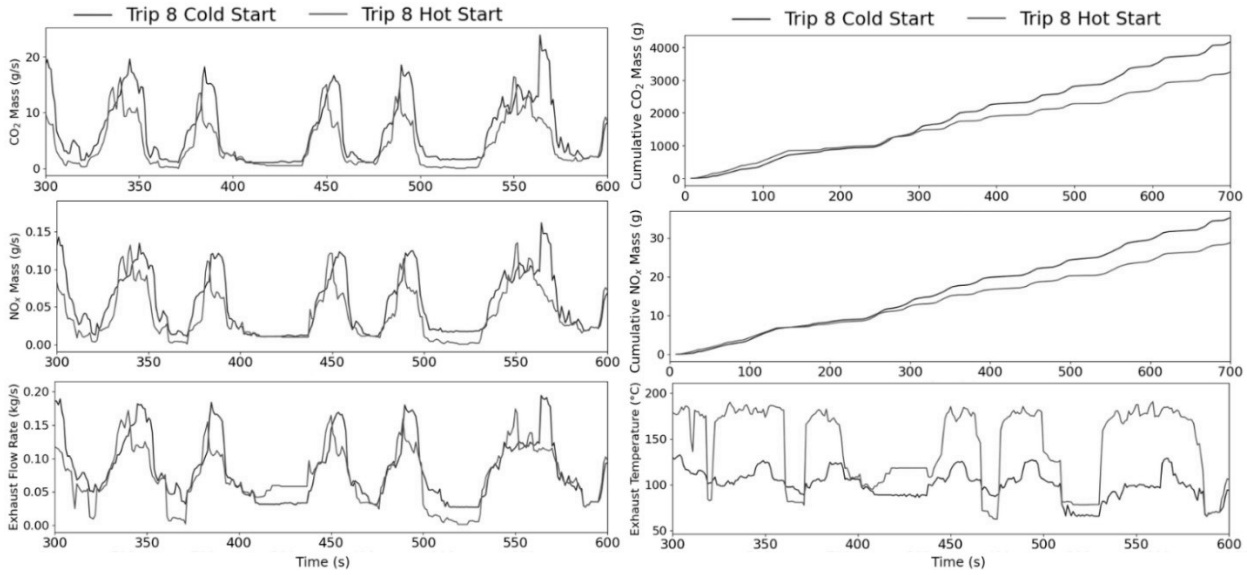

Figure 12. Trip 8: Cold start vs. hot start under identical kinematics—CO_2/NO_x rates, exhaust temperature/flow, cumulatives.

Similar trends are observed for Trips 17 and 29, confirming the generality of the patterns identified in Trips 5 and 8.

In terms of total CO_2 and NO_x emissions, all trips except 1, 11, and 13 exhibit a pronounced increase under cold-start conditions, as shown in Figure 13. The largest increases—up to approximately 28% for CO_2 and 30% for NO_x—are observed for Trip 17.

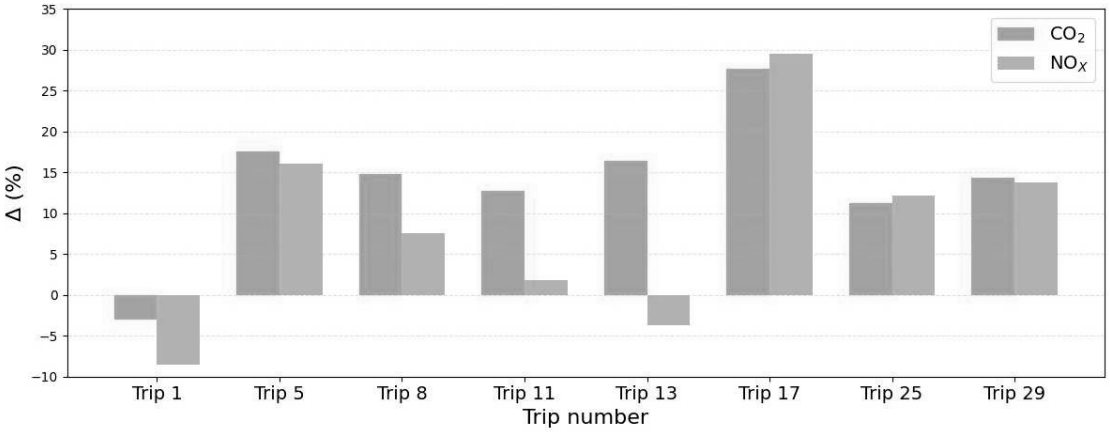

Figure 13. Relative variation (%) in cumulative CO_2 (orange) and NO_x (gray) emissions within the MCTT window, from hot-start to cold-start operation.

For Trip 1, the model predicts an increase in both CO_2 and NOx, whereas for Trip 11, NO_x emissions are largely unaffected by the initial coolant temperature. Trip 13 shows only a slight increase in NO_x. These deviations from the general trend can be explained by specific factors: Trips 1 and 11 had relatively high initial coolant temperatures (see Figure 2), which limited the thermal disadvantage of a cold start. For Trip 13, the limited MCTT combined with the competing thermal and combustion mechanisms described earlier accounts for the minor effect on NO_x emissions.

Figure 14 reports the average exhaust mass flow rate and temperature during the Matching Coolant Temperature Time (MCTT). Each value is computed over the MCTT, ensuring that cold-start and hot-start runs are compared over an equivalent thermal interval.

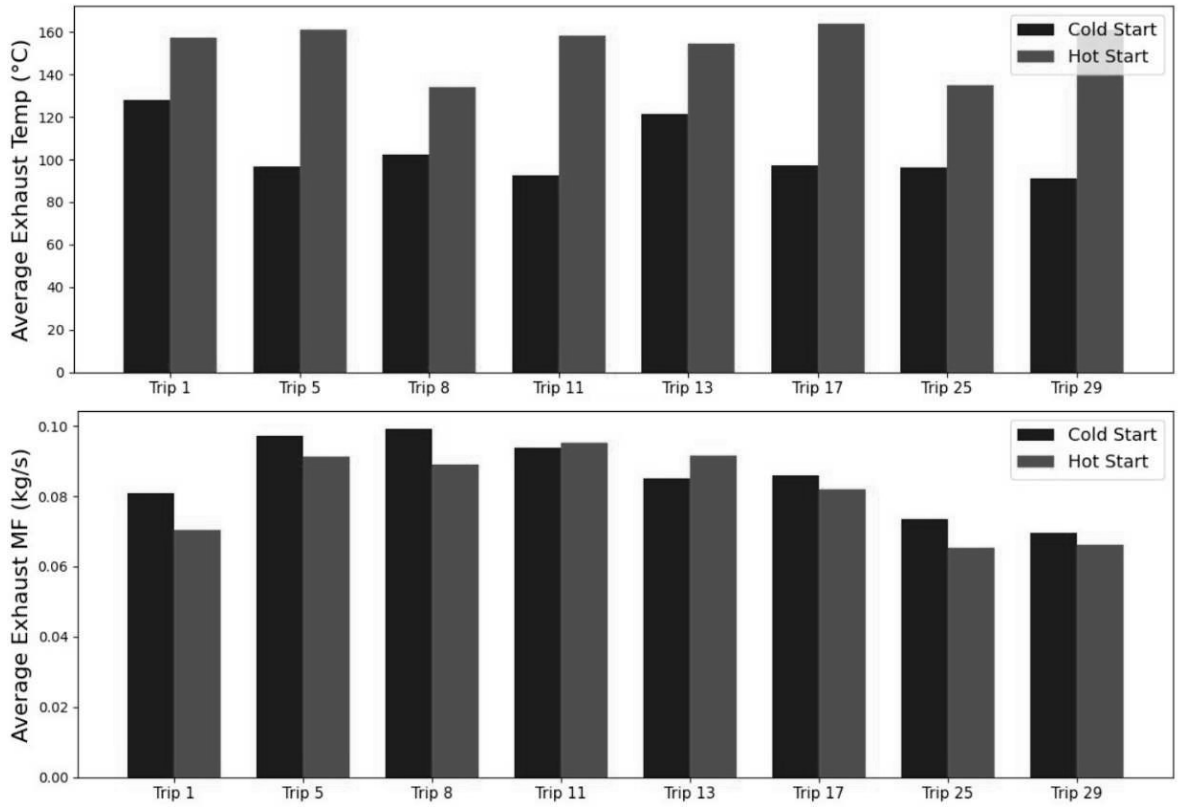

Figure 14. Average exhaust temperature/flow during MCTT for cold start (blue) vs. hot start (red) across trips.

Cold-start operation results in a reduction of exhaust temperature by 30–70 °C compared to hot-start conditions, while the average mass flow rate is only minimally affected. These findings are particularly relevant for the design of waste heat recovery (WHR) systems, as pursued in the IRIDESCENT project, since lower exhaust temperatures can directly impact the efficiency and integration of WHR technologies [28].

5.2. Hybridization of the Isuzu FTR850

In the framework of the IRIDESCENT project, the trained ML model was also used to compare the original thermal topology (TH) of the Isuzu FTR850 with a hybrid electric version inclusive of waste heat recovery that was proposed and designed in previous investigations [31].

Figure 15 illustrates the parallel hybrid electric configuration in which the electric machine operates at the same rotational speed as the engine. The speeds of both machines are assumed to match those measured onboard in the TH configuration. The Waste Heat Recovery (WHR) unit consists of an Organic Rankine Cycle (ORC) system equipped with a shell-and-tube heat exchanger (HEX). A battery is used to supply power to the electric machine when it functions as a motor and to generate power for the auxiliaries together with the WHR. The battery is recharged onboard when the electric machine operates as a generator, such as during regenerative braking. This configuration is referred to as "HEV-W."

The performance of the hybrid electric power system was benchmarked against the original configuration over a representative daily working cycle, obtained by combining four trips from the database, as described in Table 3. The battery is assumed to be recharged overnight.

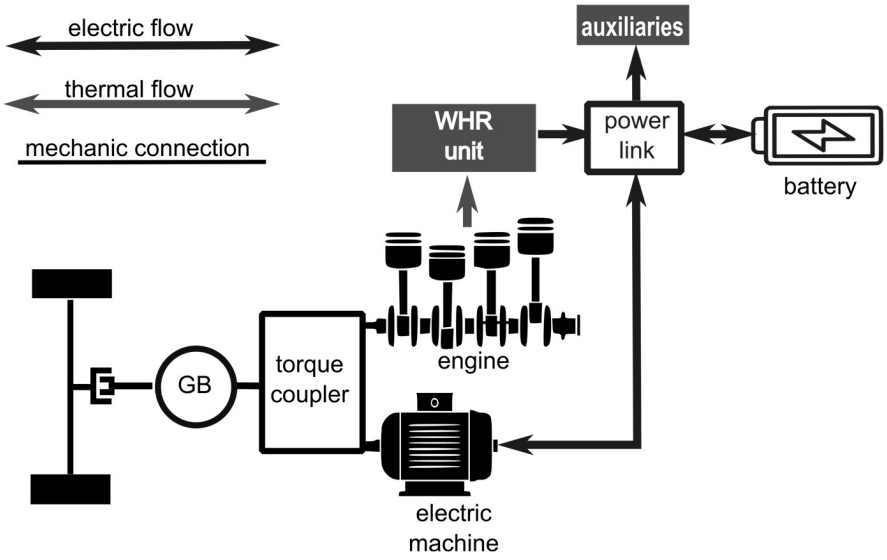

Figure 15. Parallel hybrid-electric powertrain with ORC-based WHR for Isuzu FTR850 (GB: gearbox).

Table 3. Daily working cycle hybridization: four concatenated trips with payloads/speed profiles.

Daily Working Cycle	Speed Profile	Payload
Section 1	Trip 4	0 kg
Section 2	Trip 14	3000 kg
Section 3	Trip 23	1500 kg
Section 4	Trip 4	0 kg

Two heuristic energy management strategies were proposed and compared in Ref. [32]: a constant power strategy (CPEMS) and a fuzzy logic-based strategy (FLEMS). The first strategy aims to maintain the exhaust gas mass flow rate and temperature as constant as possible to maximize the energy recovered by the WHR unit, whereas the second strategy focuses on preserving the battery state of charge (SOC) and state of health (SOH).

For both strategies, the component sizes and EMS thresholds were optimized using a quasi-static approach based on efficiency maps for the engine and electric motor, together with an equivalent electrical circuit model for the battery. The WHR system was also modeled using quasi-static performance maps.

The results of the analysis presented in Ref. [32] are reported in Table 4, along with the corresponding battery and engine sizes. The results include the battery depth of discharge (DOD), fuel consumption, and CO_2 emissions evaluated using a tank-to-wheel approach. The well-to-wheel approach is not considered relevant in this context, as the emissions associated with battery charging are outside the scope of the analysis. In fact, the present study focuses exclusively on the thermal pathway, and therefore only the emissions directly related to the onboard energy conversion processes are considered.

The emissions of NO_x were not included in the previous analysis because the quasi-static approach is not well suited for predicting phenomena strongly influenced by the transient behavior of the engine [14]. The reliable estimation of NO_x emissions typically requires dynamic engine models capable of representing transient operating conditions. In fact, NO_x formation is highly sensitive to rapid variations in engine load, temperature, air–fuel ratio, and combustion dynamics, which cannot be accurately captured using steady-state efficiency maps. Unlike the quasi-static approach, which relies on stationary maps and cannot capture pollutant emissions, the ML model accounts for history-dependent dynamics, as demonstrated in the validation phase. Therefore, it can be used to estimate NO_x emissions.

Table 4. Specifications of the optimized hybrid configurations vs. the original thermal powertrain and results obtained with the previous model.

	Thermal	CPEMS	FLEMS
P_e^{SP} (kW)	6-cyl	4-cyl	6-cyl
Battery size (kWh)	-	48	20
battery DOD (%)	-	61.8	61.7
FC (kg)	60.9	34.0	45.3
FC (L/100 km)	28.8	16.7	23.1
$CO_{2,TTW}$ (kg)	191.8	107.2	135.6
$CO_{2,TTW}$ (g/km)	781.8	436.8	537.6

In this investigation, the model developed and validated above was used to compare the TH and HEV-W configurations with the two energy management strategies. Note that in terms of inputs for the ML model, the three cases of Table 4 differ only for the engine torque signal, being the same as the other signals (ambient and coolant temperature, engine speed, and vehicle payload). The outputs of the model in terms of exhaust gas properties and engine emissions are discussed below.

Using the same speed profiles as the original powertrain in the dataset was a deliberate choice to isolate the comparison from the influence of driver behavior and transmission control strategies. This assumption ensures consistency across simulations and allows the analysis to focus specifically on the effects of powertrain architecture and hybridization. However, this simplification may lead to an underestimation of the potential efficiency gains that are achievable with hybrid systems. In particular, the benefits associated with optimized gear-shifting strategies, adaptive gear ratios, and decoupled engine operation are not fully captured under this constraint. In real-world hybrid applications, the ability to operate the engine in more efficient regions—independently of wheel speed—can yield additional fuel economy and emission improvements. The development and integration of advanced gearbox control strategies, including predictive and optimization-based approaches, are therefore identified as key areas for future work, as they are expected to further enhance the performance and realism of the hybrid powertrain model.

Using the cumulative CO_2 values predicted by the ML model, the estimated fuel consumption for the hybrid and non-hybrid configurations was calculated as reported in Table 5 for the TH and HEV-W configurations with CPEMS and FLEMS strategies.

Table 5. Estimated emissions obtained with proposed ML mode.

Configuration	CO_2 (kg)	CO (kg)	NO_x (kg)
TH	223.12	0.772	1.824
HEV-W + CPEMS	198.20	1.133	1.743
HEV-W + FLEMS	199.97	0.748	1.688

In terms of emissions, the ML mode estimated a reduction of CO_2 and NO_x in the HEV-W with both EMSs compared with the original thermal case TH, the reduction stemming from engine efficiency improvement. Particularly, the cumulative CO_2 mass reached 198.20 kg and 199.97 kg for the hybrid with CPEEMS and FLEMS, respectively, and 223.12 kg for the non-hybrid configuration, representing an absolute reduction of about 11% in the CPEMS case. Compared with the map-based model, the advantage of hybridization in terms of total CO_2 estimated with the ML approach is less conspicuous, as can be noticed from Figure 16 where the two approaches are compared. In both cases, the CPEMS strategy allows for the highest reduction of emissions since the engine is operated at constant power.

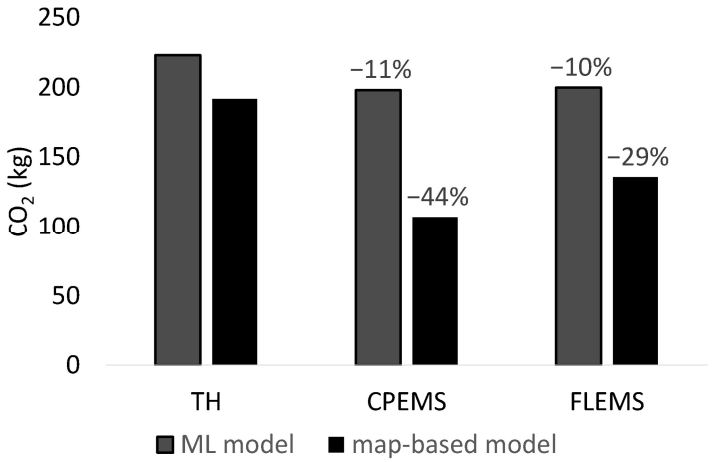

Figure 16. Cumulative CO_2 emissions: comparison of ML (red) and map-based (black) models across the thermal configuration (TH) and hybrid control strategies (CPEMS and FLEMS).

The NO_x mass exhibits the same directional behavior but with a smaller magnitude; the hybrid configuration attains cumulative values of 1.69 kg and 1.74 kg for the FLEMS and CPEMS strategies, compared with 1.82 kg for the non-hybrid case. In this case, the best result (−7% than TH case) is obtained with the FLEMS control, while with the CPMES strategy, the emissions of NO_x are only −4% lower than the non-hybrid case. This behavior is expected because the fuzzy-logic controller implemented in the FLEMS strategy promotes smoother engine operation, with reduced rapid fluctuations in engine load, unlike the CPEMS approach, where torque varies continuously with engine speed(see Figure 17).

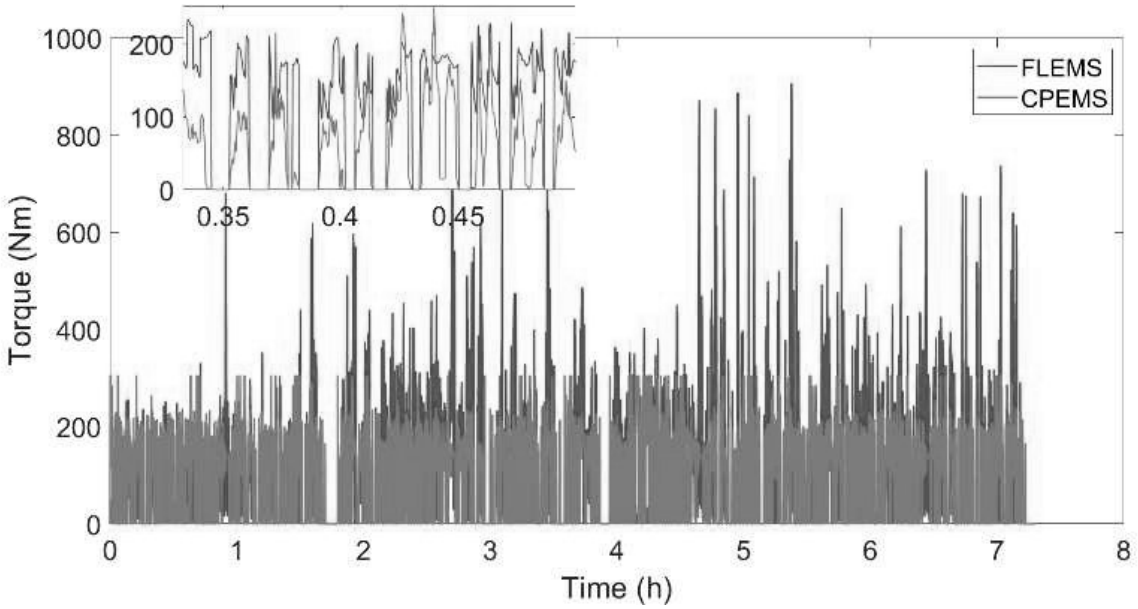

Figure 17. Engine torque time series for FLEMS vs. CPEMS (25 kW, six-cylinder).

The average exhaust temperature and mass flow rate predicted by the ML model for the hybrid and non-hybrid configurations are summarized in Figure 18. In the hybrid electric cases, the model estimates slightly lower mean exhaust temperature and exhaust mass flow rate compared with the TH case. These reductions reflect the lower combustion load and thermal output associated with partial electric torque support during propulsion. Larger variations are obtained in terms of standard deviation. In this case, the CPEMS strategy presents the lowest values of standard deviation in both exhaust mass flow rate

and temperature. In fact, the rationale behind the CPEMS is that iso-power curves in engine maps nearly coincide with curves of constant exhaust temperature and mass flow rate, as demonstrated in Figure 19.

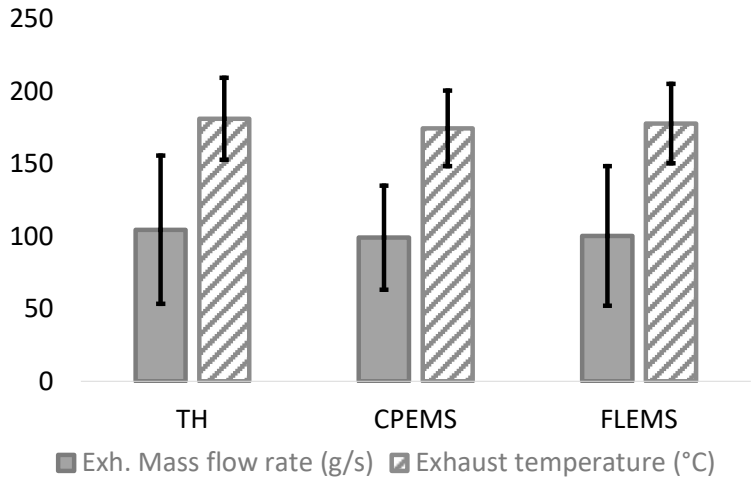

Figure 18. Average exhaust temperature/flow (ML model) for TH, CPEMS/FLEMS hybrids; error bars = standard deviation.

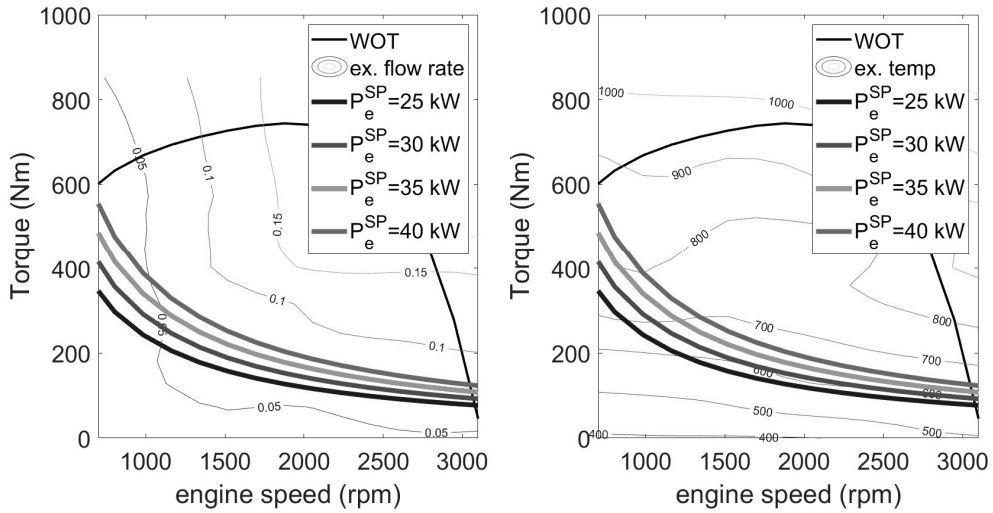

Figure 19. Engine maps with iso-power curves overlaid on exhaust temperature (**left**) and mass flow (**right**); constant power ≈ constant exhaust conditions.

Taken together, these results reinforce that the trained ML model plausibly captures the effect of hybridization and the differences in the energy management strategies. The slightly lower exhaust temperature and flow levels predicted for the hybrid configuration also highlight an important consideration for the design of waste-heat recovery and aftertreatment systems, which must account for reduced thermal energy availability under hybrid operating conditions [28]. However, the validation of the results is not possible because hybridization is not actually implemented in the vehicle. Moreover, due to the limitations of the available dataset—which does not include measurements of particulate matter (PM/soot)—we could not evaluate the effect of hybridization on the NO_x–PM trade-off, a fundamental issue in diesel combustion [20]. Therefore, this section should be regarded as a use-case demonstration of the integration between map-based and ML-based models.

6. Limitation of the Present Work

This study is subject to several limitations that should be considered when interpreting the results and assessing the general applicability of the proposed framework.

First, the dataset comprises 26 trips collected from a single instrumented vehicle operating on a fixed route and driven by the same driver. The current model implicitly captures driving style and gear-shifting effects through input variables such as vehicle speed, acceleration, engine speed, and load-related quantities, which reflect the combined outcome of driver behavior and transmission control. Therefore, the model does not explicitly generalize across different driving styles. While this ensures consistency and reduces variability due to external factors, it inherently limits the representativeness of the dataset. Consequently, the trained model cannot be assumed to generalize across different vehicle classes, engine technologies, aftertreatment systems, driving styles, or duty cycles. Extending the applicability of the proposed methodology requires validation on multi-vehicle, multi-route datasets encompassing a broader range of real-world operating conditions.

Second, although the model demonstrates excellent predictive accuracy on held-out trips, its performance is conditioned on the statistical domain of the training data. As observed for specific cases (e.g., low-load cold-start conditions), distributional shifts in key input features may degrade prediction fidelity. This highlights the need for domain adaptation strategies or enriched training datasets to improve robustness under extrapolation scenarios.

Third, the study relies on a data-driven (black-box) modeling approach, which, despite offering high accuracy, provides limited direct physical interpretability compared with physics-based models. While feature importance and ablation analyses partially address this issue, the lack of explicit physical constraints may limit confidence in predictions outside the observed data range.

Fourth, the analysis of hybridization effects is based on a virtual implementation derived from previously developed quasi-static models, and the resulting benefits are estimated through the ML surrogate rather than measured experimentally. Therefore, the reported reductions in CO_2 and NO_x emissions should be interpreted as model-based estimates, and experimental validation on a real hybridized platform is required to confirm these findings.

Fifth, the absence of particulate matter (PM/soot) measurements in the dataset prevented the inclusion of one of the most critical diesel pollutants in the analysis. This limitation reduces the comprehensiveness of the emissions assessment, particularly under cold-start conditions where PM formation can be significant, and restricts the evaluation of hybridization benefits, as it precludes discussion of the NO_x–PM trade-off typical of diesel engines.

Finally, although the framework is conceptually extendable to alternative fuels (e.g., biofuels or e-fuels), no such generalization is demonstrated in this work, and additional datasets would be required to ensure validity across different fuel chemistries and combustion behaviors.

7. Conclusions

This study presented a data-driven, machine-learning framework for the prediction of transient emissions and exhaust thermal behavior in a heavy-duty diesel vehicle under real driving conditions, with specific focus on cold-start operation and hybridization scenarios.

The proposed Random Forest-based model demonstrated high predictive accuracy across multiple outputs, including CO_2, NO_x, exhaust temperature, and exhaust mass flow rate, with coefficients of determination approaching unity for most variables. The adoption of causal temporal features and trip-level validation ensured methodological

robustness and avoided data leakage, supporting the reliability of the reported performance. Compared with a previous quasi-static approach, the model significantly improved the prediction of exhaust thermal variables, which are critical for both emissions analysis and waste heat recovery applications.

The analysis of cold-start operation confirmed that emission penalties are concentrated in the early phase of the driving cycle, typically within the first few minutes, where suboptimal thermal conditions lead to increased fuel consumption and pollutant formation. The results quantified increases of up to approximately 28% in CO_2 and 30% in NO_x in the first minutes of the trip, highlighting the importance of thermal management strategies in reducing real-world emissions.

The application of the model to a virtual hybrid-electric configuration confirmed the reduction of CO_2 emissions and allowed for the estimation of NO_x savings with hybridization, primarily through reduced engine load variations. However, the model also predicted lower exhaust temperature and mass flow rates, indicating reduced thermal energy availability. This finding has important implications for the design and control of waste heat recovery and aftertreatment systems, which must be adapted to hybrid operating conditions.

Overall, the proposed framework provides a computationally efficient and deployment-oriented tool for estimating the real-time emissions and exhaust properties of a specific vehicle using only standard onboard signals. Its ability to capture non-linear and history-dependent dynamics makes it particularly suitable for hybrid powertrain design, thermal management optimization, and onboard emission monitoring.

Future work should focus on:

- Extending the dataset to multiple vehicles and operating conditions.
- Exploring hybrid modeling approaches that combine data-driven and physics-based methods.
- Developing a gearbox control strategy to enhance the performance and realism of the hybrid powertrain model.

These developments would further enhance the robustness, interpretability, and practical applicability of the proposed methodology for next-generation low-emission heavy-duty vehicles.

Author Contributions: Conceptualization, T.D. and T.M.; methodology, T.D. and T.M.; software, T.M.; validation, T.D., T.M. and P.M.; formal analysis, T.D.; investigation, T.D., T.M. and P.M.; resources, T.D.; data curation, T.M.; writing—original draft preparation, T.D. and T.M.; writing—review and editing, T.D., T.M. and P.M.; visualization, T.M.; supervision, T.D. and P.M.; project administration, T.D. and P.M.; funding acquisition, T.D. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 1409, published on 14 September 2022 by the Italian Ministry of University and Research (MUR) and funded by the European Union NextGenerationEU Project Title IRIDESCENT—P2022YB8HY CUP F53D23009880001 (Grant Assignment Decree No. 5) adopted on 1 September 2023 by the Italian Ministry of University and Research (MUR).

Data Availability Statement: Data will be made available on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
BMEP	Brake Mean Effective Pressure
BMIFS	Backward Maximum Information Feature Selection
BP	Backpropagation (neural network training)
CNN	Convolutional Neural Network
CNG	Compressed Natural Gas
CO	Carbon Monoxide
CO2	Carbon Dioxide
CR	Correlation Coefficient
DOC	Diesel Oxidation Catalyst
DNN	Deep Neural Network
ECU	Engine Control Unit
EGR	Exhaust Gas Recirculation
EHC	Electrically Heated Catalyst
EWMA	Exponentially Weighted Moving Average
FTR	Forward Tilted Radiator
GB	Gear Box
GBRT	Gradient Boosted Regression Trees
GCV	Generalized Cross-Validation
GRU	Gated Recurrent Unit
GWO	Grey Wolf Optimization
HC	Hydrocarbons
HDDV	Heavy-Duty Diesel Vehicle
HDDE	Heavy-Duty Diesel Engine
IEA	International Energy Agency
LASSO	Least Absolute Shrinkage and Selection Operator
LSTM	Long Short-Term Memory Network
MAE	Mean Absolute Error
MAP	Manifold Absolute Pressure
MAPE	Mean Absolute Percentage Error
MCTT	Matching Coolant Temperature Time
MLP	Multilayer Perceptron
ML	Machine Learning
MoE	Mixture of Experts
MPC	Model Predictive Control
MSE	Mean Squared Error
nRMSE	Normalized Root Mean Squared Error
NO	Nitric Oxide
NO_2	Nitrogen Dioxide
NO_x	Nitrogen Oxides
OBD	On-Board Diagnostics
PCA	Principal Component Analysis
PEMS	Portable Emissions Measurement System
PM	Particulate Matter
PSO	Particle Swarm Optimization
R^2	Coefficient of Determination
RDE	Real Driving Emissions
RF	Random Forest
RMSE	Root Mean Squared Error

SCR	Selective Catalytic Reduction
SHAP	Shapley Additive Explanations
SVM	Support Vector Machine
TCN	Temporal Convolutional Network
TFT	Temporal Fusion Transformer
TSF	Time-Series Forecasting (Transformer variant)
VVA	Variable Valve Actuation
VVT	Variable Valve Timing
WHR	Waste Heat Recovery
WLTP	Worldwide Harmonized Light Vehicles Test Procedure
WHTC	World Harmonized Transient Cycle

List of Symbols

The following symbols are used in this manuscript:

t	Time (s)
$P_{\mathrm{w}}(t)$	Wheel power demand (W)
v	Vehicle speed (m·s−1)
ρ	Air density (kg·m−3)
C_d	Drag coefficient (–)
A_f	Frontal area (m2)
m	Vehicle mass (kg)
g	Gravitational acceleration (m·s−2)
α	Road gradient (–)
a	Acceleration (m·s−2)
T_e	Engine torque (N·m)
$\omega_e(i)$	Engine speed (rad·s−1)
$\eta_{driveline}$	Driveline efficiency (–)
$P_{aux}(i)$	Auxiliary power (W)
$\dot{m}_{fuel}$	Fuel mass flow rate (g·s−1)
$\dot{m}_{\mathrm{exh}}$	Exhaust mass flow rate (kg·s−1)
T_{exh}	Exhaust temperature (°C)
$\dot{m}_{\mathrm{CO_2}}$	CO2 mass rate (g·s−1)
$\dot{m}_{\mathrm{NO}_x}$	NOx mass rate (g·s−1)
$\dot{m}_{\mathrm{CO}}$	CO mass rate (g·s−1)
ϵ	Prediction error (–)
$\hat{y}$	Model prediction (–)
$\overline{y}$	Mean of measured value (–)

Appendix A. Detailed Analysis of the Scientific Literature

Data-driven models have often been used as surrogate estimators of pollutant mass rates and exhaust thermal variables when synchronized measurements are available. Recent studies on diesel engines show that sequence models can capture time dependence in transient NO_x prediction, including CNN–LSTM variants and other recurrent architectures trained on dynamometer or transient-cycle data [18,33,34]. These approaches typically improve the fit by encoding short-term memory and non-linear interactions among operating variables.

For heavy-duty applications, deep learning has also been applied to engine-out and tailpipe NO_x prediction using richer sensing that includes engine and aftertreatment variables [23]. When PEMS data are available, dimensionality reduction and recurrent models have been proposed for heavy-duty diesel vehicles under regulatory transient tests, with reported performance dependent on the measured signal set and the evaluated cycles [21,35]. In parallel, tree ensembles such as Random Forest and gradient-boosted trees are widely used as strong baselines for emissions regression because they handle non-linearities and

mixed-scale predictors with limited tuning effort, which is beneficial when datasets are moderate in size [36]. A complementary line of work focuses on deployment-oriented "virtual sensing" using OBD-accessible predictors. For example, ML-aided remote monitoring has been explored using OBD data streams to support high-emitter identification, and XGBoost-based NO_x estimators have been demonstrated under controlled conditions [36,37]. Virtual sensors can achieve high agreement when additional instrumentation is available (for example, in-cylinder pressure), but such signals reduce applicability at fleet scale [24]. Beyond pollutants, data-driven time-series models have also been proposed for exhaust gas temperature estimation under transients, supporting the feasibility of predicting exhaust thermal state variables when history dependence is represented explicitly [38].

Table A1 summarizes representative cold-start and transient-emission models and highlights that reported metrics are not directly comparable across studies because targets, units, sensors, and duty cycles differ.

Table A1. Data-driven models for cold-start and transient emission prediction (reported results as stated in the cited studies; metrics and units are study-specific and not directly comparable).

Model Type	Engine/Context	Target Pollutants/Variables	Accuracy/Performance Metrics	Ref.
ANN (MLP)	Euro 6 diesel, real-world cold start	CO_2, NO_x, CO, THC	MAE, RMSE not explicitly stated; good cold-start alignment	[39]
ANN (MLP)	Euro 6 diesel, RDE cycles (cold + warm)	NO_x	MSE = 0.00043; RMSE = 0.02084; CR = 0.73	[40]
DNN	Heavy-duty diesel, FTP/RMC cycles	Engine-out and tailpipe NO_x	$R^2 > 0.99$ (engine − out and tailpipe, engine dyno); $R^2 = 0.97$ (engine − out chassis); $R^2 = 0.93$ (tailpipe chassis)	[23]
LightGBM	Light-duty diesel, RDE-type transient	Engine-out NO_x	$R^2 = 0.95$ (transient); $R^2 = 0.91$ (steady − state map); n RMSE = 0.04	[41]
CNN–LSTM	Turbocharged diesel, transient cycles	Engine-out NO_x	$R^2 = 0.977$; RMSE = 33.495	[18]

Two methodological implications follow from this literature. First, many high-performing models depend on high-dimensional sensing (injection parameters, aftertreatment internal states, or specialized measurements), whereas engineering use cases often require estimators based on a compact set of deployable signals. Second, time-series validation must be designed to avoid leakage, which can occur if feature construction uses future samples or if training and testing interleave samples from the same trip. These constraints motivate the present study's focus on (i) a deliberately restricted predictor set available from OBD streams and (ii) a trip-grouped evaluation protocol described later to obtain evidence-bounded performance on unseen trips from the same measurement campaign.

Table A2 shows that hybrid architectures and attention-based models have been applied to CO_2 and multi-pollutant prediction, again with performance strongly conditioned on the data-generating process and the available predictors.

Table A2. Advanced ML architectures for CO2 and multi-pollutant emission prediction (reported results as stated in the cited studies; targets and units vary by study).

Model Architecture	Target Variable(s)	Accuracy/Performance Metrics	Ref.
MSCL-Attention (multi-scale CNN–LSTM with self-attention)	Road-traffic CO_2 emissions (vehicles)	Error metric = 0.01172; coefficient of determination $R^2 = 0.99714$	[42]
CarbonMLP (deep fully connected neural network)	Vehicle CO_2 emissions (g/km)	MSE = 0.0002, RMSE = 0.0141, $R^2 = 0.9938$, MAPE ≈ 2.5%	[43]
LSTM (univariate daily CO_2 emissions)	Near-real-time daily CO_2 emissions	MSE = 3.5179×10^{-4}, RMSE = 0.0187, $MAE = 0.0140$, MAPE = 14.8291%, $R^2 = 0.9844$	[44]
LSTM (RNN for off-road machinery emissions)	Multi-pollutant: CO, NO, NO_2, NO_x (off-road diesel machinery)	RMSE for NO emission forecasting ranges from 56.37 (loader) to 143.66 (forklift)	[45]

A compact overview of temporal modeling families and typical performance ranges reported in the cited works can be found in Table A3.

Table A3. Overview of temporal modeling techniques for emissions prediction (numerical values taken from the cited studies; units depend on the study and target definition).

Model Name	Key Features	Performance	Numerical Results	Ref.
CNN–LSTM for transient NOx prediction	CNN captures spatial patterns; LSTM models temporal dynamics; trained on WHTC diesel data.	Outperforms LSTM, CNN, and BP models.	$MAE \approx 23.98$ mg; RMSE ≈ 33.50 mg; MAPE $\approx 18.4\%$; $R^2 \approx 0.977$; LSTM baseline RMSE$\approx$ 48 mg ($R^2 \approx 0.953$).	[18]
TCN–LSTM for maritime CO_2 emissions	TCN captures long-range temporal patterns; LSTM models sequential dependencies; trained on ship-engine operational data.	Outperforms standalone TCN, CNN, and LSTM models for CO_2 prediction.	$R^2 \approx 0.9726$; $MAE \approx 47.34$; RMSE ≈ 58.57; MAPE $\approx 4.23\%$; Pearson correlation ≈ 0.986.	[46]
LSTM vs. DNN (soot emission)	LSTM trained on WLTP diesel transient cycle; compared with fully connected DNN using engine and soot-sensor inputs.	LSTM captures transient soot behavior more accurately than DNN.	LSTM: $R^2 \approx 0.976$; DNN: $R^2 \approx 0.922$; MAE within 0.30–1.47% of peak soot values.	[47]
LSTM virtual NO_x sensor (physical test)	Minimal LSTM using only intake-air temperature and injection timing to replace physical NO_x sensors.	Strong agreement between predicted and measured NO_x.	$R^2 \approx 0.994$; numerical metrics otherwise not detailed.	[48]
CNN–Transformer (NH_3 emissions)	1D CNN extracts local patterns; Transformer decoder models long-range dependencies for diesel NH_3 prediction.	Outperforms LSTM, Random Forest, and standalone Transformer models.	$R^2 \approx 0.986$; $MAE \approx 0.663$; MSE ≈ 2.285.	[49]
BMIFS–LSTM and BMIFS–CNN–LSTM (SCR-inlet NO_x)	BMIFS removes redundant inputs; LSTM and CNN–LSTM model the remaining emission sequence for SCR-inlet NO_x.	BMIFS–LSTM reduces error by half vs. standard LSTM; BMIFS–CNN–LSTM gives the highest accuracy.	BMIFS–LSTM: MRE $\approx 0.0297\%$, RMSE ≈ 1.52 mg. LSTM baseline: MRE $\approx 0.0643\%$, RMSE ≈ 3.03 mg. BMIFS–CNN–LSTM: MRE $\approx 0.0126\%$, RMSE ≈ 1.01 mg.	[50]
Stacked ensemble (RF + XGBoost + GRU + LSTM) for hybrid-bus NO_x	Stacking integrates RF, XGBoost, GRU, and LSTM predictions using hybrid-bus OBD data.	Ensemble outperforms all individual temporal models.	Stacking: $MAE \approx 0.0068$ g/s, RMSE≈ 0.0283 g/s, $R^2 \approx 0.956$. LSTM: $MAE \approx 0.0128$, RMSE ≈ 0.0577, $R^2 \approx 0.747$. GRU: $MAE \approx 0.0126$, RMSE≈ 0.0516, $R^2\approx 0.798$.	[51]
TSF–Transformer (truck exhaust emissions)	Transformer architecture enhanced with a time-series forecasting head; compared with LSTM, TFT, GBRT, and XGBoost.	Achieves lowest error among all tested models.	Average RMSE ≈ 0.65; lower than TFT ($\approx$0.67) and LSTM ($\approx$1.27).	[29]

Table A3. *Cont.*

Model Name	Key Features	Performance	Numerical Results	Ref.
Transformer–XGBoost hybrid (ship-engine emissions)	TSF–Transformer extracts temporal features; XGBoost performs regression; LASSO used for feature selection.	Higher accuracy than bottom-up factors and standalone XGBoost.	CO_2: $R^2 \approx 0.994$, RMSE ≈ 43.5, $MAE \approx 35.8$. CO: $R^2 \approx 0.632$, RMSE ≈ 0.94, $MAE \approx 0.78$. NO_x: $R^2 \approx 0.972$, RMSE≈ 1.60, $MAE \approx 1.27$. Improvements $\approx$34–40% vs. bottom-up; $\approx$6–10% vs. XGBoost.	[52]
Phase-specific MoE + XGBoost (NO_x)	Separate XGBoost models for each engine phase; gating network selects expert; autoencoder supports MPC.	Outperforms single-model XGBoost; real-time capable.	$R^2 \approx 0.918$; RMSE reduced $\approx$ 58%; MPC cuts NO_x by 11–13%.	[53]
GRU vs. LSTM (diesel, biodiesel, CNG)	Speed, BMEP and fuel-type inputs; models trained per pollutant.	GRU is better for CO and NO_x; LSTM is better for CO_2.	CO: $R^2 \approx 0.968$ (RMSE $\approx$ 0.047–0.048). CO_2: LSTM $R^2 \approx 0.994$ (0.022), GRU $R^2 \approx 0.987$ (0.035). NO_x: GRU $R^2\approx 0.988$ (0.028), LSTM $R^2 \approx 0.984$ (0.034).	[30]
GWO–LSTM (transient NO_x)	GWO tunes LSTM parameters; PCA reduces inputs.	Higher accuracy than PSO-BP and static maps.	$R^2 \approx 0.987$; $MAE\approx 18.75 \times 10^{-6}$ g; MAPE $\approx$ 3.23%; RMSE $\approx 20.29 \times 10^{-6}$ g.	[34]
Explainable RF (diesel NO_x)	RF with SHAP for variable attribution.	Competitive accuracy and interpretable output.	$R^2 \approx 0.965$; RMSE $\approx$ 15.76 mg.	[54]
DNN for SCR emissions	Fully connected DNN for NO_x, exhaust temperature and de-NO_x efficiency.	Strong correlation across all outputs.	NO_x: $R^2 \approx 0.983$ ($MAE \approx 3.04$ ppm). T: $R^2 \approx 0.912$ ($MAE \approx 3.1$ °C). De-NO_x: $R^2 \approx 0.905$ ($MAE \approx 3.65$%).	[55]

Appendix B

This appendix details the data-cleaning process and the sensitivity analysis performed to select the trips for model development and validation.

Appendix B.1. Missing Data Imputation/Data Cleaning

For each trip, signals from the PEMS, OBD-II, GPS, and ambient probe were aligned to a common time base and screened for completeness. Coarse clock offsets were removed from log headers, and a fine time shift was obtained by maximizing the cross-correlation between GPS and OBD vehicle speeds. Channels were then merged on a common timeline using nearest-neighbor association for discrete events and linear interpolation for continuous channels when the time gap satisfied $|\Delta t| \leq \varepsilon$, where $\varepsilon = 0.5\ \Delta t_{nom}$ and Δt_{nom} are the nominal sampling interval. No channel was up-sampled beyond 1 Hz to avoid artificial high-frequency content. When a uniform grid was needed for model training, quasi-static channels were held piecewise and constant, and sample weights $w_k \propto\ \Delta t_{nom}$ were used so 0.2 Hz trips did not under-contribute.

Imputation depends on gap length $g = \ t_b - t_a$, where t_a is the last available sample before the gap and t_b is the first available sample after the gap. For quasi-static channels (ambient temperature/pressure/humidity; coolant temperature near steady state) with $g \leq 5$ s, missing values were filled by last-observation carry-forward:

$$\hat{x}\ (t) = x(t_a),\ t\ \in (t_a, t_b] \tag{A1}$$

For continuous dynamic channels (GPS–OBD vehicle speed, engine speed, fuel flow, exhaust temperature) with $5 < g \leq 10$ s, linear interpolation preserved first-order continuity:

$$\hat{x}\,(t) = x(t_a) + \frac{t - t_a}{t_b - t_a}[x(t_b) - x(t_a)],\ t\ \in (t_a, t_b] \tag{A2}$$

Here, t_a is the last available sample before the gap, t_b is the first available sample after the gap, and $g = t_b - t_a$.

$$\hat{x}(t_k) = median\left\{x(t_i) : t_i \in \left[t_k - \frac{W}{2}, t_k + \frac{W}{2}\right], x(t_i)\ available\right\} \tag{A3}$$

For $g > 10$ s, quasi-stationary channels used a robust local median over a two-sided window of duration $W = 30$ s, where t_k is an imputed timestamp and t_i are timestamps with available samples.

Strongly transient channels and emission targets (exhaust flow and pollutant mass rates) were not imputed across gaps; affected intervals were masked and excluded from training and evaluation. Targets (pollutant mass rates) were never imputed. After imputation, domain consistency and cross-sensor agreement were enforced; non-negativity and instrument limits were respected. GPS and OBD speeds, after alignment, were required to agree within a data-driven envelope:

$$|v_{GPS}(t) - v_{OBD}(t)| \leq \max\left(3\ \text{km h}^{-1}, 0.03\ v(t)\right). \tag{A4}$$

Univariate envelopes and channel-specific slope bounds were derived from the empirical extrema of the present dataset, computed over all 26 valid trips; for speed, samples were additionally bounded by the per-trip 99.9th percentile plus a small tolerance and by the route limit $\left(120\ \text{km h}^{-1}\right) +\ 5\ \text{km h}^{-1}$:

$$0 \leq v(t) \leq min(p_{99.9}\ (v) + 5)\ \text{kmh}^{-1} \tag{A5}$$

Mechanical coherence was verified from the longitudinal power balance. Sustained violations over moving windows were flagged; affected segments were reviewed and, if necessary, masked. Where available, emission mass-rate self-consistency was assessed using wet molar fractions x_i, mixture mean molar mass $\overline{M} = \sum_j x_j\ M_j$, and exhaust mass flow $\dot{m_{exh}}$:

$$\dot{m}_i^{\text{est}}(t) = \frac{x_i(t) M_i}{\overline{M}(t)} \dot{m}_{\text{exh}}(t) \tag{A6}$$

Compared with the reported $\dot{m_i}$ after reference corrections, persistent discrepancies were treated as calibration issues rather than being imputed. Cumulative fuel $\int \dot{m_f}\ dt$ and cumulative pollutants $\int \dot{m_i}\ dt$ were required to be non-decreasing; any negative increments were zero-clipped, and the underlying samples inspected. If a predictor was masked at time t, all features derived from it at t were also masked, and the sample was excluded from supervised training to prevent the leakage of the fabricated structure; sample weights w_k were retained after masking.

For deployment and streaming evaluation, outliers were handled causally with bounded latency using the offline envelopes but without retrospective smoothing. Let x_t be the sample at step t with interval Δt. A sample was flagged if it violated the instantaneous envelope $x_t \notin [L_x, U_x]$ bounds derived from the dataset and route constraints (Table A4) and if a causal Hampel filter on the last w samples detected an impulsive spike:

$$m_t = \text{median}\{x_{t-i}\}_{i=0}^{w-1},\ s_t = 1.4826\ \text{median}\{|x_{t-i} - m_t|\}_{i=0}^{w-1},\ |x_t - m_t| > k s_t (k = 3) \tag{A7}$$

Or, if the bounded-slope test failed:

$$\left|\frac{x_t - x_{t-1}}{\Delta t}\right| \leq S_x \tag{A8}$$

with S_x estimated from training as the 99.9th-percentile slope plus a small margin. For exhaust temperature and flow, an optional curvature bound

$$\left|\frac{x_t - 2x_{t-1} + x_{t-2}}{\Delta t^2}\right| \leq C_x \tag{A9}$$

was used to capture zig-zag artifacts. Actions were channel-dependent: quasi-static channels were replaced by m_t (or held last); dynamic channels (rpm, fuel flow, exhaust temperature/flow) and all targets were masked rather than replaced to avoid inventing gradients. The model emitted an input vector only when all mandatory channels passed these tests; otherwise, the timestep was skipped to prevent feature–target desynchronization. With w = 11 at 1 Hz and w = 5 at 0.2 Hz, detection latency remained below 6–10 s. Univariate envelopes and key limits used by these procedures are summarized in Table A4, computed from the 26 valid trips.

Table A4. Limits (empirical absolute minima and maxima used for envelopes and bounds).

Variable	Absolute Minimum	Absolute Maximum
Ambient Temp(°C)	15.1	28.6
Coolant Temp (°C)	18.3	99.3
Engine Speed (rpm)	0	1986.5
Engine Torque (Nm)	336.3	700

Appendix B.2. Sensitivity Analysis

A sensitivity analysis was performed to evaluate the robustness of the Random Forest model with respect to data partitioning, feature selection, and hyperparameter tuning. Trip-level train/test splits of 70/30, 80/20, and 90/10 were examined, along with five random allocations of test trips under the same 80/20 ratio to assess the influence of trip composition. The variation in model performance across all configurations remained within $\pm 3\%$ for R^2, confirming that the predictor is largely insensitive to the specific partitioning scheme. Additional sensitivity checks involving incremental feature exclusion and adjustment of Random Forest hyperparameters demonstrated that coolant temperature is the most influential input variable, while increasing the number of estimators beyond 500 trees or altering maximum depth yielded negligible changes in accuracy, indicating stable and well-generalized model behavior.

References

1. Yuan, Q. The Evolution and Future Trajectory of Diesel Engine Technology: Applications, Environmental Challenges, and Innovative Solutions for Sustainability. *Highlights Sci. Eng. Technol.* **2024**, *119*, 829–834. [CrossRef]
2. Cunanan, C.; Tran, M.-K.; Lee, Y.; Kwok, S.; Leung, V.; Fowler, M. A Review of Heavy-Duty Vehicle Powertrain Technologies: Diesel Engine Vehicles, Battery Electric Vehicles, and Hydrogen Fuel Cell Electric Vehicles. *Clean Technol.* **2021**, *3*, 474–489. [CrossRef]

3. Greim, H. Diesel Engine Emissions: Are They No Longer Tolerable? *Arch. Toxicol.* **2019**, *93*, 2483–2490. [CrossRef]
4. Liu, X. Research on Emission Standards and Control Technology of Heavy Commercial Diesel Vehicles. Available online: https://scispace.com/papers/research-on-emission-standards-and-control-technology-of-2rxans9v4k (accessed on 26 June 2025).
5. Shane, M. Tracking Trucks and Buses. Available online: https://www.iea.org/energy-system/transport/trucks-and-buses (accessed on 15 April 2025).
6. European Union, *Regulation (EU) 2019/1242 of the European Parliament and of the Council of 20 June 2019 Setting CO2 Emission Performance Standards for New Heavy-Duty Vehicles and Amending Regulations (EC) No 595/2009 and (EU) 2018/956 of the European Parliament and of the Council and Council Directive 96/53/EC.*; European Union: Luxembourg, 2019.
7. Frekers, D.Y.; Weber, D.J.; Balaji, A.; Schatorjé, J.; Queck, D.; Herrmann, D.O.; Yoshida, S.; Tanaka, K.; Tomida, Y.; Ono, T. *The Role of the Heavy-Duty Diesel Engine Towards a Sustainable Mobility Future*; Society of Automotive Engineers of Japan: Tokyo, Japan, 2023.
8. Demuynck, J.; Kufferath, A.; Kastner, O.; Brauer, M.; Fiebig, M. Improving Air Quality and Climate through Modern Diesel Vehicles. *MTZ Worldw.* **2020**, *81*, 52–59. [CrossRef]
9. Jatana, G.S.; Chuahy, F.D.F.; Szybist, J. *Effect of Split-Injection Strategies on Engine Performance and Emissions Under Cold-Start Operation*; SAE International: Warrendale, PA, USA, 2023.
10. Ban, Z.; Guan, W.; Wang, X.; Zhao, H.; Lin, T.; Zheng, Z. Effective Engine Technologies for Optimum Efficiency and Emission Control of the Heavy-Duty Diesel Engine. In *Internal Combustion Engines and Powertrain Systems for Future Transport 2019*; CRC Press: Boca Raton, FL, USA, 2020; ISBN 978-1-003-02398-2.
11. Li, D.; Hu, J.; Hall, M.; Matthews, R. A Detailed Multidimensional Simulation for the Cold Start Process of a Gasoline Direct Injection Engine Using a Fractal Engine Simulation Model. *Int. J. Engine Res.* **2024**, *25*, 111–124. [CrossRef]
12. Siddareddy, R.B.; Franken, T.; Pasternak, M.; de Syniawa, L.L.; Oder, J.; Rottengruber, H.; Mauss, F. Real-Time Simulation of CNG Engine and After-Treatment System Cold Start Part 1: Transient Engine-Out Emission Prediction Using a Stochastic Reactor Model. *SAE Int. J. Adv. Curr. Pract. Mobil.* **2023**, *5*, 2421–2443. [CrossRef]
13. Laskowski, P.; Zimakowska-Laskowska, M.; Matej, J.; Wiśniowski, P. The Problem of Cold Start Emissions from Vehicles. *Combust. Engines* **2024**, *199*, 43–51. [CrossRef]
14. Guzzella, L.; Sciarretta, A. *Vehicle Propulsion Systems: Introduction to Modeling and Optimization*; Springer: Berlin/Heidelberg, Germany, 2013; ISBN 978-3-642-35912-5.
15. *Diesel Engine Transient Operation*; Springer: London, UK, 2009; ISBN 978-1-84882-374-7.
16. Lock, J.; Clasén, K.; Sjöblom, J.; McKelvey, T. Cold-Start Modeling and On-Line Optimal Control of the Three-Way Catalyst. *Emiss. Control Sci. Technol.* **2021**, *7*, 321–347. [CrossRef]
17. Li, Y.; Liao, J.; Xu, H.; Wu, Y.; Cai, Z.; Hu, J. Comprehensive Experimental Analysis of a Diesel Engine with an Electrically Heated Catalyst and Close-Coupled SCR during Cold/Hot WHTC. *Appl. Therm. Eng.* **2025**, *262*, 125277. [CrossRef]
18. Shen, Q.; Wang, G.; Wang, Y.; Zeng, B.; Yu, X.; He, S. Prediction Model for Transient NOx Emission of Diesel Engine Based on CNN-LSTM Network. *Energies* **2023**, *16*, 5347. [CrossRef]
19. Pina, N.; Tchepel, O. A Bottom-up Modeling Approach to Quantify Cold Start Emissions from Urban Road Traffic. *Int. J. Sustain. Transp.* **2023**, *17*, 942–955. [CrossRef]
20. Heywood, J.B. *Internal Combustion Engine Fundamentals*, 2nd ed.; McGraw-Hill Education: New York, NY, USA, 2018; ISBN 978-1-260-11610-6.
21. Wang, Z.; Luo, K.; Yu, H.; Feng, K.; Ding, H. NOx Emission Prediction of Heavy-Duty Diesel Vehicles Based on Bayesian Optimization -Gated Recurrent Unit Algorithm. *Energy* **2024**, *292*, 130559. [CrossRef]
22. Alam, G.M.I.; Arfin Tanim, S.; Sarker, S.K.; Watanobe, Y.; Islam, R.; Mridha, M.F.; Nur, K. Deep Learning Model Based Prediction of Vehicle CO2 Emissions with eXplainable AI Integration for Sustainable Environment. *Sci. Rep.* **2025**, *15*, 3655. [CrossRef]
23. Pillai, R.; Triantopoulos, V.; Berahas, A.S.; Brusstar, M.; Sun, R.; Nevius, T.; Boehman, A.L. Modeling and Predicting Heavy-Duty Vehicle Engine-Out and Tailpipe Nitrogen Oxide (NOx) Emissions Using Deep Learning. *Front. Mech. Eng.* **2022**, *8*, 840310. [CrossRef] [PubMed]
24. Mentink, P.; Seykens, X.; Escobar Valdivieso, D. Development and Application of a Virtual NOx Sensor for Robust Heavy Duty Diesel Engine Emission Control. *SAE Int. J. Engines* **2017**, *10*, 1297–1304. [CrossRef]
25. Von Rueden, L.; Mayer, S.; Sifa, R.; Bauckhage, C.; Garcke, J. Combining Machine Learning and Simulation to a Hybrid Modelling Approach: Current and Future Directions. In *Advances in Intelligent Data Analysis XVIII*; Berthold, M.R., Feelders, A., Krempl, G., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2020; Volume 12080, pp. 548–560. ISBN 978-3-030-44583-6.
26. Donateo, T.; Mujahid, T.; Morrone, P.; Algieri, A. Analysis and Simulation of Fuel Consumption and Emissions in a Heavy-Duty Diesel Truck under Real-World Driving Conditions for Hybridization and Waste Heat Recovery. In *17th International Conference on Engines and Vehicles*; SAE Technical Paper 2025-24-0096; SAE International: Warrendale, PA, USA, 2025. [CrossRef]

27. Joubert, J.W.; Gräbe, R.J. Real Driving Emissions Data: Isuzu FTR850 AMT. *Data Brief* **2022**, *41*, 107975. [CrossRef]
28. Morrone, P.; Cutuli, P.; Algieri, A.; Falbo, L.; Donateo, T. Conventional and PCM-Based Heat Recovery Configurations for Hybrid Electric Heavy-Duty Vehicles Fuelled with Fossil or Alternative Fuels. *Energy Convers. Manag.* **2026**, *353*, 121206. [CrossRef]
29. Li, Z.; Zhang, X.; Dong, Z. TSF-Transformer: A Time Series Forecasting Model for Exhaust Gas Emission Using Transformer. *Appl. Intell.* **2023**, *53*, 17211–17225. [CrossRef] [PubMed]
30. Uluocak, İ. Comparative Study of Emission Prediction Using Deep Learning Models. *Çukurova Üniversitesi Mühendislik Fakültesi Derg.* **2025**, *40*, 337–346. [CrossRef]
31. Donateo, T.; Morrone, P. Potential of ORC-Based Waste Heat Recovery in Conventional and Hybrid Heavy-Duty Powertrains. In *Proceedings of the SAE Technical Paper 2026-37-0018*; SAE International: Warrendale, PA, USA, 2026.
32. Donateo, T.; Marti, G.; Mujahid, T.; Nuzzo, F.; Tundo, S.; Zezza, L.; Algieri, A.; Morrone, P. Preliminary Design and Assessment of a Waste Heat Recovery System for a Hybrid-Electric Heavy-Duty Vehicle. *J. Phys. Conf. Ser.* **2025**, *3143*, 012094. [CrossRef]
33. Liu, L.; Zhuang, J.; Wang, Y.; Li, P.; Guo, D.; Cheng, X. WD-PSTALSTM: A Data-Driven Hybrid Model for Prediction of Diesel Vehicle NOx Emissions. *Energy AI* **2025**, *21*, 100578. [CrossRef]
34. Lu, B.; Li, J. NOX Emission Prediction of Diesel Engine Based on GWO-LSTM. *Int. J. Automot. Technol.* **2024**, 1–10. [CrossRef]
35. Wang, Z.; Feng, K. NOx Emission Prediction for Heavy-Duty Diesel Vehicles Based on Improved GWO-BP Neural Network. *Energies* **2024**, *17*, 336. [CrossRef]
36. Falai, A.; Misul, D.A. Data-Driven Model for Real-Time Estimation of NOx in a Heavy-Duty Diesel Engine. *Energies* **2023**, *16*, 2125. [CrossRef]
37. Ge, Y.; Hou, P.; Lyu, T.; Lai, Y.; Su, S.; Luo, W.; He, M.; Xiao, L. Machine Learning-Aided Remote Monitoring of NOx Emissions from Heavy-Duty Diesel Vehicles Based on OBD Data Streams. *Atmosphere* **2023**, *14*, 651. [CrossRef]
38. Brusa, A.; Grossi, A.; Lenzi, M.; Shethia, F.P.; Cavina, N.; Kitsopanidis, I. Modeling of Exhaust Gas Temperature at the Turbine Outlet Using Neural Networks and a Physical Expansion Model. *Energies* **2025**, *18*, 1721. [CrossRef]
39. Seo, J.; Yun, B.; Kim, J.; Shin, M.; Park, S. Development of a Cold-Start Emission Model for Diesel Vehicles Using an Artificial Neural Network Trained with Real-World Driving Data. *Sci. Total Environ.* **2022**, *806*, 151347. [CrossRef] [PubMed]
40. Lee, J.; Kwon, S.; Kim, H.; Keel, J.; Yoon, T.; Lee, J. Machine Learning Applied to the NOx Prediction of Diesel Vehicle under Real Driving Cycle. *Appl. Sci.* **2021**, *11*, 3758. [CrossRef]
41. Brusa, A.; Giovannardi, E.; Barichello, M.; Cavina, N. Comparative Evaluation of Data-Driven Approaches to Develop an Engine Surrogate Model for NOx Engine-Out Emissions under Steady-State and Transient Conditions. *Energies* **2022**, *15*, 8088. [CrossRef]
42. Xie, Y.; Liu, L.; Han, Z.; Zhang, J. MSCL-Attention: A Multi-Scale Convolutional Long Short-Term Memory (LSTM) Attention Network for Predicting CO_2 Emissions from Vehicles. *Sustainability* **2024**, *16*, 8547. [CrossRef]
43. Mungan, M.S.; Arpa, O. Estimation of CO_2 Emissions from Vehicles Using Machine Learning and Multi-Model Investigation. *Bull. Pol. Acad. Sci. Tech. Sci.* **2025**, *73*, e154287. [CrossRef]
44. Li, X. A Comparative Study of Statistical and Machine Learning Models on Near-Real-Time Daily Emissions Prediction. *arXiv* **2023**, arXiv:2302.01152.
45. Li, T.; Jing, X.; Wang, F.; Wang, X.; Gao, D.; Cai, X.; Tang, B. Transient Emissions Forecasting of Off-Road Construction Machinery Based on Long Short-Term Memory Network. *Energies* **2024**, *17*, 3373. [CrossRef]
46. Lim, S.; Oh, J. Hybrid Neural Network-Based Maritime Carbon Dioxide Emission Prediction: Incorporating Dynamics for Enhanced Accuracy. *Appl. Sci.* **2025**, *15*, 4654. [CrossRef]
47. Shin, S.; Won, J.-U.; Kim, M. Comparative Research on DNN and LSTM Algorithms for Soot Emission Prediction under Transient Conditions in a Diesel Engine. *J. Mech. Sci. Technol.* **2023**, *37*, 3141–3150. [CrossRef]
48. Shin, D.; Jo, S.; Kim, H.J.; Park, S. Application of Physical Model Test-Based Long Short-Term Memory Algorithm as a Virtual Sensor for Nitrogen Oxide Prediction in Diesel Engines. *Int. J. Automot. Technol.* **2023**, *24*, 585–593. [CrossRef]
49. Bai, X.; Guo, X.; Wu, C. A Novel CNN-Transformer Fusion Framework for Accurate Prediction of NH3 Emissions from Diesel Vehicles. In *Proceedings of the Prevention of Air Pollution*; Wang, Q., Hosseini-Bandegharaei, A., Eds.; Springer Nature: Cham, Switzerland, 2025; pp. 135–146.
50. Song, M.; Xue, J.; Gao, S.; Cheng, G.; Chen, J.; Lu, H.; Dong, Z. Prediction of NOx Concentration at SCR Inlet Based on BMIFS-LSTM. *Atmosphere* **2022**, *13*, 686. [CrossRef]
51. Qi, J.; Zou, X.; He, R. Construction of a NOx Emission Prediction Model for Hybrid Electric Buses Based on Two-Layer Stacking Ensemble Learning. *Atmosphere* **2025**, *16*, 497. [CrossRef]
52. Lim, S.; Oh, J. Real-Time Emission Prediction for Ship Engine Using Stacked Time-Series Learning: A Transformer-XGBoost Hybrid Framework. *Sci. Rep.* **2025**, *15*, 40505. [CrossRef]
53. Mądziel, M. Phase-Specific Mixture of Experts Architecture for Real-Time NOx Prediction in Diesel Vehicles: Advancing Euro 7 Compliance. *Energies* **2025**, *18*, 5853. [CrossRef]

54. Park, J.J.; Lee, S.; Shin, S.; Kim, M.; Park, J. Development of a Light and Accurate Nox Prediction Model for Diesel Engines Using Machine Learning and Xai Methods. *Int. J. Automot. Technol.* **2023**, *24*, 559–571. [CrossRef]
55. Samosir, B.F.; Quach, N.Y.; Chul, O.K.; Lim, O. NOx Emissions Prediction in Diesel Engines: A Deep Neural Network Approach. *Environ. Sci. Pollut. Res.* **2024**, *31*, 713–722. [CrossRef] [PubMed]