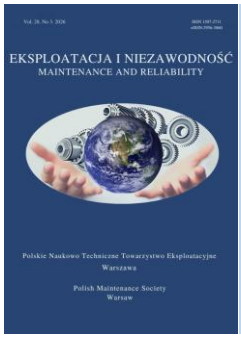


Eksploatacja i Niezawodnosc – Maintenance and Reliability

Volume 28 (2026), Issue 3

journal homepage: http://www.ein.org.pl

Article citation info:
Merkisz J, Guzanek P, Borucka A, Žvirblis T, Matijosius J, Title: Modeling of Instantaneous CO_2 Concentration under Real-World Operating Conditions of Passenger Vehicles Using RDE Data and Standard Onboard Engine Parameters, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2026: 28(3) http:/10.17531/ein/216132

Modeling of Instantaneous CO_2 Concentration under Real-World Operating Conditions of Passenger Vehicles Using RDE Data and Standard Onboard Engine Parameters

Indexed by: Web of Science Group

Jerzy Merkisz[a], Patrycja Guzanek[b,*], Anna Borucka[c], Tadas Žvirblis[d], Jonas Matijosius[d]

[a] Poznań University of Technology, Poland
[b] Doctoral School of Military University of Technology, Poland
[c] Military University of Technology, Poland
[d] Vilnius Gediminas Technical University, Lithuania

Highlights

- Instantaneous CO2 concentration reconstructed using only standard onboard engine data.
- Ensemble ML models outperform regression in reproducing RDE transient CO2 behavior.
- Exhaust mass flow is the dominant predictor of real-world CO2 concentration.
- XGBoost accurately captures rapid CO2 drops during fuel-cut and deceleration.
- Method enables low-cost CO2 monitoring without PEMS instrumentation.

Abstract

Precise evaluation of vehicle CO_2 emissions under real-world driving conditions is essential for meeting stringent regulations. While laboratory procedures such as WLTC ensure repeatability, they fail to reflect transient conditions typical of everyday driving, and PEMS-based RDE measurements remain costly. This study proposes a data-driven approach for reconstructing instantaneous CO_2 concentration using onboard engine parameters - engine speed, exhaust gas temperature and exhaust mass flow rate - without direct CO_2 sensing. Linear, nonlinear and ensemble machine learning models were evaluated using an RDE dataset of 5200 synchronized observations collected on a mixed urban-rural route. Ensemble methods, particularly Random Forest ($R^2 = 0.715$, RMSE = 15,307 ppm) and XGBoost ($R^2 = 0.669$), achieved the highest accuracy and reproduced steady-state conditions and rapid CO_2 drops during fuel-cut events. The results confirm that reliable CO_2 estimation can be achieved using a minimal OBD-based input set, enabling cost-effective emission monitoring and real-time onboard applications.

Keywords
CO_2 concentration, Real Driving Emissions (RDE), onboard diagnostics (OBD), exhaust mass flow, machine learning

1. Introduction

Road transport contributes approximately 20–25% of EU CO_2 emissions, making emission reduction a key regulatory priority. Although standardized procedures such as WLTC and RDE enable emission assessment, real-time CO_2 monitoring still relies largely on costly PEMS equipment, limiting large-scale deployment. While many studies analyze CO_2 emissions using laboratory or PEMS data, relatively little attention has been given to reconstruction based on a minimal set of onboard parameters.

Estimating CO_2 concentration under real driving conditions is challenging due to nonlinear combustion dynamics and transient operating states. Most existing approaches require

(*) Corresponding author.
E-mail addresses: J. Merkisz (ORCID: 0000-0002-1389-0503) jerzy.merkisz@put.poznan.pl, P. Guzanek (ORCID: 0000-0001-6650-7187) patrycja.guzanek@student.wat.edu.pl, A. Borucka (ORCID: 0000-0002-7892-9640) anna.borucka@wat.edu.pl, T. Žvirblis (ORCID: 0000-0002-9368-8726) tadas.zvirblis@vilniustech.lt, J. Matijosius (ORCID: 0000-0001-6006-9470) jonas.matijosius@vilniustech.lt

input variables unavailable through standard OBD systems, leaving the feasibility of predicting CO_2 concentration using only engine speed, exhaust temperature, and exhaust mass flow insufficiently explored.

This study addresses this gap by evaluating whether instantaneous CO_2 concentration can be reliably reconstructed using these three universally available parameters. Linear, nonlinear, and ensemble models are compared to assess predictive performance and practical applicability, enabling low-cost, real-time CO_2 monitoring without direct measurement.

The paper is organized as follows: Section 2 reviews related work, Section 3 describes the methodology, Sections 4 and 5 present and discuss the results, and the final section summarizes the conclusions.

2. Literature Review

Carbon dioxide emissions from road transport have been intensively studied, particularly since the introduction of RDE testing in the EU in 2016, in the context of climate change mitigation and greenhouse gas reduction strategies [1, 2]. Transport remains one of the main emission sources, with passenger vehicles contributing a substantial share of the total carbon footprint [3]. Despite rapid progress in electromobility, the continued dominance of internal combustion engines makes emission monitoring highly relevant [4].

Existing research on emission prediction can be broadly classified into four categories: (1) national or sector-level forecasting models, (2) laboratory-based studies using engine or chassis dynamometers, (3) machine learning approaches based on PEMS data, and (4) estimation methods relying on onboard diagnostics (OBD). While the first three provide valuable insights into emission behavior, their applicability to real-world driving is often limited by high instrumentation requirements and insufficient representation of transient operating conditions. OBD-based approaches offer a more scalable alternative; however, many rely on parameters such as torque, lambda, or fuel rate, which are not consistently available across vehicle platforms.

Although machine learning methods such as neural networks, gradient boosting, and ensemble models have demonstrated high predictive accuracy, most studies rely on direct CO_2 measurements or fuel-based estimations, thereby avoiding the challenge of reconstructing CO_2 concentration solely from onboard signals. Moreover, only a limited number of studies analyze the informational contribution of individual variables or identify minimal feature sets suitable for reliable prediction, restricting applicability to large-scale or real-time monitoring.

Regulatory frameworks at national and international levels [5, 6] require effective tools for monitoring and reducing CO_2 emissions under real driving conditions. Beyond advances in powertrain technologies, this necessitates reliable predictive models capable of supporting energy efficiency assessment, detection of abnormal operating states, and real-time emission monitoring [7, 8]. The growing availability of onboard sensors and vehicle communication systems enables such approaches, facilitating data-driven monitoring without the need for costly external equipment [9, 10].

Recent studies increasingly focus on real-time emission modeling using vehicle operational data [11-13]. For example, Li et al. reported high predictive accuracy using road-based datasets, although such performance is typically achievable under controlled conditions. Udoh et al. demonstrated strong results using decision tree models (MAE = 2.20, MAPE = 1.69%) [14], while Begum et al. achieved high accuracy in national-scale forecasting using hybrid machine learning approaches [15]. A comprehensive review by Jin [16] further highlighted the growing integration of predictive modeling, optimization, and feature selection. Nevertheless, many of these studies rely on aggregated or laboratory data rather than real-world RDE measurements.

An increasing number of studies incorporate vehicle-level parameters such as engine speed, exhaust temperature, and exhaust mass flow, accessible via CAN or OBD interfaces at sampling rates of 1–10 Hz [17-18]. However, the robustness of such approaches under highly dynamic real-world conditions remains insufficiently explored. Practical deployment requires models that are not only accurate but also resilient to changing loads, environmental conditions, and sensor disturbances [20, 20].

Ensuring such robustness requires adaptive mechanisms such as noise resistance, online validation, and tolerance to sensor degradation. These features significantly enhance the long-term applicability of emission models in real-world

applications, including fleet management and regulatory compliance [22-24].

Despite significant progress, a clear research gap remains: few studies have validated models capable of reconstructing CO_2 concentration using only standard onboard parameters under real RDE conditions [25]. While some investigations report promising results using real-world data, they often lack transparency regarding scalability or operational feasibility [26, 27]. Others focus primarily on laboratory or aggregated datasets rather than the limited parameter sets available during standard RDE testing [28, 28].

Consequently, further research should focus on models based exclusively on parameters natively available during RDE tests via OBD-II (e.g., engine speed, exhaust mass flow, exhaust temperature), without reliance on external PEMS equipment. Such an approach enables broader applicability of RDE data and supports scalable, cost-effective emission monitoring for both research and regulatory purposes.

3. Methodology

3.1. Research Design and Conceptual Framework

The methodological framework of this study was designed to assess whether instantaneous CO_2 concentration during Real Driving Emissions (RDE) tests can be reliably reconstructed using a minimal set of standard onboard parameters. Unlike conventional approaches relying on direct CO_2 measurements from PEMS equipment, this work investigates the predictive potential of three universally available engine signals: engine speed, exhaust gas temperature, and exhaust mass flow. These parameters are routinely recorded during RDE tests and offer the possibility of developing low-cost, scalable emission estimation methods.

The proposed framework evaluates the feasibility, accuracy, and robustness of a data-driven modeling approach under real-world operating conditions. It integrates physical understanding of combustion processes with statistical and machine learning techniques to assess whether reliable CO_2 estimation can be achieved without direct emission sensing.

The methodological structure consists of four main stages:

(1) Data acquisition and preprocessing, ensuring the alignment, quality, and analytical suitability of onboard diagnostic signals;

(2) Exploratory technical analysis, including both physical interpretation of engine–emission relationships and statistical examination of variable interactions;

(3) Model development, involving regression approaches and nonlinear machine learning algorithms;

(4) Model evaluation, consisting of error quantification, residual analysis, variable importance assessment, and comparison of predictive capabilities.

This multi-step structure was designed to integrate physical understanding of combustion processes with data-driven modeling, while ensuring methodological transparency, reproducibility, and generalizability.

3.2. Data Acquisition and Signal Overview

3.2.1. RDE Test Procedure and Data Source

The dataset used in this study originates from a certified RDE test conducted on a passenger vehicle equipped with a spark-ignition internal combustion engine. Although no new measurements were performed, a high-resolution empirical dataset obtained using a standard Portable Emission Measurement System (PEMS) was analyzed. The test route covered 6.2 km and included urban, rural, and high-speed segments, representing typical real-world driving conditions.

The dataset consisted of 5,200 synchronized samples recorded at 1 Hz, in accordance with RDE regulatory requirements (EU 2016/427, EU 2017/1154). Three engine parameters were selected as model inputs: engine speed (RPM), acquired via OBD-II (SAE J1979 PID 0C); exhaust gas temperature (°C), measured using a post-turbine thermocouple; and exhaust mass flow (g/s), calculated internally by the engine control unit based on airflow estimation and lambda control.

The target variable, CO_2 concentration (ppm), was measured using a PEMS NDIR analyzer and served as the reference signal for model training and evaluation.

3.2.2. Signal Synchronization and Downsampling

Since data originated from several independent subsystems (ECU sensors and PEMS analyzer), it was necessary to align all channels to a common time base. The PEMS raw data, originally sampled irregularly at 10–12 Hz depending on signal type, were resampled to 1 Hz, consistent with the OBD-II sampling rate. A zero-phase anti-aliasing low-pass filter was

applied to avoid information loss during downsampling. This preserves the dynamic characteristics of the CO_2 signal without introducing phase distortion - critical for instantaneous modeling.

The adopted procedure included: Application of a Butterworth-type low-pass filter with a cut-off frequency of 0.4 Hz; Cubic-spline interpolation of intermediate values; Equidistant resampling at 1 Hz; Alignment of timestamps according to GPS clock and ECU event markers.

3.2.3. Data Quality Assurance and Outlier Handling

RDE data frequently contain transient artifacts caused by: PEMS analyzer delays, sensor synchronization drift, exhaust reverse-flow events, fresh-air dilution during deceleration, CAN bus latency.

To mitigate these effects, the following procedures were applied: Outliers exceeding ±3.5σ from moving median were flagged; Short-duration dips (<1 s) in CO_2 signal were retained, as they reflect real combustion cutoff events; Missing values (<0.2% of the dataset) were imputed using forward-fill interpolation due to the temporal nature of the data; Sensor anomalies were manually inspected using sliding-window variance analysis.

The resulting dataset was validated to ensure physical consistency: exhaust mass flow ≥ 0 g/s, CO_2 concentration between 0 and 160,000 ppm, exhaust temperature ≥ ambient temperature.

3. 3. Exploratory Data Analysis

3. 3. 1. Physical Interpretation of Variables

Real-time CO_2 concentration is directly influenced by instantaneous fuel consumption, air–fuel ratio, and combustion completeness. Because no fuel mass flow sensor is present in standard passenger vehicles, the present methodology relies on variables that indirectly reflect combustion intensity. ExMassFlow is the strongest proxy for fuel burn rate. ExTemp indirectly reflects combustion efficiency and heat release. EngineSpeed provides contextual information on engine load and gear state but is not a direct measure of combustion intensity.

The raw trajectories (Figure 1) show clear correlations between engine transients and fluctuations in CO_2 concentration. The numerous short-term drops in CO_2 result from fuel-cut events during deceleration, while high plateaus correspond to high load or steady cruising.

3. 3. 2. Statistical Characterization and Correlation Analysis

To complement physical reasoning, statistical exploration was performed. Both Pearson's correlation and visual scatterplot inspection were carried out (Figure 2).

Results indicated: Strong correlation between CO_2 and ExTemp ($r = 0.80$), reflecting thermal inertia; Moderate correlation with ExMassFlow ($r = 0.71$), consistent with fuel–air mass flow proportionality; Weak direct correlation with EngineSpeed ($r = 0.23$), confirming that RPM is not a load indicator.

However, these linear correlations underestimate the true relationships, which are nonlinear and conditional. This justified the application of machine learning models capable of capturing complex dependencies.

3. 4. Model Development Strategy

To evaluate the feasibility of CO_2 reconstruction from onboard data, models of different complexity levels were tested (see Table 1).

Table 1. Categorization of Linear, Nonlinear and Machine Learning Models Applied for CO_2 Prediction.

Category	Models	Purpose
Linear	Linear Regression, Polynomial Regression	Baseline, interpretability
Nonlinear Statistical	GAM, MARS	Capturing smooth and segmented nonlinearities
Machine Learning	Random Forest, XGBoost	Capturing complex, hierarchical, high-frequency nonlinear behavior

This spectrum ensures both methodological transparency and performance benchmarking.

The dataset was randomly partitioned using 80% training and 20% test splits, preserving temporal order to prevent information leakage. Hyperparameter optimization was executed using grid search with 5-fold internal cross-validation on the training dataset.

All models were implemented using Python (NumPy, SciPy, scikit-learn, pyGAM, XGBoost libraries).

Random Forest tuning range: n_estimators ∈ [200, 500];

max_depth ∈ [6, 15]; min_samples_leaf = 1–4.

XGBoost tuning range: max_depth ∈ [3, 10]; learning_rate ∈ [0.01, 0.2]; subsample ∈ [0.6, 1.0]; colsample_bytree ∈ [0.5, 1.0]

Windows of hyperparameters were not arbitrarily chosen - they reflect the need to limit model complexity to avoid overfitting, especially given moderate dataset size (n = 5200).

Three commonly accepted performance metrics were used:

- RMSE – Root Mean Square Error : suitable for evaluating large deviations, especially important because CO_2 dips can exceed 60,000 ppm in amplitude.
- MAE – Mean Absolute Error: provides more robust average deviation measure, less sensitive to extreme outliers.
- R^2 – Coefficient of Determination: indicates how well the model explains total variance in CO_2 concentration - particularly important for RDE signals dominated by stochastic behavior.

Residual plots, time-domain overlays, and variable importance measures were also used to evaluate interpretability and robustness.

3. 5. Diagnostic Procedures and Robustness Checks

Residuals were examined for autocorrelation using the Durbin–Watson statistic (d ≈ 1.4–1.7, indicating moderate autocorrelation due to time-series nature). No systematic bias was observed.

Sensitivity tests included: ±5% noise injected into RPM and ExTemp; ±3% bias added to ExMassFlow; random dropout of individual samples.

Random Forest and XGBoost showed the highest robustness; GAM was the most sensitive.

Temporal hold-out validation confirmed that ensemble models retain performance even during segments containing strong speed/gear transitions and fuel-cut events.

3. 6. Computational Workflow

The full computational pipeline contained the following stages: Data import and raw signal inspection; Synchronization and filtering; Outlier detection and handling; Feature standardization (mean normalization for regression models); Train/test split; Hyperparameter optimization; Model training; Prediction; Evaluation (RMSE, MAE, R^2); Visualization and post-analysis.

The entire workflow was designed to be reproducible and scalable for future datasets or additional vehicles.

3. 7. Ethical, Technical, and Practical Considerations

Although no human subjects or personal data were involved in this research, certain methodological safeguards were implemented: GPS data was removed to preserve privacy; Raw dataset was anonymized; All computations relied solely on physical and operational variables.

From a practical standpoint, the methodology was intentionally designed to rely only on parameters universally available in modern vehicles, enabling future large-scale adoption.

3. 8. Analysis of uncertainty of results

Model uncertainty varies strongly with driving conditions. In stable segments (±5–7% around 100,000–110,000 ppm), all models show reduced errors, with MAE decreasing by 25–35%. For example, Random Forest MAE drops from 7,163 ppm to 4,500–5,300 ppm, while XGBoost reaches 5,000–6,000 ppm (22–30% improvement). RMSE follows a similar trend.

During rapid CO_2 drops of 60–80% within 1–3 s, errors increase markedly. Random Forest shows a 40–60% increase (up to 70–80% in extreme cases), reflecting its limited ability to capture sharp minima. XGBoost performs more robustly, with error increases limited to 15–25%, yielding two- to threefold lower uncertainty in highly dynamic phases.

GAM and MARS exhibit the highest sensitivity to abrupt changes, with error increases of 60–90%, resulting in the lowest predictive accuracy. Overall, errors decrease by 20–35% under stable conditions and increase by 40–80% during rapid transitions. Random Forest provides the lowest average error, while XGBoost offers the best balance between accuracy and responsiveness under dynamic RDE conditions.

3. 9. Limitations of the Methodology

The modeling framework is limited by: Single-vehicle dataset used for model validation; Absence of torque, fuel-rate, and vehicle load measurements; Assumption of sensor reliability; Non-use of deep learning approaches, which may outperform current models but require larger datasets.

These limitations shape recommendations for future

methodological improvements.

4. Results

4.1. Preliminary Data Analysis

The RDE dataset consisted of synchronized time series of CO_2 concentration, exhaust mass flow, exhaust temperature, and engine speed. For analysis, three onboard parameters were selected: engine speed, exhaust mass flow, and exhaust temperature.

Figure 1 illustrates the highly dynamic nature of real-world driving. Engine speed varies considerably between urban and extra-urban conditions, while exhaust mass flow closely follows these changes and reflects instantaneous combustion intensity. Exhaust temperature evolves more gradually due to thermal inertia, stabilizing after warm-up and increasing under higher load conditions.

CO_2 concentration exhibits a high baseline (~100,000 ppm) with frequent sharp drops caused primarily by transient flow effects rather than combustion losses. These dynamic characteristics strongly influence model performance and must be considered when evaluating CO_2 reconstruction accuracy.

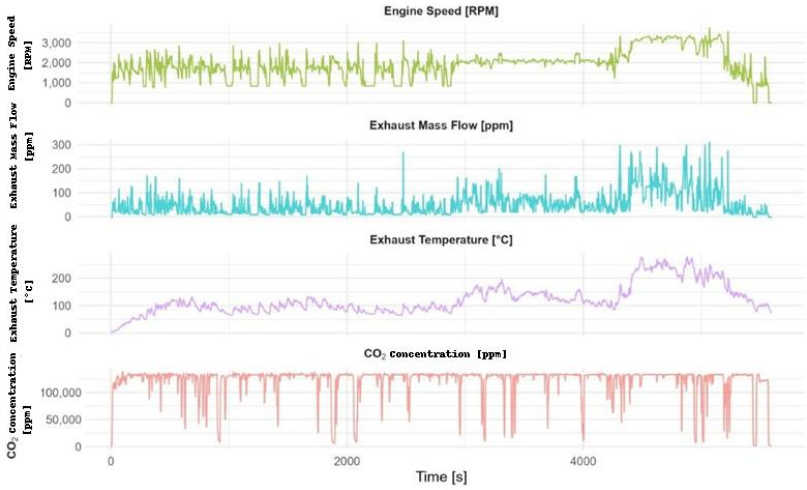
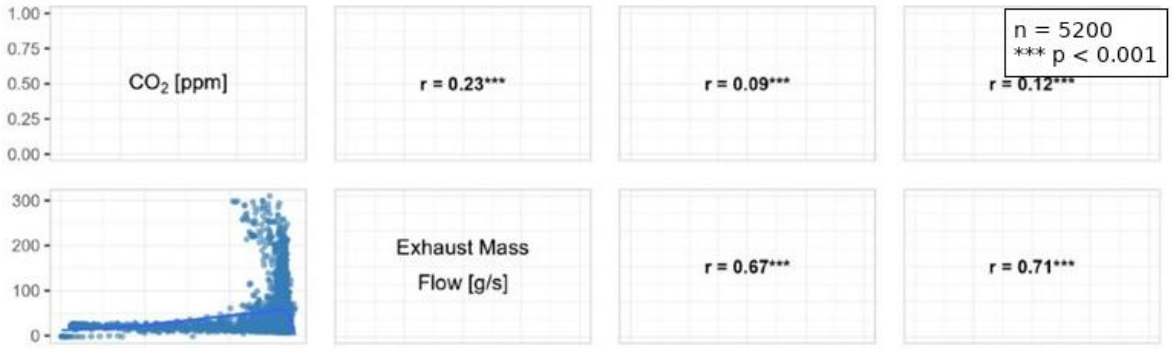


Figure 1. Time series of CO_2 concentration and key vehicle operating parameters from the RDE test;

To provide an initial assessment of relationships between the selected variables and CO_2 concentration, a linear correlation analysis was performed, acknowledging that time-series autocorrelation may affect coefficient values. All three variables showed statistically significant correlations with CO_2 concentration ($p < 0.05$), although their strength was generally moderate to weak.

Exhaust mass flow exhibited a positive correlation with CO_2 concentration, consistent with the direct relationship between fuel consumption and emissions. Exhaust temperature showed a negative correlation, indicating that higher temperatures are generally associated with more efficient combustion. Engine speed exhibited a weak and nonlinear relationship, reflecting the influence of additional factors such as load, gear selection, and transient operating conditions.

The relationships between the variables and CO_2 concentration are illustrated using correlation matrices and pairwise plots in Figure 2.



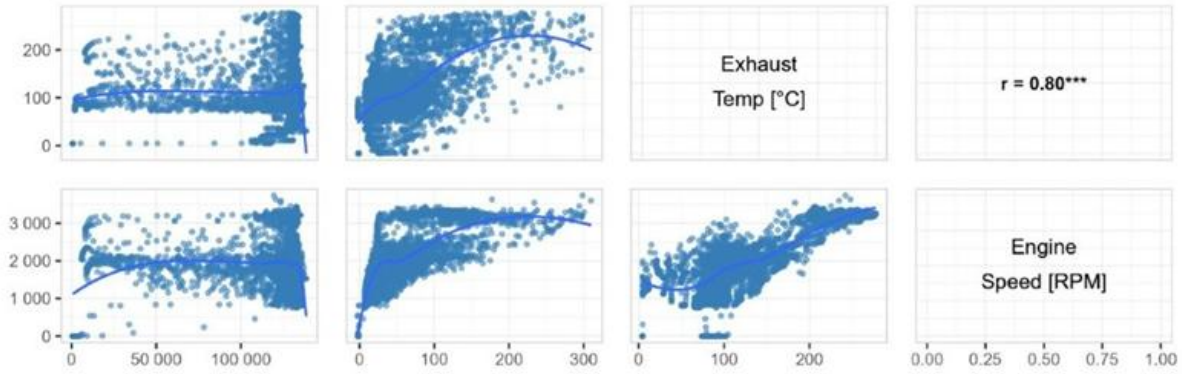


Figure 2. Correlation matrix and pair plots of variables affecting CO_2 concentration

Figure 2 illustrates the relationships between CO_2 concentration and key engine parameters. The strongest correlations occur for exhaust gas temperature (r = 0.80) and exhaust mass flow (r = 0.71), confirming that combustion intensity is the main driver of CO_2 variability. Increasing mass flow leads to higher CO_2 levels, although with dispersion caused by transient RDE conditions.

Exhaust temperature shows a nonlinear relationship with CO_2, reflecting thermal inertia and explaining why nonlinear models better capture emission dynamics. In contrast, engine speed exhibits only a weak correlation (r = 0.23), as it does not directly represent engine load under real driving conditions.

The correlation between mass flow and temperature (r = 0.67) further highlights their combined influence on combustion behavior. Overall, the analysis confirms that CO_2 emissions are primarily governed by thermal and mass-flow-related processes, while engine speed plays a secondary role. These findings support the need for nonlinear, data-driven models when analyzing emissions under RDE conditions.

4.2. Linear Regression Model

As a first step – serving as a reference point – a simple linear regression model was developed, in which the dependent variable was CO_2 concentration and the independent variables were the three aforementioned vehicle parameters. This model achieved a coefficient of determination R^2 = 0.061, which confirms the earlier considerations and indicates that it explained only 6.1% of the variability in CO_2 concentration. All variables except engine speed proved statistically significant at the significance level $\alpha = 0.05$. The strongest positive effect on CO_2 concentration was observed for exhaust mass flow (184.33), while exhaust temperature had a negative effect (-47.22). Engine speed (-1.44) had no significant effect on CO_2 concentration (p = 0.13). The estimation errors amounted to RMSE = 26,434 and MAE = 15,976, which clearly demonstrates the inadequacy of this model for predictive applications.

Therefore, an extension of the classical linear regression model was proposed, including quadratic terms and potential interactions between the predictors. This approach makes it possible to capture curvilinear relationships that could remain unnoticed in the simple model, while maintaining the interpretability of the estimated parameters. The estimated model took the following form:

$$CO_2 = 96170 + 596.2\ ExMassFlow - 1.873\ ExMassFlow^2 - 199.3\ ExTemp + 1.135\ ExTemp^2 + 32.73\ EngineSpeed - 0.01313\ EngineSpeed$$

All model components were statistically significant ($p < 0.001$). The strongest predictor of CO_2 concentration was exhaust mass flow, for which both linear and quadratic terms were significant, confirming a nonlinear relationship. The positive linear term indicates that increasing flow raises CO_2 concentration, while the negative quadratic term suggests saturation effects at higher values. A similar nonlinear behavior was observed for exhaust temperature, indicating the existence of an optimal thermal range for efficient combustion. Engine speed showed a weaker influence, with a positive linear effect and a diminishing impact at higher speeds.

Including polynomial terms improved model performance, increasing the coefficient of determination from 0.06 to 0.18, corresponding to a threefold rise in explained variance. The resulting errors (RMSE = 26,434; MAE = 15,976) indicate improved but still limited predictive capability. This confirms

that linear models are insufficient to fully capture CO_2 dynamics and that more flexible, nonlinear approaches are required. Therefore, these models serve primarily as a baseline for comparison with advanced machine learning methods applied in subsequent analyses.

4.3. Nonlinear and Ensemble Models

4.3.1. Random Forest Model

Due to the limited performance of linear regression, a Random Forest (RF) model was applied to capture nonlinear relationships between variables. The dataset was divided into training (80%) and testing (20%) subsets. The RF model achieved improved predictive performance, with RMSE = 15,306, MAE = 7,163, and R^2 = 0.715, indicating strong capability in reproducing CO_2 concentration trends.

Figure 3 compares measured and predicted CO_2 concentrations. The RF model accurately follows the overall signal structure, particularly in stable operating regions where CO_2 levels remain within 90,000–110,000 ppm. This confirms its ability to represent steady combustion conditions and long-term emission behavior.

During rapid CO_2 drops associated with transient driving events, the model correctly identifies the timing of changes but tends to underestimate their magnitude. This behavior results from the averaging nature of Random Forest, which smooths short-lived extremes. Nevertheless, the model preserves temporal consistency and avoids unrealistic oscillations, making it well suited for capturing global emission patterns under highly dynamic RDE conditions.

At higher CO_2 levels, minor prediction noise is observed, reflecting typical ensemble behavior rather than systematic error. Overall, the RF model demonstrates a strong balance between stability and responsiveness and serves as a reliable reference for comparison with more advanced modeling approaches.

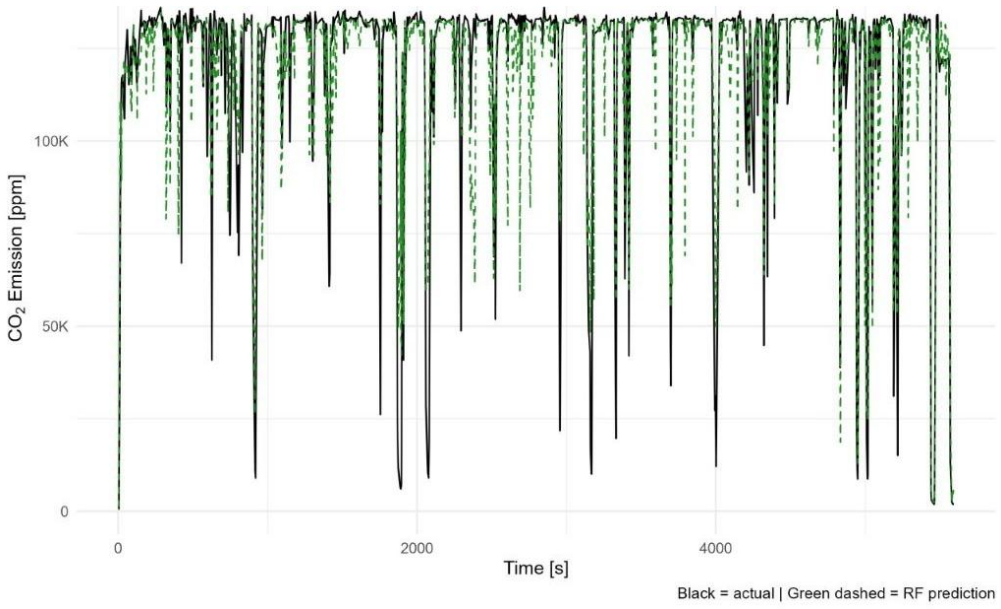


Figure 3. Random Forest model: actual CO_2 concentration (black) and predicted (green, dashed line) during the test.

Figure 3 shows that during relatively stable operating periods (e.g., 1000–2000 s and 3000–4000 s), the Random Forest (RF) model reproduces CO_2 concentration with high accuracy. This confirms that stable combustion regimes are modeled most effectively, whereas prediction errors increase mainly during rapid transients. This behavior is consistent with the correlation analysis, which indicated that CO_2 concentration is primarily governed by thermal and mass-flow-related processes.

During abrupt CO_2 drops, the RF model correctly identifies the timing of changes but underestimates their magnitude. This limitation results from the smoothing nature of decision trees, which attenuate short-lived and highly irregular extremes. Nevertheless, the model preserves the overall signal structure and avoids artificial oscillations, maintaining high robustness under dynamic RDE conditions.

Overall, the Random Forest model provides strong predictive reliability at the macro scale. Although it does not

fully reproduce extreme minima, it accurately captures global CO_2 trends and temporal patterns, confirming its suitability as a baseline model for evaluating more advanced nonlinear approaches.

Variable importance analysis (Figure 4) indicates that exhaust mass flow is the dominant predictor, followed by engine speed and exhaust gas temperature. Mass flow shows approximately 1.4–1.5 times higher importance than engine speed and about twice that of temperature, reflecting its direct relationship with fuel consumption and CO_2 formation. This result is consistent with the correlation analysis (r = 0.71).

Engine speed ranks second in importance, providing contextual information that helps distinguish operating regimes such as high load and freewheeling, despite its relatively weak linear correlation with CO_2 (r = 0.23). Exhaust gas temperature shows the lowest importance due to its delayed thermal response and partial redundancy with other predictors, limiting its contribution to short-term dynamics.

Overall, the Random Forest model effectively captures nonlinear interactions between variables, with mass flow as the primary driver, speed as contextual support, and temperature as a secondary factor. This confirms that variable importance in ensemble models reflects both physical relevance and informational uniqueness rather than simple linear correlation.

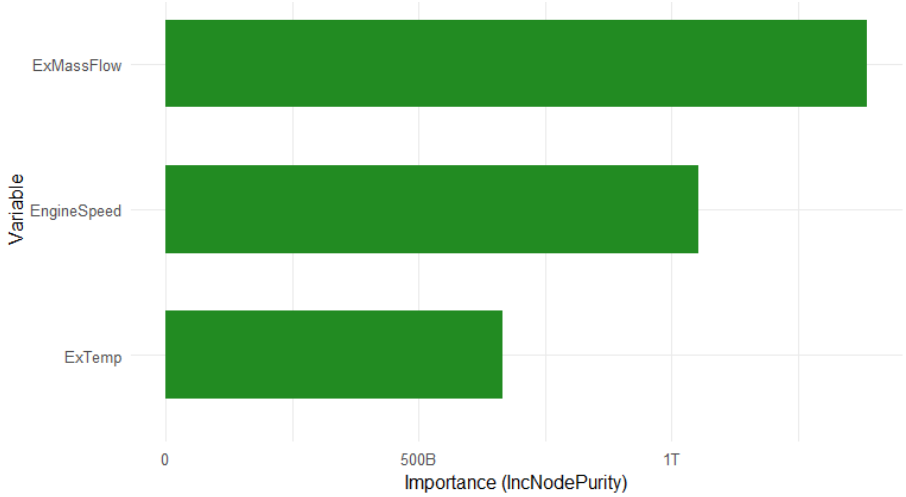


Figure 4. Predictive importance of input variables according to the Random Forest model.

The importance plot also provides insights into the model's limitations and potential improvements. One of them is the apparent lack of sensitivity to instantaneous dips observed in the prediction plot (the Random Forest model failed to accurately reproduce the depth of CO_2 dips). This shortcoming can be attributed to the model relying too heavily on mass flow and too little on temperature dynamics. Although temperature is inert, its inclusion in longer-memory models (e.g., LSTM or GRU architectures) can help to better understand systematic trends related to combustion process stability and fuel cutoff cycles.

The results confirm the earlier observations: the ExMassFlow variable most accurately reflects the variability of CO_2 concentration. In contrast, the ExTemp variable has limited informational value in the current dataset, which suggests its secondary nature relative to the other variables.

4.3.2. XGBoost Model

The XGBoost model was trained using the same 80/20 train–test split as the Random Forest model. Its predictive performance was slightly lower (RMSE = 16,498; MAE = 7,718; R^2 = 0.669), yet it demonstrated high accuracy in reproducing CO_2 dynamics.

As shown in Figure 5, XGBoost closely follows both the baseline CO_2 level and rapid transient changes, outperforming Random Forest particularly during sharp drops. The model accurately captures both the timing and depth of these events, reflecting its ability to learn complex nonlinear patterns through iterative error correction. At stable CO_2 levels (approximately 100,000–120,000 ppm), predictions remain smooth and highly consistent with measured data.

Even under highly dynamic conditions, XGBoost maintains stability and reproduces rapid fluctuations effectively. Minor overshooting may occur during extreme transitions, but this does not significantly affect overall accuracy. Overall, XGBoost provides a more precise representation of CO_2 dynamics than Random Forest, particularly for transient behavior, making it

well suited for advanced emission modeling under real-world driving conditions.

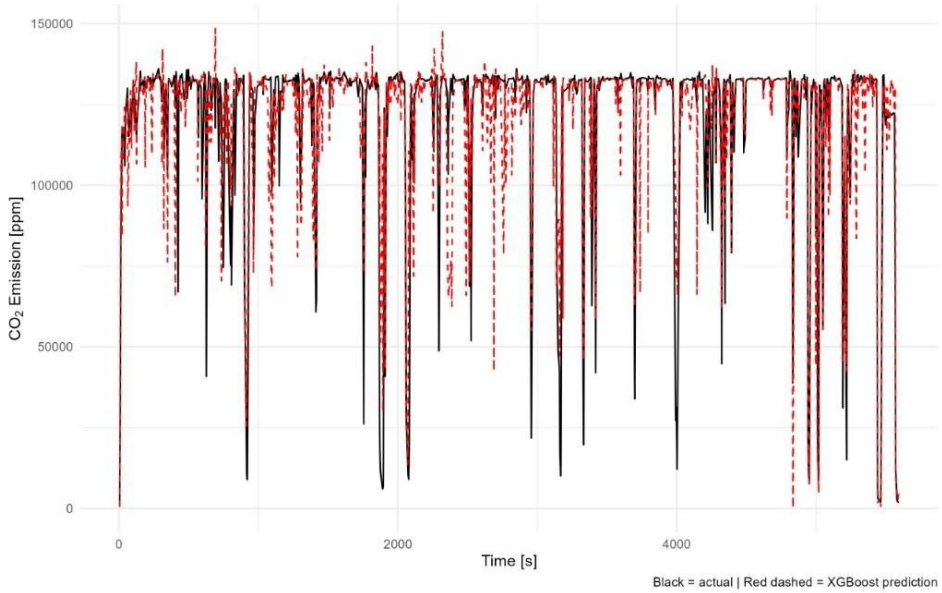


Figure 5. XGBoost model – actual and predicted values of CO_2 concentration.

Analysis of Figure 5 shows that XGBoost maintains high accuracy even during pronounced CO_2 drops, correctly reproducing both their timing and depth. This confirms its ability to capture rapid transients as well as longer-term signal structures, which is essential for reliable real-time emission prediction. Compared with Random Forest, XGBoost performs more consistently under highly dynamic RDE conditions.

Its advantage is most evident in reproducing deep CO_2 drops while maintaining stable predictions during steady operation. Although minor overshooting may occur during extreme transitions, overall signal integrity is preserved, confirming suitability for highly nonstationary emission data.

Figure 6 explains this behavior through variable importance analysis. Exhaust mass flow is the dominant predictor, followed by engine speed and exhaust gas temperature. This hierarchy is physically consistent: mass flow reflects combustion intensity, speed provides operational context, and temperature represents slower thermal dynamics. Despite its high correlation with CO_2, temperature contributes less to instantaneous prediction due to its delayed response.

Overall, XGBoost effectively exploits nonlinear interactions between variables, enabling accurate representation of both rapid transients and long-term trends. This explains its superior performance relative to Random Forest and confirms its suitability for CO_2 prediction under real-world driving conditions.

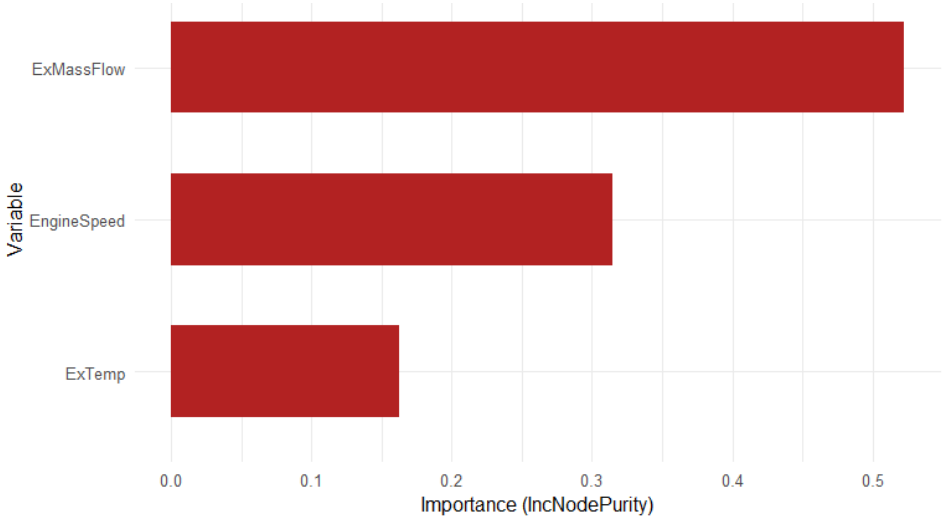


Figure 6. Informational value of variables in the XGBoost model according to the node purity index.

In conclusion, the analysis of the information value of the variables in the XGBoost model reveals a logical and physically based hierarchy of variables, in which exhaust gas mass flow dominates, while speed and temperature act as additional contextual predictors. This distribution is consistent with both the principles of engine operation and the results of the analysis of previous XGBoost predictions. This confirms that the XGBoost model correctly identified the most important predictors of CO_2 emissions and effectively used their information structure to create highly accurate predictions.

The results of the XGBoost model confirm that the structure of CO_2 concentration relationships is complex, and the best predictive performance is achieved by algorithms capable of capturing nonlinearities and conditional dependencies between predictors. Additionally, similar to the RF model, the most important information about CO_2 concentration is contained in the exhaust flow and engine speed parameters.

4.3.3. GAM Model

The Generalized Additive Model (GAM) was applied to assess its ability to model nonlinear CO_2 dynamics under RDE conditions. The model achieved moderate performance (RMSE = 20,881; MAE = 12,379; R^2 = 0.469), clearly inferior to Random Forest and XGBoost. Figure 7 shows that GAM reproduces the general CO_2 trend reasonably well, particularly at medium and high concentration levels, but performs poorly during rapid and deep concentration drops.

The model captures average emission levels but fails to reproduce sharp transitions, especially during abrupt decreases typical of real driving conditions. These limitations arise from the additive structure of GAM, which restricts its ability to represent highly nonlinear and rapidly changing interactions between engine parameters. As a result, the model tends to smooth extreme variations and underestimate the depth of CO_2 drops.

Although GAM performs adequately in stable operating regions, it lacks the responsiveness required to model highly dynamic emission behavior. This contrasts with ensemble methods, particularly XGBoost, which more effectively capture both the timing and magnitude of rapid changes. Consequently, GAM is suitable for representing long-term trends but is less effective for modeling short-term, high-intensity fluctuations typical of real-world driving conditions.

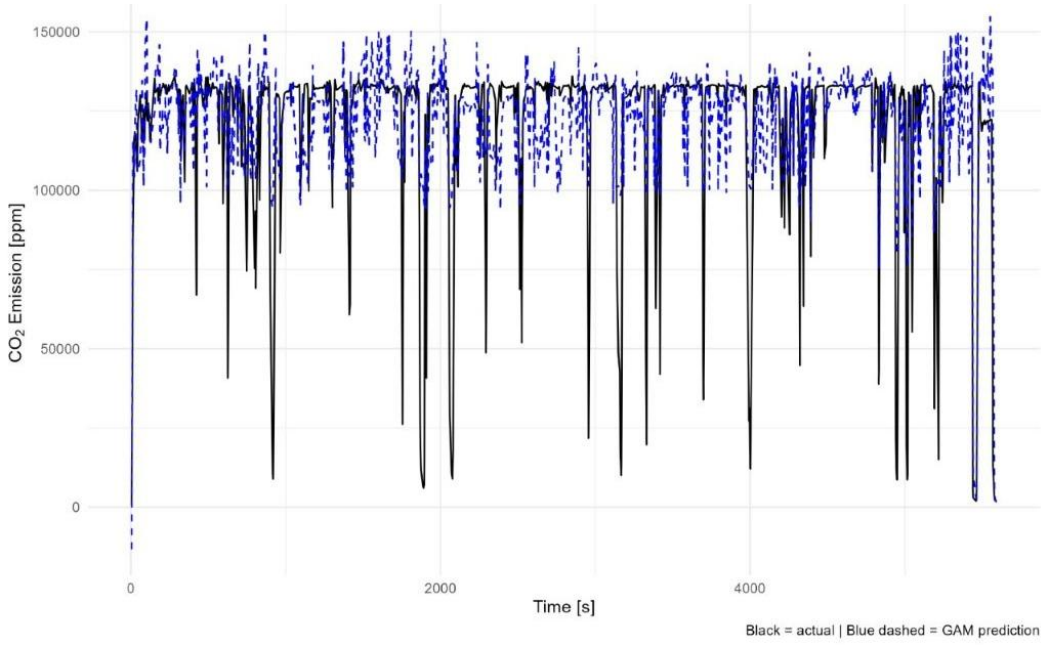


Figure 7. GAM model: actual and predicted CO_2 concentration as a function of time.

The GAM model reproduces overall CO_2 trends but performs poorly during rapid transitions, reflecting the limitations of additive modeling under highly dynamic conditions. While effective in stable regimes and useful for interpretability, it lacks the responsiveness required for accurate short-term prediction. Consequently, GAM is better suited as an explanatory rather than a predictive model for real-world driving conditions.

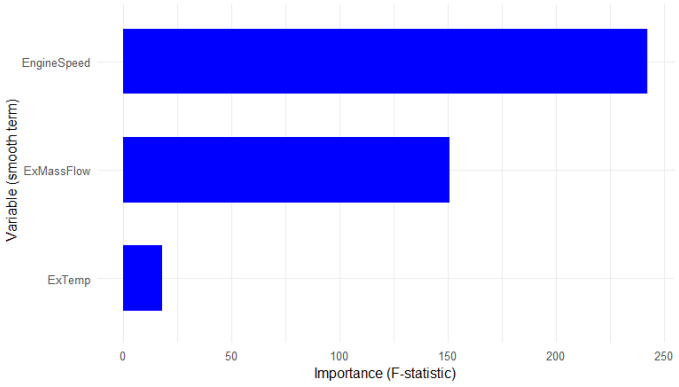


Figure 8. Predictor importance in the GAM model: assessment based on F-statistic values.

Figure 8 highlights clear differences between the GAM and ensemble models in terms of variable importance. While Random Forest and XGBoost emphasize exhaust mass flow as the dominant predictor, GAM assigns the highest importance to engine speed. This reflects fundamental differences in model structure: ensemble methods capture interactions between variables, whereas GAM relies on additive smooth functions and therefore emphasizes predictors with stable, nonlinear trends.

The dominance of speed in the GAM model explains its limited ability to reproduce sharp CO_2 drops. Although speed captures systematic operating regimes, it cannot represent rapid, interaction-driven events such as fuel cut-off, which require joint interpretation of multiple variables. As a result, GAM performs well in describing long-term emission behavior but underperforms in highly dynamic conditions.

Mass flow remains an important predictor, providing direct information on combustion intensity, while exhaust temperature plays a secondary, stabilizing role due to its thermal inertia. This hierarchy confirms that GAM prioritizes smooth, interpretable relationships over responsiveness to rapid transients.

Overall, Figure 8 demonstrates that GAM offers valuable interpretability and insight into long-term emission trends but lacks the structural flexibility required for accurate modeling of rapid CO_2 fluctuations. Consequently, it serves best as a complementary explanatory tool alongside more responsive ensemble models.

4.3.4. MARS Model

The MARS (Multivariate Adaptive Regression Splines) model shows moderate predictive performance (RMSE = 20,753; MAE = 11,926; R^2 = 0.476), clearly inferior to Random Forest and XGBoost. As shown in Figure 9, it reproduces general CO_2 trends under stable operating conditions but fails to capture rapid and deep concentration drops.

MARS performs reasonably well when CO_2 levels remain stable (≈95,000–115,000 ppm), reflecting its ability to model long-term emission behavior. However, during rapid transitions it underestimates drop magnitude due to its piecewise linear structure, which smooths abrupt changes. In highly dynamic regions, the model also shows increased noise and reduced responsiveness, limiting its effectiveness under RDE conditions.

Despite these limitations, MARS provides valuable interpretability by identifying structural changes and operating regimes. While unsuitable for accurate real-time prediction, it remains useful for analyzing long-term trends and supporting explanatory analysis alongside more advanced ensemble methods such as XGBoost.

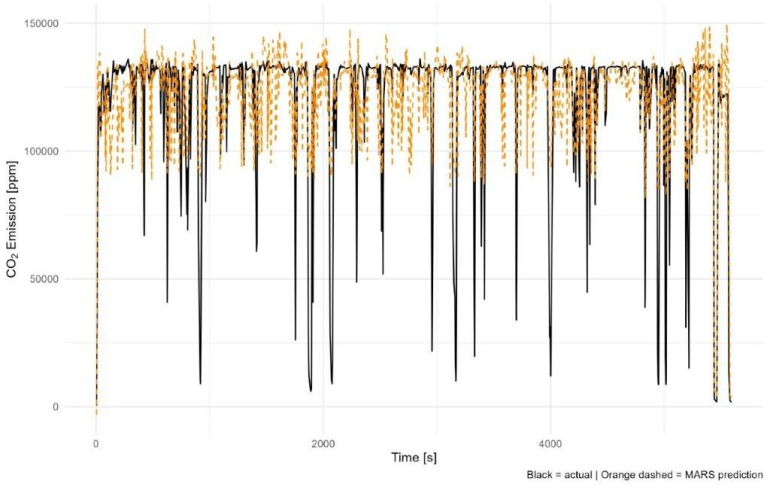


Figure 9. MARS model: actual and predicted CO_2 concentration as a function of time.

The MARS model demonstrates moderate performance in predicting CO_2 emissions under real driving conditions. It reproduces overall trends and stable operating regimes but performs poorly during rapid emission drops. Compared with Random Forest and XGBoost, MARS shows lower sensitivity to abrupt changes, making it less suitable for short-term prediction despite being more flexible than classical linear models.

MARS therefore represents an intermediate approach: more expressive than linear regression and GAM, yet less accurate and responsive than ensemble methods. Its main advantage lies in interpretability and the ability to reveal structural relationships rather than in precise dynamic prediction.

Variable importance analysis (Figure 10) indicates that engine speed is the dominant predictor, followed by exhaust mass flow and exhaust temperature. This differs from ensemble models, where mass flow dominates, and reflects the additive structure of MARS. Engine speed exhibits strong nonlinear effects across operating regimes, while mass flow contributes more smoothly and temperature plays a minor role due to its slow response.

Overall, MARS captures long-term emission behavior and structural patterns but lacks the responsiveness required for accurate modeling of rapid CO_2 fluctuations. It is therefore best suited as a complementary, interpretative tool alongside more accurate ensemble methods such as XGBoost.

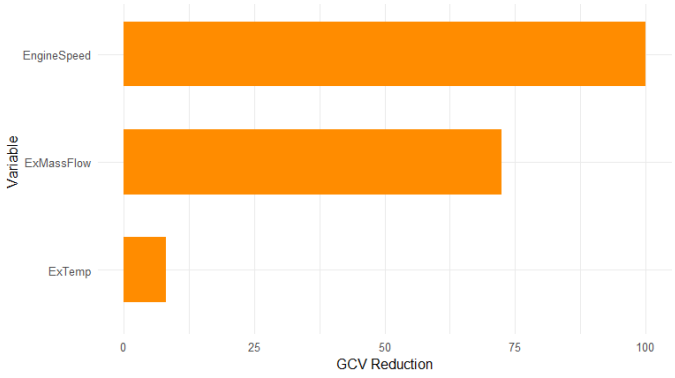


Figure 10. Informational value of variables in the MARS model according to the GCV and RSS criteria

An important insight from this analysis concerns how variable importance is interpreted in the MARS model. Unlike ensemble methods, which evaluate variables based on their contribution to error reduction, MARS assigns importance according to how effectively a variable enables local segmentation of the data. As a result, variables exhibiting strong local nonlinearities are favored. Engine speed shows the strongest such behavior, followed by exhaust mass flow, while exhaust temperature contributes the least.

This explains the behavior observed in Figure 9. The MARS model reproduces overall CO_2 trends well in stable regions but loses accuracy during sharp declines. Although engine speed provides strong nonlinear structure, it is insufficient to capture abrupt CO_2 drops, which are primarily driven by rapid changes in mass flow. Since MARS relies on segmented approximations, it struggles to represent such sudden transitions.

Overall, the variable importance ranking derived from GCV and RSS confirms a clear hierarchy: engine speed dominates nonlinear structure, mass flow provides direct physical relevance, and temperature acts as a slow background variable. This structure aligns with earlier model analyses and helps explain both the strengths and limitations of MARS in modeling RDE emission dynamics.

5. Discussion

To comprehensively compare model performance, RMSE, MAE, and R^2 metrics were analyzed (Table 2, Figures 11–12). The results clearly demonstrate a performance hierarchy among the tested models.

Simple linear and quadratic regressions performed poorly (RMSE > 24,000 ppm; $R^2 \leq 0.18$), confirming that linear approaches are inadequate for modeling highly nonlinear and nonstationary RDE data. GAM and MARS achieved moderate improvements ($R^2 \approx 0.47$), capturing general emission trends but failing to reproduce sharp CO_2 drops. Their additive structures limit responsiveness to abrupt signal changes, despite improved interpretability.

The best performance was achieved by ensemble methods. XGBoost reached RMSE ≈ 16,499 ppm, MAE ≈ 7,718 ppm and $R^2 = 0.669$, accurately reproducing both baseline levels and rapid emission drops. Random Forest achieved the lowest overall error (RMSE ≈ 15,307 ppm, MAE ≈ 7,163 ppm, $R^2 = 0.715$), demonstrating strong robustness to noise and high predictive stability, although it slightly underestimates extreme minima.

Overall, ensemble methods clearly outperform regression-based approaches under RDE conditions. Random Forest provides the highest overall accuracy, while XGBoost better

captures sharp transient events. These results confirm that modeling real-world CO_2 emissions requires nonlinear, ensemble-based methods capable of handling highly dynamic and chaotic signal structures.

Table 2. Summary of results for all models.

Model	RMSE	MAE	R^2
Linear regression	26 435	15 977	0.061
Quadratic regression	24 734	15 809	0.180
GAM	20 882	12 379	0.469
MARS	20 753	11 926	0.476
XGBoost	16 499	7 718	0.669
Random Forest	15 307	7 163	0.715

Figure 11 compares model performance using RMSE and MAE, highlighting clear differences in predictive accuracy. Linear and quadratic regressions perform worst (RMSE > 24,000 ppm, $R^2 \leq 0.18$), confirming their inability to represent the nonlinear and nonstationary nature of RDE CO_2 emissions.

Nonlinear methods (GAM and MARS) improve performance (RMSE ≈ 20,000 ppm, $R^2 \approx 0.47$) but remain limited in capturing sharp emission drops, which dominate real-world driving dynamics.

The best results are obtained with ensemble models. XGBoost achieves RMSE ≈ 16,500 ppm and MAE ≈ 7,700 ppm, effectively reproducing both baseline trends and rapid fluctuations. Random Forest performs best overall (RMSE ≈ 15,300 ppm, MAE ≈ 7,160 ppm), offering the highest robustness and lowest global error, although it slightly smooths extreme events.

Overall, accurate CO_2 modeling under RDE conditions requires flexible ensemble methods. Random Forest provides the most stable overall performance, while XGBoost better captures short-term transients. Simpler models remain useful for interpretation but are insufficient for reliable real-world prediction.

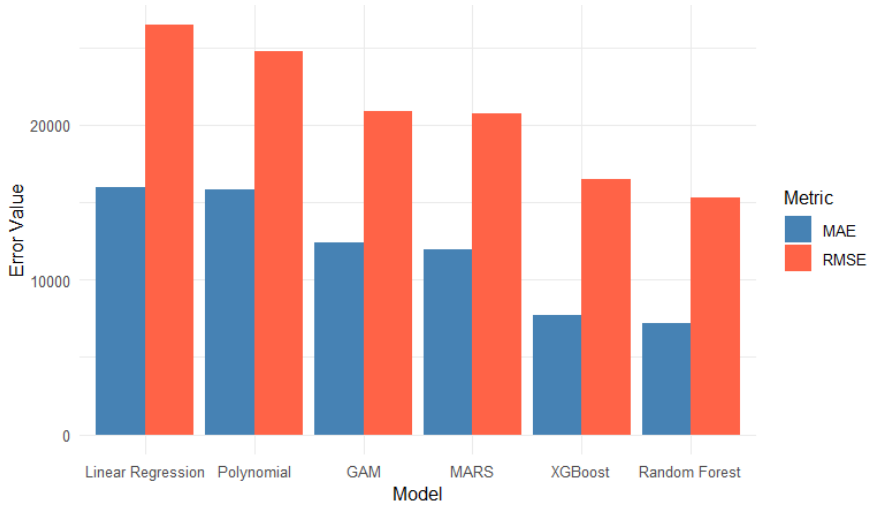


Figure 11. Comparison of RMSE and MAE errors depending on the applied model.

Figure 12 compares the coefficient of determination (R^2) for all analyzed models, highlighting their ability to explain variability in CO_2 emissions under RDE conditions. Linear and quadratic regressions perform poorly (R^2 = 0.061 and 0.178), confirming that simple models are inadequate for highly nonlinear and chaotic emission processes.

Nonlinear methods such as GAM and MARS achieve moderate performance ($R^2 \approx 0.47$), capturing general emission trends but failing to reproduce sharp and rapid CO_2 fluctuations typical of real-world driving. Their additive and segment-based structures limit their ability to model abrupt signal changes.

The best results are achieved by ensemble models. XGBoost reaches R^2 = 0.669, demonstrating strong capability to capture complex, highly dynamic emission patterns, particularly sharp drops. Random Forest achieves the highest performance (R^2 = 0.715), indicating superior overall stability and robustness. While XGBoost better reproduces extreme local variations, Random Forest provides a more balanced representation of global signal behavior.

Overall, Figure 12 confirms a clear hierarchy of modeling performance: linear models are inadequate, nonlinear additive models offer moderate improvements, and ensemble methods - especially Random Forest and XGBoost - are most suitable for accurate CO_2 emission modeling under real driving conditions.

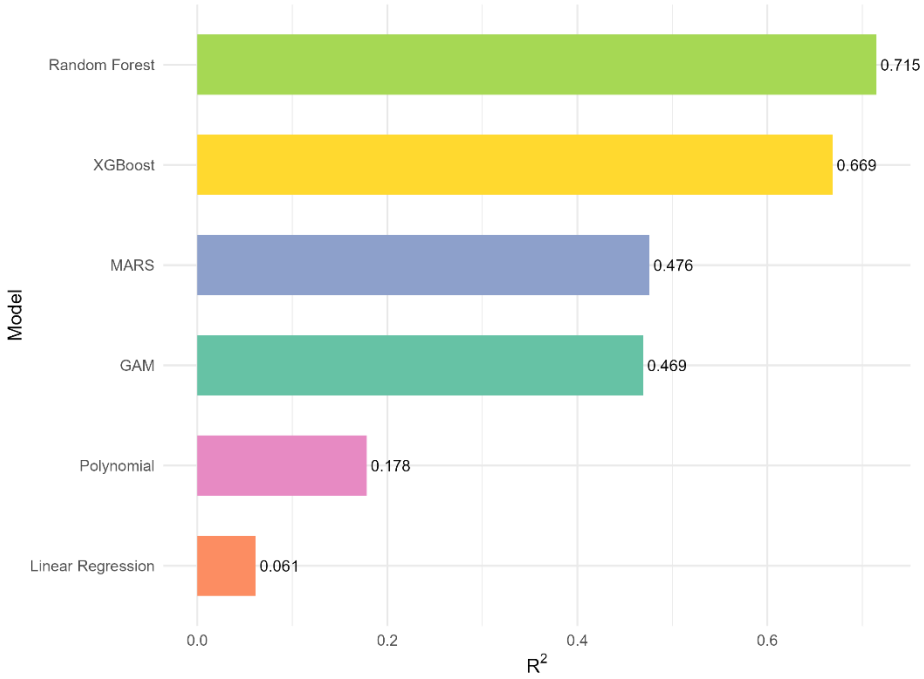


Figure 12. Comparison of R^2 coefficients of determination obtained for all models.

This analysis also allows us to draw a generalized methodological conclusion: R^2 values very clearly reflect the general applicability of the modeling methods. If the goal is structural and interpretive analysis, GAM may be a suitable choice. If the goal is to identify segmental nonlinearities, MARS is a useful tool. However, if the goal is maximum accuracy and explanation of variation, especially in the case of non-stationary RDE signals, ensemble tree methods are incomparably superior.

Finally, Figure 12 once again confirms that the comparison and analysis of methods performed in our study are methodologically sound – CO_2 emission modeling requires advanced methods capable of handling high noise and nonlinear relationships, and Random Forest and XGBoost remain the undisputed leaders in terms of R^2.

6. Conclusions

1. The study demonstrates that CO_2 concentration in passenger vehicles can be accurately predicted using only three standard engine parameters - engine speed, exhaust gas mass flow, and exhaust temperature - without direct CO_2 measurement. This confirms the feasibility of low-cost prediction systems that do not require PEMS equipment.
2. Exhaust gas mass flow proved to be the most informative and physically meaningful predictor, consistently dominating variable importance rankings and reflecting its direct relationship with combustion intensity under real RDE conditions.
3. Ensemble methods (Random Forest and XGBoost) clearly outperformed regression-based approaches, explaining over 66–71% of CO_2 variability ($R^2 = 0.669$ and 0.715). This confirms their suitability for modeling the highly nonlinear and transient nature of RDE signals.
4. Random Forest achieved the best overall predictive performance (RMSE = 15,307 ppm, MAE = 7,163 ppm), providing the most stable and accurate results and making it the most suitable method for systematic CO_2 monitoring applications.
5. XGBoost most accurately reproduced rapid CO_2 drops, which account for a large share of signal variability in RDE data. This makes it particularly well suited for modeling extreme dynamics and real-time control applications, such as fuel injection or emission regulation.
6. GAM and MARS models showed moderate performance, accurately capturing overall CO_2 levels but failing to represent rapid signal fluctuations. Their R^2 values (~0.47) indicate usefulness for structural interpretation rather than high-precision prediction.
7. Linear and quadratic regression models proved

inadequate for RDE applications, with very low explanatory power (R^2 = 0.061 and 0.180) and high errors (>24,000 ppm RMSE), rendering them unsuitable for practical CO_2 reconstruction.

8. The results confirm that standard RDE-accessible parameters (ExMassFlow, RPM, ExTemp) contain sufficient information to enable real-time CO_2 monitoring, potentially reducing dependence on costly measurement systems.
9. The modeling workflow - including data synchronization, filtering, and hyperparameter optimization - proved critical for achieving high accuracy, highlighting that data preparation is as important as model selection.
10. Future research should expand the input feature set and validate model robustness across a wider range of vehicles and driving conditions. Incorporating additional variables (e.g., load, manifold pressure, environmental conditions) is expected to further improve generalization and long-term predictive performance.

References

1. Orynycz O, Stopka O, Borucka A, Kulesza E, Merkisz J, Kolařík P. Sustainable mobility and emissions: the role of the sale structure in the automotive energy transition. Energies. 2025;18:3313. doi:10.3390/en18133313.
2. Sichigea M, Cîrciumaru D, Brabete V, Barbu CM. Sustainable transport in the European Union: exploring the net-zero transition through confirmatory factor analysis and Gaussian graphical modeling. Energies. 2024;17(11):2645. doi:10.3390/en17112645.
3. Matijošius J, Žvirblis T, Rimkus A, Stravinskas S, Kilkevičius A. Emissions, reliability and maintenance aspects of a dual-fuel engine (diesel–natural gas) using HVO additive and ANCOVA modeling. Eksploat Niezawodn. 2025. doi:10.17531/ein/208439.
4. Folęga P, Burchart D, Kubik A, Turoń K. Application of the life cycle assessment method in public bus transport. Eksploat Niezawodn. 2025;27(3). doi:10.17531/ein/204539.
5. Xiao Y, Nie Y. Technological advancement, environmental regulations and their impact on energy efficiency and CO_2 emissions. Sci Rep. 2025. doi:10.1038/s41598-025-86581-z.
6. Pata UK, Kartal MT, Ahmed Z, Sinha A. Relationship between green bonds and carbon neutrality: evidence from top five emitting countries' sectoral CO_2 emissions. Financ Innov. 2025. doi:10.1186/s40854-024-00691-7.
7. Gao H, Wang R, Wen Z, Xiong X, Zheng Z, et al. Real driving emissions characteristics of a light-duty diesel vehicle in four cities with varying altitudes in China. Case Stud Therm Eng. 2025;105831. doi:10.1016/j.csite.2025.105831.
8. Pulvirenti L, Tresca L, Rolando L, Millo F. Eco-driving optimization based on variable grid dynamic programming and vehicle connectivity in a real-world scenario. Energies. 2023;16(10). https://doi.org/10.3390/en16104121
9. Borucka A, Sobczuk S. Analysis of the relationship between energy consumption in transport, carbon dioxide emissions and state revenues: the case of Poland. Energies. 2025;18(9). https://doi.org/10.3390/en18092291
10. Alam GMI, Tanim SA, Sarker SK, et al. Deep learning model-based prediction of vehicle CO_2 emissions with explainable AI integration. Sci Rep. 2025;15:3655. https://doi.org/10.1038/s41598-025-87233-y
11. Zhang X, Wang H, Li J, Wang Y, Yu H, Liu S, et al. Effect of heavy aromatic structures on particulates emission of GDI vehicles under WLTC and RDE conditions. Fuel. 2025;401:135712.
12. Li S, Tong Z, Haroon M. Estimation of transport CO_2 emissions using machine learning algorithms. Transp Res D Transp Environ. 2024. https://doi.org/10.1016/j.trd.2024.104276
13. Ziółkowski J, Hošková-Mayerová Š, Hutsaylyuk V, Łukasiak D. Optimization of road transport within the supply network: a case study from Poland. Transport Problems. 2025;20(1):193–202. doi:10.20858/tp.2025.20.1.16.
14. Udoh J, Lu J, Xu Q. Application of machine learning to predict CO_2 emissions in light-duty vehicles. Sensors. 2024;24(24):8219. https://doi.org/10.3390/s24248219
15. Begum AM, Mobin MA. A machine learning approach to carbon emissions prediction of the top eleven emitters by 2030 and their prospects for meeting Paris Agreement targets. Sci Rep. 2025;15:4236. https://doi.org/10.1038/s41598-025-04236-5
16. Jin Y, Sharifi A, Li Z, Chen S, Zeng S, Zhao S. Carbon emission prediction models: a review. Sci Total Environ. 2024;927:172319.
17. Anjos M, Meier F. Zooming into Berlin: tracking street-scale CO_2 emissions based on high-resolution traffic modeling using machine

learning. Front Environ Sci. 2025;2:1461656.

18. Boretti A. Perspective on compressed CO_2 regenerative braking systems for passenger cars. e-Prime – Advances in Electrical Engineering, Electronics and Energy. 2025;12:100970.
19. Natarajan Y, Wadhwa G, Sri Preethaa KR, Paul A. Forecasting carbon dioxide emissions of light-duty vehicles with different machine learning algorithms. Electronics. 2023;12(10):2288. https://doi.org/10.3390/electronics12102288
20. Ahani H, Uddin M. Impact of key DMD parameters on modal analysis of high-Reynolds-number flow around an idealized ground vehicle. Appl Sci. 2025;15(2):713. https://doi.org/10.3390/app15020713
21. Alam GMI, Tanim SA, Sarker SK, Watanobe Y, Islam R, et al. Deep learning model-based prediction of vehicle CO_2 emissions with explainable AI integration. Sci Rep. 2025. doi:10.1038/s41598-025-87233-y.
22. Gładysz P, Merkisz J, Borucka A. Reliability of unmanned aerial vehicles in the context of selected factors. Eksploat Niezawodn. 2026;28(1). doi:10.17531/ein/210312.
23. Mądziel M. Modelling CO_2 emissions from vehicles fuelled with compressed natural gas based on on-road and chassis dynamometer tests. Energies. 2024;17(8):1850. doi:10.3390/en17081850.
24. Rosero F, Rosero CX, Segovia C. Simpler approaches for assessing fuel efficiency and CO_2 emissions of vehicle engines using on-board diagnostic data. Energies. 2024;17(19):4814. doi:10.3390/en17194814.
25. Ziółkowski J, Małachowski J, Oszczypała M, Szkutnik-Rogoż J, Konwerski J. Simulation model for analysis and evaluation of selected measures of the helicopter's readiness. Proc IMechE Part G: J Aerospace Eng. 2022;0(0):1–12. doi:10.1177/09544100211069180.
26. Seo J, et al. Machine learning-based estimation of gaseous and particulate emissions from internal combustion engine vehicles using ANN. Sci Total Environ. 2023;838:155949. https://doi.org/10.1016/j.uclim.2023.101734
27. Shahariar GMH, Bodisco TA, Surawski N, Rahman Komol M, Sajjad M, Chu-Van T, Ristovski Z, Brown RJ. Real-driving CO_2, NO_x and fuel consumption estimation using machine learning approaches. Next Energy. 2023;1(4):100060.
28. Vargün M, Yapmaz A, Kalender V, Yılmaz IT. Investigation of the effects of using hydrogen enriched fuel blends in a diesel engine on engine performance, combustion and exhaust emissions. Int J Hydrogen Energy. 2025;97:1399–1410. https://doi.org/10.1016/j.ijhydene.2024.11.488
29. Xu H, Lei Y, Liu M, Ge Y, Hao L, Wang X, Tan J. CO_2 emission characteristics of a light-vehicle real driving emission experiment based on vehicle-specific power distribution. Atmosphere. 2023;14(9):1467. https://doi.org/10.3390/atmos14091467