MODELING AND EVALUATION OF INSTANTANEOUS FUEL CONSUMPTION USING VEHICLE SPECIFIC POWER: A COMPARATIVE STUDY OF REGRESSION MODELS

Bui Van Chinh[1,*], **Khanh Nguyen Duc**[2]

DOI: https://doi.org/10.57001/huih5804.2026.130

ABSTRACT

Accurate prediction of instantaneous fuel consumption is essential for improving energy efficiency and reducing emissions in urban transportation systems. This study investigates the performance of three regression models including Linear Regression (LR), Optimizable Tree (OT), and Optimizable Ensemble (OE), for estimating fuel consumption using Vehicle Specific Power (VSP) as the sole input variable. The models are trained and evaluated on two datasets representing in-distribution and out-of-distribution conditions. The results show that OT achieves the best performance under in-distribution conditions, with a coefficient of determination (R^2) of approximately 0.90 and the lowest RMSE and MAE values. However, its performance deteriorates significantly when applied to out-of-distribution data, indicating limited generalization capability. In contrast, LR demonstrates more stable performance across both datasets, despite having lower accuracy under in-distribution conditions. The OE model does not show significant improvement and consistently yields higher prediction errors. These findings highlight the trade-off between model accuracy and generalization ability. While nonlinear models such as decision trees are effective in capturing complex relationships within familiar data, simpler linear models may offer better robustness under varying operating conditions. The study also suggests that the use of a single input variable (VSP) limits the effectiveness of more complex models, indicating the need for incorporating additional features or temporal dynamics to improve prediction performance.

Keywords: *Instantaneous fuel consumption; motorcycle; neural network; ensemble learning; real-world operational data.*

1. INTRODUCTION

In the context of increasingly stringent requirements on energy efficiency and emission reduction in transportation systems, accurate modeling and prediction of fuel consumption under real-world operating conditions have become essential. Such models play a key role in energy assessment, eco-driving strategy development, and real-time operational optimization [1]. For motorcycles, one of the most dominant modes of transport in urban environments, frequent stop-and-go operation, aggressive acceleration/deceleration, and varying loads result in highly dynamic and nonlinear fuel consumption behavior.

Unlike aggregated fuel consumption metrics expressed per unit distance, instantaneous fuel consumption (IFC, mg/s) directly reflects the fuel usage at each time step and is highly sensitive to transient driving phases. This makes IFC particularly suitable for micro-scale analysis of driving behavior, identification of high-consumption segments, and development of real-time energy monitoring systems. However, the relationship between IFC and vehicle operating variables is inherently nonlinear and influenced by multiple interacting factors, which limits the effectiveness of traditional linear regression models when applied across varying driving conditions [2].

With the rapid development of onboard sensors and trajectory data acquisition systems, data-driven modeling approaches have gained increasing attention. Among them, linear regression remains a widely used baseline due to its simplicity and interpretability, but it often fails to capture nonlinear dynamics in transient conditions [3]. Tree-based models, such as optimizable

---

[1] School of Mechanical and Automotive Engineering, Hanoi University of Industry, Vietnam

[2] School of Mechanical Engineering, Hanoi University of Science and Technology, Vietnam

[*] Email: chinhbv@haui.edu.vn

Received: 11/3/2026

Revised: 10/5/2026

Accepted: 29/5/2026

decision trees, provide greater flexibility by partitioning the feature space and modeling nonlinear relationships. Ensemble methods, particularly bagging-based approaches, further enhance model robustness by combining multiple learners trained on bootstrap samples, thereby reducing variance and improving generalization performance [4, 5].

Recent studies have demonstrated the effectiveness of machine learning and ensemble techniques in vehicle fuel consumption estimation using real-world driving data [6, 7]. These methods are capable of capturing complex interactions between driving variables while maintaining stability under noisy conditions. However, tree-based models may exhibit smoothing effects and may struggle to accurately reproduce sharp peaks in instantaneous signals, especially during rapid transient events [8].

In this context, a systematic comparison of different regression paradigms is necessary to better understand their strengths and limitations for instantaneous fuel consumption modeling. Therefore, this study evaluates three representative approaches: Linear Regression (LR), Optimizable Tree (OT), and Optimizable Ensemble (OE), using the same dataset, feature set, and evaluation framework. The objective is not only to compare overall error metrics such as RMSE, MAE, and R^2, but also to analyze model behavior in capturing temporal dynamics, handling extreme operating conditions, and balancing bias-variance trade-offs [8]. By providing a comprehensive comparison under consistent conditions, this study contributes to the selection of appropriate modeling techniques for real-time fuel consumption estimation and supports the development of energy-efficient transportation systems in urban environments.

2. MATERIAL AND METHOD

2.1. Training data collection

The experimental study was conducted using a Piaggio Liberty 3Vie motorcycle, which is equipped with an electronic fuel injection system. This vehicle represents a typical urban two-wheeler widely used in Vietnam. The main technical specifications of the test vehicle are summarized in Table 1.

Table 1. Main specifications of the test vehicle

Parameter	Value
Model	Piaggio Liberty 3Vie
Engine type	Single-cylinder, 4-stroke, electronic fuel injection
Production year	2013
Mileage	20,000km
Displacement	149cm^3
Maximum power	8.60kW at 8000rpm
Maximum torque	11.20Nm at 6250rpm
Fuel system	Electronic fuel injection

To collect real-world driving data, an onboard data acquisition system was installed on the vehicle. The system was designed to record instantaneous vehicle speed and IFC during operation. These measurements serve as the primary dataset for subsequent model development and validation.

The experimental driving tests were carried out on representative urban routes in Hanoi. The urban area was categorized into different traffic zones based on variations in traffic density, infrastructure conditions, and driving behavior. For each category, typical road segments were selected to capture diverse driving patterns, including congested traffic, mixed-flow conditions, and relatively free-flow segments. This approach ensures that the collected dataset reflects a wide range of real-world operating conditions.

After the data collection phase, the recorded dataset was stored and subjected to a preprocessing procedure. This includes removing abnormal or inconsistent data points that deviate significantly from the overall trend, as well as applying filtering and smoothing techniques to improve data quality. The processed dataset is then used for model training and evaluation in subsequent sections.

2.2. Selecting input variable

The selection of input variables plays a critical role in determining the accuracy and generalization capability of data-driven models for IFC. In many previous studies, multiple variables such as vehicle speed, acceleration, road gradient, and engine-related parameters have been used as model inputs to capture the complexity of fuel consumption dynamics [1, 2]. However, the inclusion of a large number of input variables may increase model complexity, introduce redundancy, and reduce robustness when applied to different driving conditions.

In this study, Vehicle Specific Power (VSP) is selected as the sole input variable for model development. VSP represents the instantaneous power demand per unit vehicle mass and integrates the effects of key driving dynamics, including speed, acceleration, rolling resistance, and aerodynamic drag. Therefore, it serves as a physically meaningful and compact descriptor of vehicle operating conditions [3].

The use of VSP as a single input variable offers several advantages. First, it inherently captures the combined influence of multiple driving factors, reducing the need for explicitly including correlated variables such as speed and acceleration. Second, VSP has been widely adopted in emission and energy modeling studies due to its strong correlation with fuel consumption and pollutant emissions under real-world driving conditions [3, 4]. Third, a single-input formulation enables a fair and consistent comparison between different modeling approaches (LR, OT, and OE), ensuring that performance differences arise from model structures rather than differences in input dimensionality.

While using a single variable may limit the ability to capture certain higher-order temporal effects, this study focuses on evaluating how different regression paradigms handle nonlinear relationships and variability in IFC when provided with the same physically representative input. This approach allows for a clearer interpretation of model behavior, particularly in terms of bias-variance trade-offs and the ability to reproduce transient consumption patterns.

The training dataset used in this study is constructed from real-world experimental measurements collected from an electric motorcycle operating under urban driving conditions. The dataset consists of time-series signals including vehicle speed $V(t)$ and instantaneous fuel consumption (IFC, mg/s), which serve as the fundamental variables for model development.

Vehicle speed is directly measured using an onboard data acquisition system, while IFC is obtained from experimental measurements reflecting the instantaneous energy consumption behavior of the vehicle. Based on the measured speed profile, vehicle acceleration is calculated using discrete differentiation between two consecutive samples:

$$a_t = \frac{V_t(i) - V_t(i-1)}{T_s} \tag{1}$$

where, T_s is the sampling time interval.

Using the calculated speed and acceleration, the instantaneous VSP is determined to represent the power demand per unit mass of the vehicle under transient operating conditions. VSP integrates the effects of acceleration, rolling resistance, road grade, and aerodynamic drag, and is computed as:

$$\begin{aligned} VSP_t = V_t \times (1.1 \times a_t + 9.81 \times \sin(\text{grade}) \\ + 0.132) + 3.02 \times 10^{-4} V_t^3 \end{aligned} \tag{2}$$

where, V_t is the instantaneous vehicle speed, a_t is the acceleration, and *grade* represents the road slope. The constant terms account for rolling resistance and aerodynamic drag contributions.

The instantaneous engine power can then be estimated from VSP as:

$$P_t = k \times m \times VSP_t \tag{3}$$

where, *m* is the total vehicle mass and *k* is the coefficient accounts for drivetrain losses [9].

In this study, VSP is selected as the sole input variable, while IFC is treated as the target output for model training. This formulation allows the model to capture the relationship between vehicle power demand and fuel consumption in a compact and physically interpretable manner.

To ensure data quality, the measured speed signals are validated based on realistic urban speed limits and the corresponding power levels derived from VSP. Data points that lead to unrealistically high power demand exceeding the physical capability of the vehicle are considered outliers and removed. Missing or corrupted data segments are subsequently reconstructed using smoothing and interpolation techniques based on least-squares estimation and time-series filtering methods [10, 11].

2.3. Training data

The training dataset used in this study is constructed from real-world experimental measurements collected from an electric motorcycle operating under urban driving conditions. The dataset consists of time-series signals including $V(t)$ and IFC which serve as the fundamental variables for model development.

Vehicle speed is directly measured using an onboard data acquisition system, while IFC is obtained from experimental measurements reflecting the instantaneous energy consumption behavior of the vehicle. Based on the measured speed profile, vehicle acceleration $a(t)$ is computed using numerical differentiation. Subsequently, VSP is calculated by combining speed, acceleration, and vehicle technical parameters, including mass, rolling resistance coefficient, and aerodynamic drag characteristics. As a result, VSP serves as a synthesized variable that captures the overall power demand per unit mass of the vehicle.

In this study, VSP is selected as the sole input variable, while IFC is treated as the target output for model training. This formulation allows the model to learn the relationship between driving power demand and instantaneous fuel consumption in a compact and physically meaningful manner.

The distribution and variability of the input and output data are illustrated in Figure 1. Specifically, Figure 1(a) presents the VSP values across the dataset, showing a wide range of operating conditions, including both positive (acceleration/load demand) and negative (deceleration or braking) regions. Figure 1(b) shows the corresponding IFC measurements, which exhibit significant dispersion and rapid fluctuations, reflecting the highly transient nature of real-world driving.

The dataset contains a large number of samples, ensuring sufficient coverage of different driving regimes. The observed variability in both VSP and IFC highlights the nonlinear and dynamic characteristics of the system, thereby providing a challenging and representative dataset for evaluating the performance of different regression models.

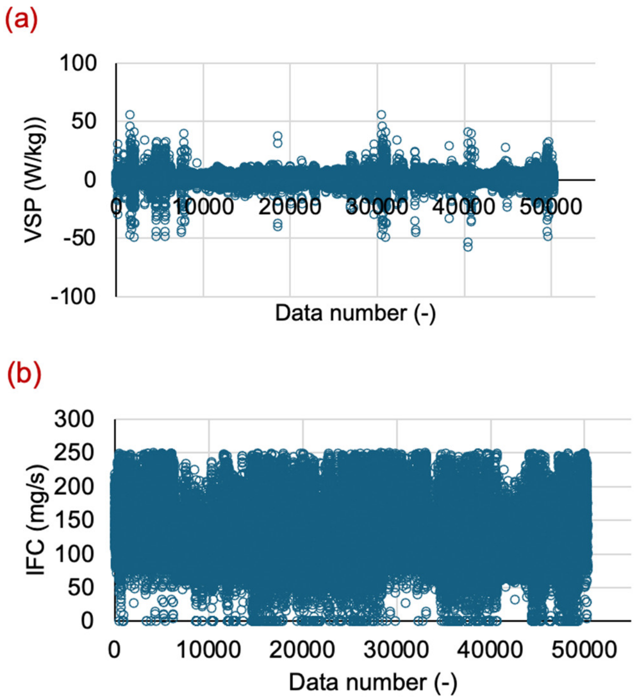

Figure 1. Distribution of training data: (a) Vehicle specific power over the dataset; (b) Instantaneous fuel consumption corresponding to the same samples

2.4. IFC prediction models

Three regression models are employed in this study, including LR, OT, and OE. LR provides a simple and interpretable baseline by fitting a linear relationship between input and output variables. In contrast, OT captures nonlinear behavior by partitioning the feature space into hierarchical regions through a tree structure. Building upon this, OE combines multiple decision trees using ensemble learning to reduce variance and improve predictive performance. As illustrated in Figure 2, LR is represented by a linear fitting trend, OT by a single decision tree structure, and OE by a collection of trees whose aggregated predictions provide more robust and stable results.

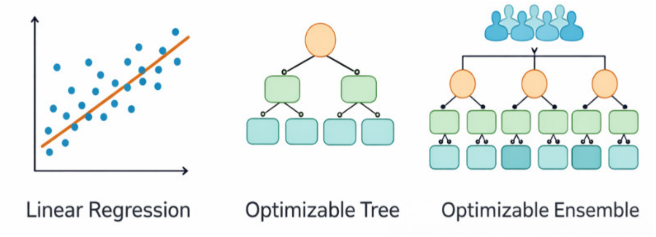

Figure 2. Slected analysis models in this study

2.5. Simulation workflow

The overall simulation and modeling procedure adopted in this study consists of three main stages: data acquisition, model development, and model evaluation. The workflow is designed to ensure consistency between experimental data processing and model validation under both in-distribution and out-of-distribution conditions.

Step 1. Data acquisition and preprocessing

Real-world driving data were collected using an onboard data logging system installed on the test vehicle. The recorded signals include vehicle speed and IFC, which serve as the primary variables for subsequent analysis. The raw data were sampled at a fixed time interval and then subjected to preprocessing steps, including noise filtering, outlier removal, and consistency checks.

The instantaneous acceleration was computed from consecutive speed measurements, and the VSP was derived using vehicle dynamic equations. VSP serves as the sole input variable of the model, representing the instantaneous power demand under varying driving conditions. To ensure data reliability, abnormal data points corresponding to unrealistic speed or excessive power demand were removed, and missing segments were reconstructed using smoothing and interpolation techniques.

Step 2. Model development

Based on the processed dataset, three regression models were developed for predicting instantaneous fuel consumption, including LR, OT, and OE. All models were trained using the same input feature VSP to ensure a fair comparison.

The dataset was divided into training, validation, and testing subsets following a standard split ratio (80% - 10% - 10%). Hyperparameters for OT and OE models were optimized using built-in optimization procedures, while LR was used as a baseline model. The training process aimed to minimize prediction errors while maintaining generalization capability.

Step 3. Model evaluation

The trained models were evaluated using two independent datasets:

- **Dataset 1 (in-distribution):** This dataset shares similar characteristics with the training data and is used to assess the interpolation capability of the models.
- **Dataset 2 (out-of-distribution):** This dataset is independent of the training data and reflects different driving conditions, allowing evaluation of model generalization.

Model performance was assessed using standard statistical metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). In addition to quantitative metrics, predicted-versus-actual plots and normalized error comparisons were used to provide visual insights into model behavior.

3. RESULTS AND DISCUSSION

3.1. Model performance evaluation on dataset 1 (In-Distribution)

The performance of the models on Dataset 1 is evaluated through both quantitative metrics as well as the predicted-versus-actual distribution (Figure 3). Notably, Dataset 1 is considered an in-distribution dataset, as it shares a similar distribution with the training data. Therefore, the results primarily reflect the interpolation capability of the models under familiar operating conditions.

As illustrated in Figure 3, the predictions of the OT model are more closely aligned with the ideal diagonal line compared to the other models, indicating higher accuracy and lower systematic bias. In contrast, the LR model shows a tendency toward deviation and compression at higher values, reflecting its limitation in capturing nonlinear relationships. The OE, on the other hand, exhibits a wider dispersion of data points, particularly in higher value regions, suggesting larger prediction errors and reduced stability. These observations are consistent with previous studies, which indicate that nonlinear models such as decision trees are more effective in capturing complex relationships in vehicle operation data [12], whereas linear models are inherently constrained by linear assumptions [13].

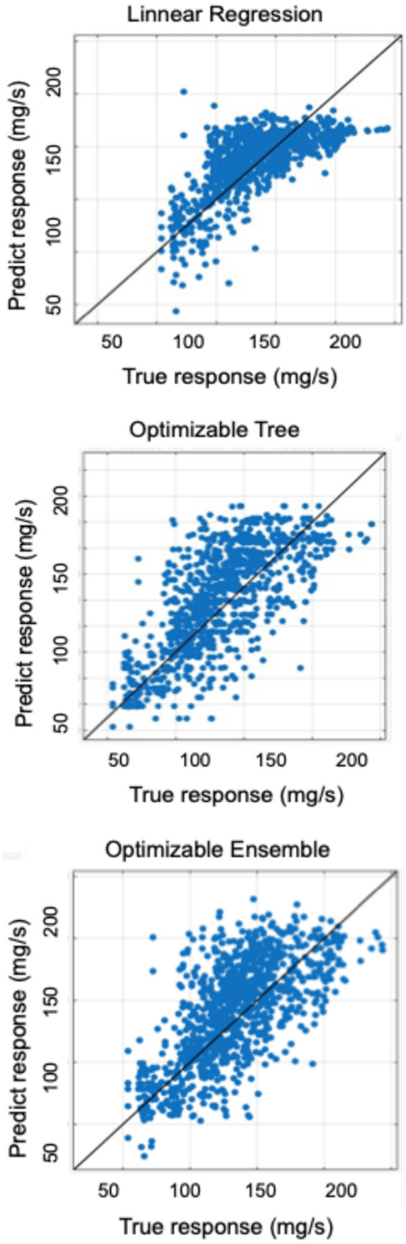

Figure 3. The distribution predicted-versus-actual IFC based on in-distribution dataset

Table 2. Performance comparison on Dataset 1

Models	RMSE	R^2	MAE	MAPE (%)
Linear Regression	25.002	0.84	16.751	16.2
Optimizable Tree	**20.135**	**0.90**	**15.485**	**15.1**
Optimizable Ensemble	31.585	0.30	19.661	20.7

Quantitative results from Table 2 and Figure 4 further confirm these observations. The OT model achieves the best performance, with RMSE reduced by approximately 19% compared to LR and 36% compared to OE. Additionally, its coefficient of determination ($R^2 = 0.90$) indicates a strong ability to explain the variance in the data. While LR provides relatively stable results, its accuracy is limited due to its inability to represent nonlinear dynamics. In contrast, OE yields the lowest performance despite the theoretical advantage of ensemble methods in reducing variance. This outcome can be attributed to the nature of ensemble learning, whose effectiveness strongly depends on the diversity of input features and data variability [14]. In this study, the use of a single input variable (VSP) restricts the ability of the ensemble model to fully exploit its advantages.

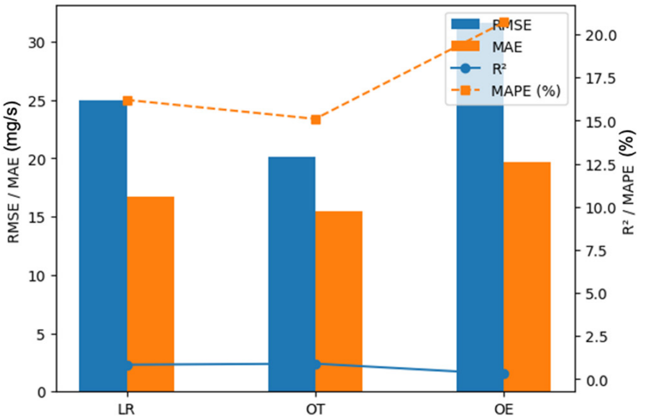

Figure 4. The statistical metrics results with dataset 1 for different models

The bar chart in Figure 4 clearly highlights the differences among the models. OT consistently achieves the lowest RMSE and MAE values, while OE exhibits the highest errors across all metrics. LR lies between the two, serving as a baseline model. This behavior aligns with the bias-variance trade-off theory, where tree-based models can achieve a better balance between flexibility and generalization under in-distribution conditions [14]. Overall, the results indicate that OT is the most suitable model for Dataset 1 due to its ability to capture nonlinear relationships, while LR remains a simple and stable baseline. In contrast, OE does not demonstrate clear advantages and may require further improvement in terms of input features or model configuration.

3.2. Model performance evaluation on dataset 2 (Out-of-Distribution)

Unlike Dataset 1, Dataset 2 is considered an out-of-distribution dataset, as it is independent from the training data and reflects different operating conditions. Therefore, the results provide insight into the generalization capability of the models.

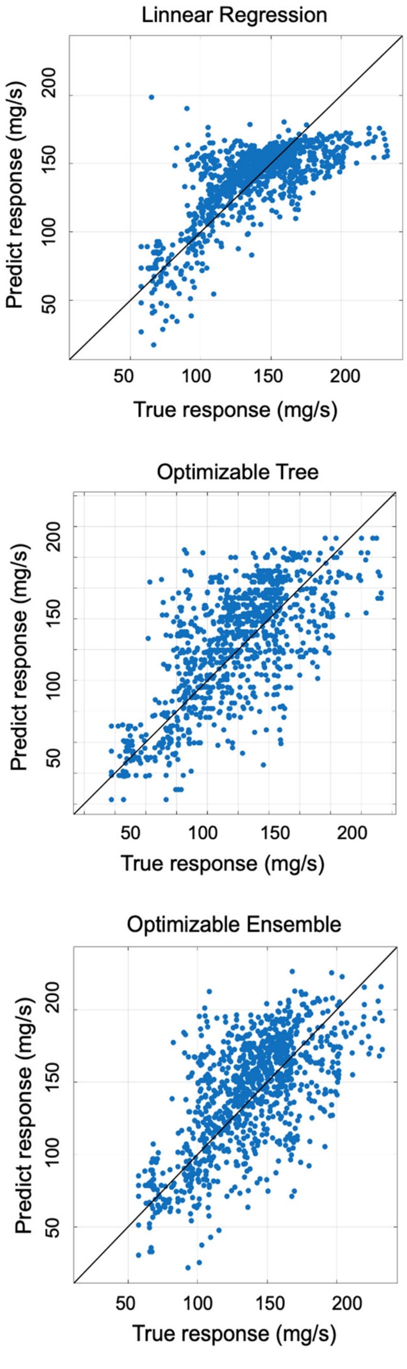

Figure 5. The distribution predicted-versus-actual IFC based on out-of-distribution dataset

As shown in Figure 5, the distribution of predicted versus actual values exhibits a noticeable degradation compared to Dataset 1. The predictions of the OT model are more dispersed and deviate significantly from the ideal diagonal line, indicating a loss of accuracy under distribution shift. The LR model, although less accurate overall, shows a relatively more consistent distribution, with fewer extreme deviations. Meanwhile, the OE model continues to exhibit a wide spread of predictions, with large variability and significant errors across the entire range of values. This behavior suggests that both OT and OE are sensitive to changes in data distribution, particularly in regions not well represented during training.

Table 3. Performance comparison on Dataset 2

Models	RMSE	R^2	MAE	MAPE (%)
Linear Regression	**24.258**	**0.55**	**18.250**	**14.6**
Optimizable Tree	28.589	0.38	22.319	17.1
Optimizable Ensemble	30.669	0.28	24.143	19.0

The quantitative results further support these observations is presented in Table 3. On Dataset 2, the LR model achieves the best performance with RMSE ≈ 24.258 and $R^2 \approx 0.55$, outperforming both OT (RMSE ≈ 28.589, $R^2 \approx 0.38$) and OE (RMSE ≈ 30.669, $R^2 \approx 0.28$). Compared to Dataset 1, the performance of OT deteriorates significantly, with RMSE increasing by approximately 42% and R^2 dropping sharply from 0.90 to 0.38. This indicates that the strong performance of OT in Dataset 1 is primarily limited to in-distribution conditions. In contrast, LR maintains relatively stable performance across both datasets, demonstrating better robustness under distribution shift. The OE model shows consistently lower performance, with only marginal improvement in RMSE compared to Dataset 1.

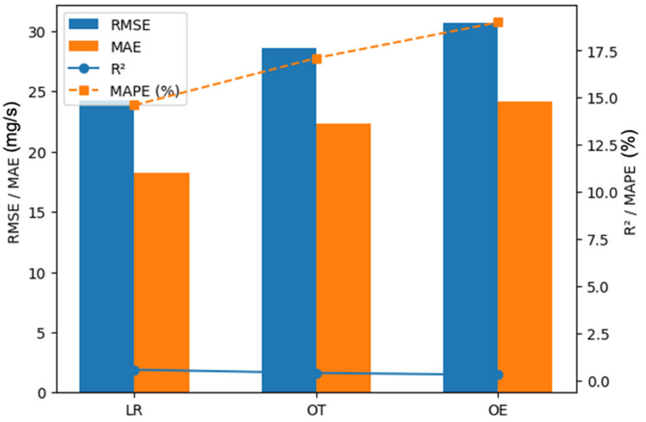

Figure 6. The statistical metrics results with dataset 2 for different models

The bar chart in Figure 6 clearly illustrates these trends. While OT achieves the lowest error in Dataset 1, it no longer maintains this advantage in Dataset 2. Instead, LR becomes the most reliable model under out-of-distribution dataset conditions. This behavior is consistent with the bias-variance trade-off theory, where more flexible models (such as tree-based methods) tend to have higher variance and are more sensitive to distribution changes, whereas simpler models (such as LR) exhibit higher bias but better generalization stability [14]. Furthermore, the limited feature space (single input variable VSP) restricts the ability of ensemble methods to leverage diversity, thereby reducing their effectiveness [14].

Overall, the results highlight a critical trade-off between accuracy and generalization. While nonlinear models such as OT can achieve superior performance under in-distribution conditions, their robustness under distribution shift is limited. In contrast, LR demonstrates more stable and reliable performance across different datasets. These findings emphasize the importance of evaluating models on independent datasets and suggest that improving generalization may require incorporating additional input features or temporal dynamics.

3.3. Comparison between the two Datasets

The comparative results between the two datasets are summarized in Table 4 and further illustrated in Figure 7, showing clear differences in both predictive accuracy and generalization ability among the three models, namely LR, OT, and OE.

As shown in Table 4, OT achieves the best performance on Dataset 1, with RMSE ≈ 20.135 and $R^2 = 0.9$, outperforming LR and OE by a considerable margin. However, when evaluated on Dataset 2, the performance of OT deteriorates substantially, with RMSE increasing to approximately 28.589 and R^2 dropping to about 0.38. In contrast, LR becomes the best-performing model on Dataset 2, with RMSE ≈ 24.258 and $R^2 = 0.55$. Meanwhile, OE remains the weakest model on both datasets and does not demonstrate the expected advantage of ensemble learning under the current feature setting.

These trends are more clearly reflected in Figure 7, where the error metrics are normalized using the best value of each metric as the reference (best = 1.0). For Dataset 1, OT stays closest to the baseline across RMSE and MAE, confirming its superior in-distribution performance. However, under Dataset 2, the normalized errors of OT increase markedly, indicating a significant loss of robustness under distribution shift. By contrast, LR maintains relatively stable error levels across the two datasets, while OE consistently remains farther from the baseline, suggesting higher error and lower reliability.

This behavior is consistent with the theory of bias-variance trade-off and model generalization. Flexible nonlinear models such as decision trees can achieve high accuracy under in-distribution conditions but are more sensitive to distribution shift and may suffer from overfitting when evaluated on unseen data [13, 14]. Simpler linear models, although more restricted in representational power, often exhibit better robustness when the test distribution changes [12, 13]. In addition, the limited feature space in this study, where only VSP is used as the input variable, may reduce the effectiveness of ensemble methods because their performance strongly depends on feature diversity and sample variability [14]. Overall, the comparison highlights a clear trade-off between accuracy and generalization, particularly, OT is more suitable for interpolation under familiar conditions, whereas LR provides a more stable choice for practical applications involving unseen operating scenarios.

Table 4. Comparison of model performance on two datasets

Models	**RMSE**	**R²**	**MAE**	**MAPE (%)**
Linear Regression	Data 1	25.002	0.56	19.281
	Data 2	24.258	0.55	18.250
Optimizable Tree	Data 1	20.135	0.90	15.485
	Data 2	28.589	0.38	22.319
Optimizable Ensemble	Data 1	31.585	0.30	25.021
	Data 2	30.669	0.28	24.143

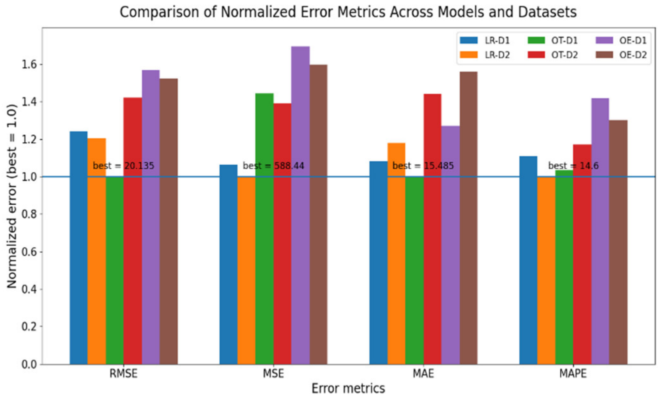

Figure 7. Normalized comparison of error metrics across models and datasets

Note: The baseline (best = 1.0) indicates the optimal performance for each metric, highlighting the degradation of OT under out-of-distribution conditions and the stability of LR.

4. CONCLUSION

This study investigated the performance of three regression models LR, OT, and OE for predicting instantaneous fuel consumption using VSP as the sole input variable. The models were evaluated under both in-distribution and out-of-distribution conditions to assess not only their prediction accuracy but also their generalization capability.

The results demonstrate that the OT model achieves the highest accuracy under in-distribution conditions, with the lowest RMSE and highest coefficient of determination $R^2 = 0.9$. However, its performance degrades significantly when applied to out-of-distribution data, indicating limited robustness under distribution shift. In contrast, the LR model, although less accurate under in-distribution conditions, shows more stable performance across both datasets, suggesting better generalization ability. The OE model does not provide noticeable improvement and consistently exhibits higher prediction errors, likely due to the limited feature space.

These findings highlight a fundamental trade-off between accuracy and generalization in data-driven fuel consumption modeling. Nonlinear models, such as decision trees, are effective in capturing complex relationships within familiar data but are more sensitive to distribution changes. Simpler linear models, despite their limited representational capacity, can offer greater robustness in practical applications where operating conditions vary.

Furthermore, the use of a single input variable VSP proves to be effective for capturing the overall trend of fuel consumption; however, it may restrict the performance of more advanced models. Future work should focus on incorporating additional features, such as temporal dynamics, driving behavior indicators, or road conditions, to enhance model generalization and predictive accuracy under real-world conditions.

REFERENCES

[1]. Sivak M., Schoettle B., "Eco-driving: Strategic, tactical, and operational decisions of the driver that influence vehicle fuel economy," *Transport Policy,* 22, 96-99, 2012. https://doi.org/10.1016/j.tranpol.2012.05.010

[2]. Zhou M., Jin H., Wang W., "A review of vehicle fuel consumption models to evaluate eco-driving and eco-routing," *Transportation Research Part D*, 49, 203-218, 2016. https://doi.org/10.1016/j.trd.2016.09.008

[3]. Bishop C. M., *Pattern Recognition and Machine Learning*. Springer, 2006.

[4]. Breiman L., "Bagging predictors," *Machine Learning*, 24(2), 123–140, 1996. https://doi.org/10.1007/BF00058655

[5]. Breiman L., "Random Forests," *Machine Learning*, 45(1), 5–32, 2001. https://doi.org/10.1023/A:1010933404324

[6]. Moradi E., Miranda-Moreno L., "Vehicular fuel consumption estimation using real-world measures through cascaded machine learning modeling," *Transportation Research Part D,* 88, 102576, 2020. https://doi.org/10.1016/j.trd.2020.102576

[7]. Fiori C., Ahn K., Rakha H., "Power-based electric vehicle energy consumption model: Model development and validation," *Applied Energy*, 168, 257-268, 2016. https://doi.org/10.1016/j.apenergy.2016.01.097

[8]. Hastie T., Tibshirani R., Friedman J., *The Elements of Statistical Learning.* Springer, 2009

[9]. Charoen-Amornkitt P., Nantasaksiri K., Ruangjirakit K., Laoonual Y., "Energy consumption and carbon emission assessment of battery swapping systems for electric motorcycles," *Heliyon,* 9(12), e22887, 2023. https://doi.org/10.1016/j.heliyon.2023.e22887

[10]. Einicke G. A., "Smoothing, filtering and prediction: Estimating the past, present and future," *InTech*, 2012. https://doi.org/10.5772/33847

[11]. Garcia D., "Robust smoothing of gridded data in one and higher dimensions with missing values," *Computational Statistics & Data Analysis*, 54(4), 1167-1178, 2010. https://doi.org/10.1016/j.csda.2009.09.020

[12]. Moradi E., Miranda-Moreno L., "Vehicular fuel consumption estimation using real-world measures," *Transportation Research Part D*, 2020.

[13]. Draper N. R., Smith H., *Applied Regression Analysis*. Wiley, 1998

[14]. Hastie T., Tibshirani R., Friedman J., *The Elements of Statistical Learning*. Springer, 2009.