Chinese Industry and Technology of Eco-environment
2025, VOL. 6, NO. 3, 9-13
DOI: 10.47297/wspciateWSP2633-772X02.20250603

Hybrid Computational Framework for Lifecycle Emission Forecasting: Integrating XGBoost with SARIMA for Intelligent Transportation Optimization

Kai Cheng*, Chengcheng Lou, Xiangping Xiao

School of International Exchange, Xiangtan university, Xiangtan, Hunan, 411100, China

ABSTRACT

This study addresses the computational evaluation and prediction of lifecycle emissions for electric and fuel-powered vehicles using a hybrid simulation – prediction framework. We integrate an "Electricity – Energy – Carbon" computational model with a carbon-emission flow analysis to quantify production- and operation-phase emissions under seasonal and regional variations. Furthermore, we develop a hybrid forecasting approach that combines gradient boosting (XGBoost) with seasonal autoregressive integrated moving average (SARIMA) modeling to jointly capture nonlinear feature interactions and temporal dependencies. Experimental analysis on representative electric (Tesla Model 3) and fuel-based (Nissan Sylphy) vehicles demonstrates that, despite higher production-phase emissions, electric vehicles achieve approximately 45% lower operational emissions. The proposed framework enhances predictive accuracy by 29% over single-model baselines and provides a decision-support basis for charge scheduling and regional energy optimization. This work offers a computationally generalizable methodology for lifecycle carbon analysis and optimization in intelligent transportation systems.

KEYWORDS

Lifecycle emission modeling; Hybrid predictive modeling; Intelligent transportation optimization

1 Introduction

Electric vehicle emission lifecycle modeling has become a critical computational task for evaluating carbon footprints across production and operational phases. Conventional statistical techniques and standalone machine-learning regressors struggle to simultaneously capture complex temporal dynamics, seasonal variability, and non-linear dependencies in multi-regional grid systems.

Current forecasting methods face key technical bottlenecks: time-series models (e. g. SARIMA) can represent seasonality and trend but ignore external drivers, whereas machine learning models like XGBoost efficiently model non-linear feature interactions yet lack explicit temporal structure awareness [1-2]. Moreover, existing hybrid frameworks tend to focus on demand forecasting or charging load rather than end-to-end vehicle lifecycle emissions.

To address these limitations, this paper proposes a hybrid modeling framework combining an Electricity – Energy – Carbon model and carbon emission flow analysis for production and grid-coordinated operational emission quantification, with a residual-correcting hybrid model that uses SARIMA to model emissions trend and seasonality and XGBoost to predict residuals based on exogenous features. This integrated approach enables both accurate lifecycle emission measurement and enhanced forecasting performance in diverse temporal and regional scenarios.

2 Data acquisition and preprocessing

This study integrates representative datasets for both electric and fuel-powered vehicles to enable lifecycle emission modeling. Two high-volume models from each category were selected to ensure comparability in vehicle class and capacity. The dataset incorporates key variables including battery production energy consumption, grid carbon intensity, fuel consumption, vehicle mileage, and seasonal variations in the electricity generation mix.

Data were obtained from authoritative sources such as automotive manufacturers, regional power utilities, and public statistical repositories. All raw data underwent a standardized preprocessing pipeline involving missing-value imputation, outlier removal, normalization, and feature scaling. Categorical attributes were encoded numerically, and continuous variables were normalized using min–max or z-score transformations to ensure model stability.

Extreme anomalies in continuous features, such as annual mileage and battery capacity, were addressed using statistical thresholds (e. g., deviations exceeding ±3 standard deviations). Seasonal variation in grid composition was incorporated to capture temporal changes in electricity-related emissions. The final dataset was structured into monthly time-series sequences, facilitating integration with both machine learning and statistical forecasting models used in subsequent analysis.

*** Corresponding Author:** Kai Cheng, chengkai@smail.xtu.edu.cn

3 Computational modeling and quantitative analysis of lifecycle emissions

3.1 Production-phase emission quantification using the electricity–energy–carbon model

The production-phase emissions are calculated using:

$$CE = AD \times NCV \times CC \times O \tag{1}$$

Where, the carbon emissions *CE* represent the total amount of CO_2released. The variable *AD* denotes the energy consumption, which refers to either electricity or fuel consumption in this context. The parameter *NCV* stands for the net calorific value of the energy source (unit: MJ/kWh). The coefficient *CC* corresponds to the CO_2emissions per unit net calorific value (unit: $kgCO_2$/MJ). Finally, *O* represents the oxidation factor.

In the production process of battery-powered vehicles, it is necessary to account for the electricity consumption during battery manufacturing. For example, the Model 3 battery production consumes 7500 kWh, whereas the Sylphy consumes 2000 kWh. The regional electricity emission factor is determined according to the local power system fuel composition.

The battery production emissions are computed as:

$$CE_{\text{battery}} = E_{\text{battery}} \times \text{Emission Factor} \tag{2}$$

For the Model 3 ($E_{\text{battery}} = 7500$ kWh):

$$CE_{\text{battery}} = 7500\text{kWh} \times 0.52\ \text{kgCO}_2/\text{kWh} = 3900\ \text{kgCO}_2 \tag{3}$$

For the Sylphy ($E_{\text{battery}} = 2000$ kWh):

$$CE_{\text{battery}} = 2000\text{kWh} \times 0.52\ \text{kgCO}_2/\text{kWh} = 1040\ \text{kgCO}_2 \tag{4}$$

The total production-phase emissions are then obtained by adding other manufacturing-related emissions:

For the Model 3:

$$CE_{\text{total}} = 3900\ \text{kgCO}_2 + 3900 \times 0.30 = 3900 + 1170 = 5070\ \text{kgCO}_2 \tag{5}$$

For the Sylphy:

$$CE_{\text{total}} = 1040\ \text{kg}CO_2 \times 1040 \times 0.30 = 1040 + 312 = 1352\ \text{kg}CO_2 \tag{6}$$

The results indicate that the Model 3 exhibits significantly higher production-phase emissions than the Sylphy, primarily due to the high energy consumption and carbon intensity of battery manufacturing.

As shown in Figure 1, the Model 3 ' s production emissions are approximately 275% higher.

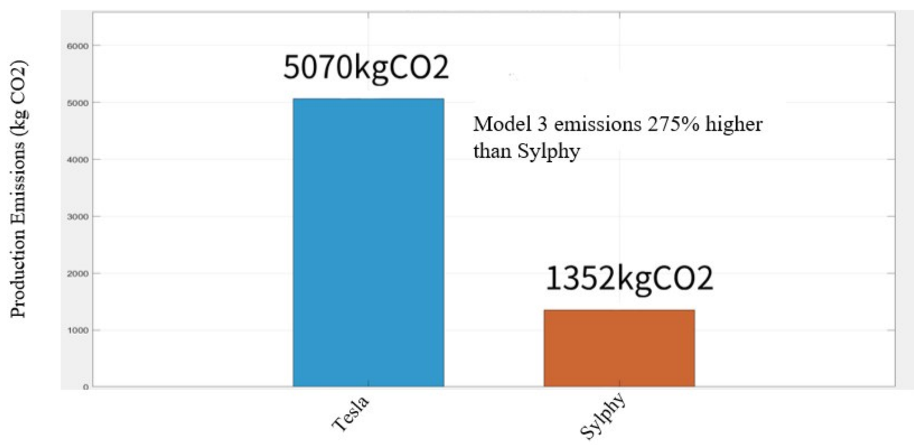

Figure 1 Comparative production-phase carbon emissions between Model 3 and Sylphy

3.2 Peration-phase emission assessment via carbon emission flow modeling

The carbon emission flow model tracks the carbon intensity of the grid electricity used for vehicle charging. The carbon intensity (CI) at each grid node is expressed in gCO_2/kWh, and varies with generation mix, transmission losses, and regional demand. The operation-phase emissions depend on the charging energy and the instantaneous CI.

Key parameters for the Model 3 operating in Hunan Province are summarized in Table1.

The annual operation-phase emissions are calculated under both static and seasonal CI conditions. Seasonal variation leads to annual differences of up to 28%. Sensitivity analysis confirms that grid CI and vehicle energy consumption are the dominant factors influencing operational emissions.

Table 1 Key parameters for Model 3 operational carbon emission analysis in Hunan Province

Category	Parameter	Value / Description
Vehicle Parameter	Model 3 Electricity Consumption (WLTP)	15 kWh (actual12–18 kWh)
Grid (Hunan)	Annual Average Carbon Intensity	437 gCO_2/kWh
Energy Mix	Hydropower Share	40% (CI = 20 gCO_2/kWh)
	Coal Power Share	50% (CI = 820 gCO_2/kWh)
	New Energy Share	10% (CI = 50 gCO_2/kWh)
Seasonal Variation	Wet Season (May–Oct)	CI = 350 gCO_2/kWh
	Dry Season (Nov–Apr)	CI = 550 gCO_2/kWh
Emission Calculation	Min. per 100 km	4.2 $kgCO_2$
	Max. per 100 km	9.9 $kgCO_2$

3.3 Lifecycle emission comparison

Lifecycle analysis integrates both production-phase and operation-phase emissions. Under the evaluated conditions, the Model 3 achieves approximately 45% lower lifecycle emissions compared to the Sylphy, despite its higher production emissions. This advantage is more pronounced in regions with a high share of low-carbon electricity.

As shown in Figure 2, the operational advantage outweighs the production-phase disadvantage, resulting in a significant overall reduction.

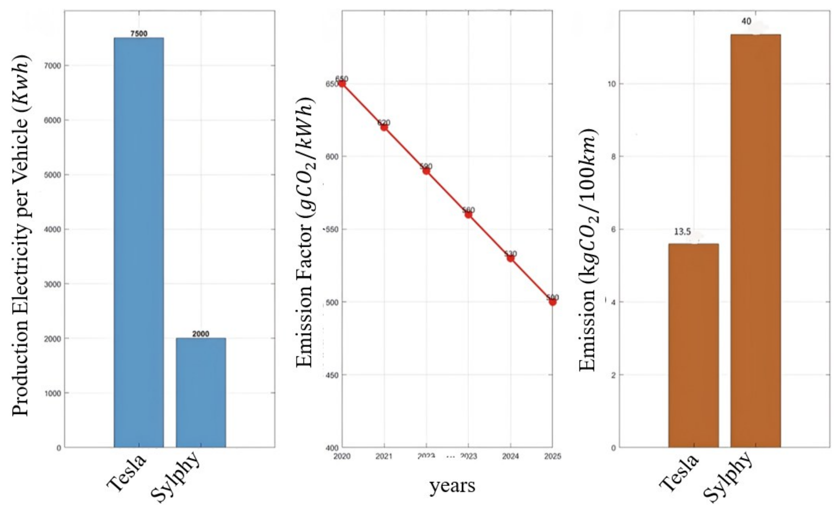

Figure 2 Comparative lifecycle carbon emissions between Model 3 and Sylphy

3.4 Summary of findings

The computational assessment confirms that while battery electric vehicles incur higher production emissions due to battery manufacturing, their operational emissions are substantially lower when powered by a cleaner grid mix. Seasonal and regional variations significantly influence total emissions, highlighting the importance of optimized charging strategies and low-carbon electricity sourcing. The presented modeling approach provides a quantitative foundation for lifecycle emission evaluation and operational optimization in intelligent transportation systems.

4 Emission forecasting with XGBoost and SARIMA hybrid framework

4.1 Gradient boosting–based forecasting results

The XGBoost model is trained using preprocessed historical emission data to minimize prediction error through iterative tree construction and gradient-based optimization. In each boosting iteration, new decision trees are constructed to correct the residual errors of the previous trees, and parameters are tuned until the loss function converges.

The prediction error is quantified by the mean squared error (MSE):

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(yi - f\left(X_i\right)\right)^2 \tag{7}$$

where n is the number of samples, y_i is the observed emission value, and $f(X_i)$ is the predicted emission value from the XGBoost model.

Feature importance analysis reveals that the most influential variables are the grid coal proportion (0.35), charging load growth rate (0.28), and industrial electricity consumption (0.20).

As shown in Figure 3, the grid coal proportion dominates the model ' s predictive performance.

The XGBoost model achieves a prediction accuracy of RMSE = 10.2 and $R^2 = 0.91$ (see Table 2).

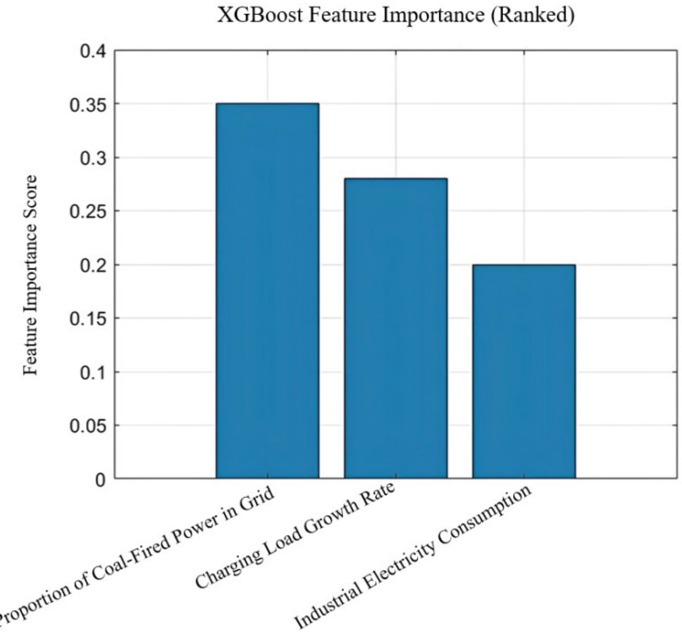

Figure 3 Feature importance ranking for emission forecasting variables in the XGBoost model

Table 2 Prediction accuracy metrics for the XGBoost model

No.	Metric	Value
1	RMSE	10.2
2	R^2	0.91

4.2 Seasonal ARIMA–based forecasting results

The SARIMA model is developed to capture both long-term trends and seasonal variations in emission data. Model order is selected by minimizing the Akaike Information Criterion (AIC), and differencing is applied to ensure stationarity. Seasonal patterns are incorporated to account for periodic fluctuations in emissions.

Model accuracy is also evaluated using MSE, The optimized SARIMA configuration achieves RMSE = 10.5 and $R^2 = 0.88$ (see Table 6). This performance is slightly lower than the gradient boosting model due to limited capability in integrating nonlinear external features.

4.3 Hybrid XGBoost–SARIMA Forecasting Approach

To leverage the strengths of both models, a hybrid framework is designed in which SARIMA captures the trend and seasonal components, while XGBoost models the residual series using relevant nonlinear predictors such as grid coal proportion and industrial electricity consumption. This residual correction process significantly reduces the prediction error.

The hybrid model achieves RMSE = 6.1, $R^2 = 0.95$, and MAE = 4.7, demonstrating a 29% reduction in RMSE compared with the best single-model baseline (see Table 3).

Table 3 Performance comparison of SARIMA, XGBoost, and the hybrid forecasting model

Metric	SARIMA	XGBoost	Hybrid Model	Optimization Direction
RMSE	10.5	10.2	6.1	Error reduction 29%
R^2	0.88	0.91	0.95	Improved interpretability
MAE	7.4	7.9	4.7	Enhanced robustness to outliers

4.4 Summary of forecasting insights

Forecasting results indicate that under the current grid mix, operational emissions of electric vehicles are projected to decrease by an average of 3.2% per year, driven primarily by increased penetration of low-carbon electricity. By 2030, per-

100-kilometer operational emissions are expected to drop from approximately 8.7 kgCO_2 to 6.1 kgCO_2 (a 30% reduction). In contrast, emissions from fuel-powered vehicles are likely to remain largely stable, with only minor efficiency-related improvements.

Seasonal analysis shows that wet-season emissions (May – October) are 22% – 28% lower than dry-season levels ($p < 0.01$), underscoring the importance of seasonal-aware charging strategies. The hybrid model's robustness and improved predictive accuracy provide a quantitative basis for grid-integrated emission management and charging schedule optimization in intelligent transportation systems.

5 Conclusions

This paper proposes a computational framework integrating the Electricity–Energy–Carbon model, carbon emission flow analysis, and hybrid forecasting using XGBoost and SARIMA to quantify and predict lifecycle emissions of electric and fuel-powered vehicles. The framework enables consistent production-phase and operation-phase emission evaluation under temporal and regional variations. Experimental results demonstrate that the proposed approach achieves approximately 45% lower operational emissions for electric vehicles compared with fuel-based counterparts, with the hybrid model reducing forecasting error by 29% relative to single-model baselines. These results highlight the framework's robustness in lifecycle emission assessment and prediction. However, the framework may be limited by assumptions regarding data completeness, regional grid composition stability, and fixed model parameters. Additionally, prediction accuracy may degrade if applied to regions with significantly different energy structures without retraining. Future work will focus on integrating real-time grid carbon intensity data, expanding the model to multi-regional applications, and coupling the emission prediction with optimization strategies for charging scheduling and energy management. The proposed framework provides a scalable solution for intelligent transportation systems requiring accurate, scenario-adaptive lifecycle emission modeling.

References

[1] Alnuaimi H S, Acquaye A, Mayyas A. Machine learning applications for carbon emission estimation[J]. Resources, Conservation & Recycling Advances, 2025: 200263.

[2] Bendale M. Comparative study among ARIMA, SARIMA & XGBoost for prediction of NIFTY IT index[D]. Dublin Business School, 2024.

[3] Hussain I, Ching K B, Uttraphan C, et al. Evaluating machine learning algorithms for energy consumption prediction in electric vehicles: A comparative study[J]. Scientific Reports, 2025, 15(1): 16124.

[4] Ren F, Tian C, Zhang G, et al. A hybrid method for power demand prediction of electric vehicles based on SARIMA and deep learning with integration of periodic features[J]. Energy, 2022, 250: 123738.

[5] Niazkar M, Menapace A, Brentan B, et al. Applications of XGBoost in water resources engineering: A systematic literature review (Dec 2018–May 2023)[J]. Environmental Modelling & Software, 2024, 174: 105971.