energies

MDPI

Article

Vehicle Type Heterogeneity in Battery Electric Vehicle Energy Consumption: A Multi-Perspective Analysis Using Real-World Fleet Data

Liangying Luo [1,2,*], Haichao Huang [3], Yizhou Wang [1] and Hongdi He [1]

[1] Center for ITS and UAV Applications Research, StateKey Laboratory of Ocean Engineering, School of Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; wang_yizhou@sjtu.edu.cn (Y.W.)
[2] Data-Driven Management Decision Making Lab, Shanghai Jiao Tong University, Shanghai 200240, China
[3] Department of Transport and Planning, Delft University of Technology, 2628 CN Delft, The Netherlands
* Correspondence: plutoo@sjtu.edu.cn

Abstract

Real-world energy consumption of battery electric vehicles (BEVs) varies with trip conditions, driving behavior, and vehicle state. Treating vehicle type merely as a control variable may mask substantial heterogeneity in energy consumption mechanisms. This study treats vehicle type as a source of mechanism heterogeneity and examines how the determinants of energy consumption rate (ECR) differ among sedans, logistics vehicles, and buses. Using 41,933 trips in China, we combined statistical analysis, machine learning, and CausalForestDML. Statistically, ECRs of sedans and logistics vehicles decrease as trip distance increases, but trip distance is nearly irrelevant for buses. Aggressive driving raises logistics vehicle ECR by 22.2% yet barely affects sedans. For prediction, permutation importance further reveals the divergence: trip distance and duration dominate sedan prediction, whereas energy recovery ratio and accelerator-pedal, which are marginally important for sedans, become primary predictors for logistics vehicles and buses. Causal estimates confirm that directional divergences persist after conditioning on covariates. For example, trip duration has opposite effects on buses versus sedans and logistics vehicles. Moreover, the causal framework decomposes each factor's effects across its value range, revealing patterns beyond correlation alone. These findings provide actionable guidance for vehicle-specific eco-driving strategies and battery management system optimization.

Keywords: battery electric vehicle; energy consumption rate; vehicle heterogeneity; permutation importance; causal forest; double machine learning

Check for updates

Academic Editor: JongHoon Kim

Received: 25 August 2026
Revised: 7 October 2026
Accepted: 9 October 2026
Published: 16 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Road transport electrification is a central strategy for reducing transport-sector carbon emissions [1]. Battery electric vehicles (BEVs) eliminate tailpipe emissions and are expected to play a key role in achieving low-carbon transportation. However, their environmental and operational benefits also depend on how efficiently electrical energy is used during real-world driving [2]. In practice, BEV energy consumption varies substantially across trips, drivers, vehicles, batteries, and operating environments [3]. This variability affects more than academic model accuracy. It influences route-level range estimation, charging demand prediction, dispatch planning, eco-driving recommendations, and the evaluation of the energy-saving benefits of electrification.

Understanding real-world energy performance requires attention to operating conditions that may not be fully represented by standardized laboratory tests. Ambient temperature and auxiliary loads can affect energy consumption and driving range, contributing to differences between rated and real-world performance [4]. This motivates the analysis of operational data to examine how energy consumption varies across driving conditions and vehicle categories.

A large body of research has examined BEV energy consumption. Trip-level variables such as distance, duration, and average speed affect energy consumption rate (ECR) because they change the relative weight of start-up losses, stop-and-go operation, and aerodynamic resistance. Driving-behavior variables, including acceleration intensity and pedal use, are also important because rapid acceleration and unstable braking patterns increase traction demand. Vehicle-state variables such as state of charge (SOC), accumulated mileage, insulation resistance, and regenerative-braking behavior can further modify energy efficiency. Environmental factors, including temperature, road gradient, and precipitation, add additional variation through battery electrochemistry, auxiliary loads, rolling resistance, and driving conditions.

Energy-consumption estimation is also central to energy-aware route planning for electric vehicles, because candidate routes must be evaluated against available battery energy and possible charging requirements. Previous work connected energy prediction with energy-efficient routing under real-world conditions [5]. Traffic simulation provides a complementary setting for evaluating routes under specified operating conditions. For example, Ben Abbes et al. used the Simulation of Urban MObility (SUMO) to validate a multi-objective route and charging planning approach [6]. In such applications, the relevance of simulated energy estimates depends on how well the driving missions and vehicle characteristics represent the intended operation.

This connection motivates comparing energy-consumption relationships across vehicle types. Trip characteristics, driving behavior, and regenerative energy recovery may have different implications for sedans, logistics vehicles, and buses. By comparing statistical associations, predictive contributions, and covariate-adjusted effect estimates, this study provides empirical guidance on which operating dimensions warrant attention in vehicle-specific energy models and simulation scenarios. These findings can inform subsequent energy-aware routing and fleet-planning studies, although operational application requires further calibration and validation at the relevant spatial and temporal scales, together with consideration of which inputs are available before departure.

Despite this progress, vehicle type remains insufficiently examined. Many empirical studies either focus on a single type of vehicle or include vehicle type as a categorical variable in a pooled model [7]. This treatment may be too limited. Sedans, logistics vehicles, and buses differ in mass, battery capacity, aerodynamic characteristics, powertrain configuration, route structure, load variation, stop frequency, and auxiliary-load demand. Because vehicle types differ in both physical characteristics and operating patterns, the same explanatory variable may capture different aspects of energy use across vehicle types. For example, a higher trip-average speed in urban bus operation may indicate less stop-related delay and more continuous movement, with lower time-dependent auxiliary energy use per kilometer. For passenger cars operating in a higher-speed regime, a further increase in speed may instead be associated with greater aerodynamic energy demand. Thus, the relationship between average speed and ECR depends on the operating regime represented within each vehicle category [8,9]. A comparative analysis is therefore needed to examine which relationships are shared and which are vehicle specific. This issue is particularly relevant when pooled analyses do not adequately represent interactions between vehicle type and operating conditions.

This study addresses this gap using 41,933 valid trips collected from sedans, logistics vehicles, and buses in China. The three vehicle types were separately analyzed to uncover type-specific energy consumption mechanisms. To facilitate the analysis, this study adopted a three-perspective analytical framework that distinguishes three related but different questions. The statistical perspective asks how ECR changes with observed variables and whether response curves differ across vehicle types. The predictive perspective asks which variables are most useful for out-of-sample ECR prediction within each vehicle type. The causal perspective asks whether selected variables retain vehicle-specific marginal effects after conditioning on observed trip, driving, and vehicle-state covariates.

This study makes three contributions.

First, rather than treating vehicle type merely as a control variable, this study explicitly conceptualizes it as a source of heterogeneity and systematically compares sedans, logistics vehicles, and buses within a unified empirical framework to uncover vehicle-specific energy consumption mechanisms.

Second, it develops an integrated framework combining statistical characterization, vehicle-specific prediction with permutation importance, and CausalForestDML estimation to compare ECR determinants across vehicle types.

Third, it identifies vehicle-type-specific energy consumption drivers, with sedans being primarily associated with trip characteristics, logistics vehicles exhibiting stronger associations with speed and energy recovery, and buses showing greater sensitivity to driving smoothness and energy recovery.

The remainder of this paper is organized as follows. Section 2 reviews related research on BEV energy-consumption factors, modeling approaches, and vehicle-type heterogeneity. Section 3 describes the data and methodology. Section 4 presents the statistical, predictive, and causal results. Section 5 concludes the paper.

2. Literature Review

Existing BEV energy-consumption models include physics-based, data-driven, and hybrid approaches. Physics-based models use vehicle longitudinal dynamics to represent traction energy requirements, and traction energy is decomposed into rolling resistance, aerodynamic drag, acceleration resistance, road-grade resistance, drivetrain losses, and auxiliary loads. A power-based model for electric vehicle energy consumption using speed and acceleration profiles was developed in previous research [10]. Physics-based models provide transparent interpretation, but they require detailed parameters such as vehicle mass, drag coefficient, drivetrain efficiency, and battery characteristics. These parameters are often unavailable in large operational datasets, particularly when multiple fleets and vehicle types are considered. As a result, recent work has increasingly used real-world operating data and data-driven models to capture differences in energy consumption [11].

At the trip level, distance, duration, average speed, acceleration and deceleration behavior, SOC, cumulative mileage, ambient temperature, road gradient, precipitation, and auxiliary loads are among the most widely studied variables. Beyond its association with total trip energy consumption (EC), trip distance is also relevant to ECR, indicating that energy use per kilometer can vary with trip length. Short trips may exhibit higher ECR when initial thermal-conditioning demand is spread over fewer kilometers, whereas the contribution of sustained auxiliary loads also depends on trip duration and operating conditions [12]. Previous research showed that distance, speed, acceleration variability, and temperature jointly affect BEV consumption, while also confirming a nonlinear relationship between speed and consumption [13]. At low speeds, congestion and repeated starts increase per-kilometer losses; at high speeds, aerodynamic drag becomes dominant [9]. In addition, heating, ventilation, air conditioning (HVAC) loads, and driving style were

shown to influence EV energy consumption, with implications for estimating route-specific energy demand [4,14]. These findings are relevant to energy-efficient route planning, as demonstrated by research linking road-segment energy consumption prediction with route selection under varying operating conditions [10].

Driving behavior is another major source of energy-consumption variation. For example, acceleration intensity, braking behavior, and speed fluctuation were shown to substantially affect electric-vehicle range [14]. Link-level analyses under real-world traffic conditions further showed that traffic state and driving dynamics interact to influence EV energy consumption [15]. These findings are important because driving behavior can be influenced through eco-driving guidance, driver training, fleet monitoring, and driver-assistance systems. However, a rapid acceleration event imposes different energy penalties on a light sedan, a delivery vehicle, and a heavy bus. This observation motivates vehicle-type-specific sensitivity analysis rather than pooled behavioral interpretation.

Vehicle and battery-state variables add another layer of heterogeneity. SOC affects battery operating efficiency and may be associated with power-limitation strategies [16]. Cumulative mileage can serve as a proxy for battery aging, drivetrain wear, or long-term usage intensity [17]. Regenerative braking can offset part of traction demand, especially in stop-and-go operation, but its effectiveness depends on both operating conditions and losses during energy conversion and storage [18]. Vehicle mass, braking frequency, motor operating limits, and battery charge acceptance influence the potential for energy recovery, while electric motor efficiency in generating mode, inverter efficiency, and battery internal resistance affect the energy ultimately stored. DC/DC conversion efficiency also contributes to these losses where a converter is included in the regenerative power path [19,20]. A logistics vehicle operating in urban pickup-and-delivery service may have frequent deceleration opportunities, while a bus may have higher braking energy but stronger constraints from passenger load, route design, and stop density. Thus, the same vehicle-state indicator can carry different meanings across vehicle categories.

Environmental and road factors have also been examined, especially ambient temperature and road gradient. Ambient temperature affects BEV energy consumption through both battery chemistry and auxiliary loads, such as HVAC [4]. Road gradient directly changes the gravitational component of traction energy, and precipitation can increase rolling resistance and alter traffic behavior. Although these variables are often included as controls, their impact can vary by vehicle type. A bus, for example, has a larger cabin and higher auxiliary-load demand than a sedan, so temperature effects may be amplified in public-transport operation.

Regression and smoothing methods help characterize relationships between operating variables and ECR, while clustering identifies groups with similar driving or operating characteristics. One relevant pattern is the U-shaped relationship between speed and ECR reported in previous studies, with lower consumption per kilometer at intermediate speeds than at low or high speeds [9,21]. This pattern can be explained by the greater contribution of time-dependent auxiliary loads per kilometer at low speeds and increasing aerodynamic resistance at high speeds [8]. Higher ECR has also been observed for short trips, indicating that normalizing energy consumption by distance does not eliminate its association with trip length [22]. These patterns motivate models that account for nonlinear relationships and multiple operating factors. Machine-learning approaches complement statistical characterization by combining these factors for energy-consumption prediction [23]. Tree-based ensemble models, including Random Forests, XGBoost, and LightGBM, can represent nonlinear relationships and interactions, with LightGBM providing an efficient gradient-boosting implementation [24]. Interpretation tools such as SHapley Additive exPlanations

(SHAP) [25] and permutation importance [26] can then be used to examine how individual variables contribute to model predictions and predictive performance.

Predictive importance does not identify the effect of an intervention. A useful predictor may also represent an unmeasured operating condition that influences both the predictor and the ECR. In causal terminology, the outcome Y is the response of interest, the treatment T is the variable whose effect is being examined, and covariates X describe observed background conditions. For continuous treatments, an effect can be expressed as the change in expected outcome associated with a small increase in T under the required identification assumptions. Double machine learning (DML) uses auxiliary, or nuisance, prediction models for the outcome and treatment to remove the components predictable from X before estimating their remaining relationship [27]. Flexible learners can represent multiple, potentially nonlinear covariate relationships; this is the purpose of the high-dimensional-control terminology, rather than a separate type of vehicle variable. Causal forests allow the estimated effect to vary with observed characteristics [28]. Previous work applied statistical, predictive, and causal perspectives to EV energy consumption across trip-distance categories [29]. Here, the comparison is organized by vehicle type.

Vehicle-type heterogeneity remains underdeveloped. Passenger-car studies often emphasize individual driving behavior, commuting patterns, speed profiles, and temperature sensitivity [11]. Electric-bus studies highlight route structure, stop spacing, passenger load, service type, schedule regularity, and HVAC demand. Previous studies estimated electric-bus energy demand using real-world public-transport data and developed energy-consumption models for strategic planning [30,31]. Backward bus models use speed profiles and passenger-dependent vehicle mass to estimate energy demand [32], providing a mechanistic perspective that complements the trip-level empirical analysis adopted here. The planning relevance of operating patterns is also evident in trolleybus grids, where the timing of electricity demand affects solar-energy utilization [33]. Electric freight and delivery-vehicle studies emphasize payload, delivery frequency, depot constraints, route planning, and stop density [34]. These operational differences motivate vehicle-specific energy estimates that, when combined with operating and charging schedules, can inform charging-demand assessment.

Taken together, the studies reviewed above establish the importance of operating conditions and vehicle context in BEV energy consumption. However, evidence from separate applications does not by itself show whether the direction, shape, and predictive relevance of an ECR relationship are comparable across sedans, logistics vehicles, and buses. Differences in datasets, variable definitions, and analytical methods can also complicate comparisons between published results. This study addressed this comparative question using common trip-level definitions and a consistent analytical design across the three categories. It separately examined observed associations, predictive contributions, and covariate-adjusted marginal-effect estimates, thereby clarifying where these forms of evidence agree and where their interpretations differ. The resulting comparisons concern the observed fleets and operating conditions; they do not isolate vehicle hardware from all differences in operating missions.

3. Data and Methodology

3.1. Data Source and Variable Construction

The study uses real-world operational data from a BEV fleet in Shanghai, China, collected between 13 July and 26 July 2023. The dataset covers 1000 BEVs, including private sedans, public buses, and freight logistics vehicles. Vehicle-side telemetry was recorded at 0.1 Hz under the Chinese national communication protocol for new energy vehicles (GB/T 32960) [35]. The raw records include timestamps, operating status, speed, cumulative

mileage, total voltage, total current, and other state variables. Before analysis, records with missing key variables or physically implausible values, including abnormal speed, voltage, or current, were removed. Valid driving records were then identified using the operating status field.

Because instantaneous energy consumption is strongly affected by transient motor response and short-term fluctuations in traffic and road gradient, the data were aggregated to the trip level. A new trip was defined when the interval between two consecutive valid driving records exceeded three minutes. Trips shorter than 10 min were excluded to reduce incomplete or unreliable driving episodes. The final analytical sample contained 41,933 valid trips: 11,308 sedan trips, 20,569 logistics vehicle trips, and 10,056 transit-bus trips. Table 1 summarizes the technical specifications and energy-consumption profiles of the fleet.

Table 1. Technical specifications and energy consumption profiles across vehicle types.

Characteristic	Sedan	Logistics Vehicle	Bus
Vehicle models	2	2	3
Sample size	385	506	109
Curb weight (kg)	1550/2000	1700/2900	11,800/17,300
Dimensions [1] (mm)	4500 × 1800 × 1500	5100–5300 × 1700–1800 × 1900–2300	10,500–12,000 × 2500–2600 × 3100–3200
Battery chemistry [2]	NMC	LFP	LFP
Rated power (kW)	50.8/61.1	50.2/41.9	324.4/303.7/248.2
Battery capacity (Ah)	174/145	150/125	528/542/480
Trajectory records	12,971,274	16,253,885	2,273,999
Valid trips	11,308	20,569	10,056
Mean EC [3] (kWh)	8.7953	6.2409	17.4552
SD EC (kWh)	8.2732	6.6097	11.8442
Mean ECR (kWh/km)	0.1697	0.1691	0.9618
SD ECR (kWh/km)	0.0635	0.0524	0.2221

[1] Dimensions are length × width × height. [2] NMC, nickel manganese cobalt; LFP, lithium iron phosphate. [3] EC, energy consumption; ECR, energy consumption rate; SD, standard deviation.

For each trip, total energy consumption (EC) was calculated by integrating the product of total voltage, total current, and the sampling interval:

$$EC_i = \sum_{t=1}^{N_i-1} \frac{V_{it} I_{it} \Delta t_{it}}{3.6 \times 10^6}, \quad (1)$$

where EC_i is the energy consumption of trip i in kWh, V_{it} and I_{it} are total voltage and total current at sampling point t, and Δt_{it} is the time interval between adjacent records in seconds. N_i denotes the number of sampling points recorded for trip i. To avoid overestimation caused by occasional long sampling gaps, intervals larger than the modal sampling interval were replaced by the modal interval. The main outcome variable was ECR:

$$ECR_i = \frac{EC_i}{d_i}, \quad (2)$$

where d_i is the trip distance in kilometers. ECR_i, measured in kWh/km, is used as the dependent variable in the statistical, predictive, and causal analyses.

The 18 trip-level explanatory variables are categorized into four groups. Trip characteristics consist of trip distance (km), trip duration (min), and mean speed (km/h). Driving-behavior indicators are represented by the mean, maximum, and standard deviation of acceleration and deceleration (m/s^2). Pedal-related variables are characterized by the mean, maximum, and standard deviation of accelerator and brake pedal positions (%). Vehicle-state variables comprise the initial state of charge (%), accumulated vehicle mileage (km), and the energy recovery ratio (%), which is defined as the proportion of regenerative

braking energy recovered during a trip relative to the total energy consumed during that trip. The detailed definitions and units of all 18 variables are provided in Table A1.

3.2. Statistical Characterization and Vehicle-Specific Prediction

To provide an intuitive statistical characterization of vehicle-specific ECR patterns, Pearson correlation analysis was first conducted separately for sedans, logistics vehicles, and buses. To further examine relationships that could not be adequately represented by linear correlation coefficients, locally weighted scatterplot smoothing (LOWESS) was applied to selected continuous variables. This non-parametric approach describes the local response of ECR to variables such as trip distance, trip duration, mean speed, battery state of charge, energy recovery ratio, and vehicle mileage without imposing a predefined linear or polynomial form. In addition, driving style and driving condition were constructed using K-means clustering within each vehicle type. The driving-style clusters were derived from accelerator- and brake-pedal variables, whereas the driving-condition clusters were derived from speed, acceleration, and deceleration variables. The resulting cluster profiles are reported in Tables A2 and A3 in the Appendix A, which present the sample size, centroid characteristics, and interpretation labels for each vehicle-type-specific cluster. Based on these centroid patterns, the driving-condition clusters were interpreted as Congestion, Moderate Traffic, and Cruising, while the driving-style clusters were interpreted as Mild, Normal, and Aggressive. This procedure transformed multidimensional operating behavior into interpretable categories, allowing ECR differences to be compared across behavioral patterns and vehicle types.

The second analytical step was vehicle-specific ECR prediction. Separate models were developed for sedans, logistics vehicles, and buses rather than fitting a pooled model. In each vehicle-specific model, ECR was predicted as a function of trip characteristics, driving-behavior variables, and vehicle-state variables, with the residual term representing unexplained variation. For each vehicle type, the dataset was randomly divided into training, validation, and test subsets using a 64/16/20 split. This ratio was obtained by first reserving 20% of the observations as an independent test set and then assigning 20% of the remaining 80% to validation, leaving 64% for model training. The design preserves a sufficiently large training sample for each vehicle-specific model, provides a separate validation subset for hyperparameter selection, and keeps the final test set untouched for unbiased out-of-sample evaluation. Five algorithms were compared: Ridge Regression, Random Forest, Extra Trees, Histogram-based gradient boosting (HistGradientBoosting), and LightGBM. Numerical missing values were imputed using the median of each variable calculated from the training data.

Hyperparameters were selected on the validation subset using the root mean squared error (RMSE) as the optimization criterion. The selected configuration for each algorithm was refitted on the combined training and validation data and evaluated on the held-out test set using RMSE, mean absolute error (MAE), coefficient of determination (R^2), and mean absolute percentage error (MAPE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{3}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{5}$$

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|. \quad (6)$$

where y_i and $\hat{y}_i$ denote the actual and predicted values for the *i*-th test sample, respectively, $\bar{y}$ denotes the mean of the actual values, and *n* is the number of test samples.

For the best-performing model of each vehicle type, feature importance was quantified using model-agnostic permutation importance. The permutation importance of feature *j* was defined as

$$PI_j = RMSE_j^{perm} - RMSE_{base}, \quad (7)$$

where $RMSE_{base}$ denotes the original test RMSE, and $RMSE_j^{perm}$ denotes the RMSE obtained after randomly permuting feature *j* while keeping all other variables unchanged.

3.3. Causal Effect Estimation

CausalForestDML was used to estimate marginal effects of selected continuous variables on energy consumption rate (ECR, kWh/km), conditional on observed covariates. For each estimation task, ECR was specified as the outcome (Y), and one selected continuous variable, such as trip distance, trip duration, or energy recovery ratio, was specified as the treatment (T). Here, "treatment" refers to the variable whose effect on ECR is being evaluated in a given model. The covariate vector X comprised the remaining observed trip characteristics, driving-behavior indicators, and vehicle-state variables listed in Table A1, excluding the current treatment. These covariates were used to adjust for observed differences between trips and to characterize treatment-effect heterogeneity. Total trip energy consumption (EC, kWh), which is directly involved in the construction of ECR, was excluded from the covariate set. Vehicle type was used to define separate analysis groups, with separate models fitted for sedans, logistics vehicles, and buses.

CausalForestDML follows the logic of double machine learning. In the first stage, LightGBM regressors were used as nuisance models to predict the outcome and treatment from X, estimating the following conditional expectations:

$$m(X) = E(Y \mid X), \quad g(X) = E(T \mid X). \quad (8)$$

Let $\hat{m}(X)$ and $\hat{g}(X)$ denote the fitted estimates of $m(X)$ and $g(X)$, respectively. The residualized outcome and treatment were obtained by subtracting these predicted values from the corresponding observed values:

$$\widetilde{Y} = Y - \hat{m}(X), \ \widetilde{T} = T - \hat{g}(X). \quad (9)$$

Thus, $\widetilde{Y}$ represents the portion of ECR not explained by the fitted outcome model using X, while $\widetilde{T}$ represents the portion of the treatment not explained by the fitted treatment model using the same covariates. In the second stage, a causal forest estimates the covariate-dependent treatment effect through local regression of the residualized outcome on the residualized treatment:

$$\widetilde{Y} = \tau(X)\widetilde{T} + \epsilon. \quad (10)$$

Here, ϵ denotes the remaining unexplained variation. For the continuous-treatment specification used here, $\tau(X)$ represents the conditional marginal effect on ECR, reported as the conditional average treatment effect (CATE) for a one-unit increase in the treatment

on the scale used in estimation. The average treatment effect (ATE) was calculated as the sample mean of these estimated marginal effects within each vehicle type:

$$\hat{\mathrm{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\tau}(X_i), \tag{11}$$

where n is the number of trips in the corresponding vehicle-type sample. Because the data are observational, a causal interpretation of the estimated effects requires the assumption that the observed covariates sufficiently control confounding. The analysis therefore examines whether the estimated factor effects remain vehicle specific after adjustment for observed trip characteristics, driving behavior, and vehicle state.

4. Results

4.1. Statistical Characterization of ECR Determinants

Figure 1 presents the Pearson correlation matrices for sedan, logistics vehicle, and bus. Overall, ECR exhibits pronounced vehicle-specific correlation structures, indicating substantial heterogeneity in energy-use mechanisms across vehicle types. Across all vehicle types, ECR is consistently associated with trip distance and trip duration, but the relationship is weaker for buses than for sedans and logistics vehicles, suggesting reduced distance dependence for buses. EC, by comparison, is strongly positively correlated with trip distance and trip duration across all three vehicle types. In contrast, ECR–behavioral correlations differ by vehicle type: they are relatively weak for sedans, while the energy recovery ratio shows consistently negative and strongest associations with ECR for buses. The internal correlation structure of behavioral variables also varies, with stronger within-group clustering observed for buses.

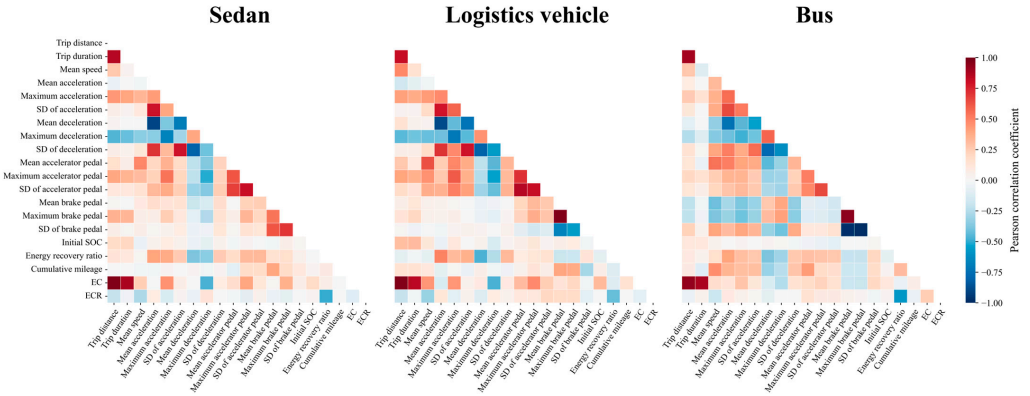


Figure 1. Lower-triangle Pearson correlation heatmaps by vehicle type.

Among the explanatory variables, trip distance and trip duration show strong positive correlations across all three vehicle types, reflecting their shared characterization of trip scale. Mean acceleration is positively correlated with the standard deviation of acceleration and negatively correlated with mean deceleration. Since deceleration is recorded as a negative quantity, the latter relationship indicates that trips with stronger acceleration tend to also involve larger deceleration magnitudes. Accelerator-pedal indicators also exhibit positive within-group correlations, while the correlation patterns among brake-pedal indicators differ across vehicle types. These associations suggest that some explanatory variables capture overlapping aspects of trip characteristics and driving behavior. Such dependencies should be considered when interpreting the subsequent ECR analyses, particularly individual permutation-importance scores, because correlated predictors can carry overlapping predictive information.

The inverse fitting between trip distance and ECR further reveals heterogeneous short-trip penalties, as shown in Figure 2, with ECR = 0.222/d + 0.160 (R^2 = 0.750) for sedans and ECR = 0.122/d + 0.154 (R^2 = 0.530) for logistics vehicles, where d denotes trip distance (km). Therefore, both types show a clear nonlinear decay pattern: ECR is high during very short trips, declines rapidly as distance increases, and converges toward approximately 0.15–0.16 kWh/km. This pattern is consistent with start-up losses, low-speed operation, repeated acceleration and deceleration, and insufficient amortization of auxiliary loads during short trips. For logistics vehicles, the further decline in the plotted ECR values beyond the relatively flat region around 150 km may reflect differences in operating conditions, including more continuous travel and fewer stop-and-go events on longer trips. However, the present analysis did not classify trips by road type or delivery service, so approximately 150 km should not be interpreted as an established boundary between last-mile delivery and long-haul transport. The fitted function for buses is ECR = 0.315/d + 0.909 (R^2 = 0.095). Therefore, trip distance alone explains little of the variation in bus ECR. The much higher asymptotic term also indicates that buses maintain a baseline ECR close to 0.91 kWh/km even over longer trips. Bus energy intensity is better interpreted as a structural consequence of heavy vehicle mass, load variation, auxiliary demand, and service operation than as a simple distance-amortization effect.

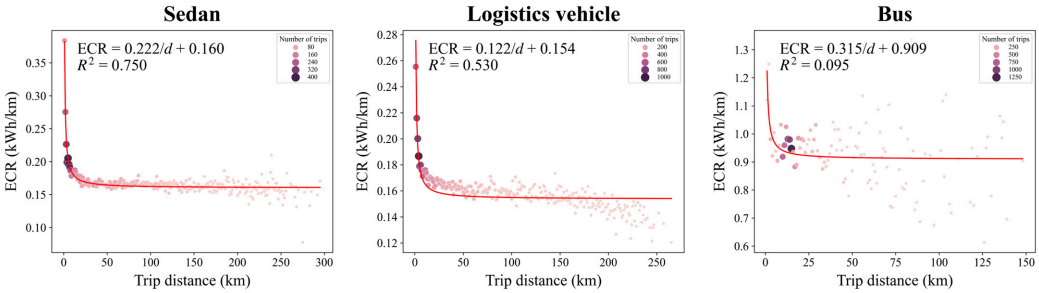


Figure 2. Inverse-distance fitting between trip distance and ECR by vehicle type.

Figure 3 indicates that trip-level operating patterns differ markedly across vehicle types, with buses showing shorter distances, shorter durations, and lower mean speeds due to their route-based and stop-intensive service mode. As trip distance increases, the ECR generally exhibits a smooth decreasing trend, particularly for sedan and logistics vehicles, suggesting that start-up energy demand and auxiliary energy consumption are progressively diluted over longer trips. In contrast, the relationship between trip distance and ECR fluctuates more for buses, indicating that energy intensity is influenced not only by trip length but also by additional factors such as route characteristics, stop frequency, and passenger load conditions. For sedans and logistics vehicles, the duration–ECR curves initially decline and then rebound modestly, whereas the bus curve generally rises with local fluctuations. These patterns may reflect variation in both trip distance and auxiliary energy demand, because additional time in slow traffic or at stops does not need to be accompanied by a proportional increase in distance. The curves describe unadjusted associations and do not establish an independent causal effect of trip duration. Mean speed shows the clearest nonlinear pattern: sedans and logistics vehicles exhibit a U-shape relationship, with the lowest ECR at around 40–45 km/h. In contrast, buses are concentrated in a lower-speed urban operating regime, where higher average speed mainly reflects smoother traffic flow, fewer stop-start events, and reduced dwell or idling time, thereby corresponding to lower ECR.

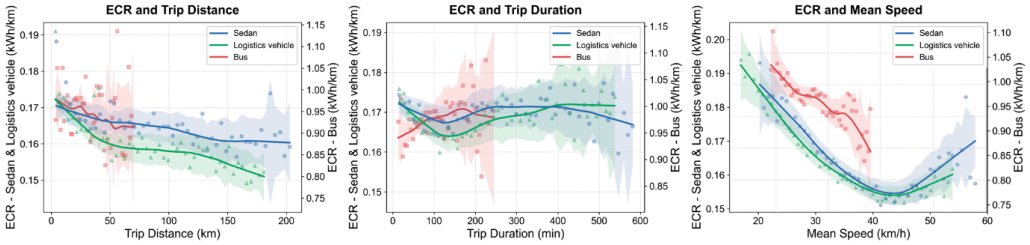


Figure 3. Associations between trip characteristics and ECR across vehicle types. Note: The dots represent binned mean ECR values, the solid lines represent LOWESS-smoothed relationships, and the shaded areas indicate the corresponding confidence intervals. The same conventions apply to Figure 4.

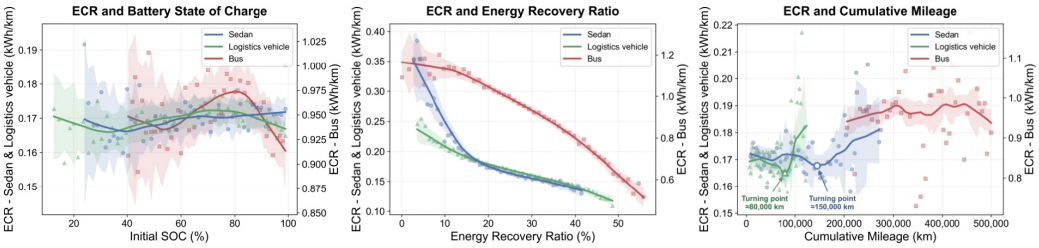


Figure 4. Associations between vehicle operating-state variables and ECR across vehicle types.

Figure 4 shows that vehicle-state variables affect ECR in different ways across vehicle types. The relationship between initial SOC and ECR is relatively weak and non-uniform: sedans show only limited variation across most SOC levels, while logistics vehicles and buses display more irregular nonlinear patterns. This suggests that SOC-related differences are likely mediated by vehicle-specific battery management and operating schedules rather than representing a simple direct effect of initial charge level. A clearer pattern appears for the energy recovery ratio. ECR decreases as the recovery ratio increases, but the sedan curve shows a visible flattening after the low-recovery range, indicating diminishing marginal returns: once a certain recovery level is reached, further improvement in the recovery ratio contributes less to reducing the ECR. The mileage-related pattern is more informative. For both sedans and logistics vehicles, ECR first declines slightly and then rises more clearly, with the turning points occurring at cumulative mileages of approximately 150,000 km for sedans and 80,000 km for logistics vehicles. These transitions suggest potential degradation-related changes associated with battery aging, increased internal resistance, drivetrain wear, or thermal-management burden. The turning point appears earlier for logistics vehicles than for sedans, which may reflect more intensive commercial operation and heavier battery-management stress. Therefore, ECR may serve not only as an energy-efficiency metric but also as a useful indicator for identifying or even predicting vehicle-health degradation points.

To improve interpretability of driving-behavior effects, K-means clustering was applied to aggregate high-dimensional pedal and speed-related variables into representative operating patterns. Figure 5 presents the ECR differences across clustered driving styles and driving conditions. The driving-style clusters are labelled as Mild, Normal, and Aggressive according to pedal-operation characteristics rather than predefined behavioral categories. Sedans show only small differences across these styles, with mean ECR remaining around 0.169–0.171 kWh/km. In contrast, logistics vehicles and buses are more sensitive to driving style: the Aggressive style corresponds to the highest ECR, reaching approximately 0.200 kWh/km for logistics vehicles and 0.975 kWh/km for buses. Relative to the Mild style, the Aggressive style therefore increases logistics-vehicle mean ECR by 22.2%, whereas the corresponding difference for sedans is minimal. This suggests that

stronger accelerator-pedal demand imposes a clearer energy penalty, especially for buses, whose larger vehicle mass requires greater traction energy during acceleration. For driving conditions, the three clusters are interpreted as Congestion, Moderate traffic, and Cruising. Across vehicle types, Cruising is associated with the lowest ECR, while the Moderate traffic condition shows the highest ECR, indicating that urban operating scenarios with more frequent speed changes and interruptions reduce energy efficiency.

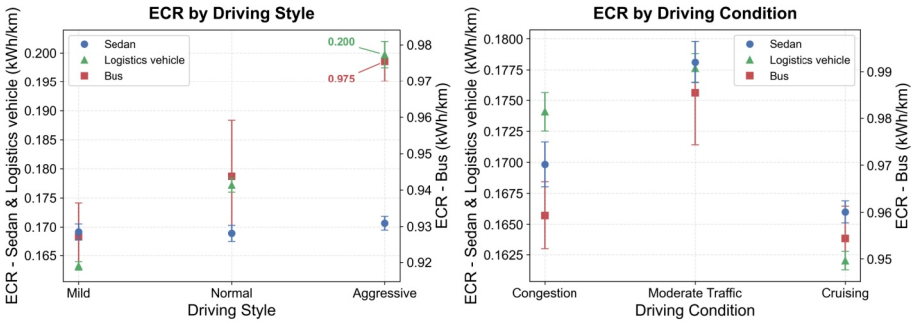


Figure 5. ECR variations across driving-style and operating-condition clusters.

4.2. Vehicle-Specific Prediction

To evaluate the predictive performance of the final variable set, separate ECR prediction models were developed for sedans, logistics vehicles, and buses. For each vehicle group, five candidate models were compared, including four nonlinear ensemble methods (LightGBM, HistGradientBoosting, Random Forest, and Extra Trees) and a linear baseline model (Ridge Regression). Model performance was assessed using validation and test metrics, including RMSE, MAE, R^2 and MAPE, to jointly evaluate prediction accuracy and explanatory capability. The comparison results are reported in Table 2. Overall, this provides a systematic basis for selecting the best-performing model for each vehicle type and for assessing whether nonlinear ensemble models offer advantages over a linear benchmark in capturing ECR variability.

Table 2. Performance of machine-learning models for vehicle-specific ECR prediction.

Vehicle Type	Model	RMSE	MAE	R^2	MAPE (%)
Sedan	LightGBM	0.0304	0.0187	0.7886	9.8832
	HistGradientBoosting	0.0309	0.0192	0.7829	10.1897
	Random Forest	0.0346	0.0218	0.7270	11.6778
	Extra Trees	0.0354	0.0223	0.7147	11.9391
	Ridge Regression	0.0474	0.0289	0.4884	15.5076
Logistics vehicle	LightGBM	0.0215	0.0140	0.8166	8.0456
	HistGradientBoosting	0.0217	0.0143	0.8141	8.2122
	Random Forest	0.0244	0.0163	0.7644	9.4253
	Extra Trees	0.0248	0.0167	0.7561	9.7387
	Ridge Regression	0.0313	0.0212	0.6118	12.4433
Bus	LightGBM	0.1102	0.0817	0.7506	9.1619
	HistGradientBoosting	0.1106	0.0828	0.7486	9.2819
	Random Forest	0.1228	0.0922	0.6900	10.2063
	Extra Trees	0.1256	0.0948	0.6756	10.5284
	Ridge Regression	0.1292	0.0979	0.6567	11.1335

Figure 6 presents the observed-versus-predicted scatter plots for the three vehicle types, and LightGBM was selected as the best-performing model for all three vehicle types, although the predictive patterns were not identical. For sedan vehicles, LightGBM achieved

RMSE = 0.0304 kWh/km, MAE = 0.0187 kWh/km, and R^2 = 0.789. The observed-versus-predicted scatter plot shows a dense concentration of trips in the low-to-medium ECR range, mainly around 0.10–0.25 kWh/km, where predictions closely follow the 1:1 reference line. At higher observed ECR values, the scatter becomes wider and several points fall below the reference line, indicating that extreme high-consumption sedan trips are partly underestimated. This tail behavior suggests that the selected predictors effectively capture the dominant ECR patterns of sedan trips but still have limited precision for less frequent high-ECR cases.

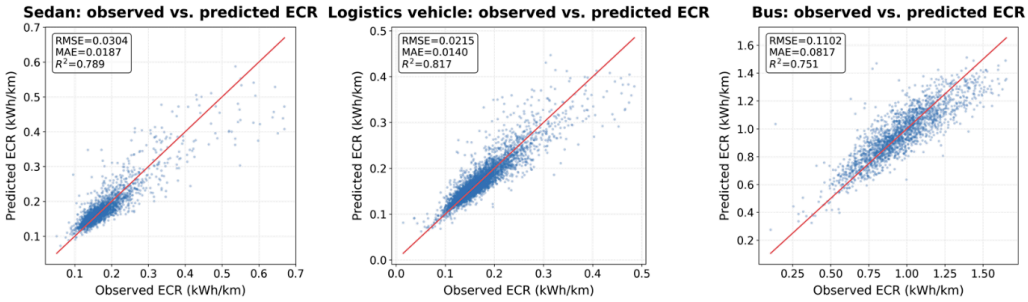


Figure 6. Observed and predicted ECR for the LightGBM model by vehicle type. Note: The red line represents the 1:1 reference line (predicted ECR = observed ECR), indicating perfect agreement between predictions and observations.

The logistics vehicle model produced the strongest overall predictive performance among the three types, with RMSE = 0.0215 kWh/km, MAE = 0.0140 kWh/km, R^2 = 0.817, and MAPE = 8.05%. The scatter plot is tighter than the sedan plot across the main ECR range, and most predictions align closely with the diagonal line between approximately 0.10 and 0.25 kWh/km. Some dispersion remains in the upper range, especially for observed ECR values above approximately 0.30 kWh/km, where several high-ECR observations are predicted at lower values. Nevertheless, the low RMSE and the highest R^2 indicate that the selected variables explain logistics-vehicle ECR more consistently than they explain the other two vehicle types.

Bus prediction exhibits a different pattern. Although LightGBM still outperformed the other candidate models, the absolute error was much larger than those of the LightGBM models for sedans and logistics vehicles, with RMSE = 0.1102 kWh/km and MAE = 0.0817 kWh/km. This larger error partly reflects the broader ECR scale of buses: the observed test values extend from approximately 0.11 to 1.65 kWh/km, far beyond the sedan and logistics-vehicle ranges. The bus scatter plot shows a clear positive relationship between observed and predicted ECR, consistent with R^2 = 0.751, but the point cloud is more dispersed. Predictions are compressed relative to the observed range, with some low observed values overestimated and some high observed values underestimated. This pattern indicates stronger residual heterogeneity in bus operation even when utilizing the comprehensive feature set that captures key operational dimensions.

Figure 7 displays the feature importance analysis, which further reveals substantial heterogeneity in the determinants of ECR across vehicle types. For sedan vehicles, trip-level descriptors dominated the prediction process. Trip distance and trip duration were the two most influential variables, followed by energy recovery ratio, indicating that passenger-vehicle ECR is mainly shaped by the overall scale and temporal structure of individual trips. In this group, driving-behavior variables such as accelerator-pedal and braking indicators contributed only marginally, consistent with the smaller ECR differences across driving-style clusters for sedans than for logistics vehicles and buses (Figure 5). This suggests that once trip distance, duration, and recovery information are known, short-term control behavior adds relatively limited predictive value.

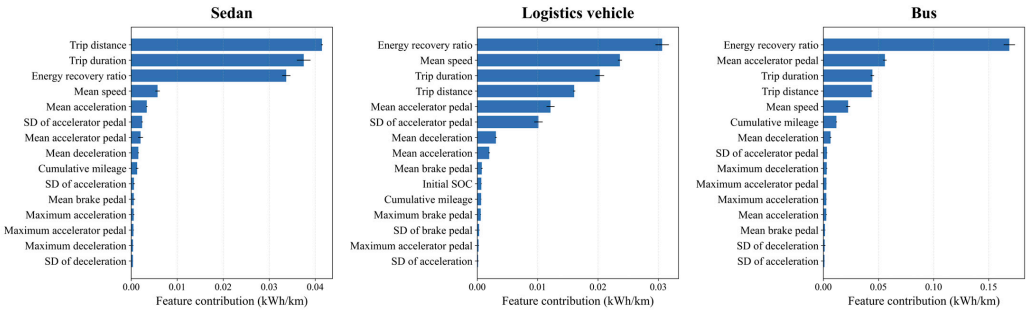


Figure 7. Permutation feature importance for the best-performing model by vehicle type.

For logistics vehicles, the contribution structure became more balanced. Energy recovery ratio ranked first, followed by mean speed, trip duration, trip distance, and accelerator-pedal variables. This suggests that logistics-vehicle ECR is not determined only by trip scale, but also by operating continuity and driving demand. Compared with sedan vehicles, the stronger contribution of speed and pedal-use variables is consistent with logistics-vehicle operation, where delivery tasks, load variation, and frequent speed changes may amplify the effect of driving behavior on unit energy consumption.

The bus model showed the most distinctive feature-importance pattern. Energy recovery ratio was by far the dominant predictor, and its contribution substantially exceeded that observed for the other two vehicle types. This indicates that bus ECR prediction is highly sensitive to regenerative braking behavior, likely because buses have larger vehicle mass and operate under frequent stop-and-go service conditions. These results suggest that future bus ECR prediction models should give particular attention to the energy recovery ratio, alongside trip-scale, speed-related, and pedal-operation variables.

4.3. Causal Analysis

Vehicle-specific CausalForestDML models were estimated to examine whether key variables retained heterogeneous conditional effects on ECR after adjustment for the observed covariates specified in Section 3.3. ECR was specified as the outcome, and each treatment was estimated separately within the sedan, logistics-vehicle, and bus samples. The top three important variables identified from the permutation importance analysis were selected for causal analysis in the main text, while the full ATE results for all treatment variables are reported in Table A4. In the following figures, each point represents an observed trip, the color scale reports the estimated CATE on ECR, and the black dashed line summarizes the binned mean CATE.

Figure 8 shows the CATE patterns for trip distance. The estimated effects are negative for all three vehicle types, which is consistent with the statistical analysis showing a declining distance–ECR relationship, especially for sedans and logistics vehicles. The estimated marginal effects of trip distance therefore remain negative after adjustment for the observed covariates specified in Section 3.3. However, the vehicle-type difference is clear. Sedans and logistics vehicles show relatively compact effect ranges, whereas buses show a wider color scale and a more irregular binned mean line in the sparse long-distance region. Distance therefore operates as a common efficiency-related variable, but its conditional effect is stronger and less stable for buses.

Figure 9 shows the CATE patterns for trip duration. Unlike the distance and energy recovery ratio, duration does not have a common negative effect across vehicle types. For sedans and logistics vehicles, the estimated marginal effects of trip duration are close to zero and slightly negative after adjustment for the observed covariates specified in Section 3.3. This is compatible with the statistical and predictive results in which distance and duration jointly describe trip scale, but distance provides the clearer efficiency gradient.

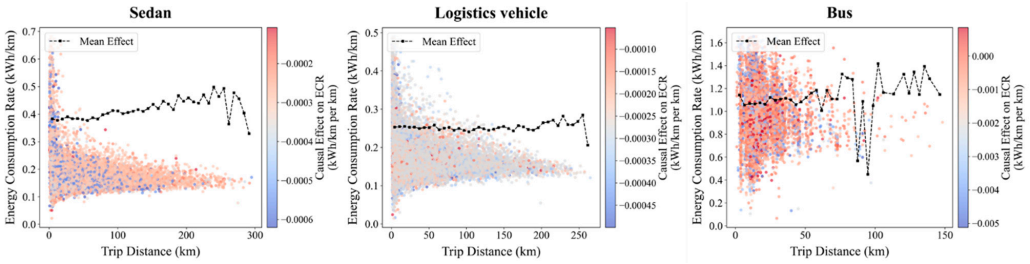


Figure 8. CATE patterns for trip distance by vehicle type.

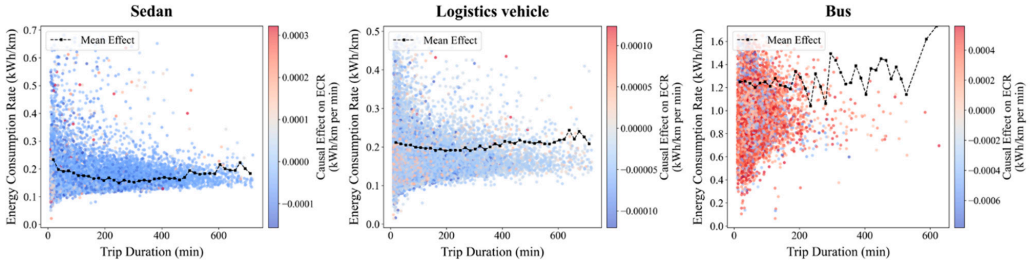


Figure 9. CATE patterns for trip duration by vehicle type.

Figure 10 shows the CATE patterns for the energy recovery ratio. All three vehicle types show negative estimated effects, indicating that higher recovery ratio is associated with lower ECR. The effect scale is largest for buses, while sedans and logistics vehicles show smaller negative ranges. This finding aligns with the earlier statistical and predictive results, which indicated that vehicle-state and recovery-related variables are more important for commercial vehicles than for sedans. The bus pattern is also consistent with the operating logic of high-mass, stop-intensive vehicles, where regenerative recovery can play a larger role in offsetting traction energy demand.

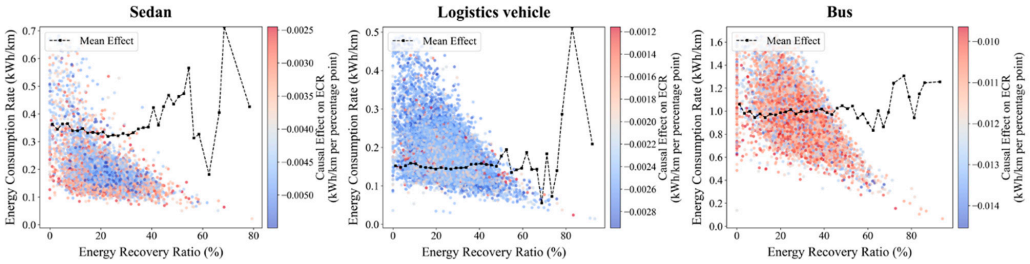


Figure 10. CATE patterns for energy recovery ratio by vehicle type.

For buses, the duration effect is positive rather than negative. The figure also shows more warm-colored observations in the dense short- and medium-duration regions, while the binned mean effect becomes volatile in the long-duration tail. Because very long bus trips are limited, the tail should be interpreted cautiously. Nevertheless, the result indicates that longer bus operation does not necessarily dilute ECR in the same way as longer distance does; it may instead reflect service exposure, dwell time, or route constraints that are not fully captured by the controls.

Taken together, the causal results reinforce the main pattern found in the statistical and predictive analyses: ECR mechanisms are vehicle specific. Trip distance represents the common trip-scale effect, energy recovery ratio captures a recovery-related mechanism that is strongest for buses, and trip duration separates bus operation from sedan and logistics-vehicle operation. These findings support a vehicle-specific interpretation rather than a pooled explanation of ECR formation.

5. Conclusions

This study investigated energy-consumption heterogeneity across sedans, logistics vehicles, and buses using statistical characterization, machine learning-based prediction, and causal-effect estimation. The results provide a comprehensive view of how energy consumption mechanisms differ across vehicle types in real-world operating conditions.

The results reveal pronounced vehicle-specific energy-consumption mechanisms. Statistically, sedan and logistics-vehicle ECR decreases as trip distance increases (inverse-fit R^2 = 0.750 and R^2 = 0.530, respectively), whereas bus ECR shows no clear distance dependence. Aggressive driving raises logistics-vehicle ECR by 22.2% but barely affects sedans. In prediction, LightGBM achieves R^2 of 0.789 for sedans, 0.817 for logistics vehicles, and 0.751 for buses, with permutation importance confirming that trip distance and duration dominate sedan prediction, energy recovery ratio, and mean speed dominate logistics-vehicle prediction, and energy recovery ratio is the single strongest predictor for buses. The causal analysis further confirms that these divergences persist after covariate adjustment: trip duration exhibits opposite directional effects for buses relative to sedans and logistics vehicles, and the magnitude of the energy recovery effect is substantially larger for buses.

These findings can inform the design of vehicle-specific driving missions for simulation-based energy assessment and route planning. For sedans, the importance of trip distance and duration suggests including trips of different lengths and durations, particularly short-distance trips, across the observed speed range. For logistics vehicles, the predictive relevance of mean speed and energy recovery ratio motivates scenarios with different degrees of speed variation, acceleration and braking demand, and delivery-related interruptions. For buses, the dominant predictive role of energy recovery ratio and the positive estimated duration effect motivate sensitivity analyses of stop-and-go operation, dwell time, and regenerative energy recovery under service constraints. Stop frequency and dwell time are proposed dimensions for subsequent simulation studies, rather than independently quantified effects in the present analysis. Simulated trips could be aggregated using consistent definitions and compared with the observed vehicle-specific ECR distributions and relationships to assess whether the simulated missions reproduce relevant operating patterns.

The findings are based on trip-level data from a two-week summer period in Shanghai, and their causal interpretation depends on observational identification assumptions. Their use in route planning requires further calibration and validation, as trip-level indicators cannot be directly translated into simulation parameters or assumed to be available before departure.

Overall, this study demonstrates that vehicle type should be treated as a core heterogeneity dimension in BEV energy-consumption analysis. Ignoring such heterogeneity may lead to biased interpretations of energy-use mechanisms and suboptimal policy recommendations. Vehicle-specific modeling therefore provides a more reliable basis for both energy-consumption prediction and system-level efficiency improvement.

Author Contributions: Conceptualization, L.L. and H.H. (Haichao Huang); Data curation, H.H. (Haichao Huang) and Y.W.; Investigation, L.L.; Formal analysis, L.L. and H.H. (Haichao Huang); Writing—original draft preparation, L.L.; Writing—review and editing, L.L., H.H. (Hongdi He) and H.H. (Haichao Huang); Supervision, H.H. (Hongdi He). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy.

Acknowledgments: The authors would like to thank Shanghai Electric Vehicle Public Data Collection Monitoring and Research Center for providing the electric vehicle data and the members of the Center for ITS and UAV Applications Research at Shanghai Jiao Tong University for their assistance in conducting the experiments.

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

Abbreviations

The following abbreviations are used in this manuscript:

ATE	average treatment effect
BEV	battery electric vehicle
CATE	conditional average treatment effect
DML	double machine learning
EC	energy consumption
ECR	energy consumption rate
HVAC	heating, ventilation, and air conditioning
LFP	lithium iron phosphate
LOWESS	locally weighted scatterplot smoothing
MAE	mean absolute error
MAPE	mean absolute percentage error
NMC	nickel manganese cobalt
RMSE	root mean squared error
SD	standard deviation
SHAP	SHapley Additive exPlanations
SOC	state of charge
SUMO	Simulation of Urban MObility
XGBoost	extreme gradient boosting

Appendix A

Table A1. Definitions of the Trip-Level Explanatory Variables.

Category	Variable	Definition	Unit
Trip characteristics	Trip distance	Total travel distance of a trip	km
	Trip duration	Total duration of a trip	min
	Mean speed	Average speed during a trip	km/h
Driving behavior (acceleration/deceleration)	Mean acceleration	Mean acceleration over the trip	m/s^2
	Max acceleration	Maximum acceleration during the trip	m/s^2
	SD of acceleration	Variability of acceleration over the trip	m/s^2
	Mean deceleration	Mean deceleration over the trip	m/s^2
	Max deceleration	Maximum deceleration during the trip	m/s^2
	SD of deceleration	Variability of deceleration over the trip	m/s^2
Driving behavior (pedal operation)	Mean accelerator pedal	Average accelerator pedal position during the trip	%
	Max accelerator pedal	Maximum accelerator pedal position during the trip	%
	SD of accelerator pedal	Variability of accelerator pedal position during the trip	%
	Mean brake pedal	Average brake pedal position during the trip	%
	Max brake pedal	Maximum brake pedal position during the trip	%
	SD of brake pedal	Variability of brake pedal position during the trip	%
Vehicle state	Initial SOC	Battery state of charge at the beginning of the trip	%
	Energy recovery ratio	Ratio of regenerative energy recovered during a trip to total trip energy consumption	%
	Cumulative mileage	Accumulated vehicle mileage	km

Table A2. Driving Condition Clusters.

Vehicle Type	Cluster	*n*	Mean Speed (km/h)	Mean Accel(m/s^2)	Mean Decel (m/s^2)	Interpretation
Sedan	Congestion	2842	29.89	0.75	−0.77	Low speed with mild dynamics; frequent low-speed stops
	Moderate traffic	2529	34.93	1.24	−1.29	Higher speed with intense speed fluctuations
	Cruising	5937	35.42	0.92	−0.97	Higher speed with smooth, steady driving
Logistics vehicle	Congestion	4980	26.02	0.71	−0.72	Low speed with mild dynamics
	Moderate traffic	5460	35.08	1.2	−1.24	Higher speed with frequent acceleration/braking
	Cruising	10,129	35.06	0.88	−0.92	Higher speed with smooth flow
Bus	Congestion	3938	24.11	0.79	−0.8	Slowest; dense-stop urban operation
	Moderate traffic	1755	31.63	1.12	−1.15	Higher speed with frequent passenger stops
	Cruising	4363	27.8	0.96	−0.98	Moderate speed with relatively steady flow

Table A3. Driving Style Clusters.

Vehicle Type	Cluster	*n*	Mean Accel Pedal (%)	Mean Brake Activation Rate (%)	Interpretation
Sedan	Mild	4074	19.04	8.13	Low pedal demand overall; conservative driving
	Normal	2955	23.85	11.32	Moderate throttle with active braking; balanced style
	Aggressive	4279	28.83	8.09	High throttle with minimal braking; acceleration-dominant
Logistics vehicle	Mild	14,531	16.77	16.12	Low throttle with moderate braking; cautious driving
	Normal	4430	26.41	17.53	High throttle with active braking; variable driving
	Aggressive	1608	24.49	0.03	High throttle with virtually no braking; coasting-free
Bus	Mild	2292	41.4	0.07	Lower pedal demand relative to bus fleet; gentle operation
	Normal	857	57.4	40.1	High throttle with heavy braking; stop-intensive pattern
	Aggressive	6907	57.22	1.04	High throttle with no braking; sustained power output

Table A4. Estimated Average Treatment Effects (ATE) on ECR by Vehicle Type.

Treatment	Sedan	Logistics Vehicle	Bus
Trip distance	−0.00033	−0.00028	−0.00139
Trip duration	−0.000026	−0.000021	0.000178
Mean speed	−0.00159	−0.00155	−0.00828
Mean acceleration	0.018806	0.024687	0.122055

Table A4. *Cont.*

Treatment	Sedan	Logistics Vehicle	Bus
Max acceleration	−0.008	−0.00416	0.01838
SD of acceleration	0.008132	0.009952	0.027988
Mean deceleration	−0.01463	−0.02105	−0.09119
Max deceleration	0.008651	0.003957	−0.00106
SD of deceleration	0.005697	0.008264	0.007
Mean accelerator pedal	−0.00023	0.000682	0.002916
Max accelerator pedal	−0.00044	0.000256	0.002237
SD of accelerator pedal	0.000847	0.002252	0.003585
Mean brake pedal	−0.00156	0.000435	0.010018
Max brake pedal	−0.00096	0.000491	−0.015064
SD of brake pedal	−0.00103	−0.00212	−0.0117
Initial SOC	0.000007	−0.000059	−0.00034
Energy recovery ratio	−0.004059	−0.002408	−0.117345
Cumulative mileage	1.54×10^{-9}	-1.17×10^{-8}	-1.39×10^{-7}

References

1. Prussi, M.; Laveneziana, L.; Testa, L.; Chiaramonti, D. Comparing e-Fuels and Electrification for Decarbonization of Heavy-Duty Transports. *Energies* **2022**, *15*, 8075. [CrossRef]
2. Boudmen, K.; El Ghazi, A.; Eddaoudi, Z.; Aarab, Z.; Rahmani, M.D. Electric Vehicles, the Future of Transportation Powered by Machine Learning: A Brief Review. *Energy Inform.* **2024**, *7*, 80. [CrossRef]
3. Castillo-Calderon, J.; Larrode-Pellicer, E. Energy Consumption Prediction in Battery Electric Vehicles: A Systematic Literature Review. *Energies* **2026**, *19*, 371. [CrossRef]
4. Liu, K.; Wang, J.; Yamamoto, T.; Morikawa, T. Exploring the Interactive Effects of Ambient Temperature and Vehicle Auxiliary Loads on Electric Vehicle Energy Consumption. *Appl. Energy* **2018**, *227*, 324–331. [CrossRef]
5. De Cauwer, C.; Verbeke, W.; Coosemans, T.; Faid, S.; Van Mierlo, J. A Data-Driven Method for Energy Consumption Prediction and Energy-Efficient Routing of Electric Vehicles in Real-World Conditions. *Energies* **2017**, *10*, 608. [CrossRef]
6. Ben Abbes, S.; Rejeb, L.; Baati, L. Route planning for electric vehicles. *IET Intell. Transp. Syst.* **2022**, *16*, 875–889. [CrossRef]
7. Chen, Y.; Wu, G.; Sun, R.; Dubey, A.; Laszka, A.; Pugliese, P. A Review and Outlook of Energy Consumption Estimation Models for Electric Vehicles. *Transp. Res. Part C Emerg. Technol.* **2020**, *120*, 102838.
8. Perugu, H.; Collier, S.; Tan, Y.; Yoon, S.; Herner, J. Characterization of Battery Electric Transit Bus Energy Consumption by Temporal and Speed Variation. *Energy* **2023**, *263*, 125914. [CrossRef]
9. Lei, X.; Lu, H.; Fan, P.; Liu, R.; Li, S.; Wu, Y.; Song, G. Comparative Sensitivity Analyses of Energy Consumption in Response to Average Speed Between Electric Vehicles and Conventional Vehicles: Case Study in Beijing, China. *Energies* **2025**, *18*, 2268. [CrossRef]
10. Fiori, C.; Ahn, K.; Rakha, H.A. Power-Based Electric Vehicle Energy Consumption Model: Model Development and Validation. *Appl. Energy* **2016**, *168*, 257–268. [CrossRef]
11. Wu, X.; Freese, D.; Cabrera, A.; Kitch, W.A. Electric Vehicles' Energy Consumption Measurement and Estimation. *Transp. Res. Part D Transp. Environ.* **2015**, *34*, 52–67. [CrossRef]
12. Al-Wreikat, Y.; Serrano, C.; Sodré, J.R. Effects of Ambient Temperature and Trip Characteristics on the Energy Consumption of an Electric Vehicle. *Energy* **2022**, *238*, 122028. [CrossRef]
13. De Cauwer, C.; Van Mierlo, J.; Coosemans, T. Energy Consumption Prediction for Electric Vehicles Based on Real-World Data. *Energies* **2015**, *8*, 8573–8593. [CrossRef]
14. Bingham, C.; Walsh, C.; Carroll, S. Impact of Driving Characteristics on Electric Vehicle Energy Consumption and Range. *IET Intell. Transp. Syst.* **2012**, *6*, 29–35. [CrossRef]
15. Qi, X.; Wu, G.; Boriboonsomsin, K.; Barth, M.J. Data-Driven Decomposition Analysis and Estimation of Link-Level Electric Vehicle Energy Consumption under Real-World Traffic Conditions. *Transp. Res. Part D Transp. Environ.* **2018**, *64*, 36–52. [CrossRef]
16. Barik, S.; Saravanan, B. Recent Developments and Challenges in State-of-Charge Estimation Techniques for Electric Vehicle Batteries: A Review. *J. Energy Storage* **2024**, *100*, 113623. [CrossRef]
17. Challoob, A.F.; Rahmat, N.A.B.; Ramachandaramurthy, V.K.; Humaidi, A.J. Energy and Battery Management Systems for Electrical Vehicles: A Comprehensive Review and Recommendations. *Energy Explor. Exploit.* **2024**, *42*, 341–372. [CrossRef]
18. Yang, C.; Sun, T.; Wang, W.; Li, Y.; Zhang, Y.; Zha, M. Regenerative Braking System Development and Perspectives for Electric Vehicles: An Overview. *Renew. Sustain. Energy Rev.* **2024**, *198*, 114389. [CrossRef]

19. Kivanc, O.C.; Ustun, O. Investigation of Regenerative Braking Performance of Brushless Direct Current Machine Drive System. *Appl. Sci.* **2021**, *11*, 1029. [CrossRef]
20. Zhao, W.; Wu, G.; Wang, C.; Yu, L.; Li, Y. Energy Transfer and Utilization Efficiency of Regenerative Braking with Hybrid Energy Storage System. *J. Power Sources* **2019**, *427*, 174–183. [CrossRef]
21. Huang, H.; Gao, K.; Wang, Y.; Najafi, A.; Zhang, Z.; He, H. Sequence-Aware Energy Consumption Prediction for Electric Vehicles Using Pre-Trip Realistically Accessible Data. *Appl. Energy* **2025**, *401*, 126673. [CrossRef]
22. Al-Wreikat, Y.; Serrano, C.; Sodré, J.R. Driving Behaviour and Trip Condition Effects on the Energy Consumption of an Electric Vehicle under Real-World Driving. *Appl. Energy* **2021**, *297*, 117096. [CrossRef]
23. Zhang, J.; Wang, Z.; Liu, P.; Zhang, Z. Energy Consumption Analysis and Prediction of Electric Vehicles Based on Real-World Driving Data. *Appl. Energy* **2020**, *275*, 115408. [CrossRef]
24. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017; pp. 3146–3154.
25. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774.
26. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
27. Chernozhukov, V.; Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W.; Robins, J. Double/Debiased Machine Learning for Treatment and Structural Parameters. *Econom. J.* **2018**, *21*, C1–C68. [CrossRef]
28. Athey, S.; Tibshirani, J.; Wager, S. Generalized Random Forests. *Ann. Stat.* **2019**, *47*, 1148–1178. [CrossRef]
29. Huang, H.; Li, B.; Wang, Y.; Zhang, Z.; He, H. Analysis of Factors Influencing Energy Consumption of Electric Vehicles: Statistical, Predictive, and Causal Perspectives. *Appl. Energy* **2024**, *375*, 124110. [CrossRef]
30. Gallet, M.; Massier, T.; Hamacher, T. Estimation of the Energy Demand of Electric Buses Based on Real-World Data for Large-Scale Public Transport Networks. *Appl. Energy* **2018**, *230*, 344–356. [CrossRef]
31. Hjelkrem, O.A.; Lervag, K.Y.; Babri, S.; Lu, C.; Sodersten, C.-J. A Battery Electric Bus Energy Consumption Model for Strategic Purposes: Validation of a Proposed Model Structure with Data from Bus Fleets in China and Norway. *Transp. Res. Part D Transp. Environ.* **2021**, *94*, 102804. [CrossRef]
32. Dabčević, Z.; Škugor, B.; Cvok, I.; Deur, J. A Trip-Based Data-Driven Model for Predicting Battery Energy Consumption of Electric City Buses. *Energies* **2024**, *17*, 911. [CrossRef]
33. Diab, I.; Saffirio, A.; Chandra-Mouli, G.R.; Bauer, P. A simple method for sizing and estimating the performance of PV systems in trolleybus grids. *J. Clean. Prod.* **2023**, *384*, 135623. [CrossRef]
34. Gillstrom, H.; Jobrant, M.; Sallnas, U. Towards Building an Understanding of Electrification of Logistics Systems: A Literature Review and a Research Agenda. *Clean. Logist. Supply Chain.* **2024**, *10*, 100134. [CrossRef]
35. *GB/T 32960.3-2016*; Technical Specifications of Remote Service and Management System for Electric Vehicles—Part 3: Communication Protocol and Data Format. Standards Press of China: Beijing, China, 2016. Available online: https://openstd.samr.gov.cn/bzgk/std/newGbInfo?hcno=674DE45C0AD3DE2CD75B9C4CD8ED57C1 (accessed on 8 October 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.