CHALMERS
UNIVERSITY OF TECHNOLOGY

Trip energy consumption estimation for electric buses

Communications in Transportation Research 2 (2022) 100069

Contents lists available at ScienceDirect

Communications in Transportation Research

journal homepage: www.journals.elsevier.com/communications-in-transportation-research

Full Length Article

Trip energy consumption estimation for electric buses

Jinhua Ji [a], Yiming Bie [a,*], Ziling Zeng [b], Linhong Wang [a,**]

[a] School of Transportation, Jilin University, 130022, Changchun, China
[b] Department of Architecture and Civil Engineering, Chalmers University of Technology, 41296, Gothenburg, Sweden

ARTICLE INFO

Keywords:
Electric bus
Trip energy consumption
Regression model
Operational data
Cold region

ABSTRACT

This study aims to develop a trip energy consumption (TEC) estimation model for the electric bus (EB) fleet planning, operation, and life-cycle assessment. Leveraging the vast variations of temperature in Jilin Province, China, real-world data of 31 EBs operating in 14 months were collected with temperatures fluctuating from −27.0 to 35.0 °C. TEC of an EB was divided into two parts, which are the energy required by the traction and battery thermal management system, and the energy required by the air conditioner (AC) system operation, respectively. The former was regressed by a logarithmic linear model with ambient temperature, curb weight, travel distance, and trip travel time as contributing factors. The optimum working temperature and regression parameters were obtained by combining Fibonacci and Weighted Least Square. The latter was estimated by the operation time of the AC system in cooling mode or heating mode. Model evaluation and sensitivity analysis were conducted. The results show that: (i) the mean absolute percentage error (*MAPE*) of the proposed model is 12.108%; (ii) the estimation accuracy of the model has a probability of 99.7814% meeting the requirements of EB fleet scheduling; (iii) the *MAPE* has a 1.746% reduction if considering passengers' boarding and alighting.

1. Introduction

1.1. Background

To combat climate change, many countries have claimed to have CO_2 emissions peak before 2030 and achieve carbon neutrality by 2050 or 2060 (Huang and Zhai, 2021). Electric buses (EBs) have the advantages of low emissions, noise, and operational costs. In recent years, the rapid deployment of EBs worldwide is of great significance to improve the traffic environment and is deemed to be a must to achieve the goal of carbon neutrality. In a study published in 2019, the Union of Concerned Scientists reported that while the 40-foot diesel bus emits 2,680 g of CO_2 per mile (g/mi), an electric bus charged on the average U.S. energy mix emits 1,078 g/mi, roughly a 60% reduction (Massoli, 2020). As the leading country in urban transit electrification, China has the world's largest EB fleet. From 2016 to 2020, the number of EBs in China has increased from 94,000 to 379,000. In the US, new city buses will only be electric ones starting from 2025 in San Francisco (Marcacci, 2018). In Europe, EBs have been running at Gothenburg, Sweden for six years. By the end of 2020, about 210 EBs were in operation, accounting for 35% of the city bus fleet (Sustainable, 2020).

Despite the rapid expansion, the deployment of EBs still faces challenges of short driving range and long charging time compared to fuel buses. Accurately estimating the trip energy consumption (TEC) for the EBs is essential for multiple planning and operation tasks. Firstly, in EB scheduling and charge scheduling, mathematical programming models are often developed to generate optimal plans. Minimizing the total energy consumption of EBs is one of the most common optimization objectives (Leou and Hung, 2017; Bie et al., 2021; He et al., 2022; Liu et al., 2022). In terms of the constraints, the range of battery state of charge (SOC) is limited to prevent service interruptions due to battery energy exhaustion (Li et al., 2019; Perumal et al., 2021). Thus, the TEC estimation is the basis for calculating the objectives and constraints of mathematical programming models. Secondly, deploying the optimal EB type is one main operation task faced by the bus company. Different EB types have different battery capacities, passenger capacities, and curb weights, resulting in different vehicle scheduling plans, energy consumption, and charging demands. The TEC estimation can investigate the impacts of different EB types on daily operation costs and help the bus company select the optimal EB type (Kunith et al., 2017; Rogge et al., 2018; Yao et al., 2020; Zhang et al., 2021b; Qu et al., 2022). Thirdly, the battery life is affected by the depth of discharge and the number of charge

* Corresponding author.
** Corresponding author.
E-mail addresses: jinhua_ji@126.com (J. Ji), yimingbie@126.com (Y. Bie), ziling@chalmers.se (Z. Zeng), wanghonglin0520@126.com (L. Wang).

https://doi.org/10.1016/j.commtr.2022.100069
Received 25 April 2022; Received in revised form 30 May 2022; Accepted 30 May 2022
Available online 22 June 2022

cycles. Given the daily EB scheduling plan and TEC estimation model, the operator can calculate the depth of discharge and the number of charge cycles of each bus. Thus, the battery life can be estimated in advance (Lajunen, 2018; Ritari et al., 2020; Zhang et al., 2021a).

The TEC of an EB in operation is affected by multiple factors such as ambient temperature, travel distance, trip travel time, curb weight, and state of the Air Conditioner (AC). Black-box methods (i.e., artificial neural network, random forest, support vector machine) are usually employed in the estimation and they perform well in exploring relationships between input and output variables. These methods obtain estimators with high accuracy but without closed-formed. In fact, there are random fluctuations in some contributing factors, such as trip travel time, passenger demand, etc. The random fluctuations not only affect the TEC estimation precision, but also the EB operation plan. It is impossible to develop a robust EB operation optimization model based on the TEC produced by black-box methods. In such a condition, it is not suitable to estimate the TEC based on the black-box methods. Besides, the black-box methods usually require a large amount of EB operational data to improve the estimation accuracy and do not perform well with a small amount of data. Therefore, this paper utilizes the regression method to establish a relationship model between TEC and contributing factors at the aggregated level. Then a sensitivity analysis is performed for each factor on the energy consumption. The proposed estimation model provides a prerequisite for EB fleet planning, operation, and life-cycle benefit assessment.

1.2. Literature review

In previous studies, most research efforts in energy consumption concentrated on electric vehicles (Wu et al., 2015; Fiori et al., 2016; Liu et al., 2017; Qi et al., 2018). However, EBs have a significant difference from light-duty electric vehicles. An EB running on a fixed route at a relatively low speed and stopping frequently at bus stations make it not applicable to use energy consumption estimation methods of EVs in the case of EBs. Current studies on EB energy consumption can be roughly categorized into four groups in terms of research approaches, which are the empirical method, physics-based method, deep learning method, and regression method.

(i) Empirical method

An empirical method is to estimate the energy consumption of an EB using an average electricity consumption per hundred kilometers so as to reduce the complexity of bus operation optimization problems. The average electricity consumption per hundred kilometers applied in previous studies includes 124–248 kWh/100 km (Gao et al., 2017), 120–290 kWh/100 km for 10m length EBs (He et al., 2018), 120 kWh/100 km (Li et al., 2019), 150 kWh/100 km in Stockholm, Sweden (Xylia et al., 2017), 180–300 kWh/100 km for plug-in hybrid buses and 320–500 kWh/100 km for hybrid buses (Stempien and Chan, 2017). Here the average electricity consumption per hundred kilometers is to divide the rated capacity by the maximum driving range provided by vehicle manufacturers or to divide the total energy consumption of an EB accumulated from its entering into service by the accumulated operation distance.

Because stochastic volatility in energy consumption induced by differences in trip travel times, average vehicle speed, and ambient temperature among various trips is not included, these empirical approaches for TEC calculations are poor in terms of accuracy.

(ii) Physics-based method

A physics-based method is to calculate the instantaneous energy consumption utilizing the dynamics model or the vehicle-specific power model with input data including vehicular instantaneous velocity, acceleration, weight, and road gradient (Hjelkrem et al., 2021). Galleta et al. (2018) presented a longitudinal dynamics model to calculate the energy demand for EBs, which took the details of the operational characteristics of the transportation network into account and required no high-resolution velocity profiles. Kivekäs et al. (2018) developed a driving cycle synthetization method that could generate a large number of varying cycles and passenger numbers based on only a handful of measured cycles, then the impact on energy consumption variations of an EB was analyzed using a suburban route in Espoo, Finland. Łebkowski, 2019 established an energy consumption model for EB based on the vehicle dynamics model and verified it with the actual bus line data in Poland. Al-Ogaili et al. (2020) developed a longitudinal dynamic model with a spatial version of a digital elevation model to determine the energy demand of a large-scale bus network. El-Taweel et al. (2021) proposed a generic model to calculate the EB energy consumption and generated a set of speed profiles using the basic information of the bus trip: trip time, trip length, and distances between successive bus stops, to reflect various traffic conditions and speed behaviors of real-world situations. Beckers et al. (2020) developed a nonlinear steady-state cornering model to establish the additional energy losses during cornering. Fiori et al. (2021) developed a microscopic energy consumption model of electric buses, which estimated the instantaneous power required for traction and the average power required by the auxiliary systems, leveraging on a vast data collection campaign of EB daily operations.

Compared to the empirical method, the physics-based method incorporates instantaneous EB movement parameters and leads to an improvement in estimation accuracy. However, at the stage of EB scheduling plan optimization, it is impossible to accurately predict the speed of each bus per second. In addition, the impacts of ambient temperature on the battery heat loss and on-board auxiliary system are often ignored, which leads to the degradation of prediction performance of physics-based models.

(iii) Deep learning method

Gao et al. (2018) established a power consumption prediction model with multiple factors and investigated the related situation of EBs in Baoding under the temperature range of −9 °C–25 °C. The wavelet neural network was used to train the power consumption factors together with power consumption data in the feature library, which was established by the grey relational analysis method. Pamuła and Pamuła (2020) obtained an energy consumption prediction model between consecutive stations with deep learning networks, leveraging real-world data of EB fleets in the town of Jaworzno in Poland under the temperature of −5 °C–23 °C. Chen et al. (2021) developed the long short-term memory and artificial neural network models to estimate instantaneous energy consumption based on continuous monitoring data of electric buses in Chattanooga. Li et al. (2021) proposed a trip-based electricity consumption prediction model using the random forest algorithm. With variables of dynamic traffic conditions, route characteristics, and environmental conditions as input, the impacts of these three categories of variables on EB electricity consumption were explored.

The deep learning-based method can perform TEC estimation well and achieve promising performance in accuracy. However, the relationship between TEC and contributing factors is not presented in explicit functions. It is impossible to quantify the impacts of random fluctuations in contributing factors and develop a robust EB operation optimization model based on the TEC produced by the deep learning method.

(iv) Regression method

Wang et al. (2017a) developed an estimation model for residual energy of pure EBs using the time-series algorithm, and the model performance was evaluated and analyzed using real data. Abdelaty et al. (2021) developed a multiple linear regression model to predict the EB energy consumption considering the impacts of road grade, the initial SOC, road condition, passenger loading, driver aggressiveness, average speed,

HVAC, and stop density. Abdelaty and Mohamed (2021) developed a multivariate multiple linear regression model for EB energy consumption considering the impacts of vehicular, operational, topological, and external parameters. However, the ambient temperature was not taken into the above estimation models. Based on our field investigation, ambient temperature is one important influencing factor to the EB energy consumption. In wintertime, the electricity consumption per kilometer of an EB sharply increases and subsequently the driving range fiercely shrinks.

1.3. Objective and contributions

The objective of this study is to develop an EB energy consumption estimation model from the perspective of EB fleet planning, operation, and evaluation. In such a condition, transit operators are more concerned with the total energy consumption of each trip, rather than the instantaneous energy consumption. Hence, the trip level estimation is selected in this study. In addition, the proposed model should have an explicit relationship function between TEC and contributing factors, which is helpful in developing mathematical programming models for robust EB scheduling and charge scheduling.

The contributions of this study are threefold: (i) the impacts of travel distance, curb weight, travel time, ambient temperature, and the duration of the AC system in the energized state on TEC are analyzed, a multivariable nonlinear regression model is used to estimate the TEC accordingly, and sensitivity analysis is performed on trip travel time, curb weight, and ambient temperature; (ii) in the real-world data used in this study, the minimum and maximum temperatures are −27.0 and 35.0 °C, respectively, showing a huge divergence of 62.0 °C, which expands the application range of the proposed estimation model, especially under low-temperature conditions; (iii) an evaluation method for model adaptability is proposed, to determine whether the estimation accuracy of the proposed model meets the requirement of bus scheduling.

The rest of this paper is organized as follows. Section 2 introduces and depicts the collected real-world operational data of EBs in a cold region. Section 3 formulates the trip-based energy consumption model of EBs based on the multivariable regression method, incorporating the energy consumptions respectively for the vehicle motion and AC system requirements. Section 4 evaluates the model from three aspects, including estimation accuracy, sensitivity analysis, and model adaptability. Finally, some concluding remarks and possible future works are given in Section 5.

2. Data collection and description

The real-world operational data of EB fleets are collected in Meihekou City, Jilin Province, China. Meihekou is located at 125.6°E, 42.5°N. The annual average temperature is 5.5 °C. The lowest and highest recorded temperatures in history are −33 and 36 °C, respectively. We gathered the operational data of 31 EBs from 6 January 2020 to 1 March 2021, involving a total of 4,360 trips through the EB monitoring platform. Note here that a bus trip is defined as a bus running from the start station to the terminal station. The gathered information of each trip includes the date, license plate number, traveling direction, trip distance, curb weight, trip departure time, trip arrival time, TEC, SOC at the departure time, trip travel time, average speed, time duration of AC system in the on-state and the route number. In addition, we also collected the data for average ambient temperature per hour from the meteorological department.

These 31 EBs are deployed on three routes, including Route 108, Route 103, and Route 106 (see Fig. 1 for layout). The EBs have three bus models, denoted respectively by Model A, Model B, and Model C. All EBs are powered by the $LiFePO_4$ battery purchased in June 2018 with an 8-year guarantee period. Basic information on the above three routes and three bus models are illustrated in Tables 1 and 2.

The analysis of the collected data is shown in Fig. 2. Trip travel time ranges from 14 to 42 min, and its average value is 29 min. The average TEC of all gathered trips is 5.8 kWh. Considering the route operating hours, the lowest temperature of our collected trips was recorded as −27.0 °C, and the highest temperature was 35.0 °C. The variations of daily highest temperature, daily lowest temperature, and daily average temperature are plotted in Fig. 3 based on historical weather information. The minimum operational time of the AC system is 0 min, while the maximum value is the trip travel time. Fig. 4 shows a frequency histogram that displays the number of trips under different AC operation modes in different temperature ranges (5 °C intervals).

3. Methodology

In this section, we analyze how each factor affects the TEC of an EB. The TEC estimation model is further proposed based on these factors.

3.1. Selection of contributing factors

With fixed operation routes, low operational speed, and repeated stop-and-go behaviors, multiple factors have impacts on the TEC (denoted by w) of an EB.

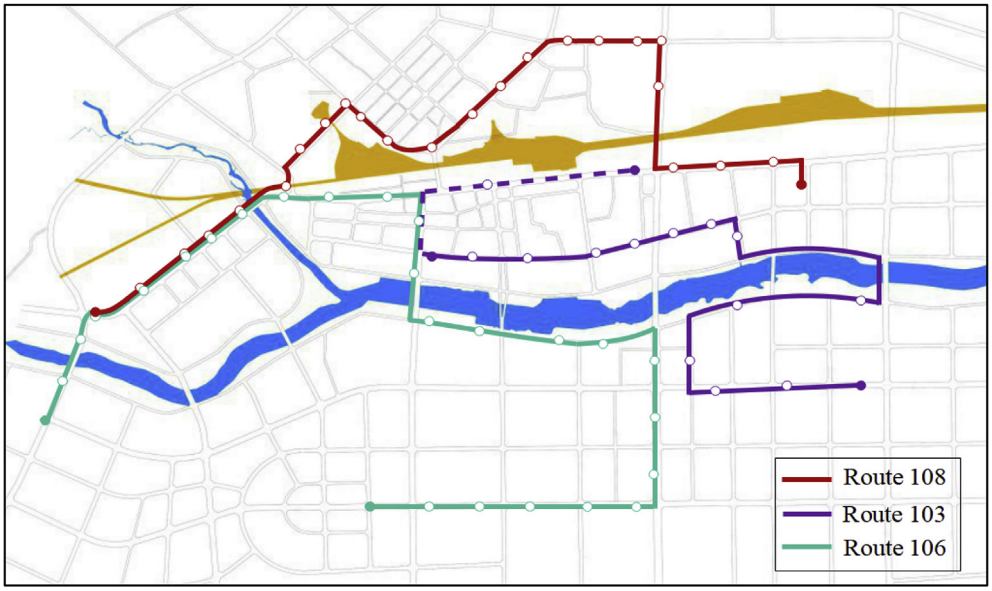


Fig. 1. Layout of three bus routes in Meihekou city.

Table 1
Basic information of the three bus routes.

Route	Distance (km)	No. of stations	No. of buses	Dispatching headway (min)	Vehicle model	No. of samples
108	7.9	24	9	5–15	A	2,032
103a/103b	7.5/10	14/16	17	6	B/C	1,823
106	12	26	5	7–10	C	505

Table 2
Parameters of the three EB models.

Model	Battery rated capacity (kWh)	Vehicle size (mm^3)	Curb weight (kg)	Load capacity (pax)	No. of samples
A	162.3	8500×2500×3215	8200	60	2032
B	199.37	10500 × 2500 × 3200	11350	76	1409
C	202.93	10500 × 2500 × 3215	10550	76	919

(i) Travel distance, trip travel time and average speed

Travel time and distance are normally proportional to TEC, although average speed is inversely proportional. When the travel distance is fixed, trip travel time is obtained by dividing the travel distance by the average speed. Travel time or average speed can also be used to indicate traffic conditions. Specifically, the service level of a bus line decreases during peak hours, mainly due to the increase in travel time or the decrease in average speed, while the level of public transport services during off-peak hours increases due to short travel time or large average speed. Let L, t and $\bar{v}$ denote travel distance, travel time and average speed of each trip, respectively.

(ii) Bus models

Different bus models have different energy consumption rates. A vehicle with a larger battery capacity, heavier curb weight, and higher traction force lead to higher energy consumption. Here we use the curb weight M of an EB to represent model difference and investigate its impact on TEC.

(iii) Ambient temperature

Low temperature causes an increase in the internal resistance of a battery, leading to higher heat evolution of the battery and increased electricity consumption rate. Furthermore, the battery thermal management system would spend additional energy to keep the battery temperature within a proper range around the optimum operating temperature ($\overline{T}_g^*$) in both high and low ambient temperatures. In general, the greater the difference between ambient temperature and $\overline{T}_g^*$, the more energy is required (Wang et al., 2017b; Liu et al., 2018). In this study, the average ambient temperature of each trip is denoted by $\overline{T}$.

(iv) Air conditioner system

During summer or winter, the increment in TEC is also resulted from the operation of the AC system. We use t_a to represent the operation time of the AC system of each trip.

(v) Other factors

In addition to the above factors, there are still other factors that may affect TEC, such as driving behavior, number of stations and intersections, road grade, battery SOC at the departure time, etc. These factors are not analyzed in detail due to their minor impacts on TEC (please refer to Section 5 for details).

The selection of independent variables is the basis of regression model formulation. We have analyzed the variables influencing w, which are all potential independent variables. Among these variables, however, some of them might have no significant impact, or have low quality, or have dependent correlations with others. These deficiencies increase the computational burden and further jeopardize the stability of regression models, making it impossible to directly apply all potential independent variables.

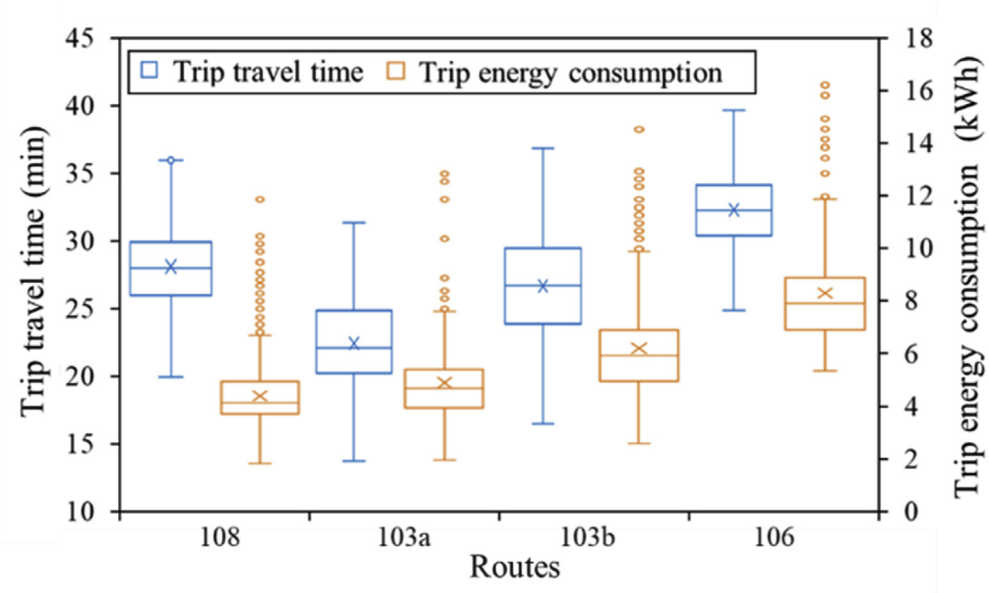


Fig. 2. Distribution of trip travel time and TEC.

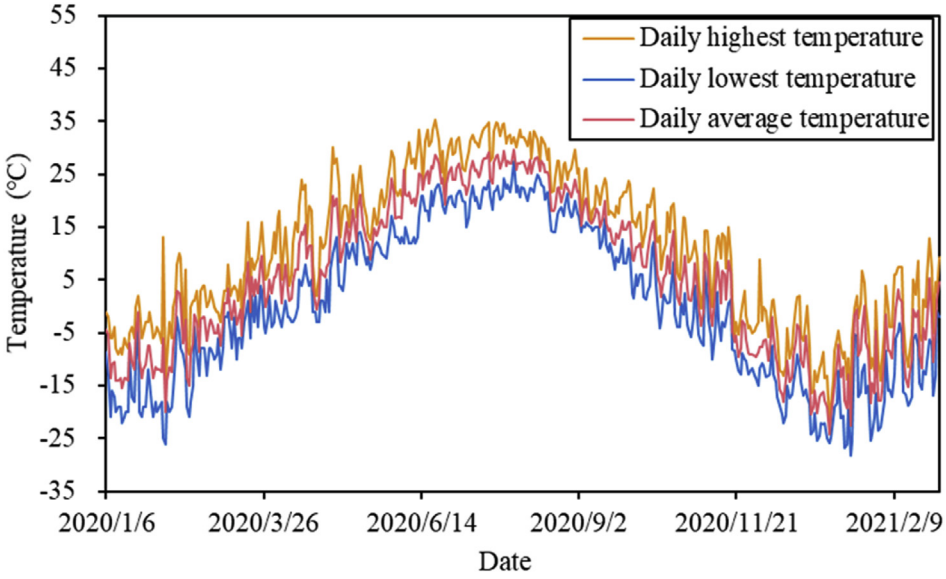


Fig. 3. Variations of daily temperature during the data collection period.

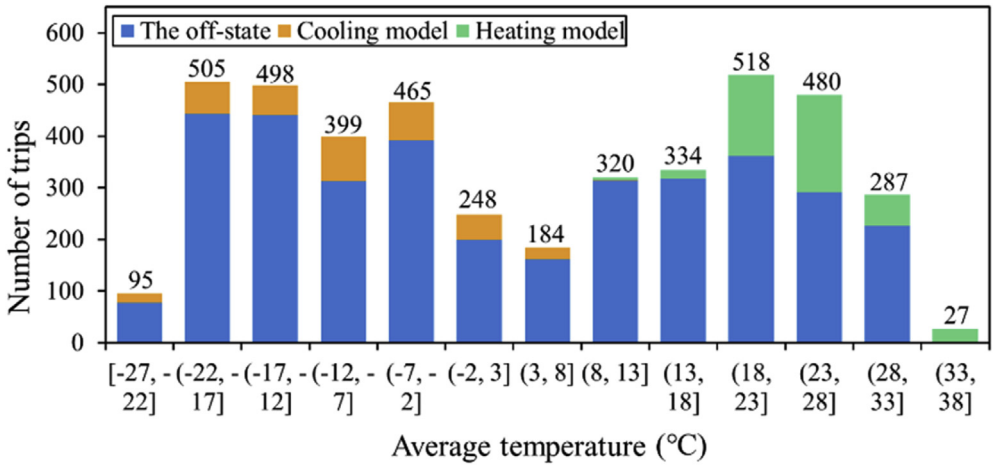


Fig. 4. Number of trips under different AC operation modes in different temperature ranges.

When the absolute value of the correlation coefficient is greater than 0.5, the contributing factor is generally considered to have a strong correlation with TEC. When the value is between 0.3 and 0.5, the correlation is medium. When the value is between 0.1 and 0.3, the correlation is weak. When the value is less than 0.1, no correlation exists. We first conduct the correlation analysis on each factor using Spearman's rank correlation coefficient. Table 3 displays Spearman's rank correlation analysis among TEC and contributing factors. The Spearman's rank

Table 3
Spearman's rank correlation analysis among TEC and contributing factors.

Factors	w (kWh)	L (km)	M (kg)	t (min)	$\bar{v}$ (km/h)	$\overline{T}$ (°C)	t_a (min)
w (kWh)	1.000						
	—						
L (km)	0.596	1.000					
	0.000	—					
M (kg)	0.418	0.130	1.000				
	0.000	0.000	—				
t (min)	0.423	0.560	−0.087	1.000			
	0.000	0.000	0.000	—			
$\bar{v}$ (km/h)	0.359	0.350	0.440	−0.337	1.000		
	0.000	0.000	0.000	0.000	—		
$\overline{T}$ (°C)	−0.524	−0.367	−0.043	−0.271	−0.226	1.000	
	0.000	0.000	0.002	0.000	0.000	—	
t_a (min)	0.289	0.066	−0.004	0.021	0.037	0.136	1.000
	0.000	0.001	0.392	0.077	0.007	0.000	—

Note: In each cell, the upper value represents the Spearman's rank correlation coefficient, and the lower value represents the P-value.

correlation coefficient between L and w is 0.596 while that between t and w is 0.423, which indicate the strong and medium correlations. The correlation coefficients between w and L, t and $\bar{v}$ are all greater than 0.3, and the correlation coefficient between w and $\bar{v}$ is 0.359, which is smaller than 0.596 and 0.423. Therefore, we exclude $\bar{v}$ and keep L and t for further analysis. L, M, t, $\overline{T}$ and t_a are selected as the explanatory variables and w is selected as the explained variable.

Then the partial correlation analysis is applied to eliminate effects from other variables so as to investigate the pure impact of each variable on w. Results are shown in Table 4. The partial correlation coefficient between t_a and w is 0.541, representing a strong linear correlation between them. In addition, medium or weak linear correlations are observed from the following pairs: L and w, M and w, t and w, $\overline{T}$ and w, from their partial correlation coefficients as 0.319, 0.346, 0.247 and −0.451. Thus, there may be nonlinear correlations in these pairs.

3.2. Model structure

TEC of an EB mainly contains three components including traction energy, the energy required by the battery thermal management system (BTMS), and the energy required by the AC system operation. The vehicle traction and BTMS require the battery energy during the EB running, and their consumptions are affected by some similar factors, such as ambient temperature, trip travel distance, travel time, etc. Unlike the above two, the AC system is controlled by the driver and not always in working condition during the vehicle operation. To actually capture the consumption, the traction energy and the energy required by the BTMS are combined in the regression and we try to establish a TEC estimated regression model in the following form:

$$\widehat{w}_i = \widehat{w}_{1i} + \widehat{w}_{2i} \tag{1}$$

where i is the serial number of trips and $i = 1, 2, \ldots, N$, N is the sample size; $\widehat{w}_i$ is the estimated TEC of an EB, kWh; $\widehat{w}_{1i}$ is the estimated energy consumption required by traction and BTMS (T-BTMS), kWh; $\widehat{w}_{2i}$ is the estimated energy consumption for the AC system operation, kWh. If the AC system remains in the off state during the whole trip, then $\widehat{w}_{2i} = 0$ and $\widehat{w}_i = \widehat{w}_{1i}$.

As described in Section 3.1, there might exist a nonlinear relationship between $\widehat{w}_{1i}$ and contributing factors. We plotted the scatter diagrams among TEC and travel distance, curb weight, and travel time after logarithm, respectively. Linear relationships among the independent variable and dependent variables were found, which indicated that the double logarithm model was suitable for estimating the TEC. In previous studies (Wang et al., 2017b; Liu et al., 2018), energy efficiency and ambient temperature approximately presented a "U" shape with inconsonant lowest values. In order to increase the stability of the regression model, reduce the impacts of multicollinearity, and simplify the following curve fitting, we build a logarithmic linear model, as expressed by Eq. (2):

$$\ln\widehat{w}_{1i} = \widehat{\alpha}_0 + \widehat{\alpha}_1 \ln L_i + \widehat{\alpha}_2 \ln M_i + \widehat{\alpha}_3 \ln t_i + \widehat{\alpha}_4 \left|\overline{T}_i - \overline{T}_g^*\right| \tag{2}$$

In Eq. (2), L_i is the travel distance of trip i, km. M_i is the curb weight of the EB serving trip i, kg. t_i is the actual travel time of trip i, min. $\overline{T}_i$ is the average ambient temperature during trip i, °C. $\overline{T}_g^*$ is the optimum working temperature of EB, °C. $\widehat{\alpha}_0$, $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ are estimated regression parameters. The number of samples is represented by $i = 1, 2, \ldots, N_1$, where N_1 is the size of samples with $t_a = 0$ and here $N_1 = 3{,}536$.

Table 4
Partial correlation analysis among TEC and contributing factors.

w (kWh)	L (km)	M (kg)	t (min)	$\overline{T}$ (°C)	t_a (min)
Partial correlation	0.319	0.346	0.247	−0.451	0.617
Significance	0.000	0.000	0.000	0.000	0.000

A dummy variable D^1 is introduced to represent the state of the AC system, $D^1 = 1$ when the AC system is in cooling mode while $D^1 = 0$ when the AC system is in heating mode. Accordingly, the energy consumption for the AC system operation is modeled as

$$\widehat{w}_{2i} = \begin{cases} \widehat{\zeta}_1 t_{ai} & D^1 = 1 \\ \widehat{\zeta}_2 t_{ai} & D^1 = 0 \end{cases} \tag{3}$$

where t_{ai} is the recorded operation time of the AC system during trip i, min; $\widehat{\zeta}_1$ and $\widehat{\zeta}_2$ are estimated values of regression parameters; $i = 1, 2, \ldots,455$ when $D^1 = 1$ and $i = 1, 2, \ldots,369$ when $D^1 = 0$.

3.3. Regression model calibration

Determination of $\overline{T}_g^*$ for an EB has a significant impact on the goodness-of-fit of $\widehat{w}_{1i}$ estimation. Most previous studies took 20 or 25 °C as the optimum working temperature of the $LiFePO_4$ battery based on experience (Wang et al., 2011; Hu et al., 2020; Zhang et al., 2020). In our research, with actual EB operational data of trips where $t_a = 0$, estimations are conducted for the optimum working temperature and regression parameters of the $\widehat{w}_{1i}$ estimation model by combining Fibonacci and ordinary least squares (OLS). The detailed estimation procedure is illustrated as follows:

Step 1. Observe the TEC distribution at different temperatures to determine the temperature search range $[c, d]$. The objective function at the *j-th* working temperature denoted by $F(\overline{T}_{gj})$ and its loss function denoted by $\psi(\overline{T}_{gj})$, are calculated as follows:

$$F(\overline{T}_{gj}) = \frac{1}{N_1}\sum_{i=1}^{N_1}\left|\frac{\widehat{w}_{1i} - w_{1i}}{w_{1i}}\right| \tag{4}$$

$$\psi(\overline{T}_{gj}) = \sum_{i=1}^{N_1}(\widehat{u}_{1i})^2 = \sum_{i=1}^{N_1}\left[\ln w_{1i} - \left(\widehat{\alpha}_{0j} + \widehat{\alpha}_{1j}\ln L_i + \widehat{\alpha}_{2j}\ln M_i + \widehat{\alpha}_{3j}\ln t_i + \widehat{\alpha}_{4j}\left|\overline{T}_i - \overline{T}_{gj}\right|\right)\right]^2 \tag{5}$$

where w_{1i} is the observed T-BTMS energy consumption, kWh; $\widehat{u}_{1i}$ is the residual error; $\widehat{\alpha}_{0j}$, $\widehat{\alpha}_{1j}$, $\widehat{\alpha}_{2j}$, $\widehat{\alpha}_{3j}$, $\widehat{\alpha}_{4j}$ are coefficients for estimation.

Step 2. Select two initial points $\overline{T}_{g1} \in [c, d]$ and $\overline{T}_{g2} \in [c, d]$ which are in symmetric positions in the interval so that $\overline{T}_{g2} = c + d - \overline{T}_{g1}$ and $\overline{T}_{g1} < \overline{T}_{g2}$.

Step 3. Based on OLS, the necessary conditions of minimizing the loss function are $\frac{\partial\psi}{\partial\widehat{\alpha}_{0j}} = 0$, $\frac{\partial\psi}{\partial\widehat{\alpha}_{1j}} = 0$, $\frac{\partial\psi}{\partial\widehat{\alpha}_{2j}} = 0$, $\frac{\partial\psi}{\partial\widehat{\alpha}_{3j}} = 0$ and $\frac{\partial\psi}{\partial\widehat{\alpha}_{4j}} = 0$. Two sets of linear equations, including all these conditions derived by substituting $\overline{T}_{gj}$ with $\overline{T}_{g1}$ and $\overline{T}_{g2}$ are solved respectively to obtain the estimated coefficients $\widehat{\alpha}_{01} \sim \widehat{\alpha}_{41}$ and $\widehat{\alpha}_{02} \sim \widehat{\alpha}_{42}$, as well as the functions $F(\overline{T}_{g1})$ and $F(\overline{T}_{g2})$.

Step 4. Compare values of objective functions $F(\overline{T}_{g1})$ and $F(\overline{T}_{g2})$. If $F(\overline{T}_{g1}) \le F(\overline{T}_{g2})$, $\overline{T}_{g2}$ is identified as a bad point. The interval outside the bad point is abandoned, and $[c, \overline{T}_{g2}]$ is kept, which means that $\overline{T}_g^* \in [c, \overline{T}_{g2}]$. Otherwise, $[\overline{T}_{g1}, d]$ is preserved and $\overline{T}_g^* \in [\overline{T}_{g1}, d]$.

Step 5. In the narrowed interval, take a point $\overline{T}_{g3}$ which is in a symmetric position to the kept point. The estimated regression parameters $\widehat{\alpha}_{03} \sim \widehat{\alpha}_{43}$ and the objective function $F(\overline{T}_{g3})$ are calculated by minimizing the loss function $\psi(\overline{T}_{g3})$ using OLS. The objective function values at the new point and kept point are compared. The interval outside the bad point is cut, and the search interval is narrowed further.

Step 6. Repeat the above steps until the deviation between the two

holding points is not greater than the preset allowable error ε. Derive $\overline{T}_g^*$ and optimum estimated regression parameters $\widehat{\alpha}_{0j}^* \sim \widehat{\alpha}_{4j}^*$, corresponding to the minimum loss function value.

For our research scenario, the temperature search range [15 °C, 30 °C], $N_1 = 3{,}536$, and permissible error $\varepsilon = 0.1$ is imported to the above procedure. The optimum working temperature of EBs $\overline{T}_g^* = 23.3$ °C is exported, and the estimation results of regression parameters of the logarithmic linear model, which are demonstrated in Table 5.

The *F*-value of the regression model is 3,453.408, and the regression model has overall significance at a given significance level $\alpha = 0.05$. The *t* statistic of each coefficient satisfies $|t| > t_{0.025}(N_1 - 5)$, indicating the significant impact on $\widehat{w}_{1i}$. The coefficient of determination R^2 and Adjusted R^2 are 0.796 and 0.796, respectively. In order to ensure that the best estimation can be obtained using OLS, the regression model should satisfy two assumptions: (i) there is no correlation between any two random error terms, (ii) random error term has homoscedasticity, which is denoted by $\mathrm{Cov}(u_{1i},u_{1j}) = 0$ and $\mathrm{Var}(u_{1i}) = \sigma^2$, $(i \neq j, i, j = 1, 2, \ldots, N_1)$, respectively. If any of the above hypotheses are not true, the regression parameters are neither valid estimators nor asymptotically valid ones, resulting in the meaningless *F*-value and R^2 from the significant test.

According to the Durbin Watson (DW) test, the DW statistic of the regression model is 2.047, indicating that there is no correlation among u_{1i}. The white test is performed to test whether heteroscedasticity exists among u_{1i}. In the white test, an auxiliary regression model is established as shown in Eq. (6). The explained variable is the square of residual error $\widehat{u}_{1i}^2$ and the explanatory variables are the combination of the standard term, square term, and cross term of each explanatory variable in the regression model.

$$\begin{aligned}\widehat{u}_{1i}^2 = {} & \beta_0 + \beta_1 \ln L_i + \beta_2 \ln M_i + \beta_3 \ln t_i + \beta_4 \left|\overline{T}_i - \overline{T}_g^*\right| + \beta_5 (\ln L_i)^2 + \beta_6 (\ln M_i)^2 + \beta_7 (\ln t_i)^2 + \beta_8 \left|\overline{T}_i - \overline{T}_g^*\right|^2 \\ & + \beta_9 (\ln L_i)\ln M_i + \beta_{10} (\ln L_i)\ln t_i + \beta_{11} (\ln L_i)\left|\overline{T}_i - \overline{T}_g^*\right| + \beta_{12} (\ln M_i)\ln t_i + \beta_{13} (\ln M_i)\left|\overline{T}_i - \overline{T}_g^*\right| + \beta_{14}\left|\overline{T}_i - \overline{T}_g^*\right| \ln t_i + v_i\end{aligned} \tag{6}$$

Under the homoscedasticity hypothesis H_0: $\beta_0 = \ldots = \beta_{14}$, parameters of the auxiliary regression model are estimated using OLS, and the coefficient of determination $R^2 = 0.081$ is obtained. The statistic $N_1 R^2$ is formulated and tested under a given significance level $\alpha = 0.05$. The hypothesis is rejected because $N_1 R^2 > \chi_\alpha^2(14)$. Hence, there is heteroscedasticity in the T-BTMS energy consumption estimation model.

Weighted Least Square (WLS) is the special case of Generalized Least Square. It is usually used to deal with heteroscedasticity by assigning different weights to different observations. To be specific, an observation with a small variance has a large weight, and an observation with a large variance has a small weight. This method aims to minimize the sum of the square of residual errors and formulate a new model eliminating heteroscedasticity to further reduce the effect on parameter estimation of inaccurate observations. Since u_{1i} approximately satisfies $\mathrm{Var}(u_{1i}) \approx \widehat{u}_{1i}^2$, we use $1/\sqrt{\widehat{u}_{1i}^2} = 1/|\widehat{u}_{1i}|$ as the weight and multiply it with both sides of the original model, as shown in Eqs. (7) and (8).

$$\begin{aligned}\frac{1}{|\widehat{u}_{1i}|}\ln w_{1i} = {} & \alpha_0^* \frac{1}{|\widehat{u}_{1i}|} + \alpha_1^* \frac{1}{|\widehat{u}_{1i}|}\ln L_i + \alpha_2^* \frac{1}{|\widehat{u}_{1i}|}\ln M_i + \alpha_3^* \frac{1}{|\widehat{u}_{1i}|}\ln t_i + \alpha_4^* \frac{1}{|\widehat{u}_{1i}|}\left|\overline{T}_i - \overline{T}_g^*\right| \\ & + \frac{1}{|\widehat{u}_{1i}|}u_{1i}\end{aligned} \tag{7}$$

$$\mathrm{Var}\left(\frac{1}{|\widehat{u}_{1i}|}u_{1i}\right) = \left(\frac{1}{|\widehat{u}_{1i}|}\right)^2 \mathrm{Var}(u_{1i}) = \frac{1}{\widehat{u}_{1i}^2}\widehat{u}_{1i}^2 = 1 \tag{8}$$

where $\beta_0 \sim \beta_{14}$ are regression parameters and v_i is a random error term. This new model satisfies the hypothesis of homoscedasticity, and the modified T-BTMS energy consumption regression model is expressed in Eq. (9):

$$\ln \widehat{w}_{1i} = \widehat{\gamma}_0 + \widehat{\gamma}_1 \ln L_i + \widehat{\gamma}_2 \ln M_i + \widehat{\gamma}_3 \ln t_i + \widehat{\gamma}_4 \left|\overline{T}_i - \overline{T}_g^*\right| \tag{9}$$

Estimated values of regression parameters in Eq. (9) are shown in Table 6. The optimum working temperature $\overline{T}_g^*$ is 23.7 °C. The modified regression model is of overall significance at a given significance level $\alpha = 0.05$. The significance test is conducted for the coefficient of each explanatory variable at the significance level $\alpha = 0.05$. Results reveal that $\ln L_i$, $\ln M_i$, $\ln t_i$ and $\left|\overline{T}_i - \overline{T}_g^*\right|$ are all significant explanatory variables to $\ln \widehat{w}_{1i}$. The standardized coefficients (betas) are used to compare the impact of each independent parameter on $\ln \widehat{w}_{1i}$. Accordingly, $\ln L_i$ has a substantial impact on the $\ln \widehat{w}_{1i}$ (Std. beta = 0.360), followed by $\ln M_i$ (Std. beta = 0.352) and $\left|\overline{T}_i - \overline{T}_g^*\right|$ (Std. beta = 0.343), while $\ln t_i$ (Std. beta = 0.234) has the lowest weight in affecting $\ln \widehat{w}_{1i}$.

Similarly, the estimation of regression parameters in the estimation model of EB energy consumption caused by the AC system operation is demonstrated in Table 7. At a given significance level $\alpha = 0.05$, operation times under cooling and heating modes are significant explanatory variables of TEC. 1-min operation time changes in cooling mode and heating mode resulted in average energy consumption changes of 0.053 kWh and 0.110 kWh variations, respectively. t_{ai} has a greater impact on the $\widehat{w}_{2i}$ when $D^1 = 0$ (Std. beta = 0.435) than $D^1 = 1$ (Std. beta = 0.278). A more significant increment in energy consumption is observed under the heating mode since the heating power is larger than the cooling power of the AC system.

4. Model evaluation and analysis

4.1. Estimation accuracy

Root mean squared error (*RMSE*), mean absolute error (*MAE*), mean

Table 5
Regression parameter estimation results based on OLS.

Parameters	$\widehat{\alpha}_0^*$	$\widehat{\alpha}_1^*$	$\widehat{\alpha}_2^*$	$\widehat{\alpha}_3^*$	$\widehat{\alpha}_4^*$
Estimated values	−8.108	0.553	0.781	0.354	0.008
Std. Error	0.182	0.019	0.020	0.017	0.000
t statistic	−44.433	28.802	40.041	21.170	41.658
P value	0.000	0.000	0.000	0.000	0.000

F-value = 3,453.408 (*P*-value = 0.0000); $R^2 = 0.796$; Adjusted $R^2 = 0.796$.

Table 6
Regression parameter estimation results based on WLS for $\widehat{w}_{1i}$.

Parameters	$\widehat{\gamma}_0$	$\widehat{\gamma}_1$	$\widehat{\gamma}_2$	$\widehat{\gamma}_3$	$\widehat{\gamma}_4$
Estimated values	−8.091	0.553	0.780	0.353	0.008
Std. Error	0.006	0.001	0.001	0.000	0.000
t statistic	−1,251.469	867.665	1,046.558	932.796	2,259.494
P value	0.000	0.000	0.000	0.000	0.000

F-value = 8,449,689 (*P*-value = 0.000).

Table 7
Regression parameter estimation results based on WLS for $\widehat{w}_{2i}$.

Parameters	Estimated values	Std. Error	*t* statistic	*P* value
$\widehat{\zeta}_1$	0.053	0.000	1,088.091	0.000
$\widehat{\zeta}_2$	0.110	0.000	1,006.768	0.000

absolute percentage error (*MAPE*), and Theil coefficient (*Theil*) are selected as evaluation indicators of the estimation model and expressed by Eqs. (10)–(13). The smaller *RMSE*, *MAE*, and *MAPE* are, the higher estimation accuracy the model has. The scale range of Theil's coefficient is [0, 1]. A smaller Theil's coefficient indicates a better estimator.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{w}_i - w_i)^2} \tag{10}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\widehat{w}_i - w_i| \tag{11}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\widehat{w}_i - w_i}{w_i}\right| \times 100\% \tag{12}$$

$$Theil = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{w}_i - w_i)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{w}_i)^2} + \sqrt{\frac{1}{N}\sum_{i=1}^{N}(w_i)^2}} \tag{13}$$

The proposed TEC estimation model is evaluated under four scenarios: (i) trips with AC off ($t_a = 0$); (ii) trips with AC in cooling mode ($D^1 = 1$); (iii) trips with AC in heating mode ($D^1 = 0$); and iv) all trips. The evaluation results are demonstrated in Table 8. Model evaluations are also performed when applied to different routes and bus models, as illustrated in Table 9.

Considering all samples, *RMSE*, *MAE*, *MAPE*, and Theil's coefficient of the TEC estimation model are 0.984 kWh, 0.696 kWh, 12.108%, and 0.082, respectively. As demonstrated by Table 9, when the model is applied to different routes and bus models, *MAPE* under four scenarios are 11.707%, 11.874%, 14.721%, and 12.229%, respectively. Theil's coefficients are all less than 0.1, indicating a reliable accuracy of the estimation model.

Table 8
Evaluation results of the trip energy consumption estimation model under different scenarios.

Scenario	*N*	*RMSE* (kWh)	*MAE* (kWh)	*MAPE* (%)	*Theil* coefficient
$t_{a=0}$	3536	0.882	0.628	11.727	0.077
$D^1{}_{=1}$	455	0.841	0.644	12.284	0.077
$D^1{}_{=0}$	369	1.761	1.405	15.538	0.107
All trips	4360	0.984	0.696	12.108	0.082

Table 9
Evaluation results of the trip energy consumption estimation model on different routes and bus models.

Bus model	*N*	*RMSE* (kWh)	*MAE* (kWh)	*MAPE* (%)	*Theil* coefficient
108-bus model A	2032	0.705	0.520	11.707	0.078
103-bus model B	1409	0.992	0.722	11.874	0.078
103-bus model C	414	1.324	1.026	14.721	0.086
106-bus model C	505	1.473	1.059	12.229	0.092

The estimation accuracy of the model under the scenario with AC in cooling mode ($D^1 = 1$) is higher than that in heating mode ($D^1 = 0$). *RMSE*, *MAE*, and *MAPE* under the former scenario are 0.841 kWh, 0.644 kWh, and 12.284%, respectively and Theil's coefficient is 0.077. For the latter scenario, these indicators are 1.761, 1.405 kWh, 15.538%, and 0.107. The possible reason for the difference in estimation accuracy is that we gathered data from a cold region where cold weather lasts for a long time each year. The heating mode might be turned on when the ambient temperature is within [−27.0 °C, 5.0 °C]. The heating power may differ for each trip and may vary along the trip based on multiple factors such as passenger numbers, onboard temperature, and ambient temperature. We only use the AC operation time of a trip to simulate the impact energy consumption without capturing the AC power changes during the trip. Therefore, the estimation error of the model when AC in heating mode is relatively high.

4.2. Comparisons

In this section, the proposed TEC estimation model is compared with a regression model developed by Abdelaty and Mohamed (2021), which is shown in Eq. (14):

$$E_i = \theta_0 + \theta_1 GR_i + \theta_2 D_i^{Agg} + \theta_3 RC_i + \theta_4 AC_i + \theta_5 PL_i + \theta_6 SD_i + \theta_7 \overline{v}_i + \theta_8 SOC_i \tag{14}$$

where E_i is the energy consumption rate of trip i, kWh/km; GR_i is the road grade of trip i, %; D_i^{Agg} is the driver aggressiveness of trip i (three levels: slow driving behavior, normal driving behavior and aggressive driving behavior); RC_i is the road condition of trip i (three levels: a good dry road condition, a fair wet road condition and a poor icy road condition); AC_i is the consumed energy of AC system of trip i, kWh; PL_i is the passenger loading of trip i, pax; SD_i is the stop density ratio, stops/km; $\overline{v}_i$ is the average speed of trip i, km/h; SOC_i is initial battery SOC at the departure time of trip i, %; $\theta_0 \sim \theta_8$ are estimated regression parameters, and their values are −0.782, 0.380, 0.065, 0.260, 0.036, 0.005, 0.128, 0.007 and 0.0124, respectively.

From Eq. (14) we can find that the energy consumption rate estimation needs instantaneous EB movement data to judge the driver aggressiveness level. In addition, the number of boarding and alighting passengers at stops should also be collected to calculate the passenger loading at each road section. Hence, the data presented in Section 2 cannot be used for model comparison. To conduct the comparison, we collected the second-by-second operational data of two EBs on Route 108 on May 11 and 12, June 22 and 23, 2021, involving a total of 150 trips. The data of each trip includes velocity, longitude, latitude, altitude, road grade, travel distance and SOC per second, the trip energy consumption and ambient temperature. Via the onboard monitoring video, the numbers of boarding and alighting passengers at each station were collected.

During the data collection periods, the AC system was off for all trips, thus $AC_i = 0$ in Eq. (14). There was no rainfall or snowfall, road condition was a good dry condition and RC_i refer to Level I. The acceleration rate was between 0.25 and 0.5 m/s^2, and deceleration rate was between 1

Table 10
Comparison results of the two estimation models.

Models	Model structure	R^2	*RMSE*
Model I	$E_i = \theta_0 + \theta_1 GR_i + \theta_2 D_i^{Agg} + \theta_3 RC_i + \theta_4 AC_i + \theta_5 PL_i + \theta_6 SD_i + \theta_7 \overline{v}_i + \theta_8 SOC_i$	0.961	0.691
Model II	$\ln \widehat{w}_{1i} = \widehat{\alpha}_0 + \widehat{\alpha}_1 \ln L_i + \widehat{\alpha}_2 \ln M_i + \widehat{\alpha}_3 \ln t_i + \widehat{\alpha}_4 \left\lvert \overline{T}_i - \overline{T}_g^* \right\rvert$	0.986	0.336

and 1.5 m/s^2. Thus, D_i^{Agg} belongs to Level I. SD_i can calculated by dividing the number of stations by route length. GR_i and PL_i take the average road grade and passenger loading for all segments of trip *i*.

The comparison results of the two models are displayed in Table 10. Model I refer to the model developed by Abdelaty and Mohamed (2021) and there are 8 contributing factors. Model II is the model proposed in this study and there are 4 contributing factors.

The R^2 of Model II is 0.986, which is larger than that of Model I. Compared with Model I, the *RMSE* of Model II decreases by 51.4%. Overall, Model II outperforms Model I in the estimation accuracy because of the two following reasons: (i) The curb weight of the EB is considered in Model II. As shown in Table 2, the curb weights of the three types of EBs are all larger than 8,200 kg, which is not a negligible factor to the TEC. (ii) We quantified the effect of ambient temperature on TEC. As the ambient temperature increases or decreases, the battery thermal management system would consume more additional energy to maintain the battery temperature within a proper range.

Passenger loading is not considered in our model (Model II) because the numbers of boarding and alighting passengers at stations are hard to be collected. Its impact on the TEC will be explained in the next section.

4.3. Sensitivity analysis

Based on the proposed estimation model, we perform the sensitivity analysis to explore the effects of ambient temperature, curb weight, passenger loading, travel distance, and trip travel time on the TEC.

(i) Effect of ambient temperature on TEC

As described in Section 3.3, the optimum working temperature of EBs is 23.7 °C. However, in most previous studies, 25 °C is taken as the optimum working temperature of EBs (Wang et al., 2011). For all trips, the *MAPE* of the estimation model is 14.478% when applying 25 °C as the optimum working temperature, demonstrating a 2.37% increase compared to the *MAPE* when applying 23.7 °C.

The estimation accuracy of the formulated estimation model slightly fluctuates among different temperature ranges. Here, evaluations are performed on the model among different temperature ranges with the 10 °C intervals, as shown in Table 11.

We further analyzed the relationship between TEC and ambient temperature under 6 combinations of different travel distances and curb weight. We assume that the trip travel time is fixed (average value of 29min) and the AC system is off. The result is shown in Fig. 5. Ambient temperature, especially extreme cold weather in cold regions, is the main contributor to the increase in TEC, the reduced driving range, and the shrinking in battery performance. Compared with bus running under the optimum working temperature 23.7 °C, TEC of an EB under the temperature of −27.0, −20.0, −10.0, 0.0, 10.0, 20.0, 30.0 and 35.0 °C increases by 47.105%, 39.471%, 29.247%, 19.773%, 10.993%, 2.857%, 4.913%, and 8.984%, respectively.

(ii) Effect of curb weight on TEC

Assuming a fixed trip travel time (29 min), ambient temperature (23.7 °C), and inactive AC system, the TEC varying curb weight under different travel distances are shown in Fig. 6. An approximately positive linear relationship can be observed between TEC and curb weight. TEC increases with the increase of curb weight, but the increment continues to decrease. Route 108 with the 8.1 km travel distance (trip length) is taken as an example. When curb weight increases within the range [8,200 kg, 18,000 kg], TEC rises from 3.461 to 6.444 kWh, but the changing rate reduces from 0.033 kWh/100 kg to 0.028 kWh/100 kg.

Table 11
Evaluation results of the estimation model under different temperature ranges.

Temperature (°C)	*N*	*RMSE* (kWh)	*MAE* (kWh)	*MAPE* (%)	*Theil* coefficient
(-30, −20]	198	1.088	0.772	10.548	0.076
(-20, −10]	1,061	1.173	0.826	11.250	0.081
(-10, 0]	849	1.264	0.950	15.173	0.093
(0, 10]	403	1.000	0.748	13.348	0.089
(10, 20]	739	0.544	0.418	9.993	0.063
(20, 30]	963	0.702	0.528	11.772	0.076
(30, 40]	147	0.700	0.533	12.124	0.115

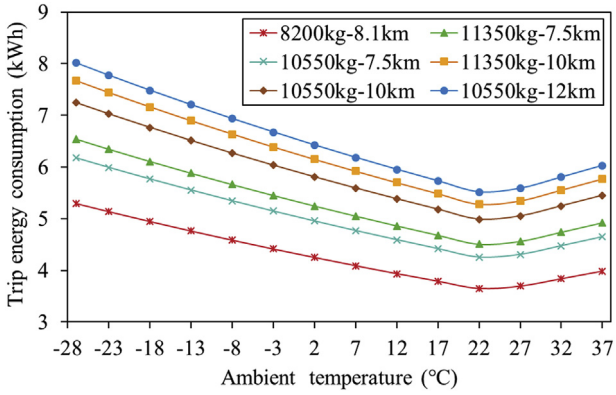


Fig. 5. Relationship curve between TEC and ambient temperature.

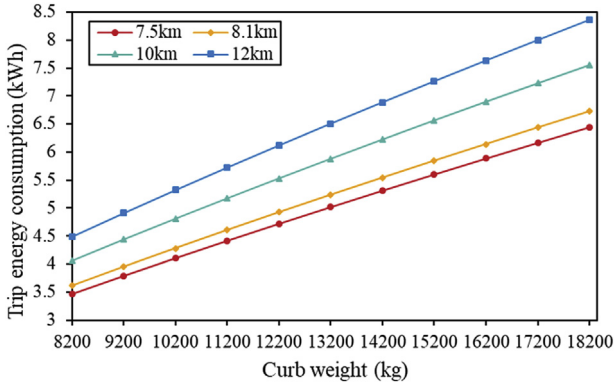


Fig. 6. Relationship curve between TEC and curb weight.

The weight of the battery system is an essential part of the EB curb weight. For example, the three bus models in Table 2 are all equipped with heavy battery packs to provide a larger rated capacity. The weight of the battery system accounts for 14.7%–15.30% of the curb weights of the three bus models. The rated capacity of common EB models in most cities is between 50 and 400 kWh. To investigate the effect of battery-rated capacity on TEC, we take bus model A on Route 108 as an instance of which battery energy density is 140.40 Wh/kg. We analyze the impact of battery-rated capacity on TEC under the scenario of fixed trip travel time (29 min), ambient temperature (23.7 °C), and the inactive AC system, as demonstrated in Fig. 7.

It is observed that when battery rated capacity rises from 50 to 400 kWh, the weight of the battery system increases from 405 to 2,892 kg, and curb weight increases from 7,399 to 9,887 kg. As a result, the ratio of battery system weight to curb weight changes from 5.470% to 29.254%. Energy consumption rate increases from 0.410 to 0.514 kWh/km by 25.366% simultaneously. In recent years, rapid charging and wireless charging technologies continue to advance, which can sharply reduce the charging time for the EB. Thus, it is an alternative to minimize battery rated capacity further to reduce battery system weight and the corresponding curb weight. In this case, the bus ownership cost and trip energy consumption reduce accordingly.

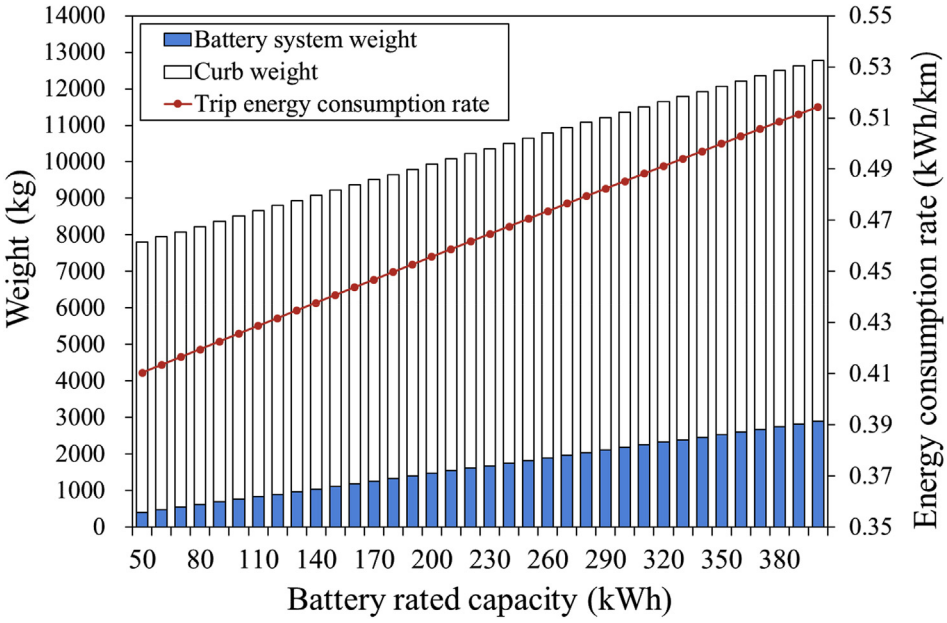


Fig. 7. Energy consumption rate of Route 108-vehicle type An under different battery rated capacities.

(iii) Effect of passenger loading on TEC

The number of passengers on board is another significant factor influencing the whole weight of an operating bus. The entire bus weight contains curb weight and the total weight of passengers onboard. However, only curb weight is considered in the formulation of the TEC estimation model. For example, for an EB with a passenger capacity of 60 persons, as the average weight of passengers assumed as 60 kg per person, the increment of 3,600 kg will be added to its curb weight and accounts for 30.51% of the whole weight of the bus during its operation, which is not negligible for TEC.

Assuming that the average passenger weight is 60 kg. We use the distance between any two consecutive stops as the weighting factor. The weighted average total passenger weight of each trip is the total passenger weight between every two consecutive stops multiplied by the corresponding weighting factor denoted by m_i, $i = 1, 2, \ldots, 150$. Here 150 is the number of trips with the AC system is off (please refer to the second-by-second operational data collection in Section 4.2). Eq. (9) is modified to be Eq. (15), and $\tilde{\gamma}_0 \sim \tilde{\gamma}_4$ are the updated fitting coefficients.

$$\ln \widehat{w}_{1i} = \tilde{\gamma}_0 + \tilde{\gamma}_1 \ln L_i + \tilde{\gamma}_2 \ln(M_i + m_i) + \tilde{\gamma}_3 \ln t_i + \tilde{\gamma}_4 \left| \overline{T}_i - \overline{T}_g^* \right| \quad (15)$$

According to Eq. (9), *RMSE*, *MAE*, and *MAPE* for these 150 samples are 0.336 kWh, 0.258 kWh, and 6.706%, respectively. However, these indicators under Eq. (15) are 0.236 kWh, 0.183 kWh, and 4.960%, revealing an improvement in estimation accuracy by considering the number of passengers on board. The improvement in *MAPE* is not significant after considering the passenger loading. This is mainly because the EB curb weight (M_i) is significantly larger than the weighted average total passenger weight (m_i), especially for the trips in non-peak operating hours. A slight increase in ($M_i + m_i$) does not result in a large increase in TEC. However, in peak hours there are more passengers in the EBs and the ratio of m_i to M_i would increase. In such a condition, the improvement in *MAPE* would be more significant.

(iv) Effects of travel distance and trip travel time on TEC

With fixed ambient temperature (23.7 °C), fixed curb weight (8,200 kg, bus model A), and inactive AC system, the relationship between the TEC and the trip travel time under different travel distances is shown in Fig. 8. Positive relationships are observed between TEC and travel distance and between TEC and trip travel time, namely, TEC rises with the increase in travel distance or trip travel time. In the 8-km route, with trip travel time increasing within the interval [14 min, 42 min], TEC grows from 2.765 to 4.076 kWh, and the changing rate drops from 0.070 to 0.034 kWh/min. Similar results are observed in another case of 29-min average trip travel time, where TEC grows from 3.451 to 4.476 kWh, and the changing rate drops from 0.254 to 0.207 kWh/km with travel distance increasing within [7.5 km, 12 km].

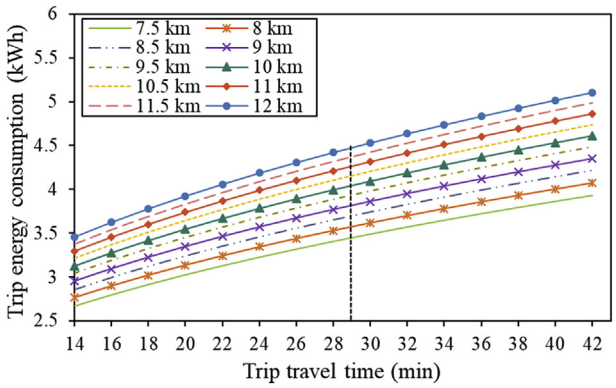


Fig. 8. Relationship curve between TEC and trip travel time.

The trip travel time of the EB is also an indicator of traffic conditions. For Route 108, the average trip travel time of an EB in off-peak hours (14:00–15:30) is 26.48 min, and the average TEC is 4.4 kWh while the average trip travel time in peak hours (17:00–18:00) is 32.37 min and the average TEC is 4.7 kWh. Thus, TEC increases by only 6.82% with the 22.24% increment in trip travel time. In a traditional fuel bus case, the average trip fuel consumptions are 2.84 L and 3.27 L in off-peak and peak hours, respectively, and the latter is 15.14% higher than the former. Therefore, compared with fuel buses, traffic conditions have a smaller impact on EB energy consumption, indicating the significant superiority in energy conversion efficiency of bus electrification. Besides, vehicle queuing time at intersections rises in peak hours. As wireless charging technology advances, EBs can get charged when waiting at the intersection. The charging time at charging stations decreases accordingly, enhancing the operational efficiency of transit systems.

4.4. Adaptability for bus scheduling

One of the most significant challenges that EBs face is the limited driving range. In EB scheduling, battery SOC should maintain within a predefined interval to prolong battery life, and ensure that bus service is not interrupted due to insufficient remaining battery electricity. Therefore, the primary objective of formulating the TEC estimation model of EBs is to provide a calculation method of TEC for bus scheduling. Hence, we analyze whether the formulated estimation model satisfies practical requirements in terms of bus scheduling.

From observing actual TEC data of EBs, it is noticed that even with the exact contributing factors like travel distance, curb weight, trip travel time, average ambient temperature, and AC system operation time, minor differences still exist in energy consumption among multiple trips. This is because TEC is also affected by some other individualized factors, such as accelerations and decelerations, driving style, and delay at intersections, which are essentially uncontrollable and not incorporated into Eq. (1). Thus, the energy consumption of any trip h and the daily energy consumption of EB k should satisfy:

$$w_h^{'} = \widehat{w}_h + \widehat{u}_h,\ h = 1, 2, ..., H \tag{16}$$

$$w_k^{'} = w_1^{'} + w_2^{'} + ...w_h^{'} + ... + w_H^{'} = \sum_{h=1}^{H} \widehat{w}_h + \sum_{h=1}^{H} \widehat{u}_h \tag{17}$$

where $w_h^{'}$ is the actual energy consumption of EB k on trip h, kWh; $w_k^{'}$ is the actual daily energy consumption of EB k, kWh; $\widehat{w}_h$ is the estimated energy consumption of EB k on trip h calculated by Eq. (1), kWh; $\widehat{u}_h$ is the residual error; H is the number of trips that EB k should serve in all-day operation time.

$w_k^{'}$ is unknown and unobservable during the bus scheduling stage. Thus, it needs to be calculated through the residual error of the estimation model. The probability distribution function of $w_k^{'}$, denoted by $G(w_k^{'})$, is determined by the characteristics of $\widehat{u}_h$. $G(w_k^{'})$ can be obtained using the probability distribution function of $\widehat{u}_h$ and Eq. (17). Battery rated capacity of EB k is denoted by Q_k. To avoid over-discharge and damage to battery life, a lower bound of battery SOC is set and denoted by λ_1 which means that the remaining battery electricity is no less than $\lambda_1 Q_k$ after EB k completing its all-day operation. That is to say, the maximum useful electricity consumption of EB k is $(1 - \lambda_1)Q_k$. As shown in Eq. (18), when the probability of $w_k^{'} \le (1-\lambda_1)Q_k$ is higher than or equal to a preset δ, it is suggested that the accuracy of the estimation model can cope with the requirements of bus scheduling.

$$P\left\{w_k^{'} \le (1-\lambda_1)Q_k\right\} = G((1-\lambda_1)Q_k) \ge \delta \tag{18}$$

As stated in Section 4.2, TEC varies with different ambient temperatures. The lower ambient temperature leads to the higher daily energy consumption of the EB, so the battery SOC is at a low level at the end of the operation. This will result in a high possibility of service interruption due to insufficient remaining battery power. Therefore, this puts forward higher requirements for the accuracy of EB's TEC estimation model. Therefore, we randomly select one EB on Route 106 and take its operational data on 18 January 2021 (the coldest day among the data collection period) to analyze whether the formulated estimation model satisfies the requirement of bus scheduling.

The battery rated capacity of the selected bus is 202.93 kWh, and 14 trips are required to serve in one day, denoted by $H = 14$. The heating mode of the AC was turned on for trip 13 and trip 14. Estimated energy consumptions of each trip $\widehat{w}_h$ ($h = 1,\ 2, ...,\ 14$) calculated by Eq. (1) and Eq. (17) are 8.4, 7.4, 7.2, 7.2, 8.4, 8.9, 8.2, 7.8, 8.5, 7.7, 8.7, 8.4, 15.6, and 13.2 kWh, respectively. Accordingly, the estimated daily energy consumption of the selected bus $\widehat{w}_k = 125.6$ kWh.

The residual error of the formulated estimation model follows the normal distribution with mean as 0.1 kWh and variance as 0.787 $(\text{kWh})^2$, denoted by $\widehat{u}_h \sim N(0.1, 0.787)$. Thus, $w_k^{'}$ follows the normal distribution with mean as $\widehat{w}_k + 0.1H$ and variance as $0.787H^2$, denoted by $w_k^{'} \sim N(127.0, 154.252)$. δ equals 0.9974 under the 3σ rule. When $\lambda_1 = 20\%$, $(1 - \lambda_1)Q_k = 162.344$ kWh. Therefore:

$$P\left\{w_k^{'} \le (1-\lambda_1)Q_k\right\} = G((1-\lambda_1)Q_k) = \Phi\left(\frac{162.344 - 127.0}{\sqrt{154.252}}\right) \approx \Phi(2.85) \tag{19}$$

where $\Phi(\cdot)$ is the standard normal distribution function. $\Phi(2.85) = 0.997814$, and then we have $P\{w_k^{'} \le (1 - \lambda_1)Q_k\} = 0.997814 > 0.9974$. Therefore, it is believed that the formulated estimation model can meet the requirement of bus scheduling.

In this case, two reasons lead to the formulated estimation model meeting the requirement. One reason is the large battery rated capacity of this bus and $(1-20\%)Q_k$ reaches to 162.344 kWh. The other reason is the small number of trips serving per day of the bus and short trip travel times. Accordingly, the estimated daily energy consumption is only 125.6 and 36.744 kWh less than $(1 - 20\%)Q_k$. However, if Q_k reduces while daily travel distance or travel time increases, then $(1-20\%)Q_k - \widehat{w}_k$ will keep reducing, which requires higher estimation accuracy.

5. Discussion

This study used the regression method to obtain the explicit relationship function between TEC and contributing factors based on real-world data. By filling the gaps in current studies, it made unique contributions to understanding the impacts of contributing factors on TEC and EB planning and operation plans.

First, some common parameters are not considered in this study, such as the number of stations and intersections (NOSI), and the road grade. The Spearman's rank correlation coefficients of the two parameters with TEC are 0.066 and 0.165, respectively. The Partial correlations are 0.085 and −0.038, which shows very weak or no correlation with TEC. The Spearman's rank correlation coefficients between NOSI and trip distance, and trip travel time are 0.744 and 0.683, respectively, which shows strong linear relationships. Thus, the impact of NOSI can be reflected by the constant term or the fitting coefficients of other contributing factors. The reason for the weak relationship between road grade and TEC is that the changes of road slopes are small and the road grades of the three routes are within [−4%, 4%]. Therefore, the proposed TEC estimation model is more suitable for plain areas.

Second, this study accounted for the influences of ambient temperatures on TEC in cold regions whereas current studies did not. In the real-world data used in the regression, the minimum and maximum temperatures were −27.0 and 35.0 °C, respectively. Due to the significant impact of the ambient environment on TEC, the difference in daily energy consumption of an EB fleet on a route could reach up to almost 50% between summer and winter. Therefore, it is necessary to deploy different numbers of EBs and charging stations for transit routes in different seasons and to adjust bus schedules accordingly, which poses a challenge for transit operation and management.

Third, this study examined the relationship between battery rated capacity and energy consumption rate. Nowadays, many transit operators like to purchase EBs with large-capacity batteries. Although this can increase the driving range, large-capacity batteries are often heavy, which will lead to an increase in the energy consumption rate, thereby increasing the operating costs of transit operators. In fact, different bus routes have different passenger demands, dispatching frequencies, and EB operating speeds. The transit operator should choose the best EB model according to the characteristics of each route, rather than the one-

sided pursuit of a large-capacity battery.

Fourth, the estimation accuracy of the proposed model was examined using different indicators. Although considering the number of passengers on board could improve the accuracy of the model, the estimation error was still inevitable. If the error is not considered when developing the EB scheduling plan, the remaining energy of an EB at the departure time would be overestimated, resulting in the service interruption. However, current studies did not consider the impacts of TEC estimation error and most of them assumed that energy consumption rates of EBs during operation remain constant. An alternative method is to determine the probability distribution function (PDF) for TEC based on the collected TEC, estimated TEC and the errors. The PDF can be used to describe the stochastic fluctuations in TEC and help improve the reliability of the planning and operation models for EBs.

6. Conclusions

Based on the real-world EB operational data collected in a cold region, a multivariable regression model was established, with multiple contributing factors as input variables and TEC of an EB as the output variable. The accuracy evaluation of the formulated model is performed and sensitivity analysis is conducted. The main conclusions of this research are emphasized as follows:

(i) The nonlinear estimation model well explains relationships between the TEC and five contributing factors. We provide a method for analyzing the adaptability of the model in terms of bus scheduling. The result of the case study shows that the estimation accuracy of the model has a probability of 99.7814% meeting the requirements of bus scheduling.

(ii) Ambient temperature is a critical contributor to TEC increments of EBs. Compared with 23.7 °C (the optimum working temperature), the TEC increases by 47.105% and 8.984%, respectively, under the conditions of the lowest temperature of −27.0 °C and the highest temperature of 35.0 °C.

(iii) When the battery rated capacity of an EB increases from 50 kWh to 400 kWh, the energy consumption rate rises from 0.410 to 0.514 kWh/km by 25.366%. With the advancement of fast charging and wireless charging technologies, it has become an alternative to equip EBs with lower rated capacity batteries, thereby reducing TEC and operating costs.

(iv) The number of onboard passengers is another important factor affecting the whole weight of an EB during its operation. *MAPE* has a 1.746% reduction by applying the modified TEC estimation model considering passengers' boarding and alighting.

Replication and data sharing

The partial data used in this research can be found at https://ets-data.sciopen.com/view/ETS2022052200002/1/3.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Yiming Bie reports financial support was provided by National Natural Science Foundation of China. Yiming Bie reports financial support was provided by China Postdoctoral Science Foundation. Yiming Bie reports financial support was provided by Fundamental Research Funds for the Central Universities.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 52131203); China Postdoctoral Science Foundation (Grant Nos. 2019M661214 & 2020T130240); and Fundamental Research Funds for the Central Universities (Grant No. 2020-JCXK-40).

References

Abdelaty, H., Al-Obaidi, A., Mohamed, M., Farag, H.E.Z., 2021. Machine learning prediction models for battery-electric bus energy consumption in transit. Transport. Res. Transport Environ. 96, 102868.

Abdelaty, H., Mohamed, M., 2021. A prediction model for battery electric bus energy consumption in transit. Energies 14 (10), 2824.

Al-Ogaili, A.S., Ramasamy, A., Hashim, T.J.T., Al-Masri, A.N., Hoon, Y., Jebur, M.N., 2020. Estimation of the energy consumption of battery driven electric buses by integrating digital elevation and longitudinal dynamic models: Malaysia as a case study. Appl. Energy 280, 115873.

Bie, Y.M., Ji, J.H., Wang, X.Y., Qu, X.B., 2021. Optimization of electric bus scheduling considering stochastic volatilities in trip travel time and energy consumption. Comput. Aided Civ. Infrastruct. Eng. 36 (12), 1530–1538. https://doi.org/10.1111/mice.12684.

Beckers, C.J.J., Besselink, I.J.M., Nijmeijer, H., 2020. Assessing the impact of cornering losses on the energy consumption of electric city buses. Transport. Res. Transport Environ. 86, 102360.

Chen, Y., Zhang, Y., Sun, R., 2021. Data-driven estimation of energy consumption for electric bus under real-world driving conditions. Transport. Res. Transport Environ. 98, 102969.

El-Taweel, N.A., Zidan, A., Farag, H.E.Z., 2021. Novel electric bus energy consumption model based on probabilistic synthetic speed profile integrated with HVAC. IEEE Trans. Intell. Transport. Syst. 22 (3), 1517–1531.

Fiori, C., Ahn, K., Rakha, H.A., 2016. Power-based electric vehicle energy consumption model: model development and validation. Appl. Energy 168, 257–268.

Fiori, C., Montanino, M., Nielsen, S., Seredynski, M., Viti, F., 2021. Microscopic energy consumption modelling of electric buses: model development, calibration, and validation. Transport. Res. Transport Environ. 98, 102978.

Galleta, M., Massiera, T., Hamacher, T., 2018. Estimation of the energy demand of electric buses based on real-world data for large-scale public transport networks. Appl. Energy 230, 344–356.

Gao, Y., Guo, S., Ren, J., Zhao, Z., Ehsan, A., Zheng, Y., 2018. An electric bus power consumption model and optimization of charging scheduling concerning multi-external factors. Energies 11 (8), 2060.

Gao, Z., Lin, Z., LaClair, T.J., Liu, C., Li, J.M., Birky, A.K., 2017. Battery capacity and recharging needs for electric buses in city transit service. Energy 122, 588e600.

He, J., Yan, N., Zhang, J., Yu, Y., Wang, T., 2022. Battery electric buses charging schedule optimization considering time-of-use electricity price. Journal of Intelligent and Connected Vehicles 5 (2), 138–145.

He, X., Zhang, S., Ke, W., Zheng, Y., Zhou, B., Liang, X., 2018. Energy consumption and well-to-wheels air pollutant emissions of battery electric buses under complex operating conditions and implications on fleet electrification. J. Clean. Prod. 171, 714e22.

Hjelkrem, O.A., Lervåg, K.Y., Babri, S., Lu, C., Södersten, C.J., 2021. 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. Transport. Res. Transport Environ. 94, 102804.

Huang, M., Zhai, P., 2021. Achieving Paris Agreement temperature goals requires carbon neutrality by middle century with far-reaching transitions in the whole society. Adv. Clim. Change Res. 12 (2), 281–286.

Hu, X., Xu, L., Lin, X., Pecht, M., 2020. Battery lifetime prognostics. Joule 4 (2), 310–346.

Kivekäs, K., Vepsäläinen, J., Tammi, K., 2018. Stochastic driving cycle synthesis for analyzing the energy consumption of a battery electric bus. IEEE Access 6, 55586–55598.

Kunith, A., Mendelevitch, R., Goehlich, D., 2017. Electrification of a city bus network-An optimization model for cost-effective placing of charging infrastructure and battery sizing of fast-charging electric bus systems. Int. J. Sustain. Transp. 11 (10), 707–720.

Lajunen, A., 2018. Lifecycle costs and charging requirements of electric buses with different charging methods. J. Clean. Prod. 172, 56–67.

Łebkowski, A., 2019. Studies of energy consumption by a city bus powered by a hybrid energy storage system in variable road conditions. Energies 12 (5), 951.

Leou, R.C., Hung, J.J., 2017. Optimal charging schedule planning and economic analysis for electric bus charging stations. Energies 10 (4), 483.

Li, L., Lo, H.K., Xiao, F., 2019. Mixed bus fleet scheduling under range and refueling constraints. Transport. Res. C Emerg. Technol. 104, 443e62.

Li, P., Zhang, Y., Zhang, Y., Zhang, K., Jiang, M., 2021. The effects of dynamic traffic conditions, route characteristics and environmental conditions on trip-based electricity consumption prediction of electric bus. Energy 218, 119437.

Liu, K., Wang, J., Yamamoto, T., Morikawa, T., 2018. Exploring the interactive effects of ambient temperature and vehicle auxiliary loads on electric vehicle energy consumption. Appl. Energy 227, 324–331.

Liu, K., Yamamoto, T., Morikawa, T., 2017. Impact of road gradient on energy consumption of electric vehicles. Transport. Res. Transport Environ. 54, 74–81.

Liu, Y., Wang, L., Zeng, Z., Bie, Y., 2022. Optimal charging plan for electric bus considering time-of-day electricity tariff. Journal of Intelligent and Connected Vehicles 5 (2), 123–137.

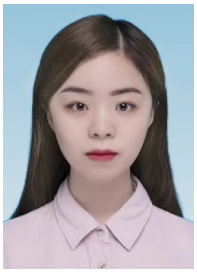
Marcacci, S., 2018. Electric buses can save local governments billions - China's showing how it's done. https://energypost.eu/electric-buses-can-save-local-u-s-governments-billions-chinas-showing-us-how -its-done/. (Accessed 5 June 2018).

Massoli, P., 2020. The road to net-zero is paved by electric buses. https://blog.greenenergyc-onsumers.org/blog/why-electric-buses-make-sense-now. (Accessed 19 May 2020).

Pamuła, T., Pamuła, W., 2020. Estimation of the energy consumption of battery electric buses for public transport networks using real-world data and deep learning. Energies 13 (9), 2340.
Perumal, S.S.G., Dollevoet, T., Huisman, D., Lusby, R.M., Larsen, J., Riis, M., 2021. Solution approaches for integrated vehicle and crew scheduling with electric buses. Comput. Oper. Res. 132, 105268.
Qi, X., Wu, G., Boriboonsomsin, K., Barth, M.J., 2018. Data-driven decomposition analysis and estimation of link-level electric vehicle energy consumption under real-world traffic conditions. Transport. Res. Transport Environ. 64, 36–52.
Qu, X., Wang, S., Niemeier, D., 2022. On the urban-rural bus transit system with passenger-freight mixed flow. Communications in Transportation Research 2, 100054.
Ritari, A., Vepsäläinen, J., Kivekäs, K., Tammi, K., Laitinen, H., 2020. Energy consumption and lifecycle cost analysis of electric city buses with multispeed gearboxes. Energies 13 (8), 2117.
Rogge, M., van der Hurk, E., Larsen, A., Sauer, D.U., 2018. Electric bus fleet size and mix problem with optimization of charging infrastructure. Appl. Energy 211, 282–295.
Stempien, J., Chan, S., 2017. Comparative study of fuel cell, battery and hybrid buses for renewable energy constrained areas. J. Power Sources 340, 347–355.
Sustainable, B.U.S., 2020. 145 electric buses are ready to start operations in Gothenburg. https://www.sustainable-bus.com/news/electric-buses-gothenburg-transdev-volvo-2020/. (Accessed 7 December 2020).
Wang, L., Yang, J., Zhang, N., Yang, Ji, Li, Y., He, J., 2017a. A spatial–temporal estimation model of residual energy for pure electric buses based on traffic performance index. Teh. Vjesn. 24 (6), 1803–1811.
Wang, J., Liu, K., Yamamoto, T., Morikawa, T., 2017b. Improving estimation accuracy for electric vehicle energy consumption considering the effects of ambient temperature. Energy Proc. 105, 2904–2909.
Wang, J., Liu, P., Hicks-Garner, J., Sherman, E., Soukiazian, S., Verbrugge, M., 2011. Cycle-life model for graphite-LiFePO4 cells. J. Power Sources 196 (8), 3942–3948.
Wu, X., Freese, D., Cabrera, A., Kitch, W.A., 2015. Electric vehicles' energy consumption measurement and estimation. Transport. Res. Transport Environ. 34, 52–67.
Xylia, M., Leduc, S., Patrizio, P., Kraxner, F., Silveira, S., 2017. Locating charging infrastructure for electric buses in Stockholm. Transport. Res. C Emerg. Technol. 78, 183e200.
Yao, E.J., Liu, T., Lu, T.W., Yang, Y., 2020. Optimization of electric vehicle scheduling with multiple vehicle types in public transport. Sustain. Cities Soc. 52, 101862.
Zhang, L., Zeng, Z.L., Qu, X.B., 2021a. On the role of battery capacity fading mechanism in the lifecycle cost of electric bus fleet. IEEE Trans. Intell. Transport. Syst. 22 (4), 2371–2380.
Zhang, J., Wang, Z., Liu, P., Zhang, Z., 2020. Energy consumption analysis and prediction of electric vehicles based on real-world driving data. Appl. Energy 275, 115408.
Zhang, W., Zhao, H., Xu, M., 2021b. Optimal operating strategy of short turning lines for the battery electric bus system. Communications in Transportation Research 1, 100023.

Jinhua Ji received her B.Eng. and M.Eng. degrees from Jilin University in 2019 and 2022, respectively. She is now a doctoral student at School of Transportation, Jilin University, China. Her research interests include public transportation operation management, and smart electric bus scheduling.

Yiming Bie received his Ph.D. degree in traffic engineering from the School of Transportation, Jilin University in 2012. He is currently a Professor at School of Transportation, Jilin University, China. His research interests include, but are not limited to advanced transit operations and adaptive traffic control.



Ziling Zeng received her B.Eng. and M.Eng. degrees from Beijing Jiaotong University, Beijing, China, in 2017 and 2019, respectively. She is currently a doctoral student at the Department of Architecture and Civil Engineering, Chalmers University of Technology. She focuses her research on smart mobility systems.

Linhong Wang received her Ph.D. degree in traffic environment and safety engineering from the School of Transportation, Jilin University in 2012. She is currently an Associate Professor at School of Transportation, Jilin University, China. Her research interest is advanced control method for electric vehicles.