

RESEARCH

Open Access



# A data-based model for driving distance estimation of battery electric logistics vehicles

Jun Bi<sup>1,2</sup>, Yongxing Wang<sup>1\*</sup> and Jiawei Zhang<sup>2</sup>

## Abstract

Battery electric logistics vehicles (BELVs) reduce transportation costs and air pollution unlike conventional logistics vehicles. However, the limited driving range of BELVs creates new problems for logistics transport. Accurate driving distance estimation of BELVs can help logistics companies determine transport strategies and alleviate the range anxiety of drivers. Based on mass data from BELVs operating in Beijing, China, this study uses a practical and effective data-based modeling method, regression analysis, to establish the data-based model of driving distance estimation. During the modeling process, a nonlinear relation between percentage of energy consumption per kilometer and driving speed is explored based on the experimental data. After determining the model variables, the model frame of driving distance in consideration of driving speed and state of charge is established. The forgetting factor recursive least-squares algorithm is applied to estimate the parameter values of the model. Verification results confirm the feasibility of the model and show that the model errors are small. The proposed model is also used to explore the economical driving speed of BELVs.

**Keywords:** Battery electric logistics vehicle, Data-based model, Driving distance estimation, Economical driving speed

## 1 Introduction

Dependence on petroleum has resulted in serious environmental and energy problems. In the field of transportation, battery electric vehicles (BEVs) are utilized to reduce pollution because they do not produce tailpipe emissions during operation. Different from conventional vehicles that use fossil fuel, BEVs convert chemical energy entirely to electricity stored in rechargeable battery packs. Therefore, BEVs are better than internal combustion engine vehicles in terms of greenhouse gas emissions and energy consumption [1]. For logistics transport, battery electric logistics vehicles (BELVs) can reduce transportation costs and air pollution. However, the driving range of BELVs is shorter than that of conventional logistics vehicles. Driving range refers to the distance that a fully charged BEV can traverse until the battery runs out of usable electricity [2]. During trips, drivers tend to exhibit range anxiety, that is, the fear of depleting battery

energy en route [3]. Drivers experiencing range anxiety feel uncomfortable during trips when the state of charge (SOC) of their BELVs is too low. SOC refers to the ratio of the remaining capacity to the nominal one [4]. Drivers can directly obtain the SOC by reading the dashboard in the vehicles. Considering the limited driving range and range anxiety, when starting transport tasks, BELV drivers desire to know the distance that a BELV can continually run from a fully charged state to a certain residual SOC value in advance because the logistics transport need to make strict plans to ensure that all the transport tasks can be fulfilled. This distance is defined as driving distance in this study. For example, if a BELV begins operating with SOC equal to 100% and stops operating with SOC equal to 20%, the driving distance is the distance that the BELV runs during this discharge process. Therefore, accurate driving distance estimation for BELVs is very important to determine transport schemes and other activities for the logistics transport. However, in practice, accurately estimating driving distance is difficult because the driving behavior of drivers and other external factors would affect the vehicle operating, besides the performance of the

\* Correspondence: [yx\\_wang@bjtu.edu.cn](mailto:yx_wang@bjtu.edu.cn)

<sup>1</sup>School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Full list of author information is available at the end of the article

vehicle itself. With the recent development of data collection techniques, a large amount of data of vehicle state and battery status of BELVs can be obtained. Besides the visible operating status, many invisible influencing factors, deriving from driving behavior and other external factors, are hidden in the data. Establishing a model by using the data is an effective method to realize the accurate driving distance estimation.

Besides BELV drivers, drivers of other types of BEVs also experience the range anxiety very often. Estimating driving range is an important issue for BEV drivers, which has similarity with the driving distance estimation. The conventional methods for solving such issues are to establish the corresponding estimation models by mechanism. The mechanism modeling methods generally adopt one or several visible vehicle states to reflect the driving range based on the mechanism analysis of the vehicle operation. Battery SOC methods are one of the most common methods to estimate the driving range of BEVs. The methods focus on determining driving range through obtaining the estimation of residual usable energy of BEVs. There are several studies using battery SOC methods to realize range estimation [5–15]. This factor, while important, is insufficient for driving range or distance estimation. It is because that the energy of a BEV can be used in many different ways depending on the driver's driving behavior. Additionally, the external factors, such as road conditions, may also potentially affect the energy consumption. However, these potential affecting factors cannot be considered in these models by mechanism since it is hard to estimate driving behavior and other external factors through mechanism analysis.

BEV operation is a complex process, which involves driving conditions and human factors besides battery status. Based on the mechanism modeling methods, reference [16] attempted to use published vehicle parameters and driving range information as a basis for establishing a simplified BEV power train model. The results of driving range estimation were compared with published manufacturer specifications under various route and driving conditions. The significant correlations among the experimental results were also examined. Reference [17] analyzed the operational mechanism regarding the complexity and nonlinearity of the BEV system. A fuzzy transform method is adopted to estimate the driving range. Reference [18] investigated the operation characteristics of inductive power transfer systems in BEV and regarded it as an important affecting factor for the driving range. A corresponding mechanism model was established to estimate the driving range. Reference [19] attempted to estimate the driving range based on energy consumption information and driving conditions. A model of driving range was established through analysis of the operational mechanism. However, the human factors, namely driving

behavior, cannot be involved in this method. Moreover, considering topography and the traffic conditions of the road network, reference [20] proposed several linear models to determine energy expenditure equations under different conditions. Dijkstra's graph search algorithm was introduced to identify the route with minimum energy consumption. The result was a routing system that could extend the driving range by calculating the minimum energy route to a destination. Once again, the human factors are not reflected in the models that are based on the mechanism analysis. When driving a vehicle, the driving behavior of the driver is not a certain factor, because it can be affected by many external and uncertain factors. Due to the human factors, the operation of BEV is a typical complex system. It is very difficult, even impossible, to predict the driving behavior through a deterministic method, such as the mechanism analysis. Therefore, the complexity of the BEV operation cannot be presented in the models as mentioned above.

The shortcoming of mechanism modeling methods can be overcome by the data-based modeling methods, which refer to a method that mines valuable information from the data of controlled systems and uses the information to formulate the relations among different variables [21]. The data-based models are more effective and practical than the mechanism models, because the large amount of data actually records the vehicle state and battery status during vehicle operating. More importantly, the actual data hide the invisible factors for BEV operating, such as driving behavior and other external factors. Although these invisible factors cannot be directly captured, they are included in the actual operational processes of vehicles. With an effective data-based modeling method, these invisible factors can be reflected in the corresponding model. Therefore, applying actual data in driving range or distance estimation is a new research focus concerning researchers and engineers. Reference [22] used the GPS travel survey data to explore the real-world driving cycles and driving range of BEVs in Beijing, China. The aim of the study is to investigate the effect of driving patterns on energy consumption and, consequently, the driving range. Reference [23] proposed a data-based modeling method to improve the accuracy of driving range estimation. The BEV operation data used in the modeling was collected from a cycle-life test and a cloud system. The growing hierarchical self-organizing map was adopted to investigate the energy consumption presented in the collected data, and then, the driving range can be estimated. Meanwhile, reference [24] adopted the BEV operation data to explore the effect of high driving speed on energy consumption and corresponding driving range. The data are collected in the different driving scenarios on the highway in Australia, and the driving range under different driving scenarios was analyzed in the study.

Reference [25] used empirical data to propose a hybrid model for BEV range estimation. The results indicated that the data-based model can significantly improve the estimation accuracy as compared to the state-of-the-art model-based range estimation. Reference [26] harnessed a lot of data from real-world driving to analyze the driving range of BEVs and provided insight into the factors that affect energy consumption. Reference [27] adopted the actual data from operational BEVs in Beijing to realize the residual range estimation for BEVs by combining the radial basis function neural network. However, these data-based methods for driving range estimation are mainly focused on private BEVs. In general, the driving behavior of drivers of BELVs is different from private BEVs, because the two types of the drivers have the different purposes of usage for the BEVs. For example, BELVs often operate more regularly following strict travel schemes as compared to the private BEVs. Therefore, the driving behavior factor hidden in the data-based model for private BEVs is not suitable for BELVs. Moreover, for BELVs, driving distance estimation would be more useful than driving range estimation. It is because that, before a transport task, a BELV driver needs to determine a travel scheme which often needs to accurately estimate the driving distance under a certain energy consumption level besides the range that a BELV can run under its current energy.

In recent years, an increasing number of logistics companies adopt BELVs to fulfill transport tasks in cities because of the energy efficiency and environmental protection. When BELVs operate in cities, a large amount of valuable data about the battery status and vehicle state are generated in real time. The information of the data usually includes SOC, driving speed, voltage, current, and driving distance. Consequently, the human factors and other external factors affecting the vehicles operating would be also hidden in the data and indirectly reflected by the information of data. Through advanced collection techniques, a large amount of data can be collected. Combining data modeling methods, the actual data can be applied to establish an effective model to estimate driving distance. In this study, we collected the data of battery status and vehicle state from 70 BELVs operating in Beijing, China. After data processing, a practical and effective data-based modeling method, regression analysis, was adopted to establish a nonlinear model for driving distance estimation in consideration of SOC and driving speed. The regression analysis is the analytic method that focuses on mining the characteristics of actual data and establishing formulation models. However, the simulation methods often explore the characteristics by using the model to simulate many possible values and then generate various statistics. Thus, the analytic methods are not applied along with the simulation ones. The analytic methods can increase the efficiency of the modeling

process and decrease the computer cost as compared to the simulation methods. Moreover, to improve the energy efficiency and reduce transport cost, the economical driving speed that contributes to the minimum energy consumption was explored by analyzing the data-based model.

The model proposed in this study may be used by logistics companies to determine transport strategies for cities or by individual drivers to estimate driving distance effectively and alleviate range anxiety.

The contributions of this study are as follows. Firstly, mass data of BELVs operating in Beijing, China, was collected, which actually record the daily operation of the vehicles. The information regarding vehicle operating state and battery status, under the regular driving behavior and conditions, is included in the data. The collected data have universal significance for the BELVs in Beijing, China, and other similar areas. Secondly, a data-based modeling method was adopted to establish the model for driving distance estimation. During the modeling process, the relation between percentage of energy consumption per kilometer (POECPK) and driving speed was explored based on the experimental data. The results indicate a novel finding about the relation between POECPK and driving speed for BELVs. Finally, the proposed model was applied to explore economical driving speed for BELVs, which can run the longest driving distance from a fully charged state to a certain residual SOC value. To the best knowledge of the authors, it is the first time that the mass data of BELVs, recording the actual operating process of vehicles and batteries, are applied to estimate the driving distance and explore the economical driving speed.

The rest of the paper is organized as follows. Section 2 describes the methods in this study. Section 3 describes the data sources and presents the relation analysis based on the experimental data. Section 4 presents the model for driving distance estimation considering SOC and driving speed. Actual data are utilized to verify the model. Moreover, the model is used to explore economical driving speed. Section 5 presents the results and discussion regarding this work. Section 6 presents the conclusions and directions for future research.

## 2 Methods

In this study, the actual data collected from 70 BELVs are applied to explore the driving distance estimation of BELVs. The information on the data includes both the battery status and vehicle operating state. After data processing, a correlation analysis is conducted to explore the relations between driving distance and other factors in the data. Based on the data analysis results, the regression analysis methods are used to establish a nonlinear model for driving distance estimation of BELVs. The

parameters of the model are determined by using the actual data. The economical driving speed is explored based on the model.

### 3 Data analysis

#### 3.1 Data collection and processing

The data used in this study are collected from the BELVs that are widely used in Beijing. The BELVs are produced by BAIC Motor Corporation, Ltd. Their vehicle type is BJ5020XXYV3R-BEV.

They are mainly used for express services in the city. Moreover, in actual operations, the loading weight of the BELVs is around 500 kg at each time. The nominally capacity of the batteries is 24 kWh, which is a common capacity level for BELV batteries in recent years. The BELVs perform their daily transport tasks following the travel schemes and operate in the road network as regular vehicles. During their operating, the data of battery status and vehicle state are obtained online through an internal controller area network bus. The data, packed as groups, are transmitted by a wireless transmission network to the vehicle service center every 5- or 10-s interval and stored in the database. To investigate the driving distance problem, we collected the data, generated from March 27, 2014, to April 30, 2016, which include 242,858 groups of original data. The data actually record the 254 discharge processes of 70 BELVs operating in Beijing. The information on the collected data includes time, total voltage, total current, SOC, temperature, cell voltage, driving distance, and driving speed.

However, during data collection, data errors and data missing would occur occasionally because the wireless transmission network is affected by the external environment during data transmission. Therefore, the original data need to be processed before being used in experiments. Deletion and interpolation processing are adopted to deal with data errors and data missing, respectively.

The repetitive and false data are the problems for the data errors. Deletion processing is a simple and effective method to solve the problems, which can search and delete repetitive and false data included in the original data. For the data missing problem, an effective method to address the problem is interpolation processing. The interpolation processing can be used to compensate for the missing data from the original data. During the interpolation processing, an effective interpolation method, Lagrange interpolation, is adopted to implement the interpolation experiments because of its ideal tradeoff between accuracy and computational cost [28]. After the deletion and interpolation processing, the complete data on each discharge process are obtained. Notably, the complete data are still unsuitable for modeling because the undesired noise may be included in the data, which is resulting from the inconsistent measurement accuracy for

different variables of the data. To remove the undesired noise in the data, a smooth processing method is used to average the variable values based on the unit change in driving distance of the data. After the smooth processing, the data can be used as the experimental data to model the driving distance estimation of BELVs.

#### 3.2 Related definitions and equations

The related equations and definitions pertaining to the battery are introduced in the following paragraphs prior to analyzing the relation between driving distance and its related factors.

SOC is the key factor of the battery. Ampere hour counting is the most commonly used method for SOC determination [29]. It was adopted to define SOC in this study. Only the SOC of the discharge process was discussed because the research purpose is to estimate driving distance rather than charge. SOC is defined as follows:

$$\text{SOC}_t = \frac{Q_m - Q_t}{Q_m} \quad (1)$$

$$Q_t = \int_{t_0}^t \eta I dt \quad (2)$$

where  $\text{SOC}_t$  is the value of SOC at time  $t$ ,  $Q_m$  refers to the nominal capacity in ampere hours,  $Q_t$  is the energy consumption from initial time  $t_0$  to time  $t$ , in ampere hours,  $\eta$  represents the Coulomb coefficient, and  $I$  is the instantaneous current in amperes.

The definition of energy consumption per kilometer is proposed to link energy consumption with driving distance, as shown in Eq. (3).

$$q_i = \frac{Q_{k+i} - Q_k}{L_{k+i} - L_k} \quad (3)$$

where  $q_i$  is the energy consumption per kilometer in ampere hours per kilometer;  $Q_k$  and  $Q_{k+i}$  represent the energy consumption from initial time to time  $k$  and  $k+i$ , respectively, in ampere hours; and  $L_k$  and  $L_{k+i}$  are the driving distance from initial time to time  $k$  and  $k+i$ , respectively, in kilometers.

The relation between energy consumption per kilometer and SOC was obtained by combining Eqs. (1) and (3). The derived equation is

$$q_i = \frac{Q_{k+i} - Q_k}{L_{k+i} - L_k} = Q_m \frac{\frac{Q_m - Q_k}{Q_m} - \frac{Q_m - Q_{k+i}}{Q_m}}{L_{k+i} - L_k} = Q_m \frac{\text{SOC}_k - \text{SOC}_{k+i}}{L_{k+i} - L_k} \quad (4)$$

where  $\text{SOC}_k$  and  $\text{SOC}_{k+i}$  represent the SOC at time  $k$  and  $k+i$ , respectively.

The ratio of energy consumption per kilometer  $q_i$  to nominal capacity  $Q_m$  must be obtained. The result is

defined as POECPK in this study. It also links SOC with driving distance, as shown in Eq. (5).

$$s_i = \frac{q_i}{Q_m} = \frac{SOC_k - SOC_{k+i}}{L_{k+i} - L_k} \tag{5}$$

where  $s_i$  refers to POECPK in percent per kilometer.

### 3.3 Correlation analysis

A correlation analysis was conducted to explore the relations between driving distance and other factors in the experimental data, as well as determine the key factors as the variables of the model. Correlation analysis is an important method of regression analysis, which is a data-based modeling method and aims to explore the correlativity between different variables in the data. Its analysis results are expressed by Pearson correlation coefficients. A Pearson correlation coefficient, ranging from  $-1$  to  $1$ , is a statistic that measures the degree of linear correlation between two variables [30]. Furthermore, adjoint probability  $P$  was adopted to test the significance of the correlation. Given that the value of adjoint probability  $P$  is lower than or equal to the designated significance level (equal to  $0.05$  generally), significance is guaranteed.

Experimental data were used for the correlation analysis. The variables in the data included SOC, driving speed, total voltage, total current, cell voltage, temperature, and driving distance. Table 1 shows the Pearson correlation coefficients and adjoint probability  $P$  between driving distance and other variables.

The Pearson correlation coefficient between driving distance and SOC is approximately  $-1$ , which indicates that a very strong negative linear relation exists between the two variables. The adjoint probability  $P$  guarantees the significance of the result. The Pearson correlation coefficients between driving distance and other variables are not close to  $1$  or  $-1$ , which indicates that the linear relation between driving distance and the other variables is not significant. Experimental data on the representative discharge processes were used to further explore the relation between SOC and driving distance, as shown in Fig. 1. As expected, driving distance presents a linear increase as SOC decreases, which also indicates a significant negative linear relation between driving distance and SOC. Moreover, it can be observed that the fluctuations within a narrow range exist in the curve. These fluctuations reflect the specific driving behavior of

drivers, such as acceleration and brake. The driving behavior cannot be captured, however, which can be indirectly reflected by other factors in collected data, as shown in the figure.

### 3.4 Effect of driving speed on POECPK

For conventional logistics vehicles, driving distance can be determined by driving speed because of the unlimited driving range. However, for BELVs, SOC needs to be considered because the driving range is limited. Therefore, given that driving distance is subject to SOC, driving speed cannot be directly used to determine driving distance. Furthermore, several studies have demonstrated that energy consumption can be influenced by driving speed [31, 32]. Considering the effect of SOC on driving distance, driving speed affects driving distance by influencing the energy consumption. Experimental data on 65 discharge processes, in which the changes in driving speed are comparatively small, were used to explore the relation between driving speed and energy consumption. The average driving speed in each discharge process was considered as driving speed of the samples, and the POECPK in each discharge process was obtained with Eq. (5). A scatter diagram of POECPK and driving speed is shown in Fig. 2.

It can be observed that the POECPK changes as driving speed changes, and a nonlinear relation exists between POECPK and driving speed. Moreover, it is noted that the driving speed, in the experimental data, ranges from  $0$  to  $90$  km/h because the BELVs operated as regular vehicles in the city and the speed limit is generally not faster than  $90$  km/h in the urban road network.

## 4 Driving distance estimation model

### 4.1 Model frame

According to the results of the correlation analysis, a monadic linear model was developed to represent the strong negative linear relation between driving distance and SOC, as shown in Eqs. (6) and (7).

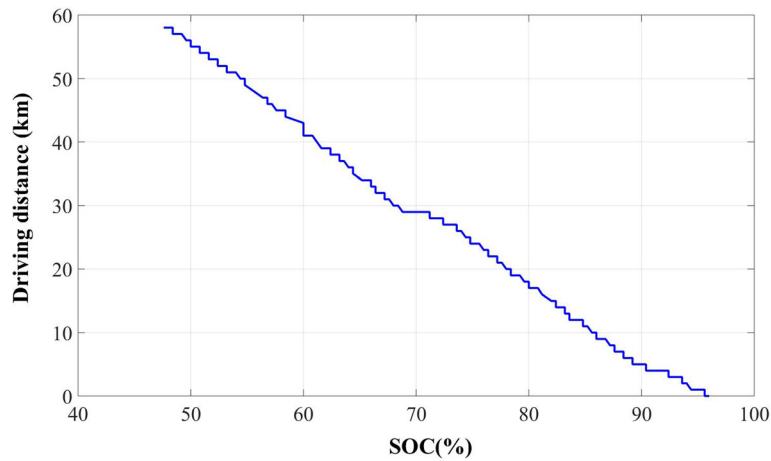
$$y = kx + b \tag{6}$$

$$x = 100SOC_t \tag{7}$$

where  $SOC_t$  refers to SOC at time  $t$  (it can be directly observed in the vehicle),  $y$  is the driving distance that the BELV can continually run from SOC equal to  $100\%$  to  $SOC_t$  in kilometers, and  $k$  and  $b$  are the undetermined parameters.

**Table 1** Results of correlation analysis

		SOC (%)	Driving speed (km/h)	Total voltage (V)	Total current (A)	Cell voltage (V)	Temperature (°C)
Driving distance (km)	Pearson	-0.976	0.274	-0.095	-0.071	-0.055	0.041
	$P$	0.000	0.000	0.041	0.130	0.239	0.377



**Fig. 1** Relation between driving distance and SOC

Equation (6) can be utilized to deduce parameter  $k$ . We suppose that  $x_1 = 100SOC_k$  and  $x_2 = 100SOC_{k+i}$  are the input variables  $x$  and the corresponding output variables  $y$  are  $L_k$  and  $L_{k+i}$  respectively, as shown in Eqs. (8) and (9).

$$L_k = kx_1 + b \tag{8}$$

$$L_{k+i} = kx_2 + b \tag{9}$$

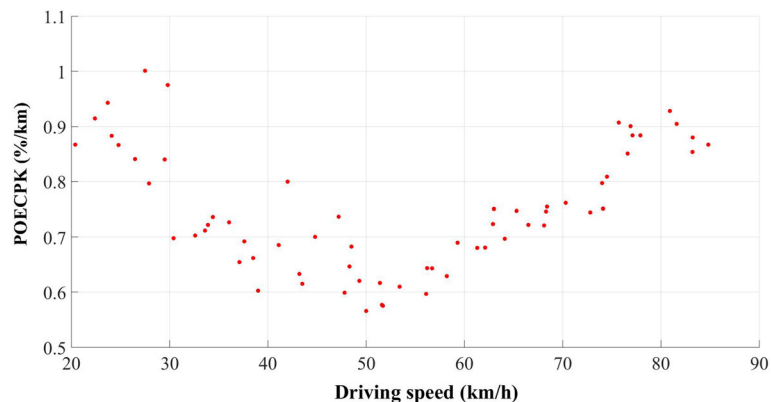
Equations (8) and (9) can be used to deduce parameter  $k$ , as shown in Eq. (10).

$$k = \frac{L_k - L_{k+i}}{x_1 - x_2} = \frac{L_k - L_{k+i}}{100(SOC_k - SOC_{k+i})} \tag{10}$$

Parameter  $k$  links SOC with driving distance. Therefore, the transition relation between parameter  $k$  and POECPK can be deduced by combining Eqs. (5) and (10) as follows:

$$k = -\frac{100}{s_i} \tag{11}$$

Given that driving speed affects POECPK, as shown in Fig. 2, parameter  $k$  is also influenced by driving speed, which is very important for driving distance estimation. In order to formulate the relation between parameter  $k$  and driving speed, based on the experimental data presented in Fig. 2, the other method of regression analysis, curve fitting method, was employed to explore this relation and further model the driving distance in consideration of SOC and driving speed. Curve fitting method is a practical data-based modeling method to formulate the relations among different variables in data. Firstly, POECPK was transformed into parameter  $k$  by using Eq. (11). Secondly, as shown in Fig. 2, it is observed that the nonlinear relation between POECPK and driving speed may be subject to a quadratic curve, which can be used as the objective for the curve fitting method. Lastly, in consideration of feasibility and effectiveness, one of the



**Fig. 2** POECPK distribution at different driving speeds

efficient curve fitting methods, curve linearization method, was adopted to implement the fitting operation for quadratic curve fitting. Curve linearization is an important curve fitting method that transforms curve fitting into linear fitting through variable substitution [33]. Figure 3 presents the fitting curve of parameter  $k$  and driving speed.

In order to test the fitting effect, several test indexes, including determinate coefficient  $R^2$  of the goodness-of-fit test,  $F$  statistic of variance analysis, and adjoint probability  $P$ , were applied in curve fitting.  $R^2$  refers to the ratio of the sum of squares for regression (SSR) to the sum of squares for error (SST), which ranges from 0 to 1 and its value increases as fitting effect improves, as shown in Eq. (12).

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{12}$$

where  $\hat{y}_i$  refers to the regression values,  $\bar{y}$  is the average value, and  $y_i$  refers to the actual observed values.

$F$  statistic refers to the ratio of the variation resulting from input variables to the random error, which is subject to the  $F$  distribution, as shown in Eq. (13).

$$F = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\left[ \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right] / (n-2)} \tag{13}$$

Adjoint probability  $P$  was adopted to test the significance of the  $F$  statistic. The test results of the fitting effect are listed in Table 2.

The fitting effect results indicate that quadratic curve fitting is suitable for use as the model of driving speed and parameter  $k$ , which is given in Eq. (14).

$$k = av^2 + cv + d \tag{14}$$

where  $v$  refers to driving speed in km/h and  $a$ ,  $c$ , and  $d$  are the parameters.

As shown in Eq. (6), when  $x$  is equal to 100,  $y$  is equal to 0. Thus, parameter  $b$  is equal to  $-100k$ . The model of driving distance and SOC can be deduced as Eq. (15).

$$y = kx - 100k \tag{15}$$

By combining Eqs. (7), (14), and (15), the model frame of driving distance in consideration of SOC and driving speed was obtained, as shown in Eqs. (7) and (16).

$$y = (av^2 + cv + d)x - 100(av^2 + cv + d) = k_1xv^2 + k_2v^2 + k_3xv + k_4x + k_5v + k_6 \tag{16}$$

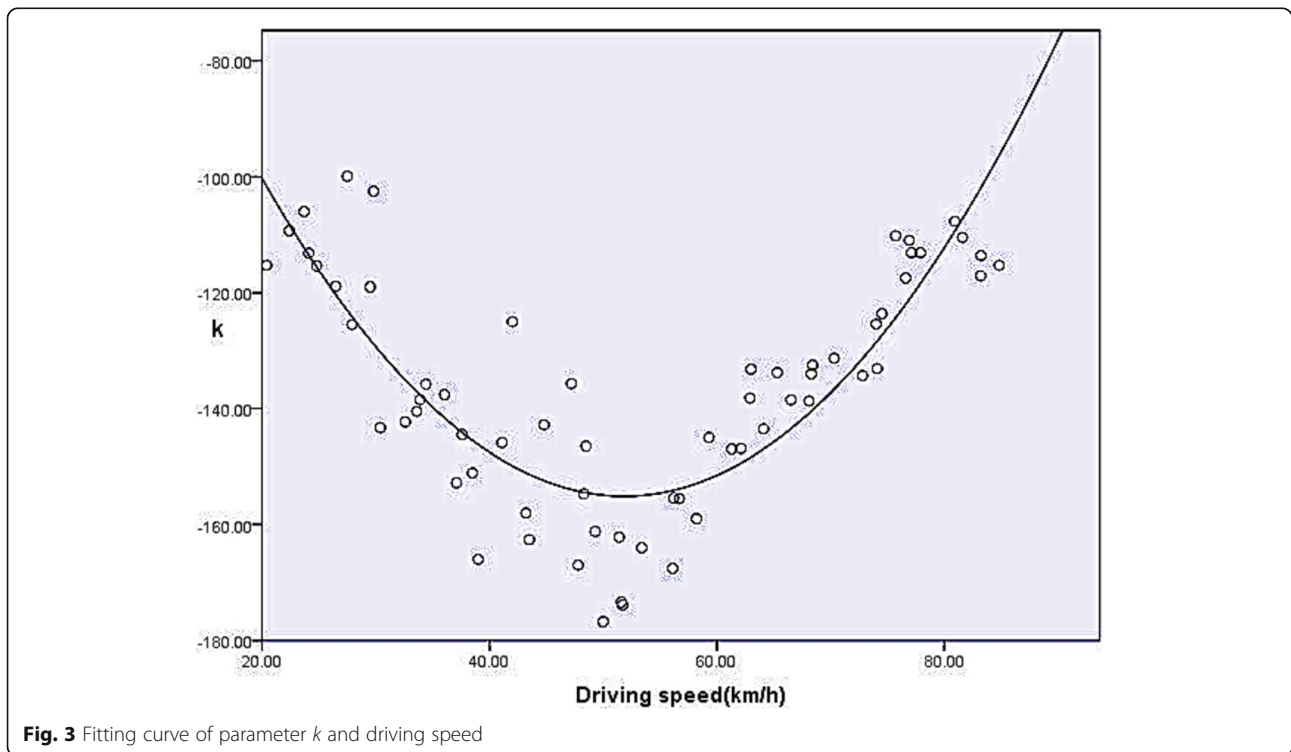


Fig. 3 Fitting curve of parameter  $k$  and driving speed

**Table 2** Test results of the fitting effect

	$R^2$	$F$	$P$
Quadratic curve	0.717	78.502	0.000

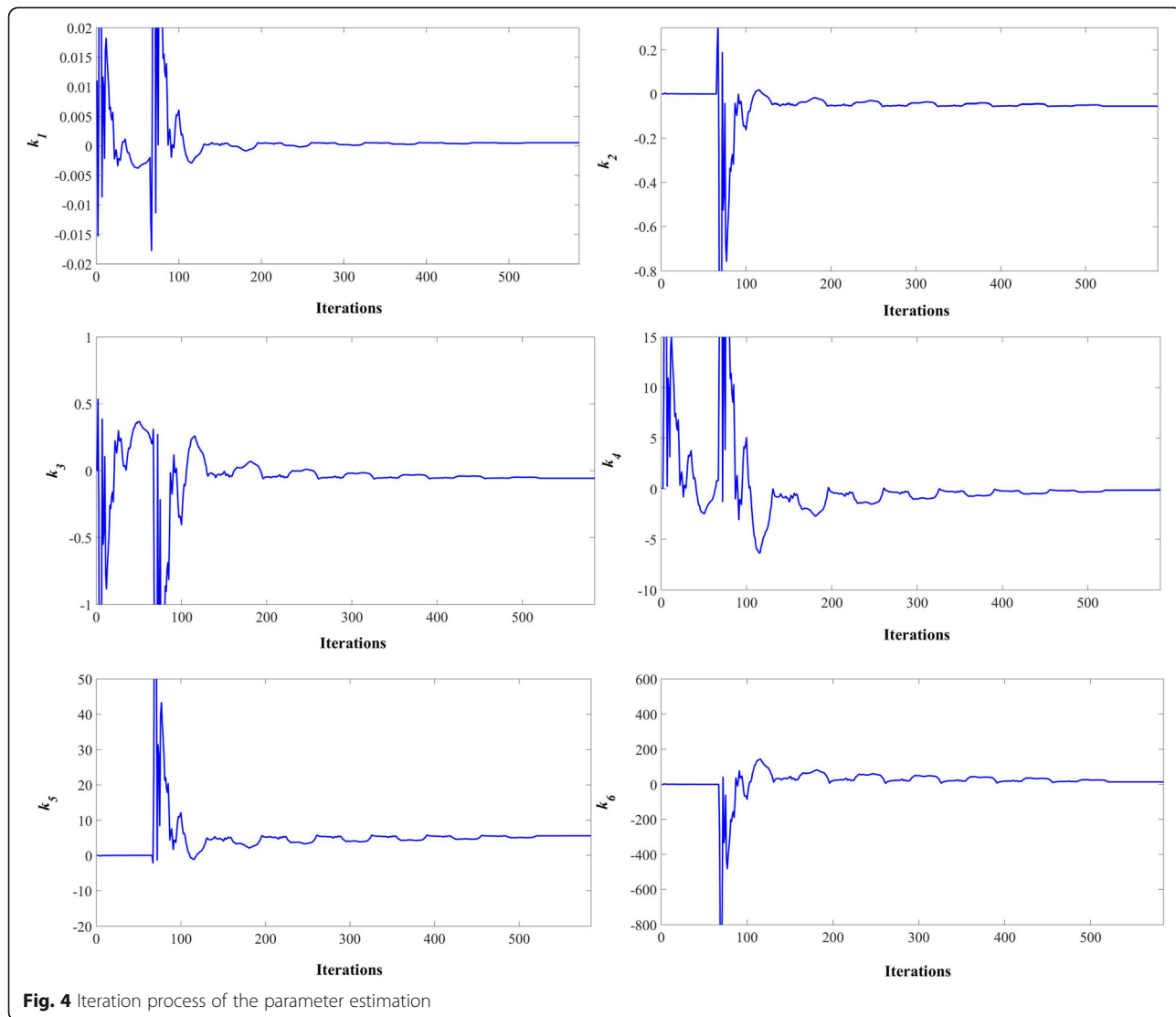
where  $k_1, k_2, k_3, k_4, k_5,$  and  $k_6$  are the undetermined parameters.

**4.2 Estimation of the model parameters**

After modeling the driving distance estimation as shown in Eqs. (7) and (16), it is indispensable to estimate the parameters of the model. Based on the variables of the model, data on the three variables (SOC<sub>t</sub>, driving speed, and driving distance) are required to realize the parameter estimation. We selected the experimental data on 65 discharge processes, which were applied to implement the fitting operation of parameter  $k$  and driving speed, as shown in Fig. 3, to estimate the parameters.

The driving speed can be directly obtained through the data. The corresponding parameter  $k$  can be used to determine the driving distance through Eq. (6) and fixed SOC<sub>t</sub>. According to Eq. (6), when  $x$  is equal to 100, the driving distance  $y$  is equal to 0. Therefore, the parameter  $b$  is equal to  $-100 k$ . The SOC<sub>t</sub> is set as 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100%. Based on the values of parameter  $k$  in the transformed experimental data on 65 discharge processes, fixed SOC<sub>t</sub>, Eqs. (6) and (7), the 585 groups of data on 65 discharge processes are obtained, which were used to estimate the parameters of the proposed model.

Considering the time variation of parameters, caused by human factors and other external factors during vehicle operating, the forgetting factor recursive least-squares (FFRLS) algorithm was adopted to estimate the parameters based on the 585 groups of data. The advantage of FFRLS algorithm is that it can effectively address the



**Fig. 4** Iteration process of the parameter estimation



problems of parameter changes compared with other parameter identification methods [34]. The iteration processes of estimation of the parameters, based on the FFRLS algorithm, are illustrated in Fig. 4.

It can be observed that the parameter estimation curves of  $k_1$ – $k_6$  presented relatively flat trends over the iterations. The curves converged completely after 400 iterations, and the parameter values were obtained (shown in Table 3).

Subsequently, the nonlinear model for driving distance estimation in consideration of SOC and driving speed was obtained, as shown in Eqs. (7) and (17).

$$y = 0.000542xv^2 - 0.0542v^2 - 0.0556xv - 0.1399x + 5.5568v + 13.9854 \tag{17}$$

It is noted that when  $v$  is equal to 0 km/h and  $SOC_t$  is equal to 100%, driving distance  $y$  is equal to -0.0046 km, i.e., approximately 0 km. This result is acceptable. Moreover, when  $v$  is equal to 0 km/h and  $SOC_t$  is equal to 0, driving distance  $y$  is equal to 13.9854 km. The result is acceptable and reasonable because that, in actual operation of BELVs, driving speed and  $SOC_t$  do not simultaneously equal to 0 and the vehicles must be recharged before its energy being depleted. Consequently, the allowable application conditions of the driving distance estimation model are as follows. Firstly, driving speed ranges from 0 to 90 km/h, which can meet the requirements of BELVs in urban road traffic operation. Secondly,  $SOC_t$  and driving speed are not equal to 0 simultaneously, which is in line with the actual operation scenario. The corresponding orthographic views were plotted based on Eqs. (7) and (17), as shown in Fig. 5.

It is observed that, in the front view, when  $SOC_t$  is fixed, driving distance initially increases and then decreases as driving speed increases. The change trend conforms to the relationship between POECPK and driving speed as shown in Fig. 2. As POECPK increases, the driving distance decreases. The left view presents the linear relation between  $SOC_t$  and driving distance when driving speed is fixed. This change trend conforms to the relationship between SOC and driving distance as shown in Fig. 1. In the vertical view, there presents the significant symmetry in the figure. The symmetry results from the impacts of driving speed on the energy consumption. Furthermore, the graphic view presents that both  $SOC_t$  and driving speed have significant impacts on driving distance. Specially, driving distance has significant symmetry regarding the driving speed, and driving speed equaling to approximately 50 km/h is the

symmetrical plane, which indicates that the BELVs can reach their maximum driving distance as the driving speed equals to approximately 50 km/h.

### 4.3 Model verification

The collected data on the other BELVs and discharge processes that were not used for modeling were adopted as experimental data to verify the proposed model. However, the variables of the data cannot be directly applied to estimate the driving distance through the model, due to the constantly changing speed in the actual operation of BELVs. The current driving speed can only represent the current driving condition, not the dynamic driving conditions. To solve the problem, a piecewise driving distance and superposition method is applied to estimate the driving distance. The continuous discharge process is divided into  $n$  samples according to the changes in driving speed. The flow of the method for driving distance estimation is shown as Fig. 6.

Through data processing, data in the continuous discharge processes were used for the verification experiment. Based on the changes in driving speed, 58 samples were obtained from the discharge processes. The values of  $SOC_t$  and driving speed in each sample were inputted to the model to calculate the estimation values of driving distance. The estimation results were compared with the actual driving distance in the data. Root-mean-square error (RMSE) and root-mean-square relative error (RMSRE) were used as performance indexes to analyze the accuracy of the model scientifically [35]. RMSE and RMSRE are given in Eqs. (18) and (19).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \tag{18}$$

$$RMSRE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{\hat{y}_i - y_i}{y_i} \right)^2} \tag{19}$$

where  $\hat{y}_i$  is the estimation value and  $y_i$  is the actual value.

Figure 7 presents the actual and estimation values of driving distance. It is observed that, as the sample number increases, the driving distance increases for both the actual and estimation driving distance values. Each sample represents the operating state under the specific driving speed during the discharge processes. In the figure, the curves of actual and estimated driving distance have the similar change trends. The results indicate that the estimated driving distance obtained from the model is close to the actual one in the data.

Moreover, RMSE is equal to 0.5986, which is lower than 1. RMSRE is equal to 0.00007, which is lower than 0.001. These results indicate that the accuracy of the

**Table 3** Final estimated parameter values

Parameters	$k_1$	$k_2$	$k_3$	$k_4$	$k_5$	$k_6$
Values	0.000542	-0.0542	-0.0556	-0.1399	5.5568	13.9854

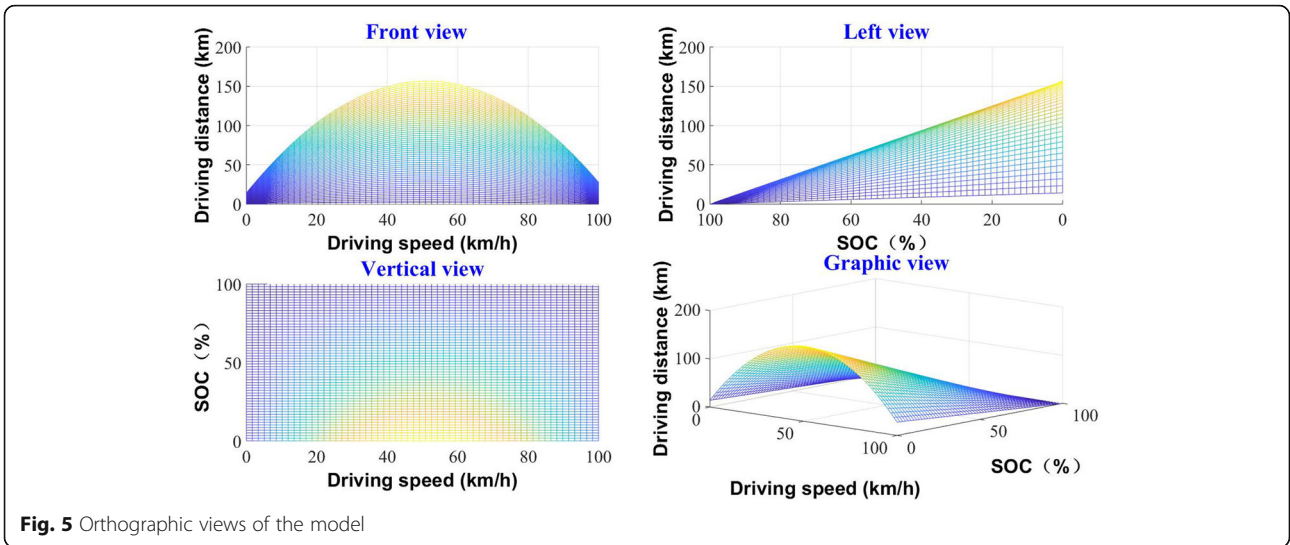


Fig. 5 Orthographic views of the model

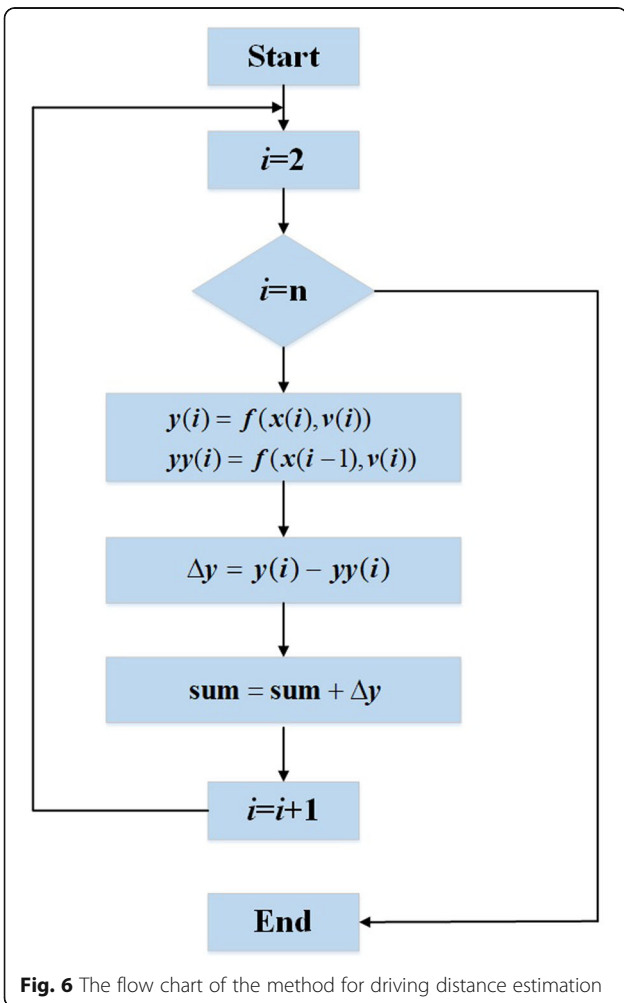


Fig. 6 The flow chart of the method for driving distance estimation

model is high. The model has an acceptable estimation effect. Notably, the driving distance estimation model is calibrated based on the experimental data from the overall BELVs operating at different times. Since the different vehicles and times may have different driving behavior and external environment, the experimental data can reflect the differences of the driving distances under different conditions. The calibrated model represents the comprehensive consideration for the operation of the overall BELVs. Therefore, although the driving distance of an individual BELV may have differences with that obtained from the model calibrated by the overall BELVs, the calibrated model can significantly reflect the operation and driving distances of individual BELVs.

#### 4.4 Economical driving speed

Economical driving speed refers to the driving speed that causes driving distance to reach its maximum value, subject to the SOC. The model was used to analyze the relations between driving distance and driving speed under different SOC<sub>t</sub> values to obtain the economical speed. Moreover, considering the range anxiety of BELV drivers, the BELV was recharged before its battery was depleted. Combining the information of experimental data, we set the lower limit of SOC<sub>t</sub> as 20%, which is a reasonable value for actual operation of BELVs in the city. Figure 8 presents the relations between driving distance and driving speed under different SOC<sub>t</sub> values.

It is observed that the curves under different SOC<sub>t</sub> values present almost the same change trends. Driving distance initially increases and then decreases as driving speed increases. The crest values represent the maximum driving distance under economical driving speed. Furthermore, all the curves have similar change trends and reach the crest value when driving speed is equal to about 50 km/h. For example, when setting SOC<sub>t</sub> as 20%

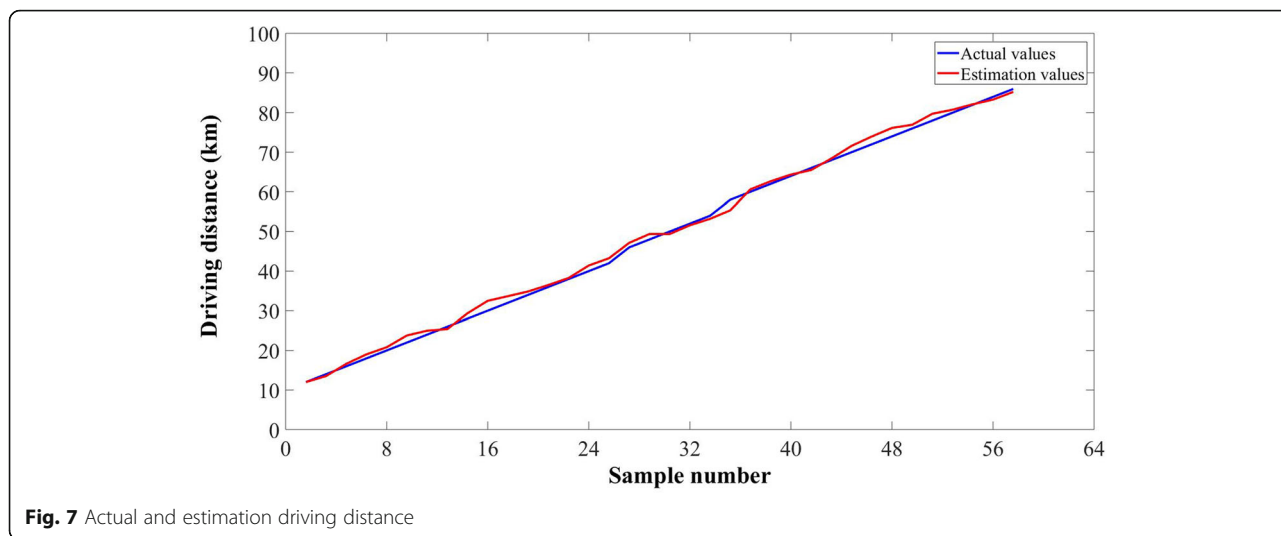


Fig. 7 Actual and estimation driving distance

and inputting it into the model, the corresponding crest value and economical driving speed can be obtained by Eq. (20).

$$y = -0.0434v^2 + 4.4448v + 13.9854 \tag{20}$$

Through calculation, when  $SOC_t$  equaling to 20%, the crest value, namely maximum driving distance, is 127.7887 km and corresponding economical driving speed is 51.2074 km/h. Similarly, when  $SOC_t$  is set as 40%, 60%, and 80%, the economical driving speeds are 51.2423, 51.2177, and 51.1439 km/h and the driving distances are 93.7796, 62.4635, and 31.1476 km, respectively. It is noted that the values of economical driving speed under different  $SOC_t$  values are very close. Therefore, the average value of the economical driving speed, equaling to 51.2028 km/h, can be considered as the overall economical driving speed.

### 5 Results and discussion

Based on the correlation analysis of actual data, it is observed that SOC and driving speed have significant impacts on driving distance of BELVs. A nonlinear relation between POECPK and driving speed was explored based on the data. Furthermore, a data-based model was established to estimate driving distance of BELVs in consideration of SOC and driving speed. The verification results indicate that the model has the good estimation accuracy. By using the model, the economical driving speed of BELVs is explored, which indicates that the overall economical driving speed is 51.2028 km/h. In addition to battery status and vehicle operating state, the loading weight and ambient temperature may also have influence on driving distance. In the future work, the driving distance model can be further improved by considering the impacts of loading weight and ambient temperature on driving distance of BELVs. The other factors regarding battery status, such as battery state of health (SOH) [36],

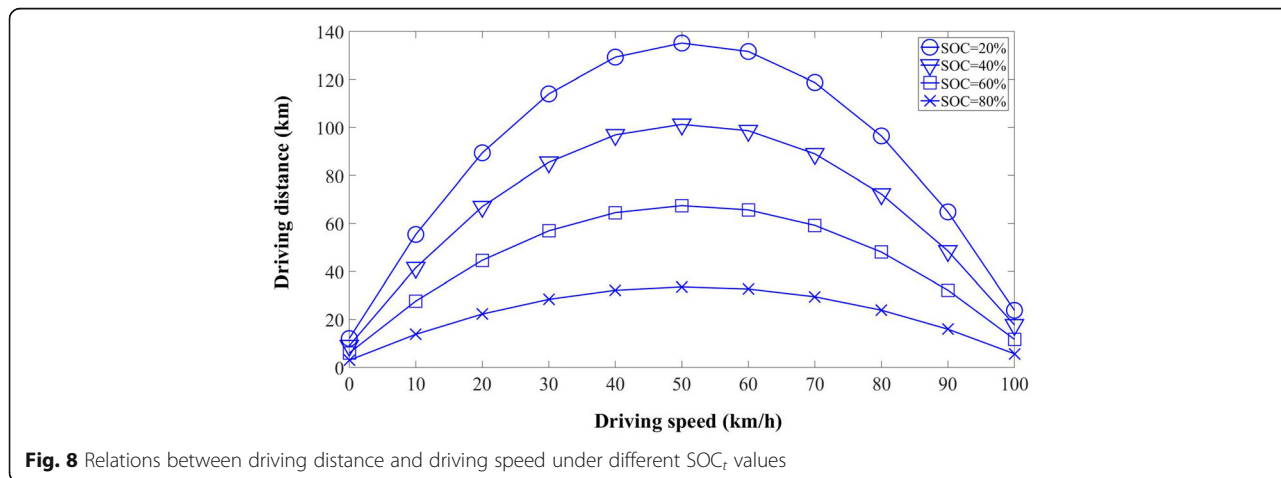


Fig. 8 Relations between driving distance and driving speed under different  $SOC_t$  values

also have impacts on driving distance, which will be considered in the future research.

## 6 Conclusions

Accurately estimating driving distance is an important issue for drivers of BELVs, which is very useful to determine transport strategies and alleviate range anxiety during trips. However, conventional mechanism modeling methods are unable to accurately estimate driving distance because the invisible affecting factors, such as human factors and other external factors, cannot be investigated by the methods. Data-based modeling methods are the effective methods to improve the driving distance estimation, which aims to apply actually collected data to establish the data-based models. Actually collected data from BELVs record the vehicle state and battery status, additionally, and hide the human factors and other external factors. In this study, a large amount of actual data that was collected from 70 BELVs operating in Beijing, China, were used to explore the driving distance estimation. After data processing, the experimental data was applied to perform the relation analysis. The results show that, besides the expected strong negative linear relation between SOC and driving distance, the driving speed has a nonlinear relation with POECPK. Furthermore, a practical and effective data-based modeling method, regression analysis, was applied to establish the model of driving distance in consideration of SOC and driving speed. The FFRLS algorithm was adopted to estimate the parameters of the model based on the experimental data. Moreover, a data-based method for model verification was employed to validate the driving distance estimation model. The results showed that all errors are comparatively small, thereby confirming that the proposed model possesses good estimation accuracy.

The proposed model was subsequently used to estimate driving distance and determine the economical driving speed. The results showed that the economical driving speeds under different  $SOC_t$  values are very close, and the average value of the economical driving speeds at different  $SOC_t$  values (i.e., 51.2028 km/h) was considered as the overall economical driving speed of BLEVs. Estimation of driving distance under different  $SOC_t$  values was also performed. Moreover, it is noted that, as compared to other models that are based on mechanism analysis, the driving distance and economical driving speed obtained by the proposed data-based model have a higher practical significance for BELV drivers, because the proposed model was established by using the actual data which truly present the vehicle state and battery status during vehicle operation and reflect the visible and invisible affecting factors on driving distance. Therefore, the model and corresponding results can provide effective guidance to BELV drivers.

Notably, all experimental data were collected online from BELVs operating in Beijing, China. Given that the loading weight of the BELVs is around 500 kg in actual operation, the change in loading weight is not clearly reflected in the data. However, the driving distance of BELV may be influenced by loading weight. Therefore, future research will consider the effect of loading weight on the driving distance model to further improve model performance. Moreover, the impact of other factors, such as temperature and the battery SOH, will be considered to improve the robustness of the method in the future research.

### Abbreviations

BELV: Battery electric logistics vehicle; BEV: Battery electric vehicle; FFRLS: Forgetting factor recursive least-squares; POECPK: Percentage of energy consumption per kilometer; RMSE: Root-mean-square error; RMSRE: Root-mean-square relative error; SOC: State of charge

### Acknowledgements

Not applicable.

### Funding

This research is supported by the National Key R&D Program of China (2018YFC0706005, 2018YFC0706000) and Fundamental Research Funds for the Central Universities (2018YJS093).

### Availability of data and materials

Not applicable.

### Authors' contributions

JB contributed in the conception of the study and design of the study. YW wrote the manuscript and carried out the simulation. JZ helped to process the data. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China. <sup>2</sup>MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China.

Received: 3 January 2018 Accepted: 11 October 2018

Published online: 24 October 2018

### References

1. M. NC Onat, O.T. Kucukvar, Conventional, hybrid, plug-in hybrid or electric vehicles? State-based comparative carbon and energy footprint analysis in the United States. *Appl. Energy* **150**, 36–49 (2015)
2. Z.H. Lin, Optimizing and diversifying electric vehicle driving range for U.S. drivers. *Transp. Sci.* **48**(4), 635–650 (2014)
3. N.S. Pearre, W. Kempton, R.L. Guensler, V.V. Elango, Electric vehicles, how much range is required for a day's driving? *Transportation Research Part C, Emerging Technologies* **19**(19), 1171–1184 (2011)
4. F.L. Zhong, H. Li, S.M. Zhong, Q.S. Zhong, C. Yin, An SOC estimation approach based on adaptive sliding mode observer and fractional order equivalent circuit model for lithium-ion batteries. *Commun. Nonlinear. Sci. Numer. Simulat.* **24**(1–3), 127–144 (2015)
5. J.C. Alvarez Anton, P.J. Garcia Nieto, C. Blanco Viejo, Support vector machines used to estimate the battery state of charge. *IEEE Trans. Power. Electron* **28**(12), 5919–5926 (2013)

6. L.L. Du, B. Li, H.D. Zhang, Estimation on state of charge of power battery based on the grey neural network model. *Appl. Mech. & Mater.* **427**, 1158–1162 (2013)
7. H. Dai, X. Wei, Z. Sun, J. Wang, W. Gu, Online cell SOC estimation of Li-ion battery packs using a dual time-scale Kalman filtering for EV applications. *Appl. Energy* **95**(2), 227–237 (2012)
8. S. Shao, J. Bi, F. Yang, W. Guan, On-line estimation of state-of-charge of Li-ion batteries in electric vehicle using the resampling particle filter. *Transp. Res. Part D: Transp. Environ.* **32**, 207–217 (2014)
9. X. Liu, Z. Chen, C. Zhang, W. J, A novel temperature-compensated model for power Li-ion batteries with dual-particle-filter state of charge estimation. *Appl. Energy* **123**(3), 263–272 (2014)
10. Y. Li, P. Chattopadhyay, S. Xiong, A. Ray, C.D. Rahn, Dynamic data-driven and model-based recursive analysis for estimation of battery state-of-charge. *Appl. Energy* **184**, 266–275 (2016)
11. T. Hansen, C.J. Wang, Support vector based battery state of charge estimator. *J. Power Sources* **141**(2), 351–358 (2005)
12. H. Sheng, J. Xiao, Electric vehicle state of charge estimation: nonlinear correlation and fuzzy support vector machine. *J. Power Sources* **281**, 131–137 (2015)
13. W. He, N. Williard, C. Chen, M. Pecht, State of charge estimation for electric vehicle batteries using unscented Kalman filtering. *Microelectron. Reliab.* **53**(6), 840–847 (2013)
14. F. Baronti, W. Zamboni, N. Femia, Parameter identification of Li-Po batteries in electric vehicles: a comparative study. *IEEE Int. Symp. Ind. Electron.* **2013**, 1–7 (2013)
15. F.C. Sun, X.S. Hu, Y. Zou, S.G. Li, Adaptive unscented Kalman filtering for state of charge estimation of a lithium-ion battery for electric vehicles. *Energy* **36**(5), 3531–3540 (2011)
16. J.G. Hayes, R.P.R. De Oliveira, S. Vaughan, M.G. Egan, Simplified electric vehicle power train models and range estimation. *Vehicle Power and Propulsion Conference*. **2011**, 1–5 9 (2011)
17. I. Cunningham, K. Burnham, Online use of the fuzzy transform in the estimation of electric vehicle range. *Meas. Control.* **46**(9), 277–282 (2013)
18. A. Shekhar, V. Prasanth, P. Bauer, M. Bolech, Generic methodology for driving range estimation of electric vehicle with on-road charging. *Transportation Electrification Conference and Expo. IEEE* **2015**(1–8) (2015)
19. E. Kim, J. Lee, G.S. Kang, Real-time prediction of battery power requirements for electric vehicles. *ACM/IEEE International Conference on Cyber-Physical Systems. IEEE*, 11–20 (2013)
20. M. Neaimah, G.A. Hill, Y. Hübner, P.T. Blythe, Routing systems to extend the driving range of electric vehicles. *IET Intell. Transp. Syst.* **7**(3), 327–336 (2013)
21. C. Li, X. Xie, Y. Huang, H. Wang, C. Niu, Distributed data mining based on deep neural network for wireless sensor network. *Int. J. Distrib. Sens. Netw.* **2015**, 1–7 (2015)
22. H. Wang, X. Zhang, M. Ouyang, Energy consumption of electric vehicles based on real-world driving patterns: a case study of Beijing. *Appl. Energy* **157**, 710–719 (2015)
23. W. X, J. Shen, Y. Li, K.Y. Lee, Data-driven modeling and predictive control for boiler-turbine unit. *IEEE Trans. Energy. Convers.* **28**(3), 470–481 (2013)
24. G. Wager, J. Whale, T. Braunl, Driving electric vehicles at highway speeds: The effect of higher driving speeds on energy consumption and driving range for electric vehicles in Australia. *Renew. Sustain. Energy Rev.* **63**, 158–165 (2016)
25. J. Hong, S. Park, N. Chang, Accurate remaining range estimation for electric vehicles. *Asia and South Pacific Design Automation Conference. IEEE*, 781–786 (2016)
26. S.K. GM Fetene, S.L. Mabit, A.F. Jensen, C.G. Prato, Harnessing big data for estimating the energy consumption and driving range of electric vehicles. *Transportation Research Part D: Transport & Environment* **54**, 1–11 (2017)
27. J. Bi, Y. Wang, S. Shao, Y. Cheng, Residual range estimation for battery electric vehicle based on radial basis function neural network. *Measurement* **128**, 197–203 (2018)
28. J.P. Calvi, P. Manh, Lagrange interpolation at real projections of Leja sequences for the unit disk. *Proc. Am. Math. Soc.* **140**(12), 4271–4284 (2012)
29. M. Gholizadeh, F.R. Salmasi, Estimation of state of charge, unknown nonlinearities, and state of health of a lithium-ion battery based on a comprehensive unobservable model. *IEEE Trans. Ind. Electron.* **61**(3), 1335–1344 (2014)
30. J. Adler, I. Parmryd, Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry Part A.* **77A**(8), 733–742 (2010)
31. M. Doudnikoff, R. Lacoste, Effect of a speed reduction of containerships in response to higher energy costs in sulphur emission control areas. *Transp. Res. Part D: Transp. Environ.* **28**(5), 51–61 (2014)
32. L.M. Gardner, M. Duell, S.T. Waller, A framework for evaluating the role of electric vehicles in transportation network infrastructure under travel demand variability. *Transp. Res. A Policy Pract.* **49**(1), 76–90 (2013)
33. P. Laug, H. Borouchaki, Curve linearization and discretization for meshing composite parametric surfaces. *Commun. Numer. Methods. Eng.* **20**(11), 869–876 (2004)
34. Z.H. Deng, H.L. Cao, X. Li, J.H. Jiang, J. Yang, Y. Qin, Generalized predictive control for fractional order dynamic model of solid oxide fuel cell output power. *J. Power. Sources* **195**(24), 8097–8103 (2010)
35. S. Feng, C.D. Wu, Y.Z. Zhang, Z.X. Jia, Grid-based improved maximum likelihood estimation for dynamic localization of mobile robots. *Int. J. Distrib. Sens. Netw.* **2014**, 1–15 (2014)
36. J. Bi, T. Zhang, Y. H, Y. Kang, State-of-health estimation of lithium-ion battery packs in electric vehicles based on genetic resampling particle filter. *Appl. Energy* **182**, 558–568 (2016)

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---