

A multi-step fast charging-based battery capacity estimation framework of real-world electric vehicles

Zhang, Dayu; Wang, Zhenpo; Liu, Peng; She, Chengqi; Wang, Qiushi; Zhou, Litao ; Qin, Zian

DOI

[10.1016/j.energy.2024.130773](https://doi.org/10.1016/j.energy.2024.130773)

Publication date

2024

Document Version

Final published version

Published in

Energy

Citation (APA)

Zhang, D., Wang, Z., Liu, P., She, C., Wang, Q., Zhou, L., & Qin, Z. (2024). A multi-step fast charging-based battery capacity estimation framework of real-world electric vehicles. *Energy*, 294, Article 130773. <https://doi.org/10.1016/j.energy.2024.130773>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.



A multi-step fast charging-based battery capacity estimation framework of real-world electric vehicles

Dayu Zhang^{a,b}, Zhenpo Wang^a, Peng Liu^{a,*}, Chengqi She^c, Qiushi Wang^a, Litao Zhou^a, Zian Qin^b

^a National Engineering Research Center of Electric Vehicles, Beijing Institute of Technology, Beijing, 100081, China

^b Department of Electrical Sustainable Energy, Delft University of Technology, Delft, 2628 CD, The Netherlands

^c Hunan Provincial Key Laboratory of Health Maintenance for Mechanical Equipment, Hunan University of Science and Technology, Xiangtan, 411201, China

ARTICLE INFO

Keywords:

Lithium-ion battery
Capacity estimation
Multi-step fast charging
Machine learning
Real-world data

ABSTRACT

Accurately evaluating battery degradation is not only crucial for ensuring the safe and reliable operation of electric vehicles (EVs) but also fundamental for their intelligent management and maximum utilization. However, the non-linearity, non-measurability, and multi-stress coupled operating conditions have posed significant challenges for battery health prediction. This paper proposes a battery capacity estimation framework based on real-world operating data. Firstly, a comprehensive feature pool is constructed from the direct external features extracted during multi-step fast charging processes and the quantitative representation of operating conditions. Subsequently, a two-step feature engineering is introduced to select the most relevant features and eliminate the interference components. The battery capacity estimation framework is then implemented using machine learning methods. Validation results demonstrate that the proposed framework achieves superior estimation accuracy with lower computational expense compared to the modelling process without feature engineering. The MAPE and RMSE reach 1.18% and 1.98 Ah, respectively, representing reductions in errors of up to 8.53% and 11.21%. Collectively, the proposed framework paves the foundation for online health prognostics of batteries under practical operating conditions.

1. Introduction

Developing electric vehicles (EVs) has been reckoned as the inevitable path towards achieving the green and low-carbon transformation of the global automotive industry and the essential means to fulfilling a carbon-neutral vision of the road transport sector [1]. Lithium-ion batteries (LIBs) have emerged as the dominant energy storage device in EV applications due to their high energy density, reasonable cycle life, and environmental friendliness [2]. However, as is the case with machines, the LIB components, such as electrodes and separators, experience varying levels of degradation during cycling, leading to the irreversible decline of capacity and power. To ensure the safe, reliable, and efficient use of LIBs, obtaining an accurate indicator of battery health, namely, the State of health (SOH), is of vital importance.

Battery SOH can be defined in various forms. It can be defined by the service time, increased internal resistance, and loss of capacity [3]. Among these measurable variables, accurately determining capacity loss is more significant to other battery management tasks, such as driving range estimation and life prediction. Thus, defining the SOH

as the ratio between the present and initial capacity has drawn broad attention [4]. However, the capacity measurement requires completely charging or discharging the batteries with the specific protocols, which is challenging for batteries in practical use. This motivates the SOH estimation from daily operation data.

Extensive efforts have been devoted to extracting features correlated with battery degradation mechanisms and constructing the relationships between these features and SOH. The related studies are commonly categorized into model-based and data-driven methods. The model-based approach relies on the empirical model or the physics-based modelling of the degradation behaviour to describe the declining trajectory of the battery. The empirical/semi-empirical model ignores the internal reaction mechanism of the batteries and fits the degradation curve using a predefined parametric function, exhibiting remarkable performance for laboratory data generated under well-controlled conditions. However, the extrapolation ability of such models is limited in changeable working conditions. Thus, this method is frequently combined with the filtering algorithm to update the model parameters with the latest available data [5,6]. The electrochemical

* Corresponding author.

E-mail address: bitliupeng@bit.edu.cn (P. Liu).

<https://doi.org/10.1016/j.energy.2024.130773>

Received 28 October 2023; Received in revised form 11 January 2024; Accepted 20 February 2024

Available online 22 February 2024

0360-5442/© 2024 Elsevier Ltd. All rights reserved.

and equivalent circuit models (ECM) are two commonly used battery models for the physics-based modelling approach. The former simulates the whole reaction process inside the batteries under given operating conditions based on serials of complicated, mutually coupled partial differential equations [7]. Although the electrochemical model can accurately simulate the reaction process inside the battery, the high modelling complexity and computational burden have restrained its wide applications in embedded battery management systems. The ECM is more concise than the electrochemical model and is ubiquitously combined with adaptive filters, like Kalman [8,9] and particle filters [10], to realize accurate states estimation. Though the physics-based model achieves excellent progress in capturing battery dynamics, some common obstacles restrict their online application. First, as depicted in [11], battery degradation results from diverse, interlaced, non-linear degradation mechanisms. No single physics-based model can describe all the mechanisms comprehensively, nor can an effective approach quantify them separately. Second, the modelling process of physics-based models involves plenty of electrochemical states and state-dependent parameters. These states and parameters are unobservable and challenging to calibrate by existing in-vehicle sensing technology [12].

With the staggering progress of big data techniques and the unprecedented penetration of artificial intelligence in the battery community, data-driven methods have attracted tremendous attention for battery health prognostics. Unlike the model-based methods, the data-driven approach can map the relationship between several features and battery health indicators without precisely knowing the mechanisms and propagation, avoiding the complex modelling process. Moreover, the data-driven model is more flexible when applied to different systems and is more feasible in real-world applications [13,14]. Among existing data-driven techniques, it has been proven that extracting features from voltage curves can achieve effective SOH estimation. For instance, incremental capacity analysis (ICA) [15,16], differential voltage analysis (DVA) [17] and differential thermal voltammetry (DTV) [18,19] have been extensively explored for battery health prognostics. However, these models are all conducted at repeatable and well-controlled operating profiles, like constant charging current and ambient temperatures, which are impractical to achieve in many real-world applications. In addition, thanks to the significant advantages of fast charging technology in terms of energy replenishment efficiency, the multi-step fast charging technique has been gradually adopted by most EV manufacturers [20,21]. The higher rate and non-constant current have proposed inevitable challenges for traditional data-driven techniques. To address this problem, Hu et al. focused on the peculiarities of the multi-step fast charging process. They extracted twelve features strongly correlated with battery degradation mechanisms from the voltage curve. Using these features as the model inputs, a dual GPR model was constructed to realize effective battery health prognostics [22]. Although the development of the digital twin technique, the cloud system's high computation capability and enormous storage space make it possible to implement advanced high-performance algorithms online for smart battery usage and health prognostic. It is foreseeable that with the gradual growth of new energy vehicle ownership, the transmission, storage, and processing of high-dimensional features will also be expensive and energy-consuming. Thus, eliminating the redundant features and extracting effective information before data analysis and developing estimation models can help reduce the computation burden and enhance the model performance [23,24]. Zhang et al. proposed a feature dependence check-and-control scheme incorporating a series of correlation analyses to select the most relevant and independent features from the constructed feature pool. Based on the features selected by this scheme, a highly accurate and robust SOH prediction model was constructed at a reasonable computation expense [11]. Furthermore, signal decomposition methods, like empirical mode decomposition (EMD) [25] and seasonal-trend decomposition based on loess (STL) [26], have attracted tremendous attention in battery health prognostics and fault

diagnosis due to their strong adaptation to abnormal values and broad adaptability.

Indeed, most existing battery SOH estimation schemes are developed at the cell or module level, with the data samples for training/testing procedures obtained at preset loading profiles and well-controlled conditions. That means once extrapolating these models to the system-level assessment tasks under practical operating conditions, the estimation accuracy and robustness will be inevitably curtailed by the varying operating temperatures and stochastic usage behaviours. Besides, although fast charging technology, as a critical enabler of mainstream EV adoption, has been gradually adopted by manufacturers, relatively few studies are designed based on multi-step fast charging mode, particularly when considering the impact of actual sampling frequency and driving behaviour simultaneously. Therefore, it is necessary to establish a battery capacity estimation framework specific to the realistic operating conditions of EVs.

To overcome these gaps, this paper proposes a systematic battery capacity estimation framework for EVs, which fully considers the influence of operating conditions on battery degradation. The exclusive contributions of this study are summarized as follows:

1. Leveraging the partial multi-step fast charging processes obtained from real-world EVs to directly extract health features that strongly correlate with battery capacity degradation.
2. Constructing a comprehensive feature pool from extensive EV operating data, incorporating both direct external features and quantitative representations of operating conditions, ensures a realistic reflection of EVs' practical characteristics.
3. A two-step feature engineering is designed to select the most relevant features and eliminate interference components from the constructed feature pool. This enhances the accuracy and reduces the computational burden of the proposed estimation framework, even in the presence of diverse degradation patterns.

An overview of the proposed framework is demonstrated in Fig. 1. The remainder of this study is organized as follows. The detailed descriptions of the dataset used in this research and the data preprocessing are described in Section 2. After that, the comprehensive feature pool is then constructed in Section 3. Section 4 elaborates on details of the proposed feature engineering scheme and the selected machine learning methods. The systematical analysis and comparison for the proposed estimation framework is presented in Section 5.

2. Dataset description and data preprocessing

2.1. Description of dataset

The dataset used in this manuscript was directly collected from the National Big Data Alliance for New Energy Vehicles (NDANEV) (see Fig. 2), which contains the operational data and other information of eighty EVs with the exact specifications. All these vehicles are operated in Guangzhou city (southern China), each with a cumulative mileage of over 20,000 km and an operation duration of over two years. The detailed static information of studied EVs is listed in Table 1. An example of the analysed real-world operation dataset is shown in Table 2. Here, the timestamp represents the moment the data is transmitted to the cloud platform, and the sampling frequency is set as 0.1 Hz. The voltage and current of the battery pack are recorded as the total voltage and current, and the charge current is defined as negative. The highest cell voltage is the maximum value among the cell voltage list. Moreover, the battery cell temperature list presents the temperatures measured by thirty-four probes installed at specific positions inside the pack, and the average of all probes is denoted as the average charging temperature.

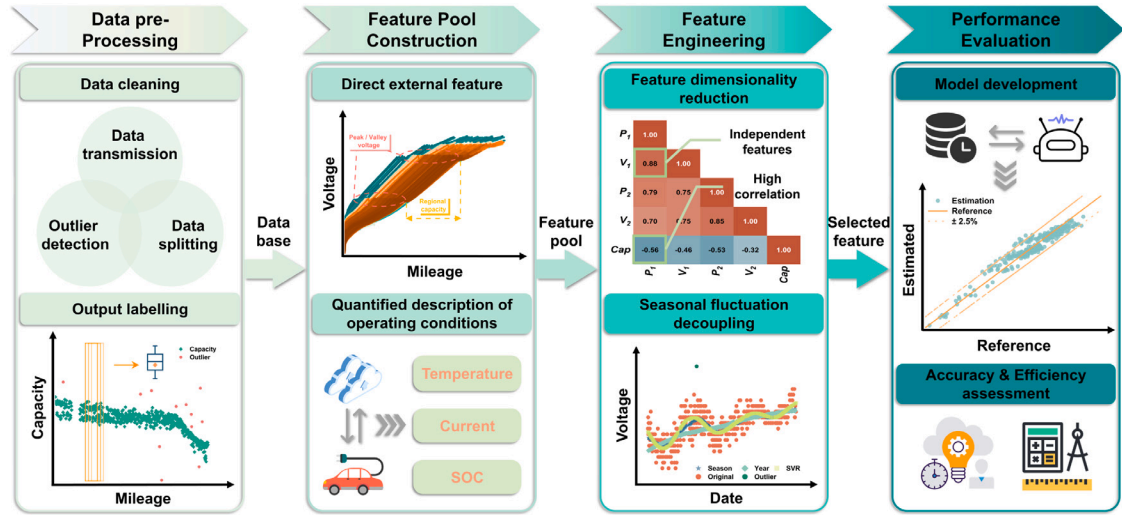


Fig. 1. The overall flowchart of the proposed battery capacity estimation framework.

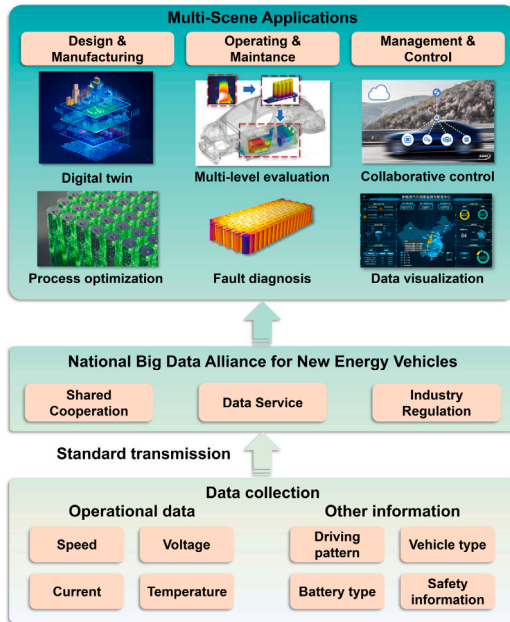


Fig. 2. The overall framework of NDANEV open lab.

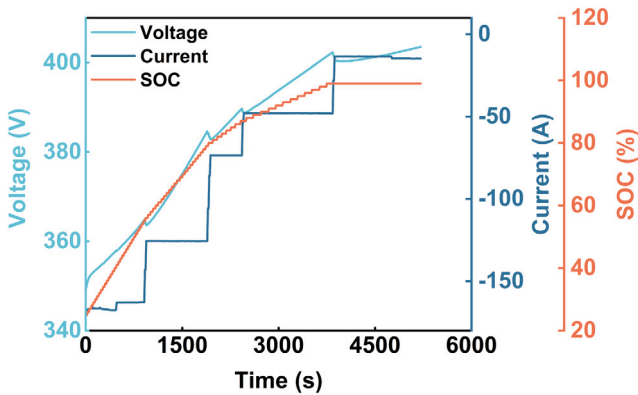


Fig. 3. Detailed description of the fast charging process.

Table 1

The detailed information of the studied electric vehicles.

Parameter	Value (Units)
Curb weight	1570 (kg)
Battery type	Ternary lithium battery
Battery capacity	153 (Ah)
Battery nominal voltage	346 (V)
Connection method	1P95S
Cell voltage range	2.8 (V)~4.3 (V)

2.2. Dataset processing

Through observing the charging data and analysing charging protocols of all eighty EVs, the minimum charging current and mean charging rate within the interval of 30%~80% SOC are utilized as criteria (the minimum charging current lower than 120 A and the mean charging rate higher than 0.75C) for fast charging segments identification. Eight vehicles with a high fast charging frequency and usage are selected from the operation dataset, and each covers the time duration from May 2019 to April 2022, with mileage ranging from 40,000 km to 400,000 km. The battery data extracted from the operation dataset is categorized into charging-related data and driving-related data. Acknowledging the stochastic nature of driving-related data and its potential to introduce more uncertainty to battery health estimation, this study employs charging-related data to calculate the system output. Consequently, 300~400 fast charging segments are extracted from the history of charging processes for each vehicle, and the detailed description of the extracted fast charging segment is shown in Fig. 3. Each segment covers the 30%~90% SOC interval, coinciding with the high-frequency charging interval of EVs [27].

Labelling the system output before constructing the estimation model is essential for supervised learning-based battery health prognosis studies. The definition of system output should also be adapted to the dataset's source. In well-controlled laboratory environments, the labelled capacity can be obtained by discharging the battery from the upper to lower cutoff voltage. However, this method is not applicable to practically operated EVs, as they are rarely fully cycled, and the discharge current is highly random. Moreover, the step-wise change and high charging current rate in the multi-step fast charging process have also restricted the application of ICA and DVA-based methods. To effectively indicate battery health in limited data conditions, we calculate the system output based on the deviation of the SOC formula [28],

Table 2

The example of the studied real-world operation dataset.

Mileage (km)	Timestamp	Velocity (km/h)	Total voltage (V)	Total current (A)	SOC (%)	Highest cell voltage (V)	Cell voltage list (V)	Cell temperature list (°C)
65 058	2019-6-10 10:55:19	0	354.60	−129.80	40	3.762	3.729_3.733_3.732_...	39_38_38_...
65 058	2019-6-10 10:55:29	0	354.90	−128.80	40	3.766	3.732_3.737_3.736_...	39_39_38_...
65 058	2019-6-10 10:55:39	0	355.20	−129.00	40	3.769	3.734_3.739_3.739_...	39_39_38_...
65 058	2019-6-10 10:55:49	0	355.50	−128.40	40	3.771	3.737_3.741_3.741_...	40_39_38_...

deduced as Eq. (1).

$$C_a = \frac{-\int_{t_1}^{t_2} I(t) \Delta t}{SOC_{t_2} - SOC_{t_1}} \quad (1)$$

where t_1 and t_2 are the start and end moments of each charging segment, $I(t)$ is the charging current at the moment t , and Δt is the fixed sample interval. The derived capacity values of the sample vehicle are depicted in Fig. 4(a), revealing a visible declining trend with increasing mileage. However, non-negligible outliers are also present, primarily arising from data transmission errors and biases introduced by in-vehicle SOC estimation methods. To address this issue, we employ a combination of the moving window and interquartile range (IQR) methods [29] to eliminate outliers while preserving the evolving nature of capacity over time. Firstly, the moving window method divides the battery's lifespan into individual intervals. Subsequently, the capacity values within each interval are sorted, and the quartiles, including Q1 (the first quartile), Q2 (the median quartile), and Q3 (the third quartile), are calculated. Then, the upper and lower bounds are established through the computation of IQR, which are derived as Eq. (2). Points that exceed the upper or lower bounds are identified as outliers and removed from this interval, and the filtered capacity values are shown in Fig. 4(b). In this research, the window size, representing the length of each data interval, is set to 2×10^4 km, while the moving step size, indicating the distance between the start points of two neighbouring intervals, is chosen as 1×10^4 km. The selection of appropriate window size and moving step size is crucial. The window size should be wide enough to characterize the capacity distribution within the mileage interval while remaining small enough to demonstrate significant degradation. Regarding the moving step size, it is implemented to generate more data samples based on insufficient datasets while complying with the progressive nature of the capacity decline.

$$\begin{cases} IQR = Q3 - Q1 \\ \text{Lower bound} = Q1 - (1.5 * IQR) \\ \text{Upper bound} = Q3 + (1.5 * IQR) \end{cases} \quad (2)$$

Moreover, the capacity calculation results at different mileage intervals for all studied vehicles are delineated through a box plot in Fig. 4(c). Notably, the mean capacity within each interval exhibits a decline commensurate with the increasing mileage, with the decline rate significantly accelerating once the capacity degrades to 80% of its initial value. This observation further verifies the rationale for establishing the retirement criteria of EV power batteries at 80% of their initial capacity. The longitudinal length of the box is positively correlated with the discrete degree of capacity. Consequently, even with the same mileage, the degradation status of capacity varies from vehicle to vehicle, thereby emphasizing the considerable impact of driving behaviours and external environments on capacity degradation.

3. Feature pool construction

Throughout the life cycle of the LIBs, irreversible internal structure and composition variations directly lead to their performance degradation. Simultaneously, the intricate operating conditions and

diverse driving behaviours further influence its degradation rate, thus contributing to significant differences in degradation trajectory from vehicle to vehicle. Moreover, the direct external features extracted from the charge or discharge curve have exhibited excellent performance and application prospects in battery health prognosis due to their easier accessibility and strong correlations with the battery degradation mechanism [30,31]. Given these considerations, the comprehensive construction of the data feature, incorporating compelling external direct features and operating condition-related features, is essential for achieving precise assessments of battery health status. In this section, we will first extract the direct external and operating condition-related features from massive real-world operation data and then comprehensively analyse these features to form the feature pool required for health estimation.

3.1. Direct external features extraction

Voltage-related features. According to the protocol of the multi-step fast charging, the charging current for each step is constant and will switch to other rates when the maximum cell voltage reaches the threshold of this step. During this switching process, the current will drop rapidly and then rise to the preset value of the next step in a few sampling intervals. Along with this current-switching process, the voltage will respond to the change simultaneously. Moreover, the side reactions such as the growth and decomposition of solid electrolyte interface (SEI), electrolyte decomposition, and lithium plating will lead to the irreversible consumption of Li-ions, the increase of impedance, and the loss of active material (LAM), which further aggravate the change of voltage [32]. Hence, the voltage sections around the current switching points are generally selected to derivate voltage-related features, such as the peak voltage (P_i), valley voltage (V_i), voltage drop (dU_i) and the slope of the voltage curve (K_i). These features are validated to have strong correlations with the decline of the battery [22]. Fig. 5 and Table 3 illustrate the detailed extraction methods for voltage-related features. It should be noted that the sampling interval (10 s) and measurement precision (one decimal place) of the system voltage signal in EVs are both significantly inferior to laboratory conditions (which are typically 0.1 s and three decimal places, respectively), thus influencing the reliability of both voltage drop and curve slope features. Consequently, considering the long duration of the first two charging steps, only the peak and valley voltage features of these two steps are extracted. Their evolution trends with accumulative mileage are shown in Figs. 5(b) and 5(c), from which we can conclude that both the peak and drop voltage exhibit a gradual ascending trend as the battery declines. This phenomenon has been proven to be associated with two battery degradation modes directly, the increase of internal resistance and the irreversible loss of lithium inventory (LLI) [33,34]. Besides, it also can be seen that both peak and valley voltage exhibit slight local fluctuations with the variation of average charging temperature, mainly attributed to the fact that the increased temperature accelerates the growth rates of the SEI layer, thus resulting in faster LLI and resistance increase [5].

Capacity-related features. Along with the increasing internal resistance during the degradation of the battery, the voltage will reach

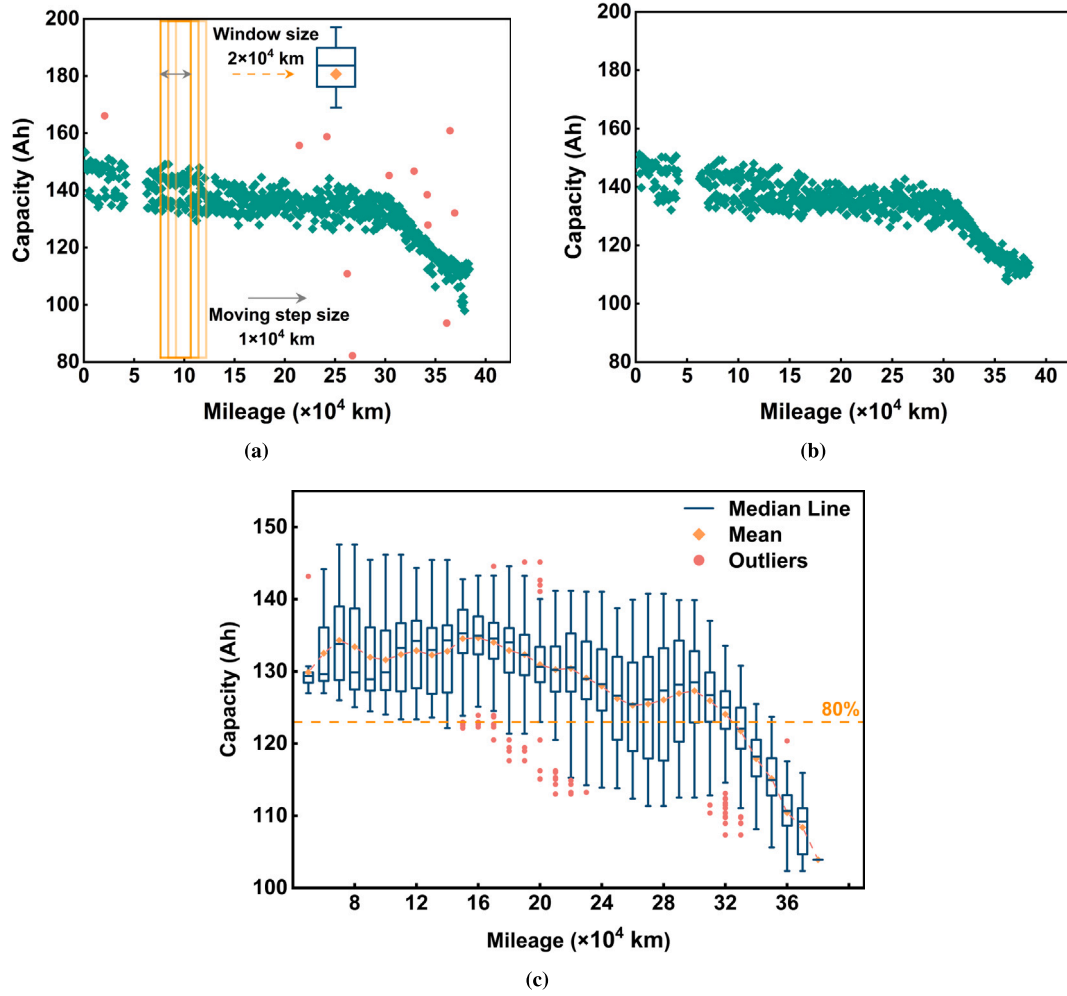


Fig. 4. The calculated system output. (a) The illustration of the filtering method for system output. (b) The smoothed system output. (c) The evolution trend of system output at different mileage for the studied vehicles. Mileage is divided at 20,000 km intervals, and each box reflects the degree of discretization of the capacity between different vehicles at the current mileage.

Table 3
The description of the voltage-related-features.

Feature	Description
P_i	The peak voltage at the current switching point. Footnote i represents the charging step index, same below.
V_i	The valley voltage at the current switching point.
dU_i	The difference between the peak and valley voltage. The definition is: $dU_i = P_i - V_i$
K_i	Slope of the charging voltage curve. The feature can be calculated as: $K = \frac{U_{i-1} - U_{i-6}}{5}$, where U_{i-1} and U_{i-6} are the voltage values at point 1 and 6 before the peak voltage, respectively.

the threshold sooner during charging, thus leading to a decrease in charging capacity. Ref. [35] has also indicated that regional charge capacity loss is the most critical factor affecting battery ageing. Taking the extracted peak voltage as the boundary, the charging capacity of the second charging step (noted as $C_{regional}$) can also be calculated by the Ampere integral formula. It can be captured from Fig. 5(d) that the calculated capacity monotonically decreases with increasing mileage. Additionally, the colour intensity of each point reflects its average charging temperature, from which we can discover that the regional capacity did not display noticeable temperature-dependent variation. This is because the intervals of the selected voltage section are not fixed, which takes into account the impact of temperature on

the peak voltage. Moreover, it further demonstrates that the influence of temperature on this feature is relatively insignificant compared to mileage.

3.2. Quantitative representation of operating conditions

As known, lithium-ion batteries' degradation results from a complex interplay of multiple internal and external stress factors. The deviations in the ageing trends and rates are mainly attributed to the following reasons: (1) The inevitable variations caused by the manufacturing and assembly process, the electrode thickness and density deviations play an essential role in the initial capacity, and the rate of capacity fade [36]. (2) The different operational environments and driving behaviours, such as the DOD, current rate and temperature, considerably complicate the ageing process and further aggravate the battery deviations. With the continuous progress in battery manufacturing and assembly technology, the initial differences in internal parameters are gradually narrowing. Consequently, accurately capturing the variations in the operating environment and driving behaviour has progressively become a necessary procedure to predict battery health status for real-world vehicles precisely. In view of this, the influence of operating conditions on battery decline is investigated in this subsection based on the massive real-world operation dataset.

Temperature-related features. As verified in many accelerated ageing tests and studies, the temperature accounts for the most significant impact on battery capacity degradation [37]. High temperatures

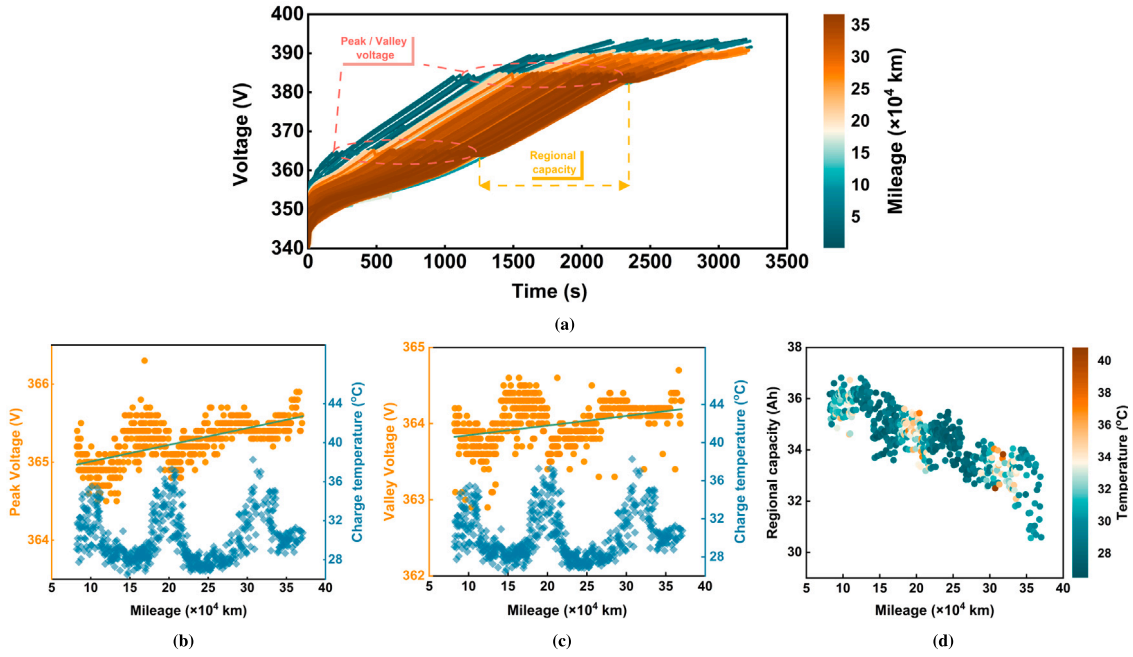


Fig. 5. Visualization of direct external features. (a) The illustration of feature extraction. (b) The evolution trend of P_1 . (c) The evolution trend of V_1 . (d) The evolution trend of $C_{regional}$.

(above 25 °C) would remarkably accelerate the growth of SEI film, degradation of the cathode and electrolyte decomposition, and the decline rates of battery increase with rising temperature [38]. On the contrary, the irreversible consumption of active lithium ions caused by lithium plating at the anode is the primary concern of low temperatures (below 25 °C), and the ageing rates of battery increase as the temperature decreases. In this study, three temperature-related features are extracted, including the average charging temperature of the current fast charging segment (noted as $T_{c,segment}$), the average charging and discharging temperature for the interval between two adjacent fast charging segments (noted as $T_{c,interval}$ and $T_{d,interval}$). It can be seen from Figs. 5(b) and 5(c) that the local fluctuations of voltage-related features have an unignorable dependency on $T_{c,segment}$, thus introducing the $T_{c,segment}$ could eliminate the influence of temperature on these features. For the $T_{c,interval}$, which can reflect the charging temperature of each period more reasonably, its influence on the capacity is demonstrated in Fig. 6(a). As shown, there is an approximately linear decrease in capacity as $T_{c,interval}$ increases, with the linear correlation coefficient exceeding 0.7, indicating a significant negative influence of temperature on battery capacity. Besides, the colour intensity of each point reflects its current mileage status, from which we can discover that the capacity of high mileage status is significantly lower than that in high temperatures, proving that the impact of mileage on battery degradation is more noticeable than the temperature for studied EVs.

Current-related features. The deposition of metallic lithium on the anode graphite surface reduces the battery life drastically and limits the fast-charging capability. In severe cases, the lithium dendrite formed by lithium plating will penetrate the separator and cause internal short [39]. Hence, significant research efforts have been devoted to understanding the lithium plating mechanisms, and the high charging rate is regarded as one of the three primary factors affecting lithium plating. Considering this, the average charging and discharging rate for the interval between two adjacent fast charging segments (noted as I_{cha} and noted as I_{dis}) are extracted to comprehensively describe the common charging and driving behaviours of the studied vehicles. The impact of the charging rate on capacity is illustrated in Fig. 6(b). It is evident that as the charging rate increases, the capacity decreases,

with a linear correlation coefficient exceeding 0.8. This observation indicates that a higher charging rate has a detrimental effect on battery degradation. Similarly, the colour intensity of each scatter reflects its mileage status. From this, we can discover that under the effects of both mileage and charging rate, the vehicle with a higher charging rate in high mileage status exhibits a more noticeable decline trend. This phenomenon further proves that operating conditions indeed influence the degradation trend of battery health.

SOC-related features. When the battery is discharged at low SOC, the abnormal increase of anode potential will lead to the anodic dissolution of the copper (Cu) current collector and the formation of Cu^{2+} ions. The reverse reaction can form copper dendrites upon recharging, possibly leading to an internal short circuit. Moreover, it has also been verified that the width of the discharge interval accelerates degradation rather than the upper and lower boundary of DOD [40]. The average DOD of the driving segments between two adjacent fast-charging processes is calculated to describe the daily driving distance. Note that during the division of driving segments, the discharging data between two neighbouring and valid charging sequences is considered one complete driving segment so that the practical driving habits of the vehicle can be reproduced as accurately as possible. The influence of DOD on capacity is shown in Fig. 6(c). It can be seen that the capacity decreases with the increasing DOD, indicating that the high DOD would exacerbate battery degradation. In addition, from the accumulative mileage distribution, the driving sequences with DOD above 40% are mainly concentrated after 250,000 km, which may also be regarded as one of the primary reasons for the rapid decline of capacity in the later stage (as shown in Fig. 4(c)).

The above feature pool construction scheme systematically extracts the direct external features strongly correlated to battery degradation mechanisms and quantitatively represents the effects of operating conditions on the capacity decline. Based on that, a comprehensive feature pool (noted as F_{All}) specific to real-world electric vehicle capacity estimation demands is preliminarily constructed, which provides the foundation for subsequent feature selection and estimation model development.

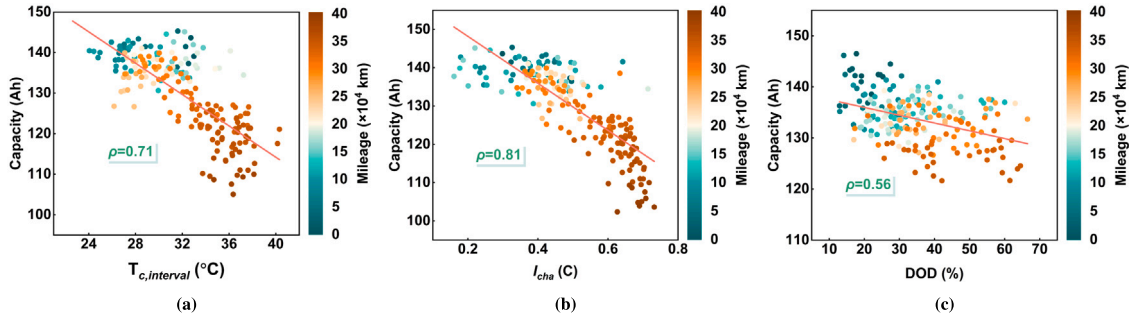


Fig. 6. The quantitative representation of operating conditions. (a) The influence of charging temperature on the capacity decline. (b) The effect of charging rate on the capacity decline. (c) The influence of DOD on the capacity decline.

4. Feature engineering and model development

4.1. Feature engineering

Based on the constructed feature pool, the principle of feature engineering is to process and select features to retain sufficient information, improving the accuracy and robustness of associated methods at the expense of acceptable computation. To achieve this, a series of correlation analyses between each feature and the system output is first conducted to select the most relevant features and avoid selecting strongly interdependent features. Following that, a time series decomposition method is employed to remove the seasonal fluctuations and outliers of the selected features.

Feature dimensionality reduction. As mentioned, the constructed data feature should render the associated methods highly accurate and robust at the expense of reasonable computation. To achieve this purpose, a series of correlation analyses are conducted, including the feature-to-feature and the feature-to-system output, paving the way for determining the most relevant features while avoiding selecting strongly interdependent components [11]. When selecting the correlation analysis method, Spearman, as a non-parametric test, does not necessitate specific assumptions about data distribution and remains insensitive to outliers [41–43]. Consequently, it serves as the primary method for correlation analysis in numerous battery health management studies based on real-world datasets [43]. Leveraging these advantages, Spearman correlation analysis is employed here to quantify the strength and direction of the monotonic association between two variables, which is computed as Eq. (3). Firstly, based on the absolute value of their relational coefficients measured by Spearman, the correlation between each feature x_i and capacity change y is assigned the score C_{x_i-y} (a value between 0 and 1). Then, the features with scores below the preset threshold will be eliminated. Based on preliminary correlation analysis, a feature dependence check-and-control scheme based on Spearman correlation analysis is again introduced to identify and discard the strongly related features. The detailed procedures are described in Table 4.

$$Spearman = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (3)$$

where x and y represents the reference and comparative sequence separately, n means the length of the sequence. In this research, the capacity and the features in the constructed feature pool are successively defined as the reference sequence.

In statistical terms, a strong correlation between variables is deemed to exist when the absolute value of the correlation coefficient exceeds 0.75. Conversely, a correlation between variables is considered weak when the absolute value of the correlation coefficient is below 0.25 [44]. Thus, the C_{low} and C_{high} in this study are set as 0.20 and 0.80 separately. The averaged correlation matrices for the constructed feature pool and the selected features are presented in Fig. 7. In

Table 4

The detailed procedure of feature dependence check-and-control scheme.

Step 1	Reference feature selection. The feature with the highest score relative to the capacity change is selected from the remaining features and denoted as the x_i (start with $i = 1$);
Step 2	Correlation score calculation. Employee the Spearman correlation analysis to calculate the correlation scores of all other features x_j ($j > i$) relative to x_i and denote them as $C_{x_j-x_i}$;
Step 3	Cross-correlation analysis. Compare the calculated correlation scores with the preset threshold C_{high} and the feature x_i will be removed whenever $C_{x_j-x_i} > C_{high}$.
Step 4	Increase the index from i to $i + 1$ and repeat from step 1 again.

comparing the two correlation matrices, the interdependence between the V_i of each charging step and the corresponding P_i is evident, as indicated by the high correlation coefficient denoted by the green square. However, the V_i of each charging step shows a lower correlation with capacity than the P_i , so we have decided to remove it from the feature pool. In addition, features like I_{dis} , $T_{c,segment}$ and $T_{d,interval}$, with low correlation to capacity, are also eliminated from the feature pool. The performance of this feature dimensionality reduction technique in capacity estimation will be discussed in Section 5.

Seasonal fluctuation decoupling. As demonstrated in Fig. 5(b), the peak voltage exhibits periodic fluctuations due to the change of seasons. To eliminate the influence of seasonality, we introduce the seasonal-trend decomposition based on the loess (STL) method to decompose the original data into additive variation components (as deduced by Eq. (4)), including seasonality, trend and remainder. Then, the P_i in data feature after dimensionality reduction (as shown in Fig. 7(b)) is replaced with its trend component, and the obtained data feature is denoted as F_{STL} . Additionally, to verify the effectiveness of the STL approach, we refer to the decoupling philosophy of conventional machine learning techniques, adding the $T_{c,segment}$ to the data feature presented in Fig. 7(b) instead of decomposing the P_i and noting the constructed data feature as F_{Tem} . The complete feature lists for F_{All} , F_{STL} and F_{Tem} are provided in Supplementary materials.

Compared to other decomposition techniques, the significant advantages of STL lie in its robust adaptation to outliers of raw data and application for large amounts of time series data without any mathematical modelling [26,45]. Generally, the iteration mechanism of STL consists of two recursive procedures, the inner and outer loops. During each inner loop iteration, the seasonal and trend components are updated by the seasonal and trend smoothing. When each inner loop is completed, the remainder component is calculated based on the updated seasonal and trend components. Then, based on the calculated remainder components, the robust weights are further computed to

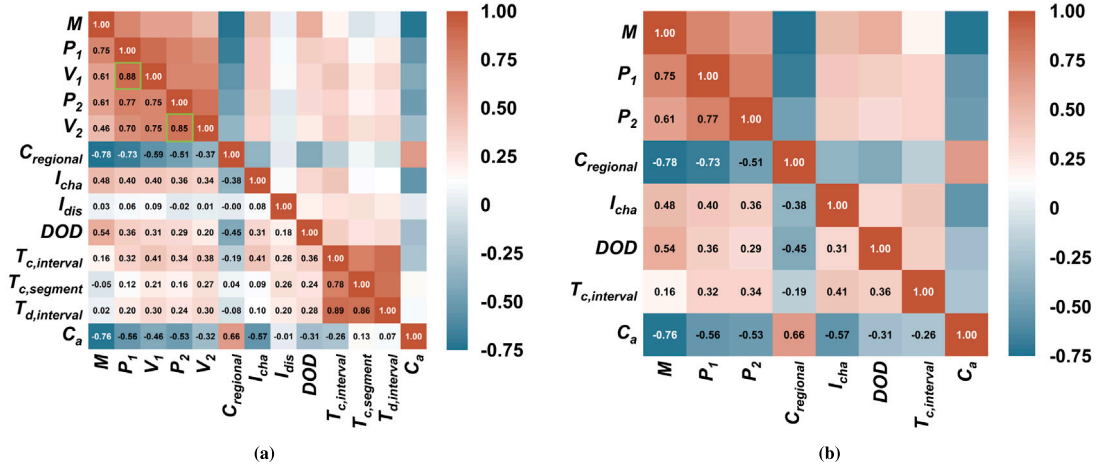


Fig. 7. The feature dimensionality reduction results and corresponding Spearman correlation matrix. (a) The correlation matrix for the constructed feature pool. (b) The correlation matrix for the selected features.

reduce the interference of outliers to the subsequent inner loop.

$$X_t = T_t + S_t + R_t \quad (4)$$

where X_t is the original time series data, T_t , S_t , and R_t represent the trend component, the seasonal component, and the remainder component after decomposition, respectively. The detailed procedures of the inner loop are described as follows:

Step 1: Detrending. Remove the trend component from the original time series data. For the $(k+1)$ th iteration of the inner loop, the trend component of k th iteration T_t^k is removed from X_t , i.e. $X_t^{detrend} \leftarrow X_t - T_t^k$.

Step 2: Seasonal smoothing. Taking advantage of Loess smoother, the cycle-subseries $X_t^{detrend}$ obtained from *Step1* is smoothed to obtain a preliminary seasonal component \tilde{S}_t^{k+1} .

Step 3: Low-pass filtering of the smoothed seasonality. The \tilde{S}_t^{k+1} obtained in *Step2* is first processed using a low-pass filter, which is comprised of three procedures: the first step is a moving average of length n , n is the number of samples in a period. Following that, a moving average of length n is implemented again, and the third step is a moving average of length 3. Then, the loess regression is applied to identify any remaining trend \tilde{T}_t^{k+1} .

Step 4: Detrending of smoothed seasonality. The additive seasonal component S_t^{k+1} is computed as the difference between the low-pass values and the preliminary seasonal component, i.e., $S_t^{k+1} = \tilde{S}_t^{k+1} - \tilde{T}_t^{k+1}$.

Step 5: Deseasonalizing. Reduce the seasonal component S_t^{k+1} from the original series X_t to obtain a seasonally adjusted series, i.e. $X_t^{deseason} = X_t - S_t^{k+1}$.

Step 6: Trend smoothing. The seasonally adjusted series $X_t^{deseason}$ obtained in *Step5* is smoothed by a Loess smoother to obtain the trend component T_t^{k+1} .

Upon the accomplishment of the inner loop, the obtained trend and seasonal components are used to compute the remaining component $R_t^{k+1} = X_t - S_t^{k+1} - T_t^{k+1}$. Additionally, any large values in R_t are regarded as outliers, whose weights are calculated in the previous iteration of the outer loop and further used to reduce the effects of outliers in the following inner loop. The original time series data is decomposed into seasonal, trend, and remainder components through

the above procedures. Generally, the decomposition window of STL should be an integer multiple of the data period to reflect the original data's trend accurately. However, for the real-world dataset used in this research, due to the differences in vehicle usage behaviour and fast charging frequency, the number of effective charging segments that can be extracted from different vehicles and their time duration also varies. However, it is unable to adjust adaptively. In view of this, the decomposition window for each vehicle is adjusted to the average number of fast charging segments extracted from its operation data per year. In addition, we also adapt the decomposition window to the average number of fast charging segments extracted from its operation data per season and smooth the peak voltage signals with support vector regression (SVR). The decomposed and smoothed results are shown in Fig. 8. It can be seen that both the smoothed voltage signals with SVR and the trend component decomposed with the seasonal benchmark still reveal significant seasonal fluctuations. In contrast, adapting the decomposition window with a yearly benchmark can effectively remove the seasonal components. In the validation section, the differences in SOH estimation performance caused by different decomposition windows will be discussed in detail.

4.2. Estimation model development

The task of battery health estimation can be formulated as a regression problem within the supervised learning framework. Among an extensive toolbox of machine learning methods for the non-linear model regression, GPR [46], random forest regression (RFR) [47], and eXtreme gradient boosting (XGBoost) [48] are very powerful and have been widely applied in various scenarios. Although these methods may have disadvantages, they also have advantages and can be considered promising candidates for battery health prognosis. During the training of the estimation models, the constructed data features F_{All} , F_{STL} , and F_{Tem} are defined as the model input, respectively. The calculated capacity is used as the model output. Taking advantage of the powerful integration of the scikit-learn library, the machine learning methods employed in this framework can be easily obtained, and their detailed descriptions are provided in Supplementary Material. During the model training process, the grid search technique is used for hyperparameter tuning, with which the optimal hyperparameter combinations for each model are identified separately. The detailed hyperparameter values of grid research used for different machine learning methods are provided in Table 5. Additionally, all the present procedures and results were accomplished using publicly available libraries integrated in Python (version 3.8.8) with an Intel i7 CUP and 32 GB RAM.

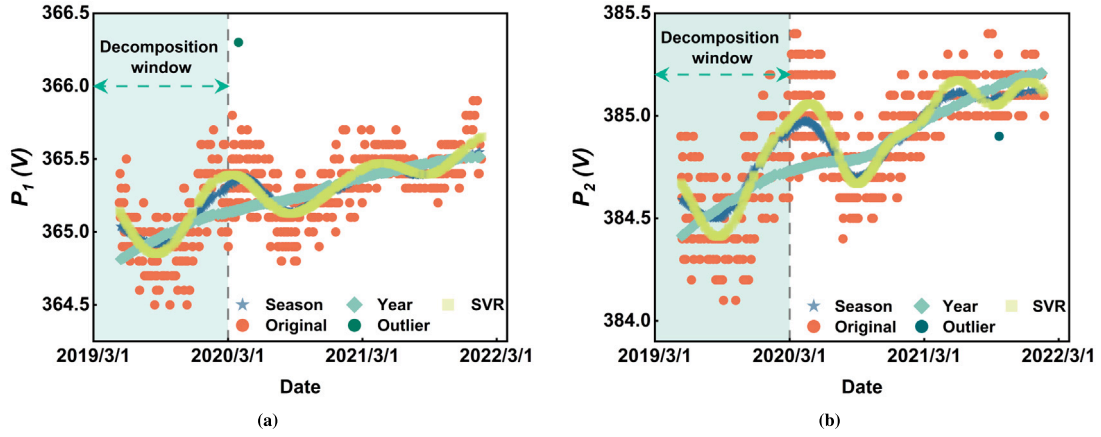


Fig. 8. Decomposing the seasonal components from the peak voltage. (a) The decomposition result of P_1 . (b) The decomposition result of P_2 .

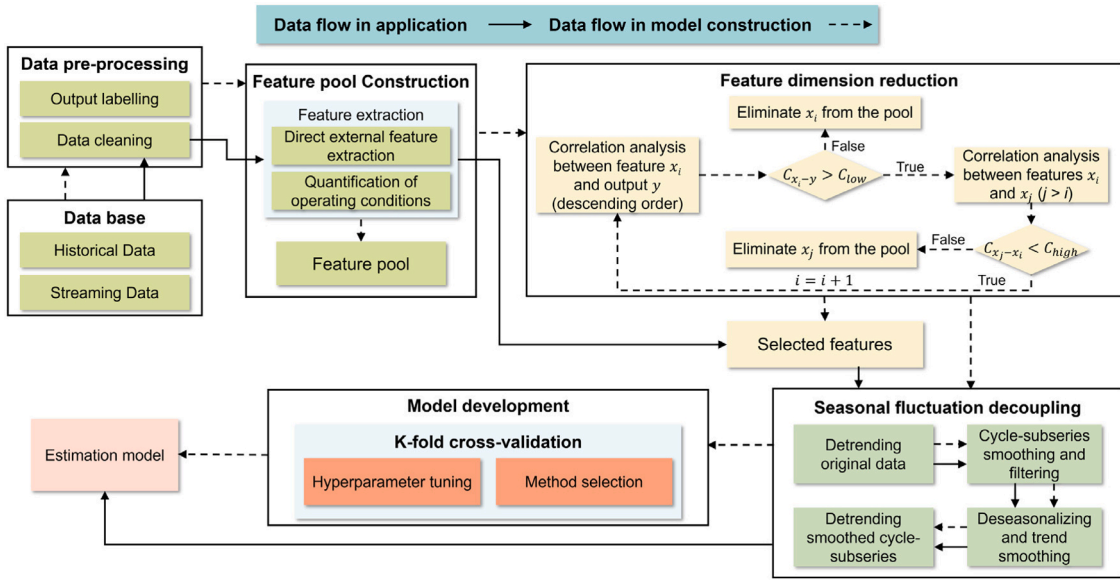


Fig. 9. The summarization of all required modules and their connections.

The required modules and their connections to establishing the framework are summarized in Fig. 9, demonstrating the offline development path based on historical data and the online application path with streaming data. In the offline development stage of the proposed capacity estimation framework, the historical operating data of EVs with the same type is used for feature pool construction, feature dimensionality reduction and decoupling, and the training of the estimation model. The historical dataset is sufficiently labelled while covering long time and mileage duration, which ensures the estimation model can be well-trained. Once the developed model is embedded into the cloud platform, it can accurately reconstruct the capacity degradation process for the same type of vehicles using the data feature F_{STL} constructed from streaming data as input. Compared to the historical data, streaming data covers a relatively short time span and is inadequately labelled, which could be obtained from BMS or cloud platforms during regular maintenance.

5. Results and discussion

The battery capacity estimation framework specific to real-world EVs is developed with the foundational works introduced in the previous section. Based on that, a systematical analysis and comparison will be conducted in this section. Here, the operation data of five

Table 5

The detailed hyperparameter values of grid research used for different machine learning methods.

Model	Hyperparameter
GPR	Kernel function: RBF kernel, 2/5 Matérn Kernel, 2/3 Matérn Kernel
RFR	Number of trees: [100, 200, 500] Maximum tree depth: [8, 9, 10]
XGBoost	Number of trees: [100, 200, 500] Maximum tree depth: [2, 3, 5]

electric passenger vehicles in the selected dataset are considered the training sample, while the rest are labelled as the test sample. The mean absolute percentage error (MAPE) and root mean squared error (RMSE) are leveraged to gauge the estimation performance of the proposed method.

5.1. Results of battery capacity estimation

In this subsection, the performance of F_{STL} in battery health assessment is analysed using the excellent non-linear regression capability of the machine learning methods. The estimated battery capacity, utilizing

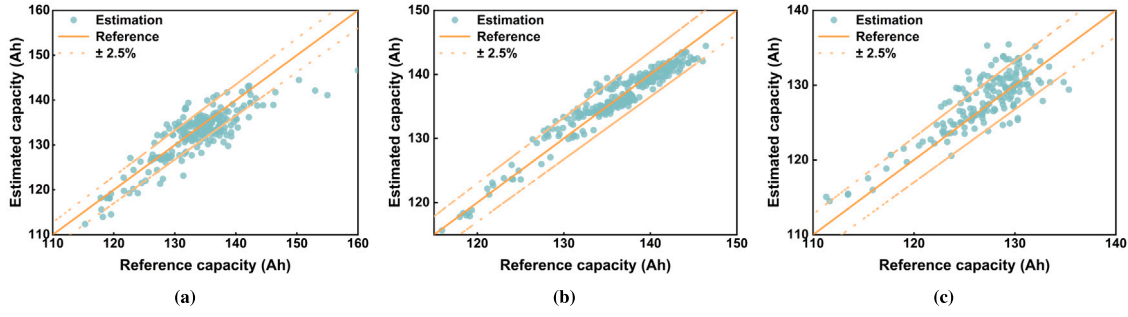


Fig. 10. Battery capacity estimation results based on the proposed framework. (a)–(c) exhibit the estimated capacity values of test vehicles, respectively. The orange dotted lines are the bounds of $\pm 2.5\%$ estimation error.

Table 6

The battery capacity estimation results of the proposed framework.

Decomposition window	Test vehicle	GPR		RFR		XGBoost	
		RMSE (Ah)	MAPE (%)	RMSE (Ah)	MAPE (%)	RMSE (Ah)	MAPE (%)
Seasonal benchmark	Vehicle No. 1	3.85	2.28	1.81	1.09	3.78	2.37
	Vehicle No. 2	2.74	1.61	1.75	1.03	2.87	1.75
	Vehicle No. 3	3.35	2.02	2.57	1.58	3.27	2.02
	Average	3.32	1.97	2.04	1.23	3.27	2.02
Yearly benchmark	Vehicle No. 1	3.26	1.96	1.74	1.02	3.12	1.88
	Vehicle No. 1	2.43	1.43	1.60	0.94	2.30	1.31
	Vehicle No. 1	3.66	2.18	2.61	1.59	3.61	0.27
	Average	3.11	1.86	1.98	1.18	3.01	1.92

the yearly benchmark with RFR as an example, is illustrated in Fig. 10. It can be observed that the vast majority of the estimated capacity values for all test vehicles fall within $\pm 2.5\%$ error bounds, indicating that the constructed estimation model can accurately capture the decline trajectories of each vehicle, even in the presence of significant variations in degradation patterns among different vehicles. To further clarify the effect of the decomposition window selection on model performance, numerical accuracy for both seasonal and yearly benchmarks is summarized in Table 6. Notably, the results indicate that decomposing the voltage signals with the yearly benchmark yields higher estimation accuracy when compared with the seasonal benchmark, both from the individual performance and the global perspective. This is attributed to the fact that, for the operation data used in this study, the alternation of seasons plays a more significant role in the evolution of features. In contrast, changes in ambient temperature within a single season may have less impact on the monotonicity of features, as shown in Fig. 8. For the results obtained from the yearly benchmark, the RMSE and MAPE of different machine learning methods are below 3.20 Ah and 2.00% separately, which satisfies the requirements of many industrial applications that typically demand an error of fewer than 5.00% [49]. Among all selected machine learning methods, RFR outperforms all its alternatives with the MAPE of 1.18% and the RMSE of 1.98 Ah, respectively. These results further corroborate the feasibility of the proposed framework for real-world battery health prognosis.

5.2. Effectiveness verification

As detailed in Section 4.1, in addition to removing the seasonal component from the peak voltage using the STL method, we decouple the temperature and voltage signal by incorporating $T_{c,segment}$ to the reduced-dimensional data feature. To further substantiate the effectiveness of F_{STL} , we also perform the battery capacity estimation case with the constructed feature pool F_{All} . Subsequently, the performance of all data features regarding estimation accuracy and efficiency is compared. The distributions of capacity error with F_{STL} , F_{Tem} and F_{All} , utilizing RFR as the estimation model, are depicted in Fig. 11. The numerical accuracy for the estimation cases based on different data features is summarized in Table 7. From a global perspective, the MAPE for all

Table 7

The numerical accuracy and computational cost of the degradation trajectory prediction based on each data feature.

Data feature	RMSE (Ah)	MAPE (%)	Computational time (s)	
			Training	Testing
F_{STL}	1.98	1.18	81.43	0.05
F_{Tem}	2.13	1.24	84.42	0.05
F_{All}	2.23	1.29	105.92	0.07

three data features is below 1.30%, and the RMSE is less than 2.30 Ah. Notably, F_{STL} exhibits the most favourable performance, achieving an RMSE of 1.98 Ah and a MAPE of 1.18%. It is closely trailed by F_{Tem} with an RMSE of 2.13 Ah and a MAPE of 1.24%. In comparison, F_{All} displays the highest estimation error among the three. The enhancement for F_{STL} compared to F_{All} amounts to 11.21% and 8.53% in terms of RMSE and MAPE, separately. When looking into the details, F_{STL} performs optimally in the vast majority, both in the maximum and mean of capacity error. Furthermore, when considering computational cost, it can be observed that the estimation models relying on F_{STL} and F_{Tem} necessitate less training and testing time in comparison to the model that is reliant on F_{All} . This reduction in training time amounts to 23.12%. Although the reduction in testing time is not noticeable with the current data scale, this strength will be more pronounced as the size of the operation dataset grows. The comparisons above illustrate that the feature dimensionality reduction significantly enhances computational efficiency by eliminating irrelevant and redundant features. Simultaneously, it optimizes model prediction accuracy to some extent. Building upon this foundation, the STL method can further enhance model estimation accuracy with superior trend component capturing and outlier elimination capabilities. Consequently, combining the dimensionality reduction technique and the STL method enhances the model estimation accuracy while reducing the computational cost.

Although the constructed feature pool and proposed feature engineering method are all conducted with the ternary lithium-ion battery, it can be confirmed that this estimation framework is also suitable for other types of lithium-ion batteries. This can be claimed as the

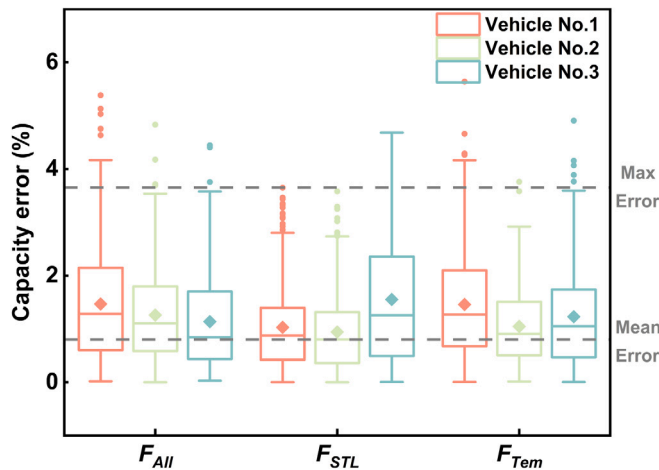


Fig. 11. The distributions of percentage error with F_{STL} , F_{Tem} and F_{All} .

similar degradation mechanisms among the commonly used lithium-ion batteries [50] and the excellent generalization of current change point-based features [22,51].

6. Conclusion

Data-driven modelling has demonstrated impressive potential in battery health prognostics. Still, most existing models are commonly developed at specific testing profiles, which may not fully capture the realistic operating characteristics of EVs. This limitation can lead to deficiencies when these models are directly migrated to real-world scenarios. To address these challenges, we propose a battery capacity estimation framework applicable to real-world EVs under multi-step fast charging scenarios. By exploring the intrinsic ageing information from the partial voltage curve, multiple health features that are strongly correlated with battery capacity are derived. Then, by integrating these direct features with quantitative representations of operating conditions, a comprehensive data feature pool is constructed to effectively reflect the ageing trajectory of the battery capacity and accurately quantify the influence of operating conditions on degradation. Additionally, a two-step feature engineering is designed to remove the irrelevant features and eliminate the interference components from the feature pool. The obtained results showcase the effectiveness of our proposed framework in achieving impressive estimation accuracy and computational efficiency, even in the presence of dramatic differences in vehicles' degradation patterns. With a MAPE of 1.18% and an RMSE of 1.98 Ah, the proposed framework effectively satisfies the safety management requirements of electric vehicles in real-world environments. Notably, this represents a relative error reduction of more than 11.21% and 8.53% compared to the traditional modelling approaches. Furthermore, there is a significant relative enhancement in both training and testing times, reaching 23.12% and 28.57%, respectively. This advantage is expected to become even more substantial as the dataset size continues to expand.

Overall, this framework paves the foundation for online health prognostics of batteries under practical operating conditions. In future work, the stability and reliability of the proposed framework need to be further verified with different usage scenarios, such as different operating areas and purposes. Besides, the adaptive modification for the decomposition window and the transfer optimization of the estimation model will also be an interesting and significant research focus.

CRedit authorship contribution statement

Dayu Zhang: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Zhenpo Wang:** Funding acquisition, Conceptualization, Project administration, Supervision, Writing – review & editing, Methodology. **Peng Liu:** Supervision, Writing – review & editing, Conceptualization, Methodology. **Chengqi She:** Conceptualization, Methodology, Supervision. **Qiushi Wang:** Conceptualization, Methodology. **Litao Zhou:** Conceptualization, Methodology, Supervision. **Zian Qin:** Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported by the National Key Research and Development Program of China under Grant 2022YFB3305400, and in part by the China Scholarship Council under Grant 202206030133.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.energy.2024.130773>.

References

- [1] Wang Z. Annual report on the big data of new energy vehicle in China (2021). Springer Nature; 2023.
- [2] Lu J, Xiong R, Tian J, Wang C, Sun F. Deep learning to estimate lithium-ion battery state of health without additional degradation experiments. Nat Commun 2023;14(1):2760. <http://dx.doi.org/10.1038/s41467-023-38458-w>.
- [3] Kamali MA, Caliwig AC, Lim W. Novel SOH estimation of lithium-ion batteries for real-time embedded applications. IEEE Embed Syst Lett 2021;13(4):206–9. <http://dx.doi.org/10.1109/LES.2021.3078443>.
- [4] Lipu MSH, Hannan MA, Karim TF, Hussain A, Saad MHM, Ayob A, Miah MS, Mahlia TMI. Intelligent algorithms and control strategies for battery management system in electric vehicles: Progress, challenges and future outlook. J Clean Prod 2021;292. <http://dx.doi.org/10.1016/j.jclepro.2021.126044>.
- [5] Xiong R, Pan Y, Shen WX, Li HL, Sun FC. Lithium-ion battery aging mechanisms and diagnosis method for automotive applications: Recent advances and perspectives. Renew Sustain Energy Rev 2020;131. <http://dx.doi.org/10.1016/j.rser.2020.110048>.
- [6] Yang XG, Leng YJ, Zhang GS, Ge SH, Wang CY. Modeling of lithium plating induced aging of lithium-ion batteries: Transition from linear to nonlinear aging. J Power Sources 2017;360:28–40. <http://dx.doi.org/10.1016/j.jpowsour.2017.05.110>.
- [7] Hu X, Xu L, Lin X, Pecht M. Battery lifetime prognostics. Joule 2020;4(2):310–46. <http://dx.doi.org/10.1016/j.joule.2019.11.018>.
- [8] Kim J, Lee S, Cho BH. Complementary cooperation algorithm based on DEKF combined with pattern recognition for SOC/Capacity estimation and SOH prediction. IEEE Trans Power Electron 2012;27(1):436–51. <http://dx.doi.org/10.1109/TPEL.2011.2158554>.
- [9] Wang L, Ma J, Zhao X, Li X, Zhang K, Jiao Z. Adaptive robust unscented Kalman filter-based state-of-charge estimation for lithium-ion batteries with multi-parameter updating. Electrochim Acta 2022;426. <http://dx.doi.org/10.1016/j.electacta.2022.140760>.
- [10] Sadabadi KK, Jin X, Rizzoni G. Prediction of remaining useful life for a composite electrode lithium ion battery cell using an electrochemical model to estimate the state of health. J Power Sources 2021;481. <http://dx.doi.org/10.1016/j.jpowsour.2020.228861>.

- [11] Zhang YZ, Wik T, Bergstrom J, Pecht M, Zou CF. A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data. *J Power Sources* 2022;526. <http://dx.doi.org/10.1016/j.jpowsour.2022.231110>.
- [12] Tanim TR, Rahn CD, Wang CY. State of charge estimation of a lithium ion cell based on a temperature dependent and electrolyte enhanced single particle model. *Energy* 2015;80:731–9. <http://dx.doi.org/10.1016/j.energy.2014.12.031>.
- [13] Yu Q, Nie Y, Liu S, Li J, Tang A. State of health estimation method for lithium-ion batteries based on multiple dynamic operating conditions. *J Power Sources* 2023;582. <http://dx.doi.org/10.1016/j.jpowsour.2023.233541>.
- [14] Tang A, Jiang Y, Yu Q, Zhang Z. A hybrid neural network model with attention mechanism for state of health estimation of lithium-ion batteries. *J Energy Storage* 2023;68. <http://dx.doi.org/10.1016/j.est.2023.107734>.
- [15] Bloom I, Christophersen J, Gering K. Differential voltage analyses of high-power lithium-ion cells 2. applications. *J Power Sources* 2005;139(1–2):304–13. <http://dx.doi.org/10.1016/j.jpowsour.2004.07.022>.
- [16] She CQ, Wang ZP, Sun FC, Liu P, Zhang L. Battery aging assessment for real-world electric buses based on incremental capacity analysis and radial basis function neural network. *IEEE Trans Ind Inf* 2020;16(5):3345–54. <http://dx.doi.org/10.1109/TII.2019.2951843>.
- [17] Mohtat P, Lee S, Siegel JB, Stefanopoulou AG. Comparison of expansion and voltage differential indicators for battery capacity fade. *J Power Sources* 2022;518. <http://dx.doi.org/10.1016/j.jpowsour.2021.230714>.
- [18] Merla Y, Wu B, Yufit V, Brandon NP, Martinez-Botas RF, Offer GJ. Novel application of differential thermal voltammetry as an in-depth state-of-health diagnosis method for lithium-ion batteries. *J Power Sources* 2016;307:308–19. <http://dx.doi.org/10.1016/j.jpowsour.2015.12.122>.
- [19] Wang ZP, Yuan CG, Li XY. Lithium battery state-of-health estimation via differential thermal voltammetry with Gaussian process regression. *IEEE Trans Transp Electr* 2021;7(1):16–25. <http://dx.doi.org/10.1109/TTE.2020.3028784>.
- [20] Zhou K, Wu Y, Wu X, Sun Y, Teng D, Liu Y. Research and development review of power converter topologies and control technology for electric vehicle fast-charging systems. *Electronics* 2023;12(7). <http://dx.doi.org/10.3390/electronics12071581>.
- [21] IONITY. Ionity (2017) joint venture of BMW, daimler, ford and volkswagen for deploying fast charging stations. 2017. <http://www.ionity.eu/ionity-en.html>.
- [22] Hu XS, Che YH, Lin XK, Deng ZW. Health prognosis for electric vehicle battery packs: A data-driven approach. *IEEE-Asme Trans Mechatronics* 2020;25(6):2622–32. <http://dx.doi.org/10.1109/TMECH.2020.2986364>.
- [23] Ayesha S, Hanif MK, Talib R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Inf Fusion* 2020;59:44–58. <http://dx.doi.org/10.1016/j.inffus.2020.01.005>.
- [24] Espadoto M, Martins RM, Kerren A, Hirata NST, Telea AC. Toward a quantitative survey of dimension reduction techniques. *IEEE Trans Vis Comput Graphics* 2021;27(3):2153–73. <http://dx.doi.org/10.1109/TVCG.2019.2944182>.
- [25] Li XY, Zhang L, Wang ZP, Dong P. Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and elman neural networks. *J Energy Storage* 2019;21:510–8. <http://dx.doi.org/10.1016/j.est.2018.12.011>.
- [26] Zhou Y, Zhao J, Song YJ, Sun JJ, Fu H, Chu M. A seasonal-trend-decomposition-based voltage-source-inverter open-circuit fault diagnosis method. *IEEE Trans Power Electron* 2022;37(12):15517–27. <http://dx.doi.org/10.1109/TPEL.2022.3190937>.
- [27] Cui DS, Wang ZP, Liu P, Wang S, Zhang ZS, Dorrell DG, Li XH. Battery electric vehicle usage pattern analysis driven by massive real-world data. *Energy* 2022;250. <http://dx.doi.org/10.1016/j.energy.2022.123837>.
- [28] Deng Z, Xu L, Liu H, Hu X, Duan Z, Xu Y. Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles. *Appl Energy* 2023;339. <http://dx.doi.org/10.1016/j.apenergy.2023.120954>.
- [29] Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol* 2014;14:1–13.
- [30] Yin AJ, Tan ZB, Tan J. Life prediction of battery using a neural Gaussian process with early discharge characteristics. *Sensors* 2021;21(4). <http://dx.doi.org/10.3390/s21041087>, 1087 33562499.
- [31] Alipour M, Tavallaey SS, Andersson AM, Brandell D. Improved battery cycle life prediction using a hybrid data-driven model incorporating linear support vector regression and Gaussian. *Chemphyschem* 2022;23(7). <http://dx.doi.org/10.1002/cphc.202100829>, e202100829 35075749.
- [32] Lin XK, Khosravinia K, Hu XS, Li J, Lu W. Lithium plating mechanism, detection, and mitigation in lithium-ion batteries. *Prog Energy Combust Sci* 2021;87:100953. <http://dx.doi.org/10.1016/j.pecs.2021.100953>.
- [33] Birkel CR, Roberts MR, McTurk E, Bruce PG, Howey DA. Degradation diagnostics for lithium ion cells. *J Power Sources* 2017;341:373–86. <http://dx.doi.org/10.1016/j.jpowsour.2016.12.011>.
- [34] Li Y, Liu KL, Foley AM, Zulke A, Berecibar M, Nanini-Maury E, Van Mierlo J, Hoster HE. Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review. *Renew Sustain Energy Rev* 2019;113:109254. <http://dx.doi.org/10.1016/j.rser.2019.109254>.
- [35] Tang XP, Zou CF, Yao K, Chen GH, Liu BY, He ZW, Gao FR. A fast estimation algorithm for lithium-ion battery state of health. *J Power Sources* 2018;396:453–8. <http://dx.doi.org/10.1016/j.jpowsour.2018.06.036>.
- [36] Feng F, Hu X, Hu L, Hu F, Li Y, Zhang L. Propagation mechanisms and diagnosis of parameter inconsistency within li-ion battery packs. *Renew Sustain Energy Rev* 2019;112:102–13. <http://dx.doi.org/10.1016/j.rser.2019.05.042>.
- [37] Ma S, Jiang M, Tao P, Song C, Wu J, Wang J, Deng T, Shang W. Temperature effect and thermal impact in lithium-ion batteries: A review. *Prog Nat Sci-Mater Int* 2018;28(6):653–66. <http://dx.doi.org/10.1016/j.pnsc.2018.11.002>.
- [38] Waldmann T, Wilka M, Kasper M, Fleischhammer M, Wohlfahrt-Mehrens M. Temperature dependent ageing mechanisms in lithium-ion batteries - a post-mortem study. *J Power Sources* 2014;262:129–35. <http://dx.doi.org/10.1016/j.jpowsour.2014.03.112>.
- [39] Guan T, Sun S, Yu FB, Gao YZ, Fan P, Zuo PJ, Du CY, Yin GP. The degradation of LiCoO₂/graphite batteries at different rates. *Electrochim Acta* 2018;279:204–12. <http://dx.doi.org/10.1016/j.electacta.2018.04.197>.
- [40] Capacity fading of LiAl_{0.5}Ni_{1-x}CoxO₂ cathode for lithium-ion batteries during accelerated calendar and cycle life tests (effect of depth of discharge in charge-discharge cycling on the suppression of the micro-crack generation of LiAl_{0.5}Ni_{1-x}CoxO₂ particle). *J Power Sources* 2014;260:50–6. <http://dx.doi.org/10.1016/j.jpowsour.2014.02.103>.
- [41] Zhou Y, Huang M, Chen Y, Tao Y. A novel health indicator for on-line lithium-ion batteries remaining useful life prediction. *J Power Sources* 2016;321:1–10. <http://dx.doi.org/10.1016/j.jpowsour.2016.04.119>.
- [42] Piao C, Sun R, Chen J, Liu M, Wang Z. A feature extraction approach for state-of-health estimation of lithium-ion battery. *J Energy Storage* 2023;73:108871. <http://dx.doi.org/10.1016/j.est.2023.108871>.
- [43] Wang Q, Wang Z, Liu P, Zhang L, Sauer DU, Li W. Large-scale field data-based battery aging prediction driven by statistical features and machine learning. *Cell Rep Phys Sci* 2023;4(12):101720. <http://dx.doi.org/10.1016/j.xcrp.2023.101720>.
- [44] Iversen GR, Gergen M. *Statistics: the conceptual approach*. Springer Science & Business Media; 2012.
- [45] Qin L, Li WD, Li SJ. Effective passenger flow forecasting using STL and ESN based on two improvement strategies. *Neurocomputing* 2019;356:244–56. <http://dx.doi.org/10.1016/j.neucom.2019.04.061>.
- [46] Deng ZW, Hu XS, Li PH, Lin XK, Bian XL. Data-driven battery state of health estimation based on random partial charging data. *IEEE Trans Power Electron* 2022;37(5):5021–31. <http://dx.doi.org/10.1109/TPEL.2021.3134701>.
- [47] Tran MK, Panchal S, Chauhan V, Brahmabhatt N, Mevawalla A, Fraser R, Fowler M. Python-based scikit-learn machine learning models for thermal and electrical performance prediction of high-capacity lithium-ion battery. *Int J Energy Res* 2022;46(2):786–94. <http://dx.doi.org/10.1002/er.7202>.
- [48] Jafari S, Byun YC. XGBoost-based remaining useful life estimation model with extended Kalman particle filter for lithium-ion batteries. *Sensors* 2022;22(23). <http://dx.doi.org/10.3390/s22239522>, 9522 36502223.
- [49] Zhang Y, Li YF. Prognostics and health management of lithium-ion battery using deep learning methods: A review. *Renew Sustain Energy Rev* 2022;161:112282. <http://dx.doi.org/10.1016/j.rser.2022.112282>.
- [50] Pelletier S, Jabali O, Laporte G, Veneroni M. Battery degradation and behaviour for electric vehicles: Review and numerical analyses of several models. *Transp Res B* 2017;103:158–87. <http://dx.doi.org/10.1016/j.trb.2017.01.020>, SI.
- [51] Zhang D, Wang Z, Peng L, Qin Z, Wang Q, She C, Bauer P. Multi-step fast charging based state of health estimation of lithium-ion batteries. *IEEE Trans Transp Electr* 2023;1. <http://dx.doi.org/10.1109/TTE.2023.3322582>.