Article

A Data-Driven LiFePO$_4$ Battery Capacity Estimation Method Based on Cloud Charging Data from Electric Vehicles

Xingyu Zhou, Xuebing Han *, Yanan Wang ⓘ, Languang Lu and Minggao Ouyang *

State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China
* Correspondence: hanxuebing@tsinghua.edu.cn (X.H.); ouymg@mail.tsinghua.edu.cn (M.O.)

Abstract: The accuracy of capacity estimation is of great importance to the safe, efficient, and reliable operation of battery systems. In recent years, data-driven methods have emerged as promising alternatives to capacity estimation due to higher estimation accuracy. Despite significant progress, data-driven methods are mainly developed by experimental data under well-controlled charge–discharge processes, which are seldom available for practical battery health monitoring under realistic conditions due to uncertainties in environmental and operational conditions. In this paper, a novel method to estimate the capacity of large-format LiFePO$_4$ batteries based on real data from electric vehicles is proposed. A comprehensive dataset consisting of 85 vehicles that has been running for around one year under diverse nominal conditions derived from a cloud platform is generated. A classification and aggregation capacity prediction method is developed, combining a battery aging experiment with big data analysis on cloud data. Based on degradation mechanisms, IC curve features are extracted, and a linear regression model is established to realize high-precision estimation for slow-charging data with constant-current charging. The selected features are highly correlated with capacity (Pearson correlation coefficient < 0.85 for all vehicles), and the MSE of the capacity estimation results is less than 1 Ah. On the basis of protocol analysis and mechanism studies, a feature set including internal resistance, temperature, and statistical characteristics of the voltage curve is constructed, and a neural network (NN) model is established for multi-stage variable-current fast-charging data. Finally, the above two models are integrated to achieve capacity prediction under complex and changeable realistic working conditions, and the relative error of the capacity estimation method is less than 0.8%. An aging experiment using the battery, which is the same as those equipped in the vehicles in the dataset, is carried out to verify the methods. To the best of the authors' knowledge, our study is the first to verify a capacity estimation model derived from field data using an aging experiment of the same type of battery.

Keywords: capacity estimation; data-driven method; real vehicle data; charging cloud data; electric vehicle

1. Introduction

To reduce greenhouse gas emissions, the transition of the automotive industry toward electric vehicles (EVs) is inevitable and needs to progress swiftly [1,2]. Due to their advantages of high energy density, high power density, and low self-discharge rate, lithium-ion batteries (LIBs) have been widely used as power sources in new energy vehicles in recent years [3]. However, the battery life is unable to satisfy the demands of users, becoming one of the bottlenecks for the further development of batteries [4,5]. Therefore, it is necessary to accurately evaluate the state of health (SOH) of batteries to ensure the safe, efficient, and reliable operation of battery systems and to optimize battery management [6]. Improvements in SOH predictive performance have contributed considerably to the higher penetration of electric vehicles in the market, as they reduce risks and costs by extending battery life. This makes SOH estimation a crucial part of vehicle development.

Capacity is an important indicator of battery SOH estimation and is used to evaluate the degree of battery aging. When the capacity drops below a certain value, normally 80%

of nominal capacity, the battery reaches its end of life (EOL) and cannot work normally anymore [5]. Battery capacity estimation methods can be divided into physical-model-based, equivalent-circuit-model-based, empirical-model-based, and data-driven [7]. Physical models are based on the electrochemical processes of complex multi-physics and multi-scale material systems. The pseudo-two-dimensional (P2D) model [8–13] or single-particle model [14–17] is combined with mathematical side reaction expressions, including the growth of the solid electrolyte interphase (SEI) film, lithium plating, loss of lithium inventory (LLI), and loss of active material (LAM), to form a capacity fade model. Although physical models can yield high accuracy, they can hardly be used in realistic applications due to model complexity and difficulties in parameter identification. Equivalent circuit models (ECMs) based on the combination of circuit elements (e.g., resistors and capacitors) describe battery dynamic response and degradation behavior with fewer model parameters and higher computational efficiency [18–21]. However, as model parameters are difficult to update under realistic conditions, the shortcomings of their limited accuracy and robustness cannot be avoided, so the update of model parameters is generally determined by electrochemical impedance spectroscopy (EIS) testing in laboratories. Instead of considering complicated physical and chemical side reactions inside the cell, the empirical and half-empirical model estimates capacity based on experimental data or publicly available cell testing datasets. Empirical models, including cycle aging models [22,23] and calendar aging models [24,25], are fit to data to separately analyze battery working operation (cycle life) and standby or storage mode (calendar life). Then, coupled models [26–30] are generated to describe aging behavior based on known impact factors such as depth of discharge (DOD), temperature [31,32], and current [33,34]. However, aging experiments are so time-consuming and costly that significant interpolation and extrapolation are generally needed in realistic aging scenarios. Moreover, the traditional empirical model is an open-loop model without considering cell-to-cell variations, and it is difficult to guarantee the accuracy of the estimation results [35].

In recent years, data-driven methods using statistical theory and machine learning methods have emerged as promising mechanism-agnostic alternatives for capacity estimation due to higher estimation accuracy. Severson et al. [36] constructed a dataset containing various fast-charging protocols and extracted several statistical features from the capacity–voltage curves of the constant-current discharging process of the first 100 cycles and modeled the relationship between these features and the remaining useful life based on an elastic net. The results demonstrated a prediction error of less than 9.1%. Su et al. [37] established a conventional neural network (CNN) model based on automatically extracted features from 40 cycles of full-discharge voltage profiles to predict cycle life using a dataset containing various experimental conditions such as temperature and charging current. The model was able to achieve a 9.28% error in the training set and a 22.73% error in the testing set. Zhu et al. [38] estimated battery capacity using the XGBoost method by features derived from 30 min relaxation voltage profiles based on a large dataset containing various currents and temperatures. The best model achieved a root-mean-square error of 1.1% in the training set, and the root-mean-square error of the transfer learning model was less than 1.7% on the datasets used for model validation. Thelen et al. [39] established a multi-output Gaussian process (MOGP) regression model and an extreme learning machine (ELM) to estimate the capacity of battery cells and to diagnose their primary degradation modes using incremental capacity (IC) data with a voltage range from 3.4 V to 4.075 V. Tian et al. [40,41] established a CNN model to estimate capacity and state of charge (SOC) or to predict the charging curve using pieces of constant-current charging data. Mohtat et al. [42] and Samad et al. [43] established a capacity estimation model using the linear regression method based on expansion and force features. The features remained observable at a current rate lower than 1C and were robust when charging commenced from different SOC. However, although data-driven approaches perform well on training datasets, the performance of such feature-based machine learning models in real-life usage scenarios under new conditions is challenging because of the limited extrapolation ability of data-

driven models. Moreover, data-driven methods developed by experimental data under well-controlled charge–discharge processes, e.g., constant current–constant voltage (CC-CV), generally require certain working environments, such as constant-current charging or long relaxation times, which are seldom available for on-board battery health monitoring under realistic conditions due to uncertainties in environmental and operational conditions, to estimate battery life.

As a result, in order to be effective under realistic operational conditions, the characteristics are supposed to be independent of instantaneous operating conditions, and the models should be appropriate for realistic operational conditions in EV applications, where random charging cycles, dynamic discharging protocols, and noisy data, as well as uncertain boundary conditions (e.g., varying SOC and temperature range in each cycle), exist [44–48]. Over the past few years, due to improvements in computing ability and the appearance of the Internet of Things (IoT), cloud-based battery management systems (BMS) have brought data-driven techniques using field data into battery capacity estimation [49–51]. Zhao et al. [44] established a stacking ensemble learning capacity estimation model based on field data. However, the model has a very strict requirement for the charging pieces, and it can only be applied in the slow-charging process with a very large DOD. This work strives to address the gap in data-driven capacity estimation methods developed by experimental data and capacity estimation under realistic conditions. A cloud-based data-driven framework for battery capacity estimation is proposed based on a sample of 85 vehicles that has been running for around one year under diverse nominal conditions, as shown in Figure 1.
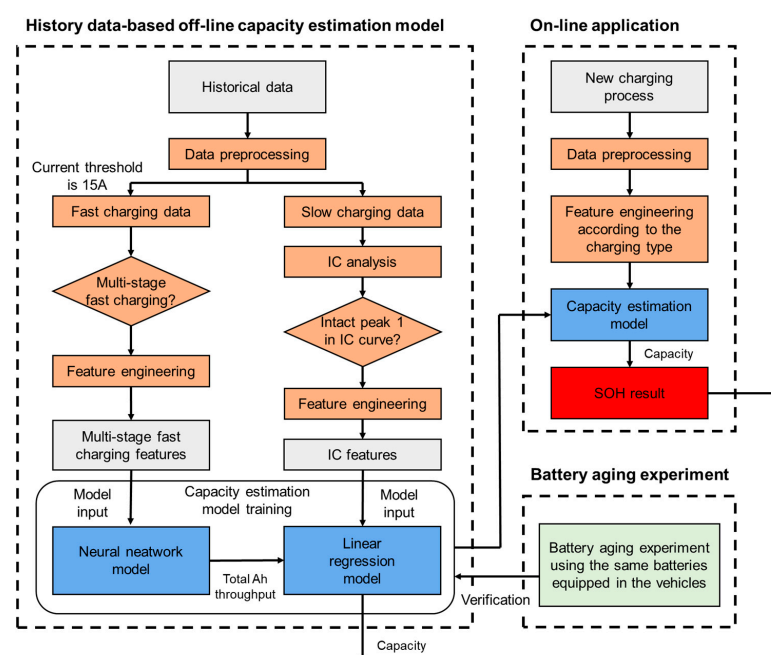
Figure 1. A cloud-based framework for battery capacity estimation in EV applications.

For slow-charging data, a linear regression capacity estimation model is established based on IC features, combining a battery aging experiment with big data analysis on cloud data. For fast-charging data, on the basis of protocol analysis and mechanism studies, a feature set including internal resistance, temperature, and statistical characteristics of the voltage curve is constructed, and a neural network model is established. A neural network is used to estimate the capacity based on the features of multi-stage fast charging. Finally, the above two models are integrated to achieve capacity prediction under complex and changeable actual working conditions. An aging experiment using the battery, which is the same as those equipped in the vehicles in the dataset, is carried out to verify the methods.

The rest of this paper is organized as follows. Section 2 briefly describes the basic information of real vehicle datasets and data preprocessing methods. Capacity estimation

using slow-charging data and capacity estimation using fast-charging data are presented in Sections 3 and 4, respectively. Section 5 summarizes the main results and draws conclusions.

2. Vehicle Data and Data Preprocessing Methods

2.1. Real Vehicle Data Overview

For this investigation, a large dataset with a total of 85 EVs with widely varying cycle numbers ranging from 115 to 1302 cycles is randomly collected from a cloud monitoring system, with around one year under diverse nominal conditions. The initial SOH of all the vehicle batteries equals 1, which means that the dataset contains the information for each vehicle from the beginning of life. The 85 vehicles are randomly selected to include as many different user behaviors as possible and to verify the robustness of the proposed method. Different user behaviors include different charging methods, different operation temperatures, different working loads, and so on. Vehicles are randomly named from vin (vehicle identification number) 1 to vin 85 for further analysis. A total of 178 cells are interconnected in a 178S1P manner in the battery pack. The cells equipped are commercial large-format LiFePO$_4$ cells and have a nominal capacity of 135 Ah and a nominal voltage of 3.2 V, with an operating voltage window between 2 V and 3.8 V. Detailed battery parameters are listed in Table 1.

Table 1. Battery nominal parameters.

| Item	Specification
Cathode material	LiFePO$_4$
Anode material	Graphite
Nominal capacity	135 Ah
Charging cutoff voltage	3.8 V
Discharging cutoff voltage	2.0 V

The dataset contains 25,031 cycles in total with widely varying operating voltage windows and charging protocols. On account of the cost of massive data storage, not all the relevant battery data measured are transmitted to the cloud nowadays, and the sampling interval of the cloud data is generally up to 30 s in the dataset. The bulk of the data available for this work includes over 49,377,239 measurements of 7 parameters or status over the life of the vehicle. These parameters or status include timestamps, voltage, SOC, discharge/charge C-rate, temperature, charging status, and vehicle status (see Table 2). It should be highlighted that the SOC values in most of the measurements are missing. Moreover, the accuracy of the SOC is not so good due to the voltage plateau in lithium-ion phosphate (LFP) batteries. As a result, traditional capacity estimation [35,52] methods based on SOC cannot be applied in real vehicle datasets. Timestamps show the datetime of every measurement in the format 'year-month-day hour-minute-second'. Although the voltage of each single cell is given, the cell that is the first to reach the upper cutoff voltage during charging is chosen for this research, considering that this cell is the limiting factor in the pack. The discharge/charge C-rate provides the discharge or charge current rate for cells, where a positive number means discharging, while a negative number means charging. Although temperatures are measured at 12 different locations in the battery pack and are saved as a sequence, the average of these values is used in further analysis on account of the lack of knowledge of the specific position of each temperature sensor. Four charging states are included in the dataset, i.e., charging finished, charging when parking, charging when driving, and uncharged, and two vehicle states are included in the dataset, i.e., power on and power off.

Table 2. An overview of the measured parameters or status available in the datasets.

Parameters or Status	Description	Note or Comment
Timestamps	Datetime of every measurement.	
Voltage	Voltage of each single cell.	
SOC	SOC value of battery packs.	Missing frequently or incorrect.
Discharge/Charge C-rate	Discharge or charge current rate for cells.	Positive number means discharging, while negative number means charging.
Temperature	Temperatures measured at 12 different locations in the batteries.	The average of these values is used in further analysis.
Charging status	Four charging states including charging finished, charging when parking, charging when driving, and uncharged.	
Vehicle status	Two vehicle states including power on and power off.	

An overview of the histograms, showing the distributions of several features collected and calculated based on the vehicle data in the dataset, is shown in Figure 2. The distribution of cycle numbers, total ampere-hour throughput, and operation time of these 85 vehicles in the dataset are shown in Figure 2a–c, respectively. As we can see, the dataset contains different aging levels with widely varying service times ranging from 331 days to 525 days and widely varying total ampere-hour throughputs ranging from 8025 Ah to 42,599 Ah, which is equivalent to 60 to 315 full cycles. The distribution of start charging voltage, end charging voltage, and charging capacity of all the 25,031 charging processes of these 85 vehicles in the dataset are shown in Figure 2d–f, respectively. With stochastic charging cycles, this real-world operational dataset reflects realistic conditions, which exhibit irregular cycling patterns and varying voltage ranges. Thus, the dataset can provide a complete picture of how cells age under realistic conditions.
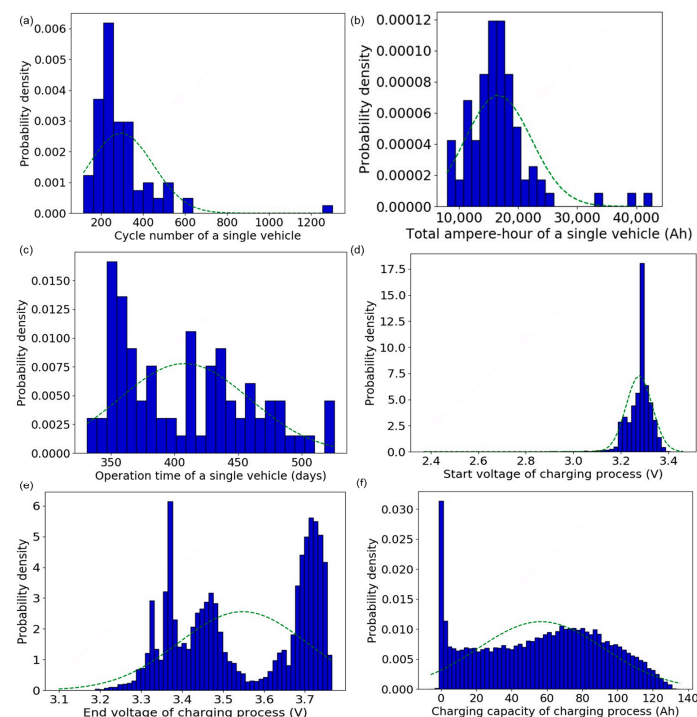
Figure 2. The histogram distribution of several features collected and calculated based on the vehicle data: (**a**) Cycle number distribution; (**b**) Total ampere-hour throughput distribution; (**c**) Operation time distribution; (**d**) Start charging voltage distribution of all the 25,031 charging processes; (**e**) End charging voltage distribution of all the charging processes; (**f**) Charging capacity distribution of all the charging processes.

Besides the start and end voltages, the charging processes are diverse, dependent on the charging protocols and station with an on-board BMS due to the dynamic operating conditions and user behavior. However, generally speaking, the charging protocols can be divided into two modes for the vehicle samples in the dataset: slow charging and fast charging. The protocol of slow charging is quite fixed, and a typical example of a slow-charging protocol is shown in Figure 3. The current is approximately constant, with a magnitude of 10.5 A, and temperature either rises slowly or fluctuates in a small range. Instead, the protocols of fast charging are highly varied, which will be discussed in detail in Section 4.
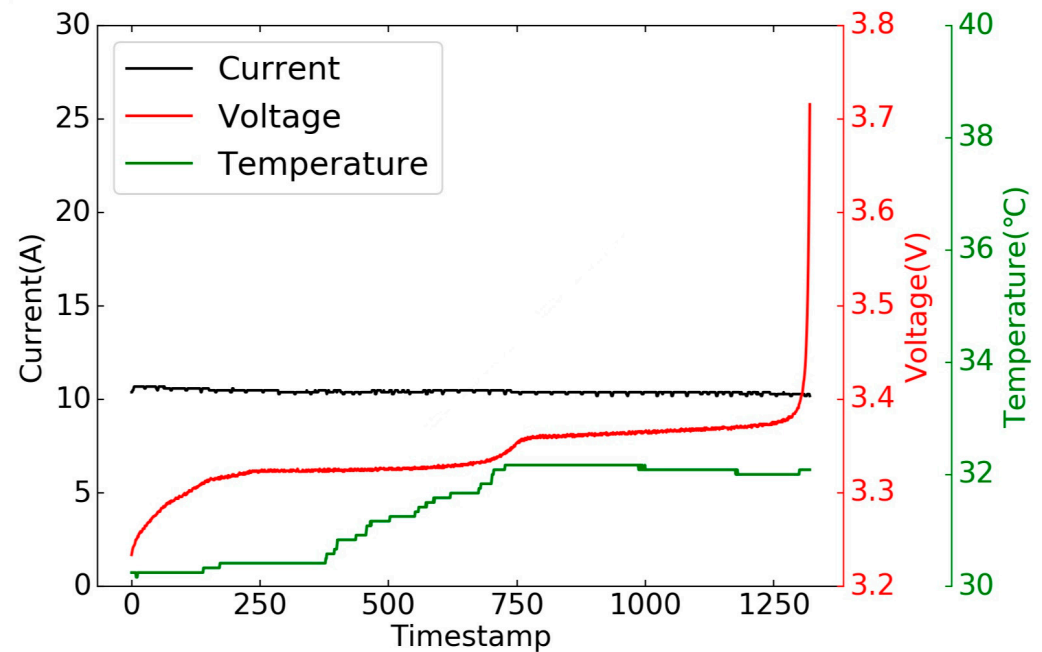
Figure 3. A typical example of a slow-charging protocol. Current, voltage, and temperature versus timestamp is given. The current is approximately constant, with a magnitude of 10.5 A, and temperature either rises slowly or fluctuates in a small range.

2.2. Data Preprocessing

In this research, the dataset is derived directly from real vehicle data, which are generated during the daily operation of EVs and transmitted to the cloud. Hence, data processing will be indispensable to select useful information and split data at a suitable charging stage for further model establishment. Figure 4 shows a typical data sample, i.e., raw data in the dataset, where some data quality problems exist and need to be dealt with as follows: (1) Partial data are unordered, and the timestamps of some variables do not match each other, resulting in misalignment. For instance, some of the switching points of current and voltage in multi-stage constant-current charging protocol do not correspond to the same timestamp, but to two adjacent timestamps. (2) Some voltage, temperature, and current information of data points is missing (NaN). (3) Time intervals between two frames in a segment are not the same as the standard sample interval (30 s), and the larger time interval leads to data discontinuity.

In view of the above data quality problems, the following data preprocessing measures are adopted: data NaN inspection, data splitting, and selection and data continuity inspection. The flowchart of the whole data preprocessing process is shown in Figure 5. First of all, the raw data are sorted in ascending order according to timestamps. Then, data NaN inspection is conducted, and the faulty data that do not contain valuable information of the voltage and the current are deleted. Otherwise, the NaN values are replaced by the mean values of two adjacent frames. In the next step, the charging data are selected and split to create the dataset for model training and validation, for the reason that charging

data are stable compared to the dynamic driving condition, i.e., discharging data, which makes it possible to estimate the battery capacity. To be specific, the raw data are selected on the basis of vehicle status value with the purpose of selecting the continuous charging segments, i.e., vehicle status equaling 'charging when parking'. In the step of data splitting, if no more than three consecutive frames of data are missing in the segment formed after the selecting process, the segment is regarded as the same charging process, with the aim of ensuring robustness to data missing. Otherwise, the segment will be split into two different charging processes when the missing data are more than three frames. Finally, data continuity inspection is carried out, and the segment, where the time intervals of data are larger than 120 s, or time intervals of data larger than 90 s are more than three, is abandoned. If the number of frames in a charging segment is less than 10, which suggests that the information extracted from this segment is not sufficient for capacity estimation, this piece of charging data will be discarded. After the data preprocessing, the dataset is further divided into two sub-datasets, i.e., slow-charging dataset and fast-charging dataset, according to the maximal current, and the threshold value is set as 15 A. The slow-charging dataset and the fast-charging dataset, including 6158 and 2640 charging pieces, respectively, are used in Sections 3 and 4 for capacity estimation.

Index	Charging status	Current	SOC	Vehicle status	Data time	Voltage	Temperature
0	NaN	NaN	NaN	NaN	2020-10-22 12:38:49	NaN	NaN
1	uncharged	1.6	NaN	power on	2020-10-22 12:39:19	NaN	NaN
2	NaN	NaN	NaN	NaN	2020-10-22 12:39:49	NaN	NaN
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
197522	charging when parking	−119.8	NaN	power off	2021-03-03 16:11:00	3.507	[26,28,28,...]
197523	charging when parking	−111.5	NaN	power off	2021-03-03 16:11:30	3.448	[26,28,28,...]
197524	charging when parking	−67.9	NaN	power off	2021-03-03 16:12:00	3.444	[26,28,28,...]
197525	charging when parking	−68.3	NaN	power off	2021-03-03 16:12:30	3.443	[26,28,28,...]
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
284109	charging when parking	−10.2	NaN	power off	2021-05-11 06:42:15	3.439	[27,28,28,...]
284110	charging finished	−10.2	NaN	power off	2021-05-11 06:42:45	3.441	[27,28,28,...]
284111	NaN	NaN	NaN	NaN	2021-05-11 06:42:57	NaN	NaN
284112	NaN	NaN	NaN	NaN	2021-05-11 11:47:47	NaN	NaN
284113	uncharged	2.8	NaN	power on	2021-05-11 11:48:29	3.359	[25,25,26,...]
284114	uncharged	2.4	NaN	power on	2021-05-11 11:48:59	3.357	[25,25,26,...]
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Some voltage, temperature and current information of data points are missing (NaN).

The timestamps of some variables do not match each other.

Data discontinuity.

Figure 4. A typical piece of data sample, i.e., raw data in the dataset. Some data quality problems including NaN, data discontinuity, and data mismatch according to timestamps appear in the raw data.
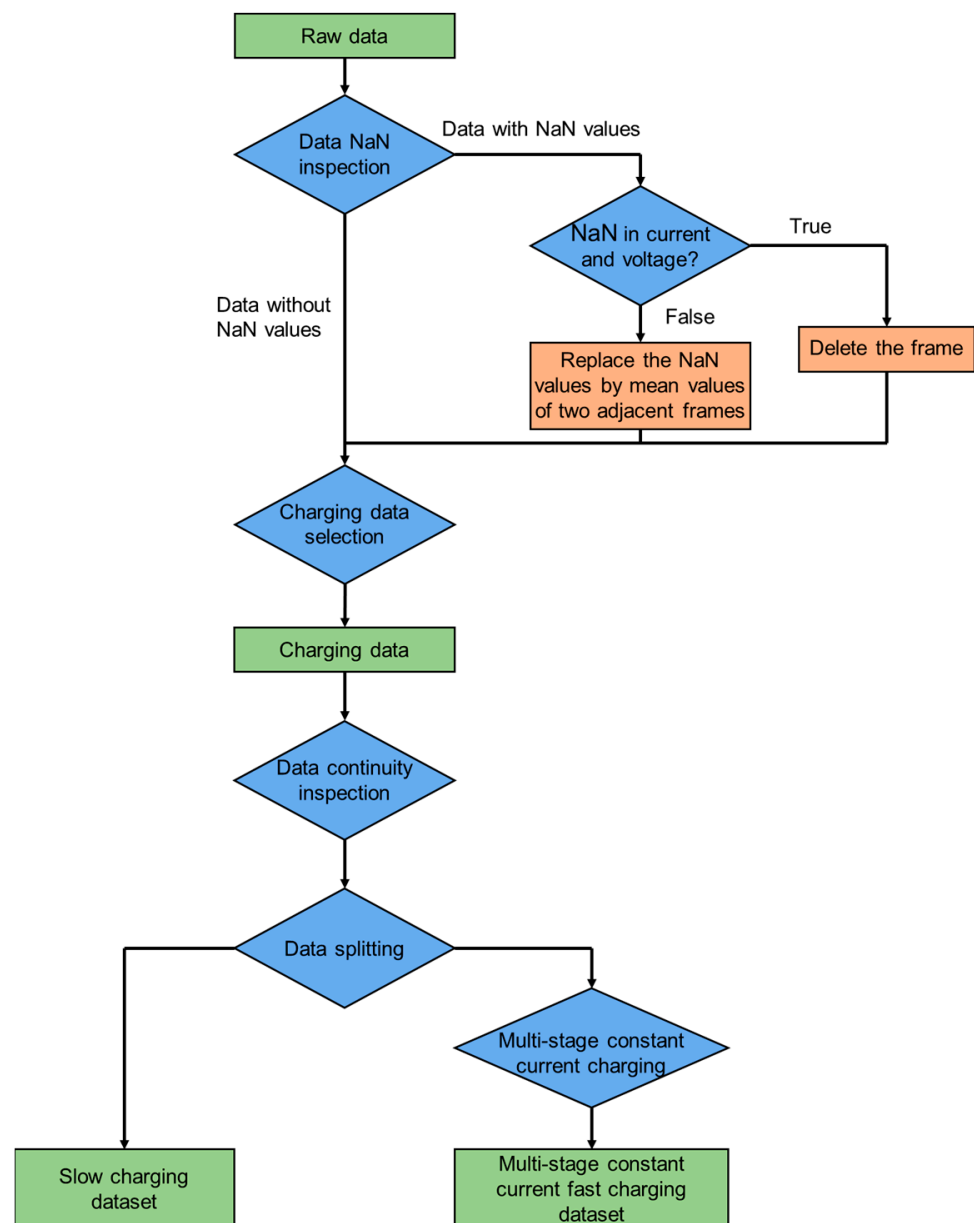
Figure 5. The flowchart of the data preprocessing process.

3. Capacity Estimation Using Slow-Charging Data

3.1. IC Analysis and Feature Selection

Over the past few years, IC and DV analyses have been extensively used for battery capacity estimation, and have demonstrated a remarkable ability to accurately locate degradation modes, which include LLI and LAM in the anode and cathode [28,53–56]. For LFP batteries, there are three obvious peaks in the IC curve, which correspond to the phase transformation processes of the graphite anode, and the three corresponding phase transformation processes are denoted as Peak 1, Peak 2, and Peak 5, respectively [56]. The voltage curve and the IC curve of the fresh cell, which is identical to the cells used in real vehicles, are shown in Figure 6. The curve is derived from the constant-current charging process with the use of cycling devices in the laboratory before the aging experiment, with a current of 10.5 A. It can be clearly seen that the three peaks of the IC curve correspond to the three plateaus of the voltage curve.
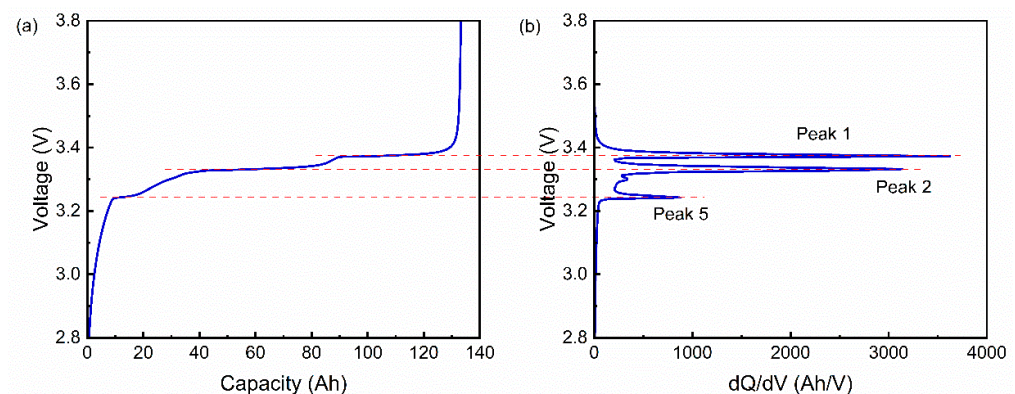
Figure 6. Voltage curve and IC curve of the fresh cell, derived from the constant-current charging process with the use of cycling devices in the laboratory, with a current of 10.5 A. There are three obvious peaks in the IC curve, which correspond to the phase transformation processes of the graphite anode. (**a**) Voltage curve of the fresh cell; (**b**) IC curve of the fresh cell.

Previous studies have shown that the height, envelope area, and location of IC peaks, which are highly related to cell degradation, are great indicators of battery capacity and degradation mechanisms [57,58]. In the traditional method, these knowledge-based features of all IC peaks are used as inputs for machine learning modeling [42,44]. However, in practice, the charging process of EVs is very stochastic and, in most cases, would not cover the SOC range from 0 to 100%, which means that all these features could be hardly simultaneously obtained in a charging process. Thus, these early efforts can be barely applied in practice. As shown in Figure 2d,e, most of the charging processes cover the voltage range from 3.34 V to 3.4 V, indicating that Peak 1 is available in the majority of charging segments. In addition, as previous studies show, the degradation of Peak 1 is expected to arise from the LLI process, which is considered to be the major degradation mode of LFP batteries [56,57]. Moreover, the position of Peak 1 implies the internal resistance change of the batteries. As a consequence, information of Peak 1 could be high-quality features for capacity estimation under real vehicle conditions. For real vehicle data with a large sampling interval and noise, the level evaluation analysis (LEAN) method [59], of which accuracy and reproducibility are proven by mathematical arguments, is used to create the IC curve.

As for the specific definition of Peak 1, the traditional method with a fixed voltage range is not applicable to vehicular applications, as the peak location and the start and the end of the peak might slightly change due to measurement noise and current fluctuation. Figure 7 demonstrates the specific definition and acquisition method of Peak 1 with an example. The IC curve is derived from a stochastically selected charging data piece from the real vehicle dataset. As we can see, the derived IC curve from practical conditions inevitably exhibits explicit ripples compared with that from laboratory conditions, as shown in Figure 6. The difference in the height of the IC peak between real vehicle data and experimental data is due to the precision of the voltage measurement and the sampling interval. Moreover, a relative degradation can be found in Peak 1 compared to the fresh cell on account of LLI. The whole peak, including the peak point and the start and the end of the peak, should be in the voltage range from 3.34 V to 3.4 V. The voltage and IC value of the peak point is defined as

$$IC_{peak} = \max(IC(n))$$
$$\text{s.t.} 3.34 \leq V(n) \leq 3.4 \tag{1}$$

$$n_{peak} = IC^{-1}\left(IC_{peak}\right) \tag{2}$$

$$V_{peak} = V\left(n_{peak}\right) \tag{3}$$

where $IC(n)$ and $V(n)$ are the IC and voltage series of a charging process, respectively. n_{peak} is the index of the peak point in the charging process. IC_{peak} is compared with the threshold, which is set as 4000, to confirm that the peak is intact. The voltage of the start and the end of Peak 1 is defined as

$$V_{start} = \max(V(n))$$
$$\text{s.t.} \; 3.34 \leq V(n) \leq V_{peak} \tag{4}$$
$$IC(n) \leq 400$$

$$V_{end} = \min(V(n))$$
$$\text{s.t.} \; V_{peak} \leq V(n) \leq 3.4 \tag{5}$$
$$IC(n) \leq 400$$
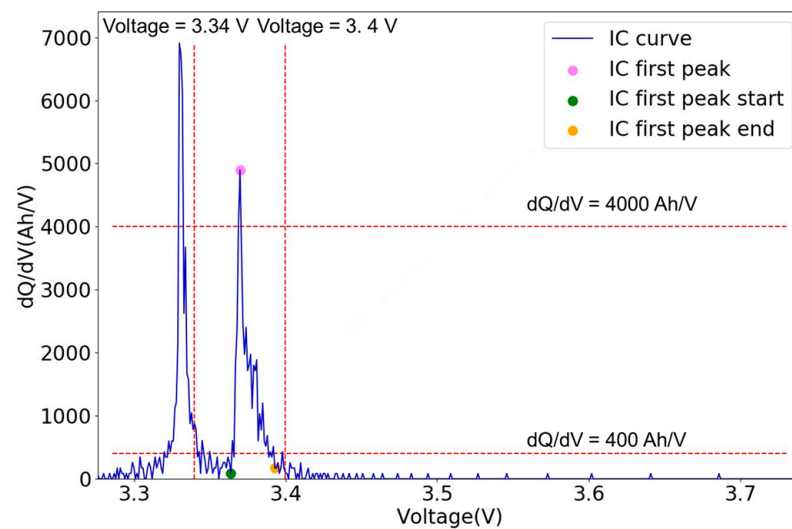
Figure 7. The specific definition and acquisition method of Peak 1. The IC curve is derived from a stochastically selected charging data piece from the real vehicle dataset. The whole peak should be in the voltage range from 3.34 V to 3.4 V. The peak point corresponds to the point with the highest IC value in the voltage range from 3.34 V to 3.4 V.

The envelope area of Peak 1 corresponds to the capacity with the voltage range from the start of Peak 1 to the end of Peak 1.

3.2. Linear Regression Model

As mentioned above, the envelope area of Peak 1, which is highly correlated with LLI and total capacity degradation, is chosen as the battery SOH indicator and model output. The correlation between the envelope area of Peak 1 and total capacity degradation will be further validated by experiments in Section 3.3. The model inputs consist of some impact factors, which have a high correlation with battery aging, i.e., total ampere-hour throughput, cycle number, average temperature, average current, and calendar life. Total ampere-hour throughput can be calculated as:

$$Ah \; throughput = \sum_{k=1}^{n} I_k \Delta t \tag{6}$$

where I_k means the current at the timestamp k, and Δt is the sampling interval, though the method is highly dependent on sensor accuracy. However, total ampere-hour throughput is calculated based on the integral of the current from the beginning of life, which means a very long time. If the error of the current is normally distributed, the errors will cancel each other out over a long period of time.

Linear regression models, with computational availability and reliability, are promising for vehicular applications. Moreover, linear equations contribute to better interpretability for battery degradation. A linear model of the form

$$\hat{y}_i = \hat{w}^T x_i \tag{7}$$

is established, where $\hat{y}_i$ means the predicted value of the envelope area of Peak 1 of the charging process i, x_i is the n-dimensional input feature vector, and $\hat{w}$ is the n-dimensional weight vector. The ordinary least squares (OLS) method is used to minimize the sum of squared residuals to find the weight $\hat{w}$, which can be calculated as:

$$\hat{w} = argmin_w \|y - Xw\|_2^2 \tag{8}$$

The mean squared error (MSE) and coefficient of determination are chosen to evaluate the model performance. The MSE is defined as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{9}$$

where y_i is the observed value of the envelope area of Peak 1 of the charging process i, and n is the total number of samples. The coefficient of determination is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{10}$$

where $\overline{y}$ is the mean value of all the observed values of the envelope area of Peak 1. R^2 indicates the fitting capacity of the model, normally ranging from 0 to 1. $R^2 = 1$ means that, in an ideal case, all the predicted values are equal to the true values, while $R^2 = 0$ means that the prediction performance equals just simply using the mean value for prediction. When the value of R^2 is greater than 0.8, we can regard the model as a good model with strong prediction ability.

Feature selection, a critical procedure in data-driven methods, is conducted to optimize the accuracy and robustness of the model. The input features are selected according to the Pearson correlation coefficient and the model performance change when including or excluding a certain feature, as shown in Table 3. The best features are selected when the model achieves the best performance of the MSE and R^2. In this study, total ampere-hour throughput is chosen as the best feature, as the model works the best with only total ampere-hour throughput as the model input. Hence, a single-factor linear regression is established for the envelope area of Peak 1.

Table 3. Pearson correlation coefficient between envelope area of Peak 1 and different features.

Feature	Pearson Coefficient
Total ampere-hour throughput	−0.88
Cycle number	−0.78
average temperature	−0.12
average current	0.02
calendar life	−0.74

3.3. Results

To show an example for the results of the linear regression model, six EVs with maximum slow-charging processes in the dataset are selected and displayed in Figure 8.
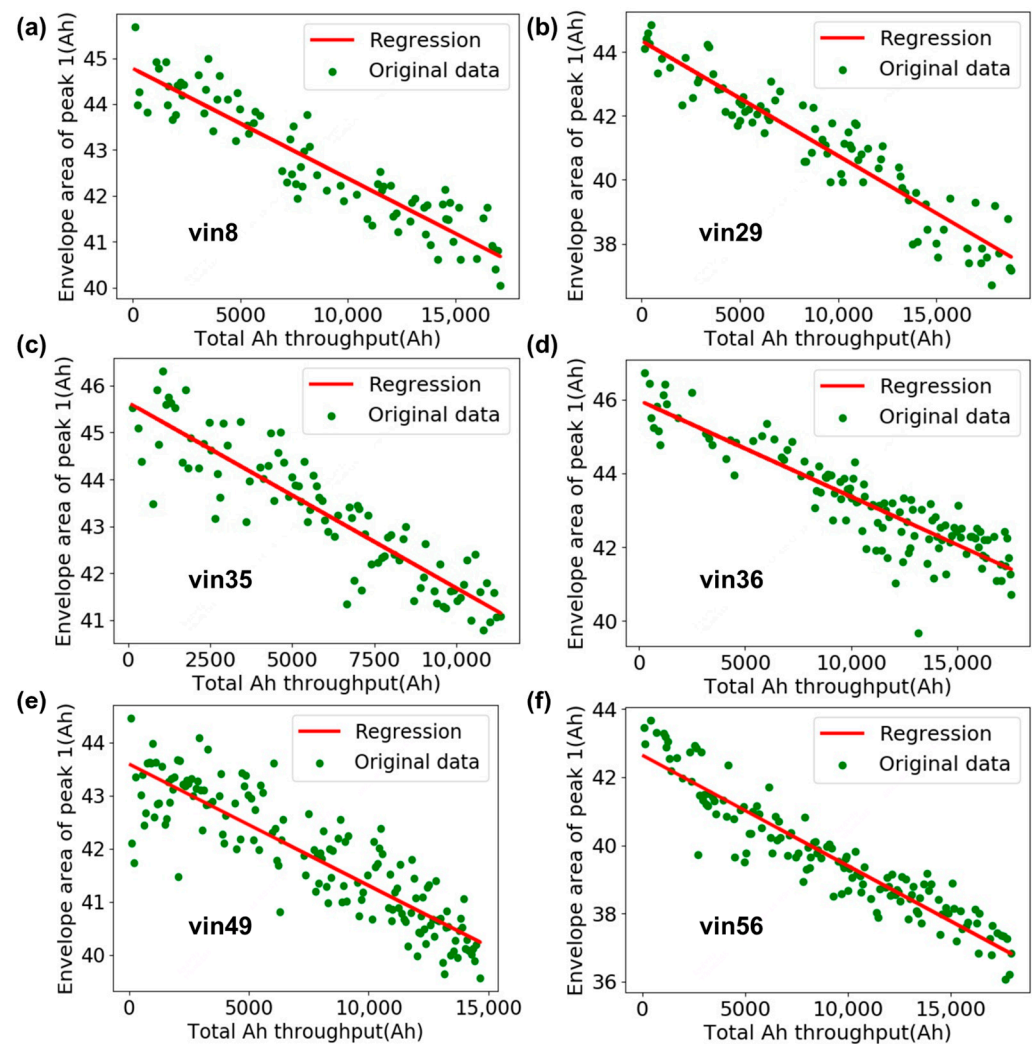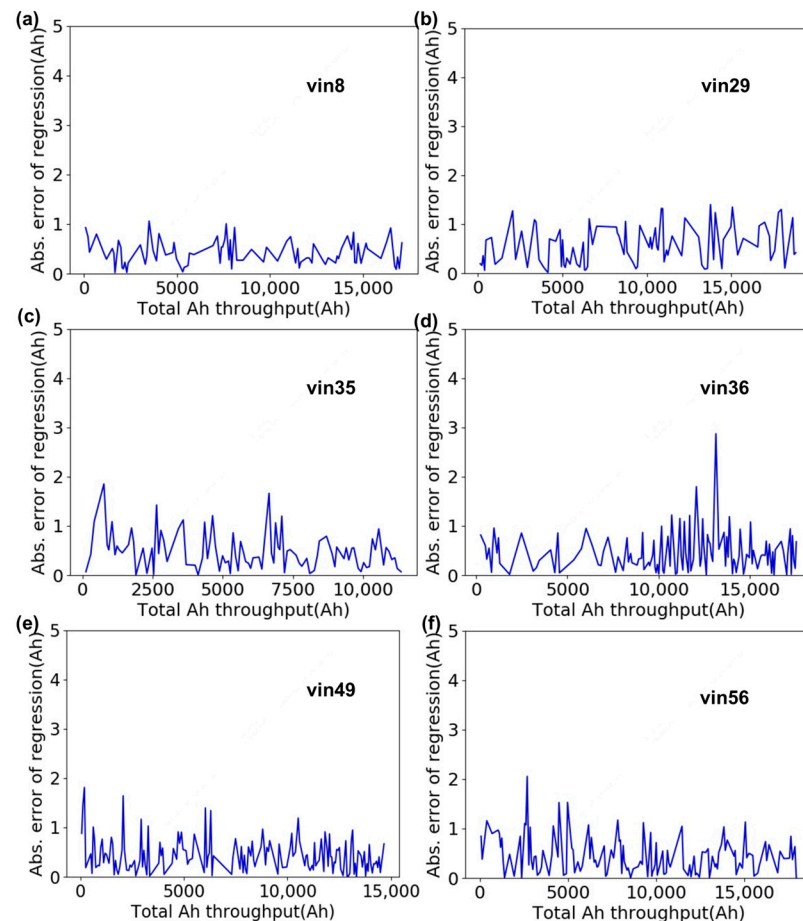
Figure 8. Health indicator estimation results based on total ampere-hour throughput of six sample EVs. (**a**) vin8. (**b**) vin29. (**c**) vin35. (**d**) vin36. (**e**) vin49. (**f**) vin56.

Each vehicle has more than 50 slow-charging processes, which cover the voltage range of Peak 1 in the IC curve. The green points represent the health indicator estimation results, and the red line represents the linear regression model. The degradation trend of the envelope area of Peak 1 can be observed, which provides the basis for capacity estimation. A clear trend emerges between the envelope area of Peak 1 and total ampere-hour throughput, showing the high predictive power of features. Table 4 shows the model performance of the six EVs. As we can see, this result is highly statistically significant ($p <$ 0.001), and the MSE for the six vehicles in the dataset varies between 0.24 and 0.88, showing that the algorithms can achieve high predictive performance. In Figure 9, the absolute model errors of the six sample vehicles are shown. In most cases, the absolute model error is smaller than 1.5 Ah, demonstrating that the model provides a simple and effective tool for capacity estimation.

Table 4. An overview of the measured parameters or status available in the datasets.

Vehicle Index	R^2	MSE	Pearson Coefficient	*p*-Value
Vin8	0.8658	0.2406	-0.93	4.46×10^{-37}
Vin29	0.8857	0.4694	-0.94	3.20×10^{-43}
Vin35	0.8144	0.3683	-0.90	2.10×10^{-35}
Vin36	0.8038	0.4745	-0.90	2.94×10^{-44}
Vin49	0.7807	0.3045	-0.88	6.22×10^{-53}
Vin56	0.8847	0.8847	-0.94	3.81×10^{-66}

Figure 9. Health indicator estimation absolute error based on IC features of six sample EVs. (**a**) vin8. (**b**) vin29. (**c**) vin35. (**d**) vin36. (**e**) vin49. (**f**) vin56.

3.4. Validation by Experiments

The linear regression model has been established between the envelope area of Peak 1 and total ampere-hour throughput in Section 3.2. However, whether the envelope area of Peak 1 is a good battery health indicator should be validated, and the correlation between the envelope area of Peak 1, and total capacity degradation should be further studied. Thus, an aging experiment is carried out using the battery, which is the same as those equipped in the vehicles in the dataset. The cell specifications are summarized in Table 1. A cell was cycled at 35 °C with a 1 C current rate for charging and a 1.5 C current rate for discharging, where cycling devices (Arbin, LBT-5V100044CH) have been employed. All charging procedures during cycling aging process and reference performance tests (RPTs) were performed with a constant-current–constant-voltage (CCCV) cycling protocol, while a constant-current (CC) protocol is used during the discharging process. The cycled cells were charged and discharged in the voltage range between 2.0 V and 3.8 V. Temperature was regulated with climate chambers (Sanwood, SC2-400-CD-3). To assess the aging of

the cell, reference performance tests were performed at the start of the aging test, at the end of the aging test, and in intervals of 100 cycles for cycle aging in between. A reference performance test included two cycles with a 1 C current rate and two cycles with a 10.5 A current at 25 °C. Experiment procedures of the aging test and the reference performance test in an aging period are shown in Table 5. The former cycles were conducted to obtain the actual capacity, and the current of the latter cycles was the same as the slow-charging process in the dataset, to derive the IC curve.

Table 5. Experiment procedures of aging test and reference performance test in an aging period.

Step	Operation	Current (A)	Termination Condition	Temperature (°C)
1	Rest	-	1 h	35
2	CC discharging	202.5	End voltage: 2.0 V	35
3	Rest	-	30 min	35
4	CC-CV charging	135	End voltage: 3.8 V Cutoff current: 1/20 C	35
5	Rest	-	30 min	35
6	CC discharging	202.5	End voltage: 2.0 V	35
7	Cycle the step from 3 to 6	-	Cycle 100 times	35
8	Rest	-	1 h	25
9	CC discharging	135	End voltage: 2.0 V	25
10	Rest	-	30 min	25
11	CC-CV charging	135	End voltage: 3.8 V Cutoff current: 1/20 C	25
12	Rest	-	30 min	25
13	CC discharging	135	End voltage: 2.0 V	25
14	Cycle the step from 10 to 13	-	Cycle twice	25
15	Rest	-	1 h	25
16	CC discharging	10.5	End voltage: 2.0 V	25
17	Rest	-	30 min	25
18	CC-CV charging	10.5	End voltage: 3.8 V Cutoff current: 1/20 C	25
19	Rest	-	30 min	25
20	CC discharging	10.5	End voltage: 2.0 V	25
21	Cycle the step from 17 to 20	-	Cycle twice	25

The results of the battery aging experiment are shown in Figure 10. Figure 10a,b demonstrates the capacity retention as a function of cycle number and total ampere-hour throughput. As we can see, in most stages of battery life before EOL, the battery is in the linear aging zone, showing a slow and stable degradation characteristic. The evolution of the IC curve is shown in Figure 10c with intervals of 100 cycles for cycle aging in between, and the inset shows a detailed view of Peak 1. The IC curves are derived from charging cycles with a 10.5 A current, the same as the slow-charging process in the dataset. It can thus be seen that only Peak 1 degrades, and Peaks 2 and 5 do not change in shape or size, showing that the envelope area of Peak 1 is a good battery health indicator. The degradation of the envelope area of Peak 1 as a function of total ampere-hour throughput is shown in Figure 10d. The linear relation acquired from real vehicle data is verified (Pearson coefficient = 0.99). As a result, the degradation of Peak 1 can be regarded as the total degradation of capacity, and the final capacity estimation results based on the linear regression model between the battery health indicator and total ampere-hour throughput are shown in Figure 11.
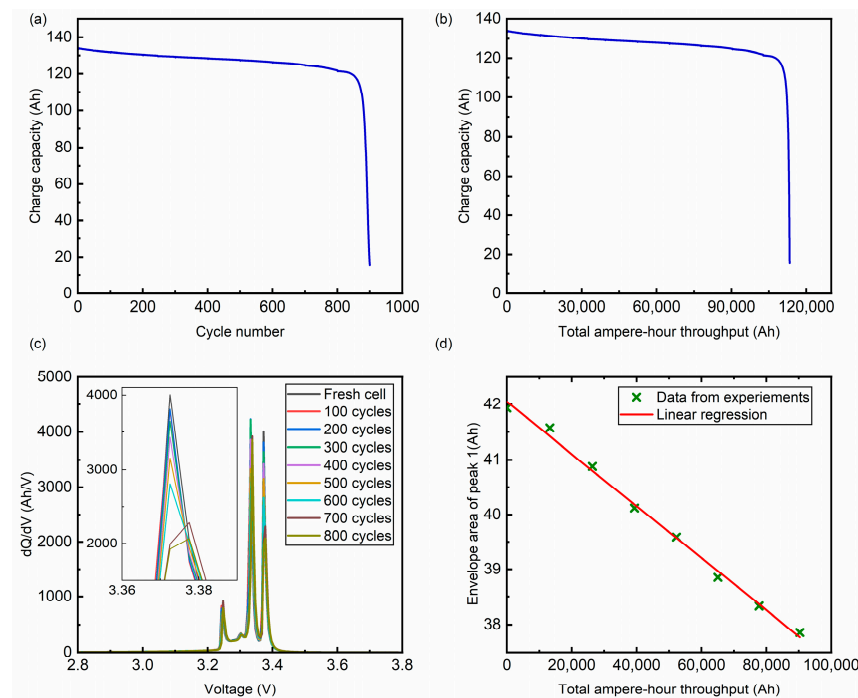
Figure 10. The results of battery aging experiment: (**a**) Capacity retention plotted as a function of cycle number; (**b**) Capacity retention plotted as a function of total ampere-hour throughput; (**c**) Evolution of IC curve in intervals of 100 cycles for cycle aging in between, where the inset shows a detailed view of Peak 1; (**d**) Degradation of the envelope area of Peak 1 as a function of total ampere-hour throughput. The linear relation acquired from real vehicle data is verified.
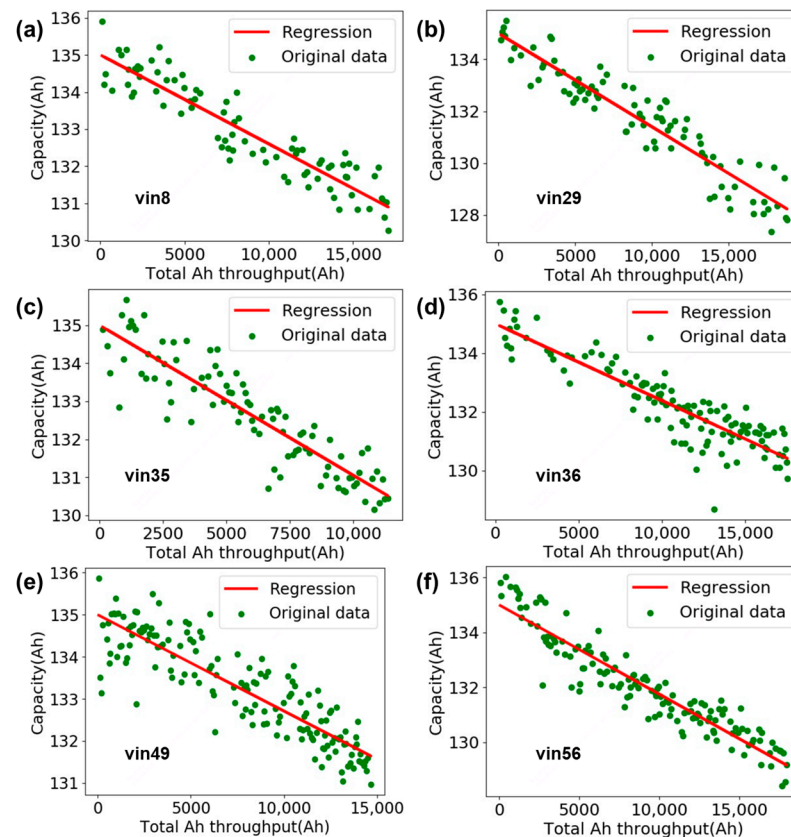

Figure 11. Capacity estimation results based on total ampere-hour throughput of six sample EVs. (**a**) vin8. (**b**) vin29. (**c**) vin35. (**d**) vin36. (**e**) vin49. (**f**) vin56.

4. Capacity Estimation Using Fast-Charging Data

4.1. Typical Fast-Charging Protocols

Contrary to the slow-charging protocols, the protocol of fast charging, which is dependent on the power output ability of chargers, is highly varied, and the current and the temperature fluctuate wildly over a wide range. The four most typical fast-charging protocols in the dataset, including the multi-stage constant-current fast-charging protocol, the current limiting at high-temperature protocol, the current limiting at low-temperature protocol, and the mild fast-charging protocol, are shown in Figure 12.
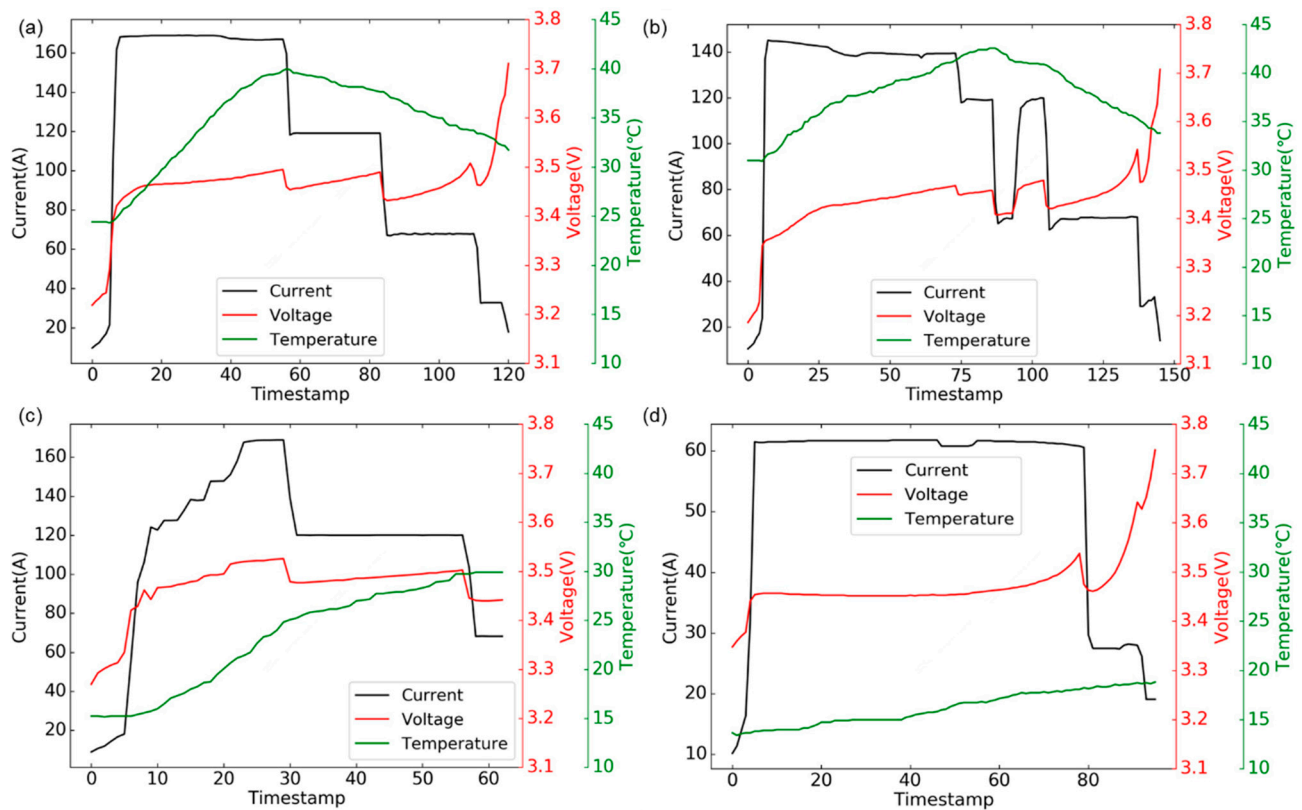
Figure 12. The four most typical fast-charging protocols: (**a**) Multi-stage constant-current fast-charging protocol; (**b**) Current limiting at high-temperature protocol; (**c**) Current limiting at low-temperature protocol; (**d**) Mild fast-charging protocol.

An example of multi-stage constant-current fast-charging protocols is shown in Figure 12a. The charging process normally contains three constant-current stages. In the first stage, the current is larger than 140 A; in the second stage, the current is in the range from 110 A to 130 A; and in the third stage, the current is in the range from 60 A to 80 A. The current limiting at high-temperature protocol and the current limiting at low-temperature protocol are shown in Figure 12b,c, respectively. Current limiting at high temperature might be attributed to safety concerns [60], and current limiting at low temperature might be attributed to lithium plating avoidance [61]. The mild fast-charging protocol is shown in Figure 12d. The current rate is much lower than that in the multi-stage constant-current fast-charging protocol, which could be ascribed to the power output ability of chargers.

4.2. Feature Engineering

As the multi-stage constant-current fast-charging protocol is adopted most frequently, and because this charging protocol is relatively regular for feature engineering, the multi-stage constant-current fast-charging protocol is selected as the main object of follow-up research. According to the current range of each stage, given in the last paragraph, 2640 fast-charging segments of 85 vehicles are filtered. In order to fully explore the information

that may reflect the physical degradation generated from the multi-stage constant-current fast-charging data, a total of 12 features are created for data-driven models, which are shown in Figure 13 and Table 6. Resistance is chosen, because resistance rise accounts for an important proportion in capacity degradation due to the large current in the fast-charging protocols. R0, R1, and R2 represent the internal resistance at different SOC, namely the internal resistance at low SOC, the internal resistance at middle SOC, and the internal resistance at high SOC. Moreover, the temperature rise rate is highly related to the internal resistance regardless of the battery thermal management system. According to [36], the statistical features of voltage series, including mean, variance, skewness, and kurtosis, are very indicative and strongly correlated with battery life. Thereinto, R0, R1, and R2 are calculated based on the voltage and current changes at the stage switch point. The temperature rise rate during Stage 1 is calculated based on the linear regression method. V_skewness can be calculated as:

$$\text{V\_skewness} = \log \left| \frac{\frac{1}{p} \sum_{i=1}^{p} (v_i - \overline{v})^3}{\left(\sqrt{\sum_{i=1}^{p} (v_i - \overline{v})^2} \right)^3} \right| \tag{11}$$

where v_i is the voltage value at timestamp i in the voltage series, $\overline{v}$ is the mean value of the voltage series, and p is the length of the voltage series. V_kurtosis can be calculated as:

$$\text{V\_kurtosis} = \log \left| \frac{\frac{1}{p} \sum_{i=1}^{p} (v_i - \overline{v})^4}{\left(\frac{1}{p} \sum_{i=1}^{p} (v_i - \overline{v})^2 \right)^2} \right| \tag{12}$$
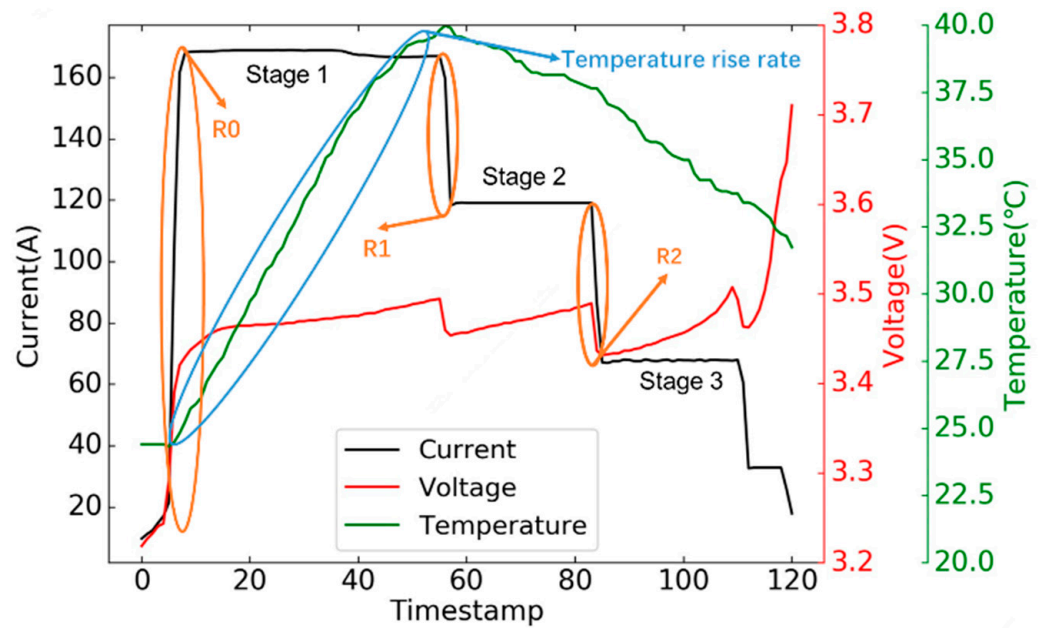
Figure 13. The features selected from multi-stage constant-current fast-charging process. The internal resistance is calculated based on the voltage and current changes at stage switch point. The temperature rise rate is calculated based on a linear regression method.

Table 6. The features selected from the multi-stage constant-current fast-charging process.

Feature	Definition
R0	Internal resistance calculated based on voltage and current changes at beginning of charging.
R1	Internal resistance calculated based on voltage and current changes at stage switch point between Stage 1 and Stage 2.
R2	Internal resistance calculated based on voltage and current changes at stage switch point between Stage 2 and Stage 3.
Temperature rise rate	Temperature rise rate during Stage 1 calculated based on linear regression method.
T1	Duration of Stage 1.
T2	Duration of Stage 2.
T3	Duration of Stage 3.
Charge number	Number of charges from beginning of life.
V_mean	Mean value of voltage series.
V_variance	Variance of voltage series.
V_skewness	Skewness of voltage series.
V_kurtosis	Kurtosis of voltage series.

4.3. Neural Network Model

For the 2640 fast-charging segments of the 85 vehicles available for training, a fully connected neural network with two hidden layers was employed for prediction. A deep learning model with far more parameters is not ideal due to overfitting. The NN model consists of 12 inputs, which were chosen due to a priori knowledge. During the development of this model, other combinations of input features were investigated but did not yield significantly improved results. On account of having no true capacity as labels for each charging process in the real vehicle dataset, total ampere-hour throughput was selected as the model output. The capacity could be estimated in the next step, based on the model established in Section 3. A group of hyperparameters were optimized with an optimization algorithm using a search matrix. The tuning parameters investigated are summarized in Table 7.

Table 7. A search matrix of hyperparameters was used to identify the optimal set for the neural network. The bolded number indicates the optimal value.

Hyperparameter	Search Range
Number of neurons (first hidden layer)	50, 100, **200**, 400
Number of neurons (second hidden layer)	50, 100, **200**, 400
Number of epochs	200, **300**, 400, 500
Batch size	16, **32**, 64

The results of the search matrix are shown in bold in Table 7, indicating that the combination of two fully connected hidden layers with 200 neurons was the most accurate. The other final optimization parameters are the number of epochs, 300, and the batch size, 32, which showed the best results from the search matrix. Relu was chosen as the activation function at each layer in the neural network to avoid the vanishing gradient problem. The neural network model was optimized using the Adagrad optimizer. The neural network model was developed in Pytorch using the Python programming language.

All the multi-stage constant-current fast-charging segments were divided into training set (80%) and test set (20%). The training set is used to choose the model features and set the values of the parameters in the neural network, and the testing dataset, generated after model development, is used to evaluate the model performance.
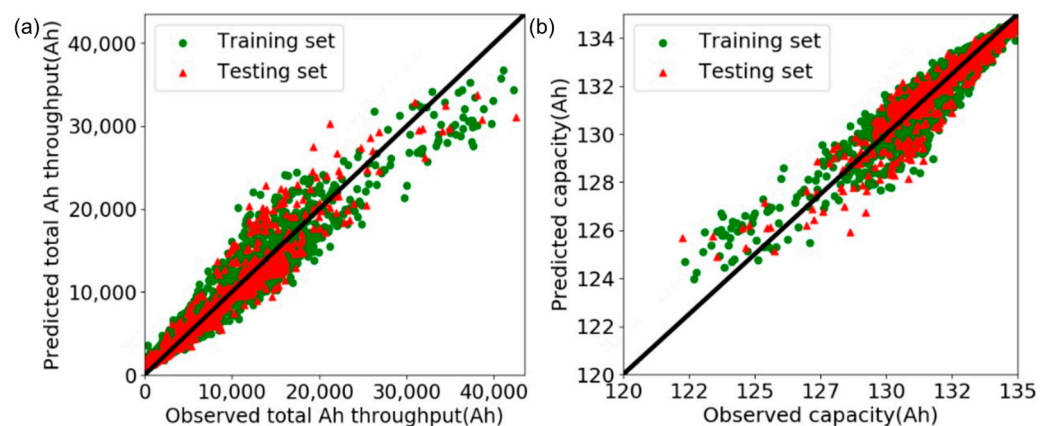
4.4. Results

The root-mean-square error (RMSE) was chosen to evaluate the model performance. The RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{13}$$

where y_i and $\hat{y}_i$ are the observed and predicted total ampere-hour throughput, respectively, of the ith samples; and n is the total number of samples in the dataset. Table 8 shows the neural network model performance with the RMSE. The RMSE of the training set and testing set is 1731 and 1986 Ah, respectively. Considering that the nominal capacity of the battery is 135 Ah, the RMSE of total ampere-hour throughput is 12.82 and 14.71 full equivalent cycles, indicating that the neural network model achieves good performance from this point of view. The performance of the neural network is displayed in Figure 14. Figure 14a shows the observed and the predicted total ampere-hour throughput, i.e., the model output and label. The point on the black lines means the predicted total ampere-hour throughput is exactly equal to the ground truth. The distance between the point and the black line reveals the error of the prediction. It can thus be seen that the model has a high accuracy in both the training set and the testing set. The output of the neural network is used as input for the capacity estimation model proposed in Section 3. As shown In Figure 11d, the linear relationship between the envelope area of Peak 1 and total ampere-hour throughput is significant, and the weight in linear regression is -4.74×10^{-5} Ah/Ah. The capacity can be calculated based on the relationship between IC features and total ampere-hour throughput. Thus, the error corresponding to the capacity is only 0.82 Ah and 0.94 Ah, respectively, for the two sub-datasets, and the observed and predicted capacities are shown in Figure 14b. As a consequence, the neural network, which is based on fast-charging data and is used to predict total ampere-hour throughput, coupled with linear regression model, provides a simple and effective tool for capacity estimation. Moreover, the neural network is trained online based on previous data. The storage requirement is low, and the computational time is short so that the proposed method can be applied for online capacity estimation.

Table 8. Model error metrics for the training set and the testing set.

Dataset	RMSE of the Dataset (Ah)	RMSE of the Dataset (Full Equivalent Cycles)
Training set	1731	12.82
Testing set	1986	14.71

Figure 14. Observed and predicted results of neural network model: (**a**) Observed and predicted total ampere-hour throughput; (**b**) Observed and predicted capacity.

5. Conclusions

In this work, in order to address the gap between data-driven capacity estimation methods developed by experimental data and capacity estimation under realistic conditions, a novel method to estimate the capacity of large-format LiFePO$_4$ batteries based on real data from electric vehicles is proposed. A comprehensive dataset consisting of 85 vehicles that has been running for around one year under diverse nominal conditions derived from a cloud platform is generated. The data are preprocessed, including data NaN inspection, data continuity inspection, and data splitting and selection. Because the charging data are stable compared to dynamic driving conditions, features from the charging curves are selected as model inputs. For slow-charging data, a linear regression capacity estimation model is established based on IC features according to a priori knowledge of the battery aging mechanism. The features are insensitive to the initial SOC at the start of charging, and appear at SOC ranges that most electric vehicles usually operate in. The selected features are highly correlated with capacity (Pearson correlation coefficient < 0.85 for all vehicles), and the MSE of the capacity estimation results is less than 1 Ah. On the basis of typical protocol analysis and mechanism studies, 12 features, including internal resistance, temperature, and statistical characteristics of the voltage curve, are extracted, and a neural network model is established for multi-stage variable-current fast-charging data. Finally, the above two models are integrated to achieve capacity prediction under complex and changeable actual working conditions, and the error of the capacity estimation method is less than 1 Ah, i.e., 0.8% SOH. An aging experiment using the battery, which is the same as those equipped in the vehicles in the dataset, is carried out to verify the methods. The results show that the proposed model, which combines degradation mechanisms and big data analysis, has high estimation accuracy and robustness, and can adapt to various practical working conditions. To the best of the authors' knowledge, our study is the first to verify a capacity estimation model derived from field data using an aging experiment of the same type of battery. Moreover, the well-established model is computationally efficient, and could be embedded in both realistic BMS and cloud platforms for real-time implementation. This work provides insights into battery capacity estimation based on massive real vehicle data.

In the future, more features will be used for modeling, including the conditions of driving, average speed, and so on. Moreover, data-driven models based on manually selected features can be upgraded to models that are able to select features automatically. The transferability of the proposed approach to other battery types or chemistries will be studied in our next work, since some statistical characteristics of the charging process unrelated to the battery material are selected as the input of the model. It should be noted that the proposed linear regression model will have poor performance in the nonlinear aging zone. A more reliable SOH estimation model, in which the full lifespan of the battery including the nonlinear aging stage is considered, will be established in our future work.

Author Contributions: Conceptualization, X.Z. and X.H.; methodology, X.Z. and X.H.; software, X.Z.; validation, Y.W., X.H. and X.Z.; formal analysis, X.Z.; investigation, X.Z.; resources, X.H.; data curation, X.Z.; writing—original draft preparation, X.Z.; writing—review and editing, Y.W., X.H. and L.L.; visualization, X.Z.; supervision, L.L., X.H. and M.O.; project administration, M.O.; funding acquisition, M.O. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

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

References

1. Axsen, J.; Plötz, P.; Wolinetz, M. Crafting Strong, Integrated Policy Mixes for Deep CO_2 Mitigation in Road Transport. *Nat. Clim. Chang.* **2020**, *10*, 809–818. [CrossRef]
2. Isik, M.; Dodder, R.; Kaplan, P.O. Transportation Emissions Scenarios for New York City Under Different Carbon Intensities of Electricity and Electric Vehicle Adoption Rates. *Nat. Energy* **2021**, *6*, 92–104. [CrossRef] [PubMed]
3. Li, Q.; Yu, X.; Li, H. Batteries: From China's 13th to 14th Five-Year Plan. *eTransportation* **2022**, *14*, 100201. [CrossRef]
4. Han, X.; Lu, L.; Zheng, Y.; Feng, X.; Li, Z.; Li, J.; Ouyang, M. A Review on the Key Issues of the Lithium Ion Battery Degradation Among the Whole Life Cycle. *eTransportation* **2019**, *1*, 100005. [CrossRef]
5. Lu, L.; Han, X.; Li, J.; Hua, J.; Ouyang, M. A Review on the Key Issues for Lithium-Ion Battery Management in Electric Vehicles. *J. Power Sources* **2013**, *226*, 272–288. [CrossRef]
6. Dai, H.; Jiang, B.; Hu, X.; Lin, X.; Wei, X.; Pecht, M. Advanced Battery Management Strategies for a Sustainable Energy Future: Multilayer Design Concepts and Research Trends. *Renew. Sustain. Energy Rev.* **2021**, *138*, 110480. [CrossRef]
7. Hu, X.; Xu, L.; Lin, X.; Pecht, M. Battery Lifetime Prognostics. *Joule* **2020**, *4*, 310–346. [CrossRef]
8. Yang, S.-C.; Hua, Y.; Qiao, D.; Lian, Y.-B.; Pan, Y.-W.; He, Y.-L. A coupled electrochemical-thermal-mechanical degradation modelling approach for lifetime assessment of lithium-ion batteries. *Electrochimica Acta* **2019**, *326*, 134928. [CrossRef]
9. Keil, J.; Jossen, A. Electrochemical Modeling of Linear and Nonlinear Aging of Lithium-Ion Cells. *J. Electrochem. Soc.* **2020**, *167*, 110535. [CrossRef]
10. Mei, W.; Zhang, L.; Sun, J.; Wang, Q. Experimental and Numerical Methods to Investigate the Overcharge Caused Lithium Plating for Lithium Ion Battery. *Energy Storage Mater.* **2020**, *32*, 91–104. [CrossRef]
11. Ren, D.; Smith, K.; Guo, D.; Han, X.; Feng, X.; Lu, L.; Ouyang, M.; Li, J. Investigation of Lithium Plating-Stripping Process in Li-Ion Batteries at Low Temperature Using an Electrochemical Model. *J. Electrochem. Soc.* **2018**, *165*, A2167–A2178. [CrossRef]
12. Yang, X.-G.; Leng, Y.; Zhang, G.; Ge, S.; Wang, C.-Y. Modeling of Lithium Plating Induced Aging of Lithium-Ion Batteries: Transition from Linear to Nonlinear Aging. *J. Power Sources* **2017**, *360*, 28–40. [CrossRef]
13. Atalay, S.; Sheikh, M.; Mariani, A.; Merla, Y.; Bower, E.; Widanage, W.D. Theory of Battery Ageing in a Lithium-Ion Battery: Capacity Fade, Nonlinear Ageing and Lifetime Prediction. *J. Power Sources* **2020**, *478*, 229026. [CrossRef]
14. Wang, S.; Wei, Y.; Han, X.; Lu, L.; Ouyang, M. A Coupled Optimization-Oriented Reduced-Order Aging Model for Graphite-LiFePO$_4$ Li-ion Batteries under Dynamic Micorgrid Conditions. In Proceedings of the 2021 4th International Conference on Energy, Electrical and Power Engineering (CEEPE), Chongqing, China, 23–25 April 2021; pp. 97–102.
15. Wei, Y.; Wang, S.; Han, X.; Lu, L.; Li, W.; Zhang, F.; Ouyang, M. Toward more realistic microgrid optimization: Experiment and high-efficient model of Li-ion battery degradation under dynamic conditions. *eTransportation* **2022**, *14*, 100200. [CrossRef]
16. Marcicki, J.; Canova, M.; Conlisk, A.T.; Rizzoni, G. Design and Parametrization Analysis of a Reduced-Order Electrochemical Model of Graphite/LiFePO$_4$ Cells for SOC/SOH Estimation. *J. Power Sources* **2013**, *237*, 310–324. [CrossRef]
17. Rechkemmer, S.K.; Zang, X.; Zhang, W.; Sawodny, O. Empirical Li-Ion Aging Model Derived from Single Particle Model. *J. Energy Storage* **2019**, *21*, 773–786. [CrossRef]
18. Wang, X.; Wei, X.; Zhu, J.; Dai, H.; Zheng, Y.; Xu, X.; Chen, Q. A Review of Modeling, Acquisition, and Application of Lithium-Ion Battery Impedance for Onboard Battery Management. *eTransportation* **2021**, *7*, 100093. [CrossRef]
19. Koleti, U.R.; Bui, T.N.M.; Dinh, T.Q.; Marco, J. The Development of Optimal Charging Protocols for Lithium-Ion Batteries to Reduce Lithium Plating. *J. Energy Storage* **2021**, *39*, 102573. [CrossRef]
20. Niri, M.F.; Dinh, T.Q.; Yu, T.F.; Marco, J.; Bui, T.M.N. State of Power Prediction for Lithium-Ion Batteries in Electric Vehicles via Wavelet-Markov Load Analysis. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 5833–5848. [CrossRef]
21. Niri, M.F.; Bui, T.M.N.; Dinh, T.Q.; Hosseinzadeh, E.; Yu, T.F.; Marco, J. Remaining Energy Estimation for Lithium-Ion Batteries Via Gaussian Mixture and Markov Models for Future Load Prediction. *J. Energy Storage* **2020**, *28*, 101271. [CrossRef]
22. Han, X.; Ouyang, M.; Lu, L.; Li, J. A Comparative Study of Commercial Lithium Ion Battery Cycle Life in Electric Vehicle: Capacity Loss Estimation. *J. Power Sources* **2014**, *268*, 658–669. [CrossRef]
23. Wang, J.; Liu, P.; Hicks-Garner, J.; Sherman, E.; Soukiazian, S.; Verbrugge, M.; Tataria, H.; Musser, J.; Finamore, P. Cycle-Life Model for Graphite-LiFePO$_4$ cells. *J. Power Sources* **2011**, *196*, 3942–3948. [CrossRef]
24. Grolleau, S.; Delaille, A.; Gualous, H.; Gyan, P.; Revel, R.; Bernard, J.; Redondo-Iglesias, E.; Peter, J. Calendar Aging of Commercial Graphite/LiFePO$_4$ cell–Predicting Capacity Fade Under Time Dependent Storage Conditions. *J. Power Sources* **2014**, *255*, 450–458. [CrossRef]
25. Naumann, M.; Schimpe, M.; Keil, P.; Hesse, H.C.; Jossen, A. Analysis and Modeling of Calendar Aging of A Commercial LiFePO$_4$/Graphite Cell. *J. Energy Storage* **2018**, *17*, 153–169. [CrossRef]
26. de Hoog, J.; Timmermans, J.-M.; Ioan-Stroe, D.; Swierczynski, M.; Jaguemont, J.; Goutam, S.; Omar, N.; Van Mierlo, J.; Van Den Bossche, P. Combined Cycling and Calendar Capacity Fade Modeling of a Nickel-Manganese-Cobalt Oxide Cell with Real-Life Profile Validation. *Appl. Energy* **2017**, *200*, 47–61. [CrossRef]
27. Schimpe, M.; von Kuepach, M.E.; Naumann, M.; Hesse, H.C.; Smith, K.; Jossen, A. Comprehensive Modeling of Temperature-Dependent Degradation Mechanisms in Lithium Iron Phosphate Batteries. *J. Electrochem. Soc.* **2018**, *165*, A181–A193. [CrossRef]
28. Sarasketa-Zabala, E.; Martinez-Laserna, E.; Berecibar, M.; Gandiaga, I.; Rodriguez-Martinez, L.M.; Villarreal, I. Realistic Lifetime Prediction Approach for Li-Ion Batteries. *Appl. Energy* **2016**, *162*, 839–852. [CrossRef]

29. Bui, T.M.N.; Sheikh, M.; Dinh, T.Q.; Gupta, A.; Widanalage, D.W.; Marco, J. A Study of Reduced Battery Degradation Through State-of-Charge Pre-Conditioning for Vehicle-to-Grid Operations. *IEEE Access* **2021**, *9*, 155871–155896. [CrossRef]

30. Bui, T.M.N.; Dinh, T.Q.; Marco, J. A Study on Electric Vehicle Battery Ageing Through Smart Charge and Vehicle-to-Grid Operation. In Proceedings of the 2021 24th International Conference on Mechatronics Technology (ICMT), Singapore, 18–22 December 2021; pp. 1–7.

31. 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*, 129–135. [CrossRef]

32. Hu, D.; Chen, G.; Tian, J.; Li, N.; Chen, L.; Su, Y.; Song, T.; Lu, Y.; Cao, D.; Chen, S.; et al. Unrevealing the Effects of Low Temperature on Cycling Life of 21700-Type Cylindrical Li-Ion Batteries. *J. Energy Chem.* **2021**, *60*, 104–110. [CrossRef]

33. Zhu, J.; Su, P.; Dewi Darma, M.S.; Hua, W.; Mereacre, L.; Liu-Théato, X.; Heere, M.; Sørensen, D.R.; Dai, H.; Wei, X.; et al. Multiscale Investigation of Discharge Rate Dependence of Capacity Fade For Lithium-Ion Battery. *J. Power Sources* **2022**, *536*, 231516. [CrossRef]

34. Xie, W.; He, R.; Gao, X.; Li, X.; Wang, H.; Liu, X.; Yan, X.; Yang, S. Degradation Identification of LiNi0.8Co0.1Mn0.1O2/Graphite Lithium-Ion Batteries Under Fast Charging Conditions. *Electrochim. Acta* **2021**, *392*, 138979. [CrossRef]

35. Li, K.; Zhou, P.; Lu, Y.; Han, X.; Li, X.; Zheng, Y. Battery Life Estimation Based on Cloud Data for Electric Vehicles. *J. Power Sources* **2020**, *468*, 228192. [CrossRef]

36. Severson, K.A.; Attia, P.M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M.H.; Aykol, M.; Herring, P.K.; Fraggedakis, D.; et al. Data-Driven Prediction of Battery Cycle Life Before Capacity Degradation. *Nat. Energy* **2019**, *4*, 383–391. [CrossRef]

37. Su, L.; Wu, M.; Li, Z.; Zhang, J. Cycle Life Prediction of Lithium-Ion Batteries Based on Data-Driven Methods. *eTransportation* **2021**, *10*, 100137. [CrossRef]

38. Zhu, J.; Wang, Y.; Huang, Y.; Bhushan Gopaluni, R.; Cao, Y.; Heere, M.; Muhlbauer, M.J.; Mereacre, L.; Dai, H.; Liu, X.; et al. Data-Driven Capacity Estimation of Commercial Lithium-Ion Batteries From Voltage Relaxation. *Nat. Commun.* **2022**, *13*, 2261. [CrossRef]

39. Thelen, A.; Lui, Y.H.; Shen, S.; Laflamme, S.; Hu, S.; Ye, H.; Hu, C. Integrating Physics-Based Modeling and Machine Learning for Degradation Diagnostics of Lithium-Ion Batteries. *Energy Storage Mater.* **2022**, *50*, 668–695. [CrossRef]

40. Tian, J.; Xiong, R.; Shen, W.; Lu, J.; Sun, F. Flexible Battery State of Health and State of Charge Estimation Using Partial Charging Data and Deep Learning. *Energy Storage Mater.* **2022**, *51*, 372–381. [CrossRef]

41. Tian, J.; Xiong, R.; Shen, W.; Lu, J.; Yang, X.-G. Deep Neural Network Battery Charging Curve Prediction Using 30 Points Collected in 10 Min. *Joule* **2021**, *5*, 1521–1534. [CrossRef]

42. Mohtat, P.; Lee, S.; Siegel, J.B.; Stefanopoulou, A.G. Comparison of Expansion and Voltage Differential Indicators for Battery Capacity Fade. *J. Power Sources* **2022**, *518*, 230714. [CrossRef]

43. Samad, N.A.; Kim, Y.; Siegel, J.B.; Stefanopoulou, A.G. Battery Capacity Fading Estimation Using a Force-Based Incremental Capacity Analysis. *J. Electrochem. Soc.* **2016**, *163*, A1584–A1594. [CrossRef]

44. Zhao, J.; Ling, H.; Liu, J.; Wang, J.; Burke, A.F.; Lian, Y. Machine Learning for Predicting Battery Capacity for Electric Vehicles. *eTransportation* **2023**, *15*, 100214. [CrossRef]

45. Li, W.; Chen, J.; Quade, K.; Luder, D.; Gong, J.; Sauer, D.U. Battery Degradation Diagnosis with Field Data, Impedance-Based Modeling And Artificial Intelligence. *Energy Storage Mater.* **2022**, *53*, 391–403. [CrossRef]

46. Zhao, J.; Burke, A.F. Electric Vehicle Batteries: Status and Perspectives of Data-Driven Diagnosis and Prognosis. *Batteries* **2022**, *8*, 142. [CrossRef]

47. Sulzer, V.; Mohtat, P.; Aitio, A.; Lee, S.; Yeh, Y.T.; Steinbacher, F.; Khan, M.U.; Lee, J.W.; Siegel, J.B.; Stefanopoulou, A.G.; et al. The Challenge and Opportunity of Battery Lifetime Prediction from Field Data. *Joule* **2021**, *5*, 1934–1955. [CrossRef]

48. Mądziel, M.; Campisi, T. Energy Consumption of Electric Vehicles: Analysis of Selected Parameters Based on Created Database. *Energies* **2023**, *16*, 1437. [CrossRef]

49. Li, W.; Rentemeister, M.; Badeda, J.; Jöst, D.; Schulte, D.; Sauer, D.U. Digital twin for Battery Systems: Cloud Battery Management System with Online State-Of-Charge and State-Of-Health Estimation. *J. Energy Storage* **2020**, *30*, 101557. [CrossRef]

50. Wang, Y.; Xu, R.; Zhou, C.; Kang, X.; Chen, Z. Digital Twin and Cloud-Side-End Collaboration for Intelligent Battery Management System. *J. Manuf. Syst.* **2022**, *62*, 124–134. [CrossRef]

51. Yang, S.; He, R.; Zhang, Z.; Cao, Y.; Gao, X.; Liu, X. CHAIN: Cyber Hierarchy and Interactional Network Enabling Digital Solution for Battery Full-Lifespan Management. *Matter* **2020**, *3*, 27–41. [CrossRef]

52. Zheng, Y.; Cui, Y.; Han, X.; Ouyang, M. A Capacity Prediction Framework for Lithium-Ion Batteries Using Fusion Prediction of Empirical Model and Data-Driven Method. *Energy* **2021**, *237*, 121556. [CrossRef]

53. Raj, T.; Wang, A.A.; Monroe, C.W.; Howey, D.A. Investigation of Path-Dependent Degradation in Lithium-Ion Batteries. *Batter. Supercaps* **2020**, *3*, 1377–1385. [CrossRef]

54. Busà, C.; Belekoukia, M.; Loveridge, M.J. The Effects of Ambient Storage Conditions on the Structural and Electrochemical Properties of NMC-811 Cathodes for Li-Ion Batteries. *Electrochim. Acta* **2021**, *366*, 137358. [CrossRef]

55. Diao, W.; Kim, J.; Azarian, M.H.; Pecht, M. Degradation Modes and Mechanisms Analysis of Lithium-Ion Batteries with Knee Points. *Electrochim. Acta* **2022**, *431*, 141143. [CrossRef]

56. Han, X.; Ouyang, M.; Lu, L.; Li, J.; Zheng, Y.; Li, Z. A Comparative Study of Commercial Lithium Ion Battery Cycle Life in Electrical Vehicle: Aging Mechanism Identification. *J. Power Sources* **2014**, *251*, 38–54. [CrossRef]

57. Dubarry, M.; Truchot, C.; Liaw, B.Y. Synthesize Battery Degradation Modes via a Diagnostic and Prognostic Model. *J. Power Sources* **2012**, *219*, 204–216. [CrossRef]

58. Dubarry, M.; Berecibar, M.; Devie, A.; Anseán, D.; Omar, N.; Villarreal, I. State of Health Battery Estimator Enabling Degradation Diagnosis: Model and Algorithm Description. *J. Power Sources* **2017**, *360*, 59–69. [CrossRef]

59. Feng, X.; Merla, Y.; Weng, C.; Ouyang, M.; He, X.; Liaw, B.Y.; Santhanagopalan, S.; Li, X.; Liu, P.; Lu, L.; et al. A Reliable Approach of Differentiating Discrete Sampled-Data for Battery Diagnosis. *eTransportation* **2020**, *3*, 100051. [CrossRef]

60. Qin, P.; Sun, J.; Yang, X.; Wang, Q. Battery Thermal Management System Based on the Forced-Air Convection: A Review. *eTransportation* **2021**, *7*, 100097. [CrossRef]

61. Liu, T.; Yang, X.-G.; Ge, S.; Leng, Y.; Wang, C.-Y. Ultrafast Charging of Energy-Dense Lithium-Ion Batteries for Urban Air Mobility. *eTransportation* **2021**, *7*, 100103. [CrossRef]