Xu, Zhicheng; Wang, Jun; Lund, Peter D.; Zhang, Yaoming

Estimation and prediction of state of health of electric vehicle batteries using discrete incremental capacity analysis based on real driving data

Published in:
Energy

DOI:
10.1016/j.energy.2021.120160

Published: 15/06/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Xu, Z., Wang, J., Lund, P. D., & Zhang, Y. (2021). Estimation and prediction of state of health of electric vehicle batteries using discrete incremental capacity analysis based on real driving data. *Energy*, *225*, Article 120160. https://doi.org/10.1016/j.energy.2021.120160

Estimation and prediction of state of health of electric vehicle batteries using discrete incremental capacity analysis based on real driving data

Zhicheng Xu[1,2], Jun Wang[1,2+], Peter D. Lund[1,3+], Yaoming Zhang[1]

[1] Key Laboratory of Solar Energy Science and Technology in Jiangsu Province, Southeast University, No. 2 Si Pai Lou, Nanjing 210096, P.R. of China

[2] Energy Storage Research Center, School of Energy and Environment, Southeast University, No. 2 Si Pai Lou, Nanjing 210096, P.R. of China

[3] School of Science, Aalto University, P.O. Box 15100, FI-00076 Aalto (Espoo), Finland

[+] Corresponding authors: 101010980@seu.edu.cn, peter.lund@aalto.fi

Abstract

The accuracy of the state of health (SoH) estimation and prediction is of great importance to the operational effectiveness and safety of electric vehicles. Present approaches mostly employ data-driven analysis with laboratory measurements to determine these parameters. Here a novel method is proposed using discrete incremental capacity analysis based on real-life driving data, which enables to estimate the battery SoH without any prior detailed knowledge of battery internal specifics such as current capacity/resistance information. The method accounts for the battery characteristics. It is robust, highly compatible, and has a short computing time and low memory requirement. It's capable to evaluate the SoH of various type of electric vehicles under different charging strategies. The short computing time and low memory needed for the SoH estimation also demonstrates its potential for practical use. Moreover, the clustering analysis is presented, which provides SoH comparison information of certain EV to that of EVs belonging to same type.

Keywords: Clustering analysis, electric vehicles, state of health, discrete incremental capacity analysis

Nomenclature

c_stat	Charging state
dq_time	Data acquisition time
General_alarm	General fault level of automobile
max_alarm_lvl	Level of automobile failure
max_cell_volt	Maximum cell voltage in battery pack
max_temp	Maximum temperature of battery
min_cell_volt	Minimum cell voltage in battery pack
min_temp	Minimum temperature of battery
mileage	Vehicle mileage
NaN	Missing data point
speed	Vehicle speed
status	Vehicle state
t_current	Battery pack current
t_volt	Battery pack voltage
vid	EV number

Abbreviations

DVA	Differential voltage analysis
EMD	Empirical Mode Decomposition
EV	Electric vehicle
IC	Incremental capacity curve
ICA	Incremental capacity analysis
IMF	Intrinsic mode function
NDANEV	National Big Data Alliance of New Energy Vehicles
NEV	New Energy Vehicle
PDE	Partial Differential Equation
RUL	Remaining useful lifetime
SoC	State of charge
SoH	State of health

1. Introduction

Electric mobility is considered a promising option to address the environmental issues and emissions in the transportation sector [1-2]. Important indicators for the performance of electric vehicles include the state-of-health (SoH) and the remaining useful life-time (RUL) of the battery, which have great importance to the operational security of EVs [3-4]. Accurate SoH estimation is important to the timing of decommissioning the battery, typically set to 80% of the rated value (SoH=80%) [5]. The RUL provides direct indication of the remaining useful lifetime or mileage, i.e., how soon the EV battery will reach the 80% SoH limit [6].

Estimation of the RUL is often based on analyzing the SoH trajectory and the historical EV driving data, for which reason estimation of the SoH plays a central role here. There are several approaches to evaluate the battery SoH, which can be divided into the following categories: Physical models, empirical models, differential voltage/incremental current analysis (DVA/ICA) and data-driven approaches [7-8]. The physical models are based on mathematical models describing the battery dynamics of internal reactions in the electrodes and on their surface yielding high accuracy [9], but often facing main challenges in practical applications due to model complexity and parameter identification [10]. Empirical models employ adaptive algorithms based on empirical battery models such as the equivalent circuit model [11-12] and reduced electrochemical model [8]. These models can easily be applied to online SoH monitoring, but the algorithm development often requires profound experimental validation and debugging to provide adequate accuracy [8]. Another type of empirical models includes using a predefined narrow set of experimental data for better model, but their accuracy is limited to similar conditions than the test data [13].

In data-driven approaches, which is also the main theme of our work, pre-knowledge of the battery chemistry or dynamics is not necessary, but just collecting aging data of the battery is required [14-16]. Deep learning methods such as the convolutional neural network for estimating the battery capacity could be used based on battery voltage, current and charge capacity during partial charge cycle [17]. In another approach, SoH was estimated based on charging voltage curves within a fixed range [18]. Gaussian process regression models have also been used for SoH evaluation [19-21]. Support vector machine and short-term current pulse have also been used for this purpose [22]. A common feature for above-described methods is building a relationship between the battery capacity information and extracted health features using machine learning or other data processing methods, which enable to determine the current battery capacity and hence also the SoH.

The main problem with the data-driven models to estimate the SoH in practical conditions is the requirement of training data on battery capacity and health feature information, which is not available in EVs as the battery is seldom fully cycled and thus the capacity status is unknown. In the DVA/ICA method, the peak position, amplitude and envelope area of the incremental capacity (IC) curve well correlate with the battery capacity for different cycles [23-27]. In the traditional incremental capacity analysis (ICA) method, the voltage, current and charging time of the charging process vis-a-vis the corresponding battery capacity are used as the training data for building the peak-capacity relationship [23-24]. Thus, the battery capacity and the SoH can be estimated as long as the latest peak information is available. However, in practice the EVs charging process is very discontinuous, and not following a 0 to 100% SoC pattern meaning that the battery current capacity is unavailable unlike the voltage, current and charging time. Moreover, as the sampling frequency is not constant this would lead to a discontinuous battery voltage. Therefore, the traditional ICA method is not applicable to real world EV driving data analysis. Our work strives to will this gap through a novel discrete incremental capacity analysis method. The physical meaning of this method is similar to that of the ICA, i.e., the voltage plateaus on the incremental capacity curves can be transformed into easily identifiable peaks, which reflects the electrochemical reaction of electrodes [7-8]. In the proposed method, the battery SoH is described by the ratio of the current charging capacity increment to that at the initial time, which would not be influenced by a discontinuous charging process and battery voltage. Also, the proposed method can evaluate the health status of various types of vehicles under different charging strategies. Moreover, a clustering analysis is presented, which provides a SoH comparison of the same type of EVs.

The rest of this paper is organized as follows. Section 2 briefly describes the research method and data processing. The discrete incremental capacity analysis is presented in Section 3. Section 4 discusses the SoH estimation and prediction as well as the calculation time and memory usage. Section 5 presents the clustering results. Section 6 summarizes the main result and provides the concluding remarks.

2. Data processing

In this paper, discrete incremental capacity analysis is employed to analyze the state of health of the EV battery. The main steps of the proposed method are shown in Fig.1. To estimate and predict the SoH at given conditions (section 4), data processing will be necessary to select and split data at suitable charging stage (section 2). In the discrete incremental capacity analysis process (section 3), new blank data will be firstly created and then NaN data point will be filled for convenience of analyzing the incremental capacity information. Finally, clustering of EVs belonging to the same type is carried out (section 5).

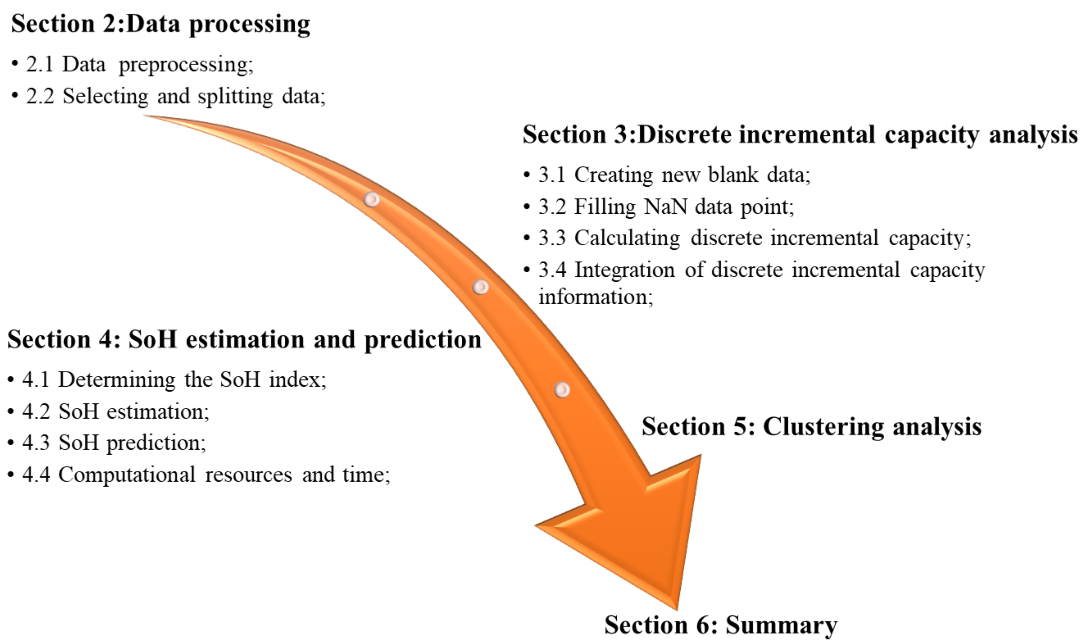

Fig.1. Flow chart of SoH estimation and prediction using discrete incremental capacity analysis.

2.1 Data preprocessing

The driving data is provided by the National Big Data Alliance of New Energy Vehicles (NDANEV) [28] and collected from the National Monitoring and Management Platform for New Energy Vehicles (NEVs). In this paper, we mainly use data from 9 electric vehicles (Car#1-Car#9) belonging to 4 EV types to illustrate how the discrete incremental capacity works. Table 1 shows a typical driving data sample (Car #1), i.e., raw data, some problems need to be mentioned as follows, making the implementation of traditional ICA unrealistic.

- Data is unordered and switches between startup (‘status’=1), flameout (‘status’=2) and other stages (‘status’=3);
- Some mileage, voltage and current data points are missing (NaN);
- The time interval between time points varies;
- The voltage is discontinuous with a minimum change of 0.1V due to the accuracy of the voltage measurement (0.1V).

Table 1. Driving data sample of Car #1.

vid	daq_time	status	c_stat	Speed ($km \cdot h^{-1}$)	Mileage (km)	t_volt/V	t_current (A)	soc	max_cell_volt (V)	min_cell_volt (V)	max_temp (°C)	min_temp (°C)	max_alarm_lvl	general_alarm
1	2018/04/29/15/34/57	1	4	1.7	8280	396.5	9.8	100	4.136	4.125	26	24	0	0
1	2018/04/29/15/36/17	1	3	10.1	8280	395.2	15.6	99	4.123	4.112	26	25	0	0
1	2018/04/29/15/38/16	1	3	44.6	8282	392.7	35.2	98	4.097	4.085	27	25	0	0
1	2018/04/29/15/40/16	1	3	22.9	8283	392.2	31.5	97	4.091	4.08	26	25	0	0
1	2018/04/29/15/40/46	1	3	45.6	8283	390	74.1	97	4.073	4.059	27	25	0	0
1	2018/04/29/15/40/26	1	2	34.7	8283	393.7	-9.1	97	4.106	4.094	26	25	0	0
1	2018/04/29/15/40/56	1	3	10.1	8283	393.2	4.4	97	4.1	4.089	27	25	0	0
1	……	……	……	……	……	……	……	……	……	……	……	……	……	……
1	2018/04/29/16/19/39	3	3	0	8289	389.7	0	94	4.065	4.054	28	26	0	0
1	2018/04/29/16/20/19	3	3	0	8289	389.7	0	94	4.065	4.054	28	26	0	0
1	……	……	……	……	……	……	……	……	……	……	……	……	……	……
1	2018/04/30/13/28/18	1	3	0	8290	388.5	8.2	93	4.051	4.043	29	28	0	0
1	2018/04/30/15/21/09	2	1	NaN	NaN	388.5	-8.2	91	4.05	4.041	31	30	0	0
1	2018/04/30/15/21/19	2	1	NaN	NaN	388.5	-8.1	91	4.05	4.041	31	30	0	0
1	……	……	……	……	……	……	……	……	……	……	……	……	……	……

Note: vid represents the EV number (Car1~9 in this study); daq_time is the data acquisition time; status is the vehicle state (1:EV startup stage, 2:EV flameout stage, 3:other stages); c_stat is the charging state (1:parking charging stage, 2:brake charging stage, 3:uncharged stage, 4:charging completed stage); t_volt is the battery pack voltage; t_current is the battery pack current; max/min_cell_volt is the maximum/minimum cell voltage in the battery pack; max/min_temp is the maximum/minimum temperature of the battery; max_alarm_lvl and general_alarm represent the battery operational state, where 0 means everything is fine.

Therefore, discrete incremental capacity analysis is proposed, but this method will require preprocessing of the data to form an appropriate data set. First, the raw data is sorted in ascending order to time (daq_time). Then, the faulty data is deleted. The data recording shown in Table 1 includes three types of voltage data: battery pack, maximum cell, and minimum cell voltage. To choose the appropriate voltage for the discrete incremental capacity analysis, a correlation analysis of the data was performed shown in Fig.2. The Pearson correlation coefficient of the three voltages is one, indicating that they are highly correlated and any of them can be used. Considering that the cell with maximum or minimum voltage change during the operational stage, the battery pack voltage was chosen for the analysis.

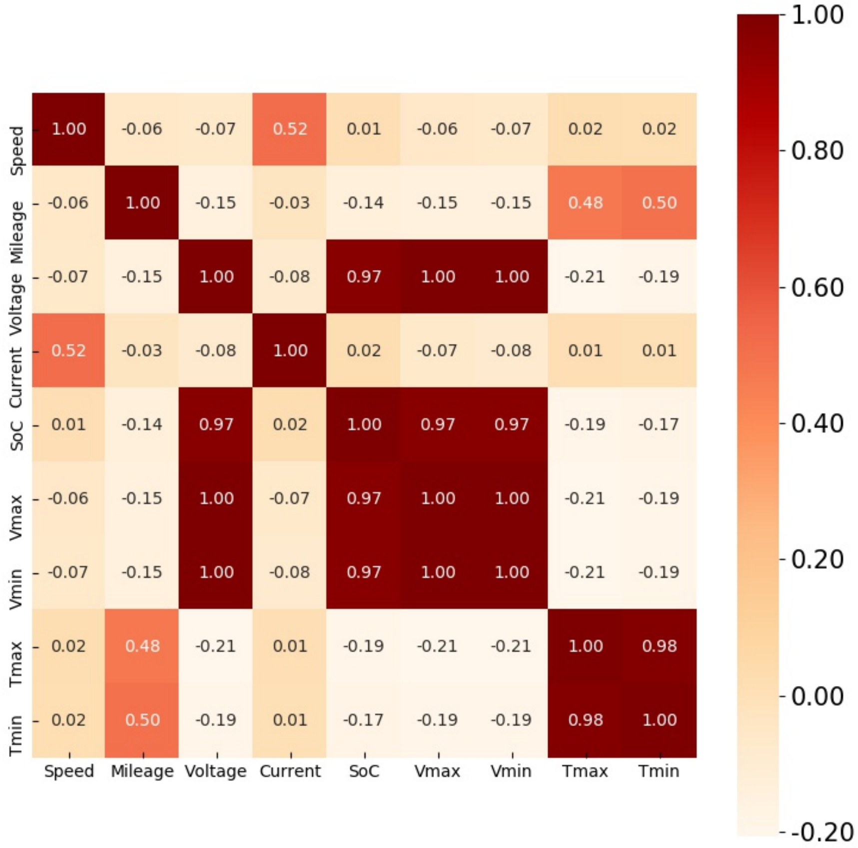

Fig.2. Parameter correlation map of measured data. Voltage= battery pack voltage, Vmax/Vmin= maximum/minimum cell voltage.

2.2. Selecting and splitting data

In this part, the raw data is selected and split with the purpose of creating an appropriate data set. First, the following clarifications for definitions are provided:

- ‘data’ represents the whole EV information recorded at a certain time, i.e., each row in Table 1;
- ‘data point’ represents the specific information involved in data such as voltage, current and mileage;
- ‘segment’ represents the continuous pieces of data formed after the selecting and splitting process;
- ‘blank data’ represents the created new blank data appearing mainly in Section 3.1;
- ‘block’ represents the continuous piece of data in the segment appearing mainly in Section 3.2.

Selecting data

After the data preprocessing, the raw data was selected according the ‘status’ value with aim of selecting the continuous charging segments, i.e., ‘status’=2, shown in Fig.3 (a). If the amount of data in a segment is less than 10, which indicates the extracting valid information from this segment is limited, this charging segment will be discarded. There is no definite limitation on the amount of data above, but 10 was found appropriate as the larger the number, the more segment will be discarded. On the other hand, the smaller the number the less data is involved in a segment meaning that limited information could be extracted in the future. Taking Car #9 as an example, the charging segments formed after selecting is 118.

Splitting data for first time

In the splitting step, if the time intervals of data in the segment formed after the selecting process are less than 300s, the segment remains the same. Otherwise, the segment will be split for data with time intervals greater than 300s shown in Fig.3 (b). This time interval threshold was set to avoid the following situations: 1) the EV has physically gone through the parking charging stage-startup-parking charging stage during this time interval and without recording data of the short startup process in which case the data in this segment is all charging data, but within two charging period; 2) the data in the segment is actually recorded during the same charging period, but the larger time interval (>300s) results in data discontinuity. After the splitting process, segments with amount of data less than 10 will be discarded. Taking Car#9 again as an example, the charging segments formed after the splitting has grown from 118 to 130.

Splitting data for second time

In the second splitting, the mode of time interval in each segment formed after first splitting is calculated. Taking the 1^{st} and 2^{nd} segment as an example, the modes of time interval for the 1^{st} and 2^{nd} segment are 10s and 2s, respectively, marked with a red circle in Fig.3(b), then each segment will be split for second time according to the algorithm given in Table 2. The reason for splitting data according this method is to ensure the amount of blank data needed to be created is less than 10, which will be discussed detailly in the section 3. Taking Car#9 for example, the charging segments formed after second splitting has grown to 150 from 130.

Table 2. Pseudocode of the second splitting of data.

Calculating the mode of time intervals in each segment:
If mode ⩾ 10: each charging segment will be split further for data with time interval > 100s; If mode < 10 : each charging segment will be split further for data with time interval ⩾ 10s;

After selecting and splitting the data, a series of successive charging segments has been forced satisfying the following criteria:

- Each segment is charging stage data, indicating status=2 and c-stat=1/4.
- The time interval between daq_time in each segment is < 10s or <100s.
- All data in the same segment has the same mileage belonging to the same charging period.

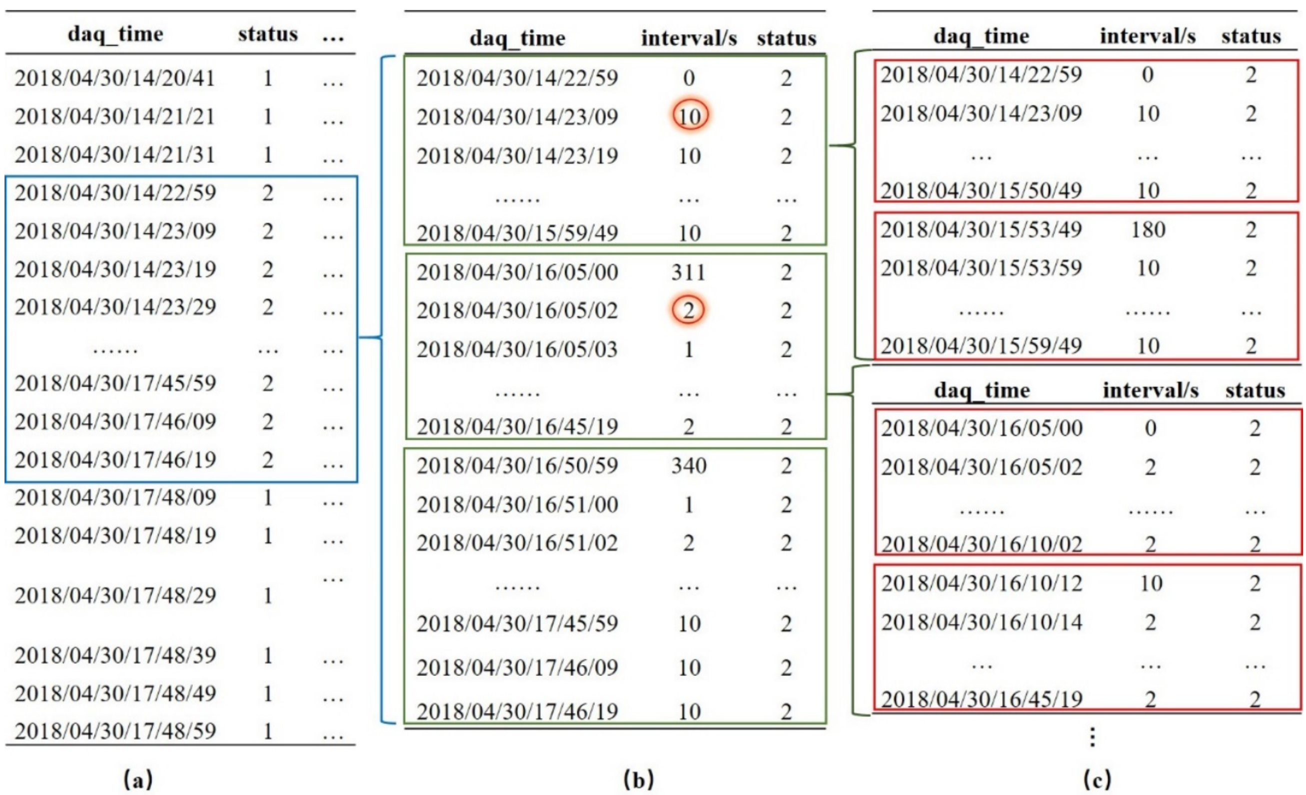

Fig.3 An example of data processing. (a) Selecting data, (b) first splitting of data, (c) second splitting of data.

3. Discrete incremental capacity analysis

3.1 Creating new blank data

After processing the raw data, a series of segments were formed. To ensure that the time intervals between the data in a segment are the same (equal to 1s or 10s), new blank data are added into these segments with the algorithm given in Table 3 and also shown in Fig.4.

Table 3. Pseudocode of creating new blank data.

Calculating the mode of time intervals in each segment:
If mode ≥ 10: Creating blank data with time interval equal to 10s and keeping the first and last data unchanged;
If mode < 10: Creating blank data with time interval equal to 1s and keeping the first and last data unchanged;
Add the segment data into the created blank data based on time proximity;

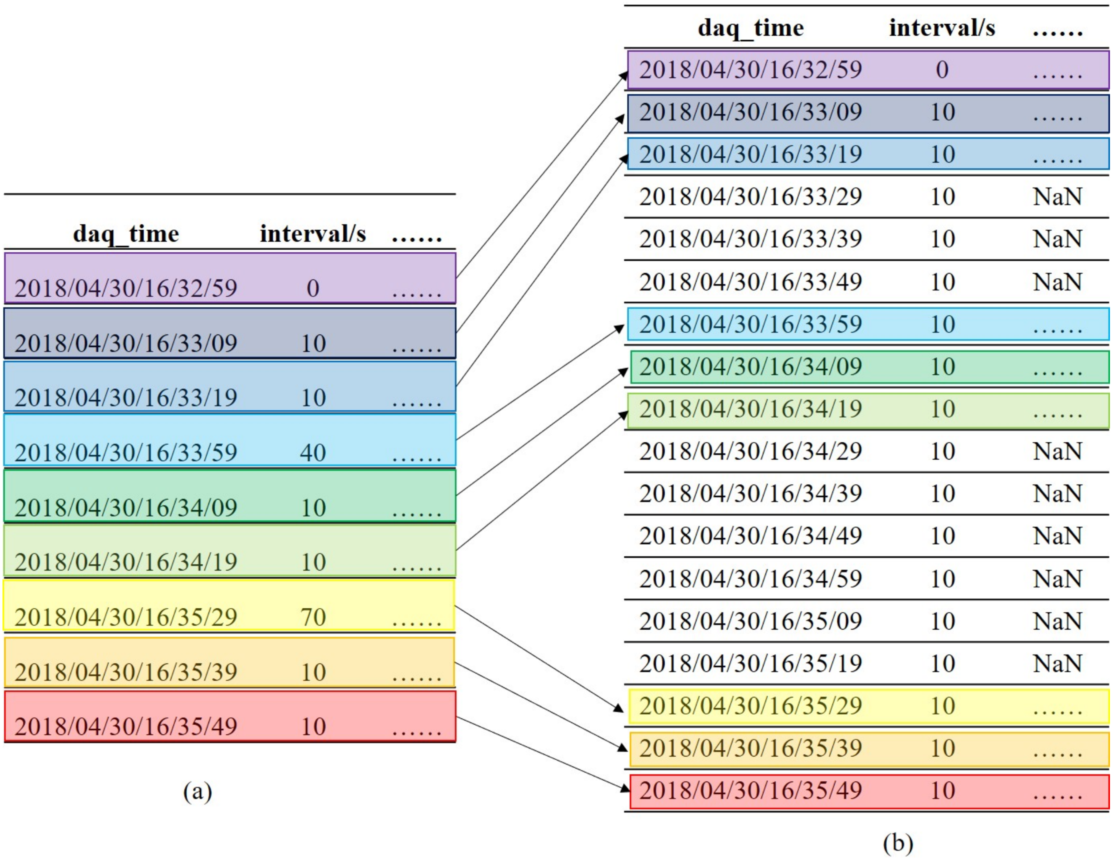

Figure 4. An example of creating new blank data with time interval = 10s. (a) Data segment achieved after data processing; (b) creating new blank data.

3.2 Filling NaN data point

In this part, the NaN data points required for the discrete incremental capacity calculation are filled, especially the mileage, current and voltage information.

Filling mileage data point

As mentioned before, some mileage, voltage and current data points are missing in the raw data. Also, the mileage of new blank data must be NaN. Therefore, it's necessary to fill the mileage with an algorithm shown in Table 4.

Filling current data point

Considering that the charging process of EVs is basically continuous and smooth, the NaN current data point follows the previous current value shown in Fig. 5.

Filling voltage data point

The voltage filling is mainly divided into two steps. The first part is the locking voltage, indicating that if the voltage before and after the NaN block is the same, the voltage of the NaN block keeps the same. The second part is the average distribution, meaning that filling the half of the NaN block with the previous voltage value and the remaining NaN with the latter voltage shown in Fig.5.

Table 4. Pseudocode of filling mileage data point

Check the mileage data point in each segment:

If at least one mileage data point is existing for the segment:
    Filling the entire segment with this mileage value;
If mileage data is NOT existing for the segment:
    Return to raw data, find the time point closest to the first count time of this segment, and fill the entire segment with the mileage of that moment in the raw data

daq_time	t_volt/V	t_current /A
20181107131950	356.3	-7
20181107132000	356.3	-7
20181107132010	NaN	-7
20181107132020	NaN	-7
20181107132030	356.3	-7
20181107132040	356.5	-7.1
20181107132050	356.5	-7.1
20181107132100	356.7	-7
20181107132110	NaN	NaN
20181107132120	NaN	NaN
20181107132130	NaN	NaN
20181107132140	NaN	NaN
20181107132150	356.9	-7
20181107132200	357.1	-7.2
20181107132210	357.1	-7.1
20181107132220	357.3	-7
20181107132230	357.3	-7
20181107132240	357.5	-7
20181107132250	357.5	-6.9

Voltage filling: Lock voltage →

Voltage filling: Average distribution →

Current filling →

daq_time	t_volt/V	t_current /A
20181107131950	356.3	-7
20181107132000	356.3	-7
20181107132010	356.3	-7
20181107132020	356.3	-7
20181107132030	356.3	-7
20181107132040	356.5	-7.1
20181107132050	356.5	-7.1
20181107132100	356.7	-7
20181107132110	356.7	-7
20181107132120	356.7	-7
20181107132130	356.9	-7
20181107132140	356.9	-7
20181107132150	356.9	-7
20181107132200	357.1	-7.2
20181107132210	357.1	-7.1
20181107132220	357.3	-7
20181107132230	357.3	-7
20181107132240	357.5	-7
20181107132250	357.5	-6.9

Fig 5. The filling strategy of NaN current and voltage data point.

3.3 Calculating discrete incremental capacity

In this part, the incremental capacity of each voltage in the segment is analyzed, before which the charging capacity is calculated according to Eq. (1) and shown in Fig. 6 (Table A):

$$\text{Charging capacity}(i) = \text{Charging capacity}(i-1) + \Delta\text{capacity} \qquad (1)$$

$$\Delta\text{capacity} = -\ (\text{t\_current}(i) + \text{t\_current}(i\text{-}1)\ )\ /\ 2 \times (\text{Charging time}(i) - \text{Charging time}(i-1)) \qquad (2)$$

where i represents the i^{th} of daq_time, Δcapacity is the charging capacity increase calculated using the average current of the i-1 and i multiplied by the time interval between i-1 and i. In the project, the charging current is negative so that the minus sign is adapted in Eq. (2) to represent the current value.

Then, the voltage is adjusted to enhance the compatibility of this method. For example, if the EV is dominated by fast charging, i.e., the voltage varies much, the voltage precision can be set to 1V. On the contrary, if the EV is dominated by slow charging, i.e., the voltage varies slowly, the voltage precision can remain unchanged (0.1V). The selection of the voltage precision has downward compatibility, i.e., the voltage precision adjustment of fast charging can also solve the problem of slow charging. Figure 6 (Table B) shows how the voltage precision is adjusted under the fast-charging condition. The charging capacity must be positive as long as the t_current is negative, i.e., the EV is charging and not discharging. The initial subtraction of 12 As in Fig. 6 (Table B) is due to the positive current of 2.4A (the third data row in t_current column in Table A). The positive current in the initial charging process is somewhat strange and could be explained by a delay or an error in the current sensor during the initial charging process when the EV is switching from the discharge state to the charging state leading to a positive current at the initial charging process. This error only occurs sporadically at the initial charging moment. Moreover, the initial subtraction capacity will be compensated in the following normal charging process, resulting in negligible calculation error in the incremental charging capacity of each voltage.

After adjusting the voltage precision, the discrete incremental capacity of each independent voltage can be analyzed by calculating the capacity difference between the first and the last capacity value in the segment with the same voltage shown in Fig.6 (Table C). Moreover, the discrete incremental capacity information of the first and last voltage in each segment is discarded because of the incomplete charging process, e.g., incremental capacity information of 322V and 328V in Fig.6 Table C.

Table A. Sample segment

daq_time	t_volt /V	t_current /A	Charging time/s	Charging capacity/As
20181207131950	321.8	0	0	0
20181207132000	321.8	0	10	0
20181207132010	321.8	2.4	20	-12
20181207132020	325.2	-52.8	30	240
20181207132030	325.7	-52.9	40	768.5
20181207132040	326.1	-53	50	1298
20181207132050	326.5	-52.8	60	1827
20181207132100	326.8	-52.8	70	2355
20181207132110	326.9	-53.1	80	2884.5
20181207132120	327	-53	90	3415
20181207132130	327.3	-53	100	3945
20181207132140	327.4	-52.8	110	4474
20181207132150	327.7	-52.9	120	5002.5
20181207132200	327.8	-53	130	5532
20181207132210	327.8	-52.8	140	6061
20181207132220	328	-53	150	6590
20181207132230	328.2	-52.9	160	7119.5
20181207132240	328.3	-52.9	170	7648.5
20181207132250	328.3	-53	180	8178

Voltage precision adjustment

Table B. Sample segment after voltage precision adjustment

t_volt/ V	Charging capacity/As
322	0
322	0
322	-12
325	240
326	768.5
326	1298
327	1827
327	2355
327	2884.5
327	3415
327	3945
327	4474
328	5002.5
328	5532
328	6061
328	6590
328	7119.5
328	7648.5
328	8178

-12-0=-12As

0As

1298-768.5=529.5As

4474-1827=2647As

8178-5002.5=3175.5As

Discrete incremental capacity calculation

Table C. Discrete incremental capacity information

voltage /V	Incremental capacity/As
322	-12
325	0
326	529.5
327	2647
328	3175.5

Fig 6. Discrete incremental capacity calculation. Sample segment (Table A), segment after voltage precision adjustment (Table B), discrete incremental capacity information of the segment (Table C).

3.4 Integration of discrete incremental capacity information

In the previous sections, the raw data was processed into many charging segments and the discrete incremental capacity information of each segment was obtained. The voltage span of the discrete incremental capacity information in each segment is different because of the different charging time, which makes it difficult to form a completed discrete incremental capacity curve. As a result, the discrete incremental capacity information of the five segments is put together to increase the voltage span.

Henceforth, the discrete incremental capacity curves of each Car are formed. Each curve represents the incremental capacity variation of each voltage during the charging process under a certain mileage condition. These curves are not smooth and also includes some sensor and data processing errors, for which reason a Gaussian filter is applied. Taking Car #9 as an example shown in Fig. 7, Fig. 7 (a) is the discrete incremental capacity curves and Fig. 7 (b) shows the Gaussian filtering result.

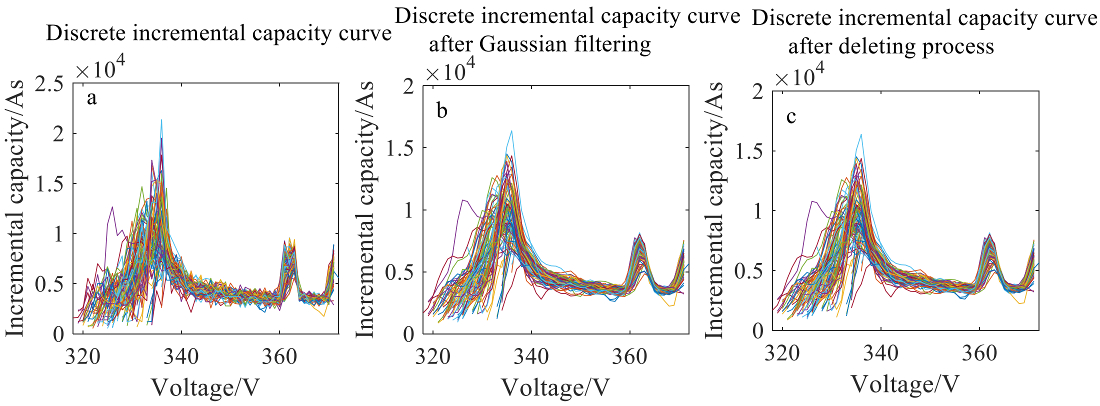

Fig 7. Discrete incremental capacity curves of Car#9. (a) The discrete incremental capacity curve; (b) the discrete incremental capacity curve after Gaussian filtering, (c) the discrete incremental capacity curve after deleting process.

In order to improve the accuracy of the SoH estimation, the curves with a voltage span less than 50% of the maximum value are deleted. The discrete incremental capacity curve after the deleting process is shown in Fig. 8. The 9 Cars are belonging to 4 different EV types, i.e., type 1-type 4. The types (1-4) indicate different features, mainly the curve peaks and scope of voltage variation, i.e., Cars with the same type have similar curves, which makes a clustering analysis possible.

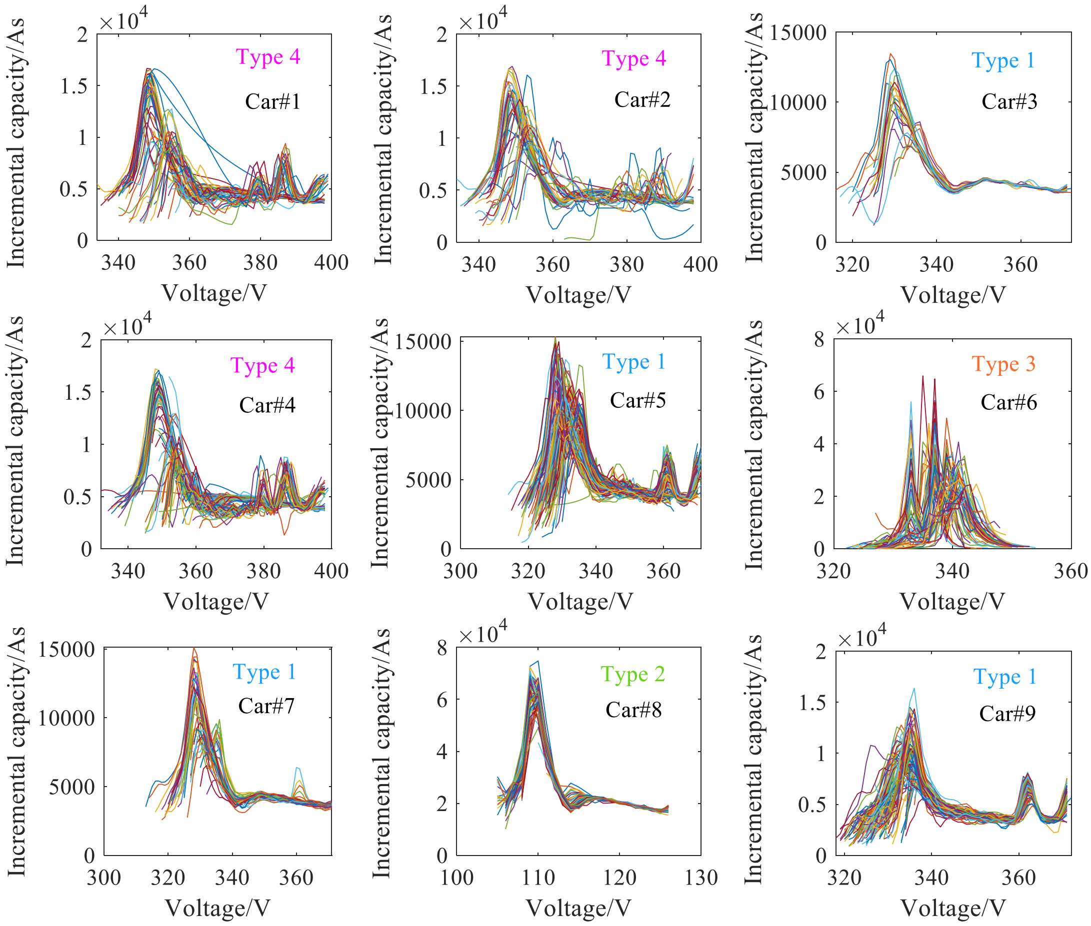

Fig 8. Discrete incremental capacity curves of Car#1-#9.

4. SoH estimation and prediction

In this section, the SoH index will firstly be determined based on the discrete incremental capacity curves formed in previous Section. Then, the SoH will be estimated based on the given data. Moreover, the SoH after six and twelve months will be predicted, indirectly representing a RUL assessment. The computational resource assessment is also presented to demonstrate the superiority and feasibility of this method.

4.1 Determining the SoH index

In the previous section, the discrete incremental capacity curves of each Car were formed. By adding the incremental capacity of each voltage points to a curve form, the battery capacity of the EV under the corresponding mileage conditions, i.e. health state, could be estimated. Taking Car # 9 as example, the 1st and 2nd curve shown in Fig. 9 (a) stand for the incremental capacity variation of voltage point during charging process with mileage of 91501 km and 92342 km, respectively. The shape of the two curves is somewhat different, i.e., the incremental capacity versus voltage and the voltage span have changed because of the mileage increase and capacity fade leading to a difference in the cumulative incremental charging capacity of each voltage point on the curve. Here, this cumulative incremental capacity difference of the EV during the charging process under different mileage conditions is used to represent the SoH variation. The SoH is determined by summing up the incremental capacity of the overlap voltages (Fig. 9b) as follows:

$$\frac{\mathrm{SoH}(i+1)}{\mathrm{SoH}(i)} = \frac{\mathrm{Capacity}(i+1)}{\mathrm{Capacity}(i)} \tag{3}$$

where Capacity = sum of incremental capacity of overlap voltage points in each curve.

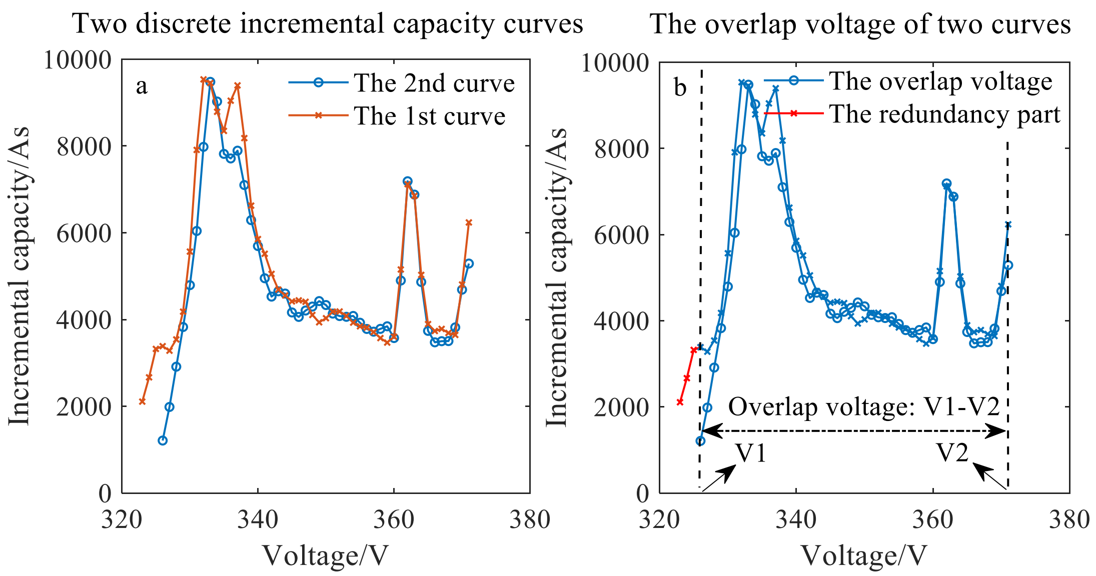

Fig 9. Overlap voltages of discrete incremental capacity curves. (a) The 1st and 2nd curve of Car#9, (b) overlap voltage of two curves.

4.2 SoH estimation

Once the discrete incremental capacity curves are formed, the relationship between the SoH and mileage can be formed through Eq. (3). Taking Car #9 as example, the SoH variation with the increase of mileage is shown in Fig. 10 (a). The SoH curve doesn't continue to fade as the mileage increases, but shows some stagnation. This phenomenon is, however, consistent with prior knowledge as the EV battery capacity must decrease with the mileage increase, which is a long-term trend, but some seasonal or cyclical factors such as seasonal temperature and driving habits will also inevitably cause small short-term fluctuations in the EV capacity. Therefore, the Empirical Mode Decomposition (EMD) method [29], which can be used for decomposing nonlinear, multicomponent signals, is employed here to decompose the SoH curve to extract the long-term trend part (Fig. 10 (b)) and short-term fluctuation part (Fig. 10 (c)). Also, the temperature variation is shown in Fig. 10 (d). The trend of the temperature curve and short-term fluctuation curve exhibit certain similarities, i.e., the temperature change may be considered as one of the main factors causing short-term fluctuation in the SoH of this Car. However, the temperature changes don't necessarily lead to the fluctuation of the SoH if the long-term trend part is strong enough.

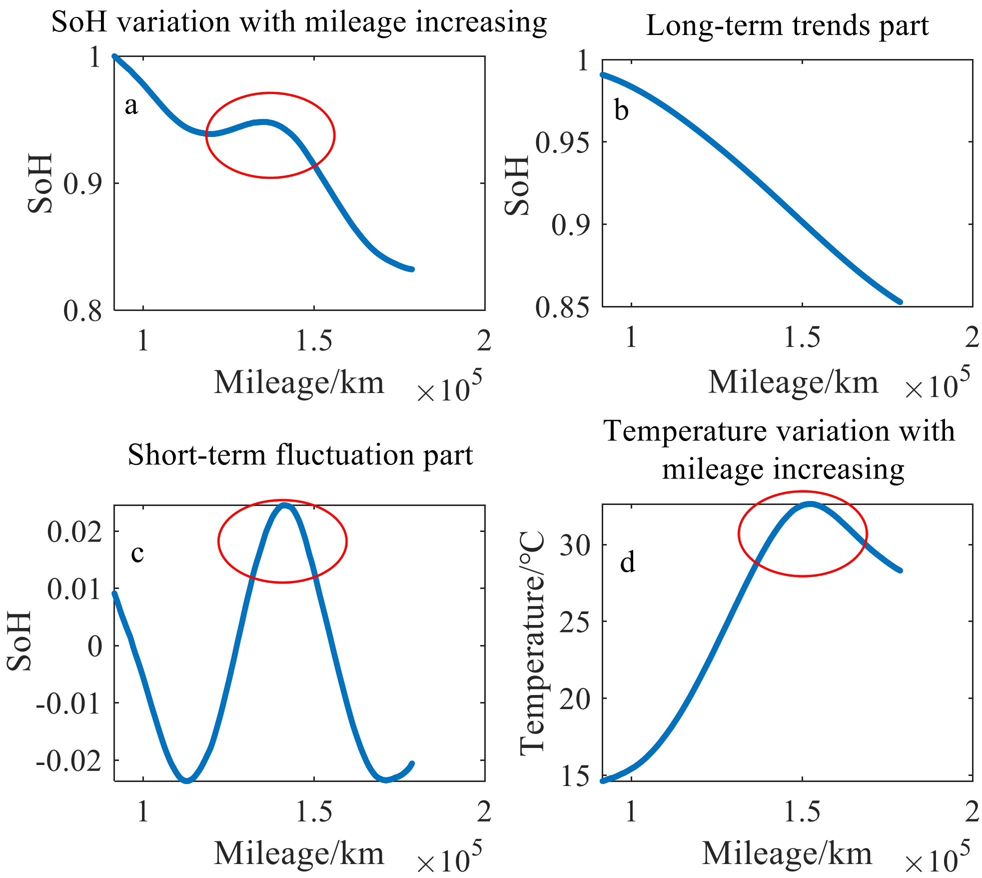

Fig 10. SoH and its decomposition for Car#9. (a) SoH curve of Car#9, (b) long-term trend part of SoH, (c) short-term fluctuation part of SoH, (d) temperature variation.

4.3 SoH prediction

Here, the SOH for the next 6 and 12 months will be predicted based on the SoH estimation results. First, the average daily mileage increase is calculated using the given data, and then the mileage after 6 and 12 months is estimated. The prediction of the SoH is based on building the SoH curve from the end mileage at 6 and 12 months.

The SoH curve was decomposed into long-term trend and short-term fluctuation part using the EMD. The numerical characteristics of these two parts are different, i.e., the long-term trend part is actually a nonstationary series with a linear trend while the short-term fluctuation part is a nonstationary series with periodic fluctuation characteristics. Therefore, the Third Exponential Smoothing Method (Holt-Winters) [30] and the Second Exponential Smoothing Method [30] are employed to predict the short-term fluctuation and long-term trend part of the SoH for the next 6 and 12 months.

Figure 11 shows the long-term and short-term part prediction result of Car#9. The prediction results of the long-term trend (Fig. 11a) and the short-term fluctuation part (Fig. 11b) show linearity and volatility, which is consistent with their own data characteristics. By adding these two parts together, the SoH estimation and prediction of Car#9 could be realized shown in Fig. 11(c). Figure 12 shows the SoH estimation and prediction results for all 9 Cars. On one hand, the SoH of all the 9 Cars shows a decreasing trend with increasing of driving mileage. On the other hand, these SoH curves show a varying degree of volatility indicating the effects of temperature, driving habits and charging habits on the SoH drop rate.

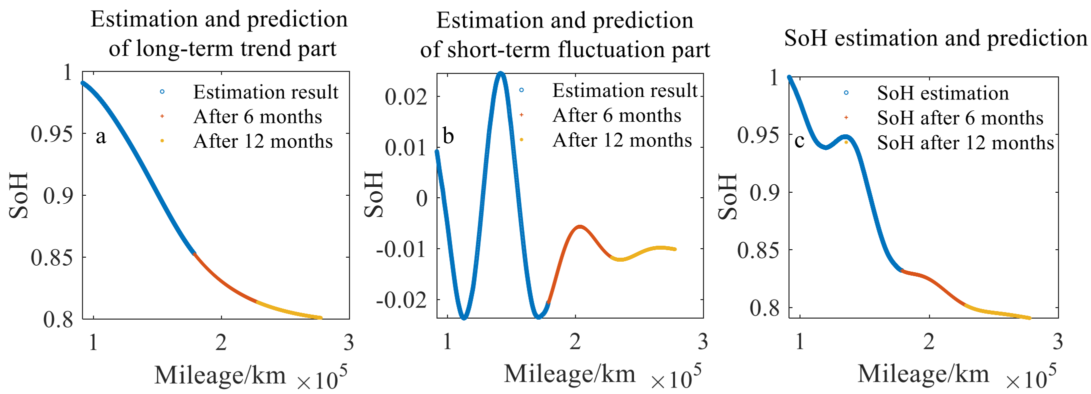

Fig 11. The SoH prediction of long-term and short-term part of Car #9. (a) Prediction of long-term trend part, (b) prediction of short-term fluctuation part, (c) prediction of SoH for Car#9.

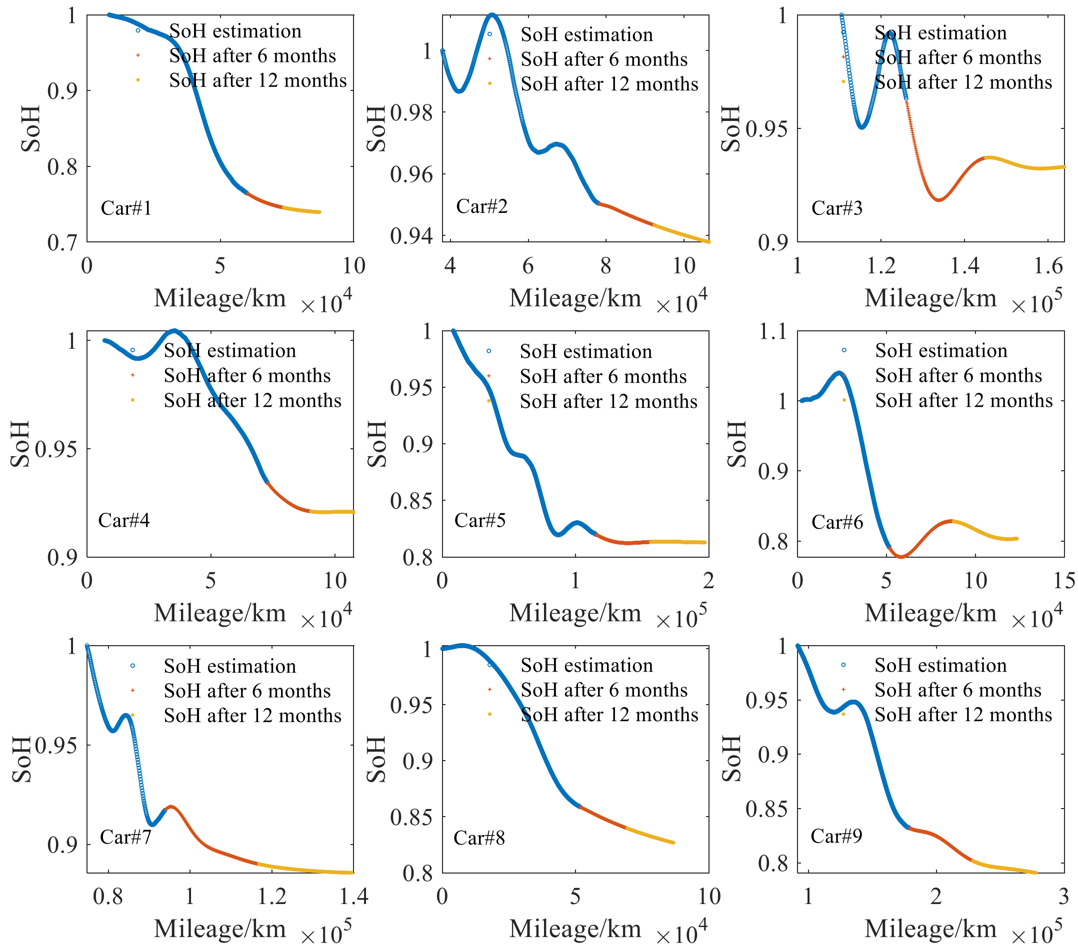

Fig 12. The SoH prediction results for Cars #1-9.

4.4 Computational resources and time

The method used in this work was built with explicit physical significance, i.e., the incremental capacity of the voltage actually reflects the electrochemical behavior of the electrode materials during the charging process yielding a high computational efficiency as complex algorithms typical for black-box modelling could be avoided. The computational time and peak memory needed for each car is given in Table 5. Matlab 2019b © was used in the calculations running on Intel(R) Core (TM)i7-9700 CPU @ 3.00GHz with RAM 16.0GB.

The SoH curve is formed by successively analyzing the sum of the charging capacity of overlap voltage of discrete incremental capacity curves. As listed in Table 5, taking Car #9 as an example, the number of discrete incremental capacity curves is 90, indicating that the SoH curve of Car#9 is formed with 90 SoH values. The total computer time was 31s and the peak memory used was 207Mb. Therefore, the time and peak memory needed for each curve, i.e., for each SoH update is 0.3s and 2.3Mb.

Taking all 9 Cars into consideration, the average time and memory for the SoH update was 1.3s and 4.3Mb. Considering the nearly 5 million EVs in China as an example, and the SoH of these EVs were updated 100 times, the total computer time and memory needed would be 130s and 208 Gb only, demonstrating the great potential of this method.

Table 5. Computational time and peak memory needed for each car.

ID	Number of data points	Number of IC curves	Computer time (s)	Peak memory (Mb)	Time for per curve (s)	Peak memory for each curve (Mb)
Car1	450596	65	30.0	55.1	0.5	0.8
Car2	374026	48	25.4	45.8	0.5	1.0
Car3	408461	20	41.1	128.3	2.1	6.4
Car4	560297	82	35.5	68.5	0.4	0.8
Car5	1644545	115	55.1	201.7	0.5	1.8
Car6	8430173	53	265.7	1031.1	5.0	19.5
Car7	605488	37	49.1	128.3	1.3	3.5
Car8	1235292	65	62.9	152.6	1.0	2.3
Car9	1688715	90	31.0	206.5	0.3	2.3
				Average	1.3	4.3

5. Clustering analysis

A clustering analysis is performed to compare the SoH of the EV to that of other EVs with same type under certain mileage conditions. The clustering result can help the driver to know if the Car is performing better or worse than the other Cars for the same mileage.

The 9 Cars represent 4 different types of EVs (Type 1-4), i.e., different EV manufacturers, different battery systems and different endurance. In the following, Type 1 (including Car#3, Car#5, Car#7, Car#9) is taken as an example to demonstrate the clustering process aiming at analyzing the SoH level of Car#3 relative to Car#5, Car#7 and Car#9 under the same mileage condition. The discrete incremental capacity curves of Car#5, Car#7 and Car#9 are shown in Fig. 13(a). Each curve represents the incremental capacity variation of voltage during the charging process at a certain mileage. It may happen that three curves in Fig. 13(a) corresponds to one certain mileage. Therefore, the average of these three curves is taken to represent the incremental capacity level of the voltage at that mileage situation regarding this type of EVs. The discrete incremental capacity curves of Car#5, Car#7 and Car#9 after the averaging process is shown in Fig. 13(b).

Taking the first curve of Car#3 as an example and considering it as the target curve with 110492 km. Then the curve with mileage closest to 110492 km is selected from Fig. 13(b) as the average curve representing the incremental capacity level of Car#5, Car#7 and Car#9 under same mileage situation. The SoH calculation is done by analyzing the target curve and average curve using Eq. (2). Then, the SoH change of Car#3 is compared with Car#5, Car#7 and Car#9 at 110492 km expressed in Eq. (4), and shown by the red dot in Fig. 13(c). The SoH change of Car#3 at 110492 km is about 0% indicating that the SoH level of Car#3 is actually similar to that of the other Cars at current mileage.

$$SoH\ change\ \ at\ current\ mileage = \frac{SoH(target\ vurve) - SoH(average\ curve)}{SoH(average\ curve)} \tag{4}$$

With increasing mileage, the SoH change of Car#3 for different mileage fluctuates around 0%, e.g., -5% at ca. 112000 km, +5% at 113000 km etc., indicating that on long term the SoH of Car#3 is close to the average SoH level of this type of Cars.

After the clustering analysis, the discrete incremental capacity curves of Car#3 are added into that of the Car#5, Car#7 and Car#9 to update the discrete incremental capacity curves of Type 1 shown in Fig. 13(d, e). The implication of Figs. 13(d, e) is similar to that of Figs. 13(a, b). The only difference is that Figs. 13(d, e) include the discrete incremental capacity information of Car#3, which can be considered as an update of Figs. 13(a, b). If more Cars of Type 1 were available, e.g., Car#10, the clustering analysis of Car#10 could easily be performed by comparing it with Car#3, Car#5, Car#7 and Car#9 based on Figs. 13(d, e).

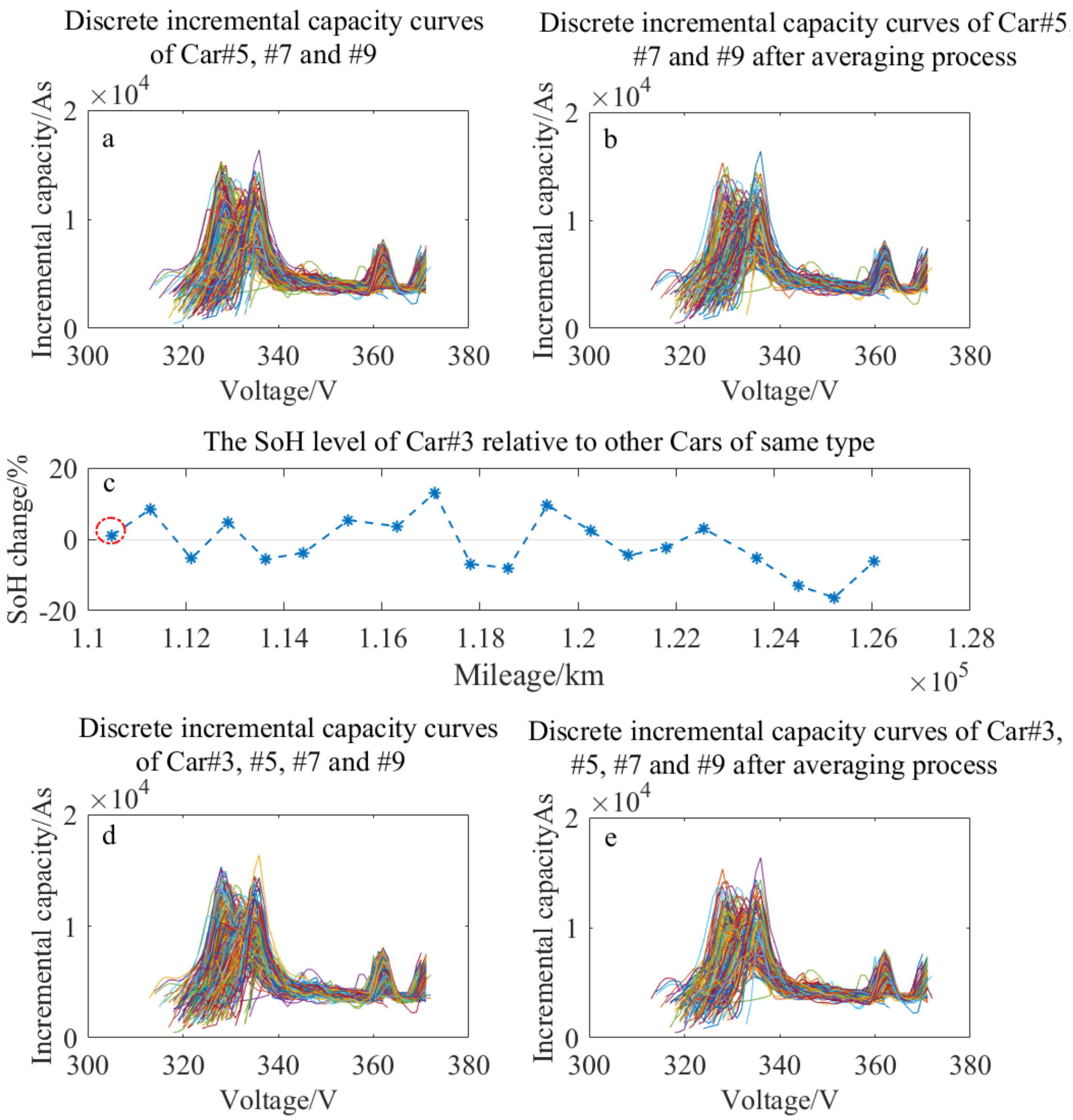

Fig 13. The clustering analysis of Car#3, #5, #7 and #9. (a) Discrete incremental capacity curves of Car#5, Car#7 and Car#9, (b) discrete incremental capacity curves of Car#5, Car#7 and Car#9 after averaging process, (c) clustering analysis result of Car#3

to Car#5, Car#7 and Car#9, (d) discrete incremental capacity curves of Car#3, Car#5, Car#7 and Car#9, (e) discrete incremental capacity curves of Car#3, Car#5, Car#7 and Car#9 after averaging process.

6. Conclusions

In this paper, a novel state of health estimation method for EV batteries is proposed based on the discrete incremental capacity analysis. This method provides a new direction for estimating the battery health with great robustness, high compatibility, short computing time requirement and low memory need.

Taking the 9 Cars in this paper as an example, it's clear that the SoH of EVs doesn't continue to fade linearly as the driving mileage increases, but shows varying degrees of stagnation and fluctuation, which could be considered as an effect from the seasonal temperature variation, driving habits and charging strategy. The computer time and memory needed for SoH estimation of 5 million EVs in China would be 130s and 208Gb only, also demonstrating the potential of commercialization of this method in the future. The clustering process presented here provides useful information to the EV driver on the SoH situation compared to other EVs with the same mileage.

Declaration of conflicting interest

There is no conflict of interest.

Acknowledgments

This work is based on a first prize competition entry by one of the authors (Xu) to the National College New Energy Vehicle Big Data Application Innovation Competition (http://www.ncbdc.top/) organized by the National Big Data Alliance of New Energy Vehicles (NDANEV: http://www.ndanev.com/). The authors thank NDANEV for providing data on the EVs. The financial support by the National Science Foundation of China (Grant number 51736006) is greatly acknowledged. This work was partially supported by Aalto University.

References

[1] Zhicheng Xu, Jun Wang, Qi Fan, Peter Lund, Jie Hong. Improving the state of charge estimation of reused lithium-ion batteries by abating hysteresis using machine learning technique. *J Energy Storage* 2020, 32:101678. https://doi.org/10.1016/j.est.2020.101678.

[2] MK Tran, M Akinsanya, S Panchal, R Fraser, M Fowler. Design of a Hybrid Electric Vehicle Powertrain for Performance Optimization Considering Various Powertrain Components and Configurations. *Vehicles* 2021, 3(1): 20-32. https://doi.org/10.3390/vehicles3010002.

[3] Hongbin Ren, Yuzhuang Zhao, Sizhong Chen, Taipeng Wang. Design and implementation of a battery management system with active charge balance based on the SOC and SOH online estimation, *Energy* 2019, 166: 908-917. https://doi.org/10.1016/j.energy.2018.10.133.

[4] Jinhao Meng, Lei Cai, Daniel-Ioan Stroe, Guangzhao Luo, Xin Sui, Remus Teodorescu. Lithium-ion battery state-of-health estimation in electric vehicle using optimized partial charging voltage profiles. *Energy* 2019, 185:1054-1062. https://doi.org/10.1016/j.energy.2019.07.127.

[5] Zhicheng Xu, Jun Wang, Peter Lund, Qi Fan, Ting Dong, Yan Liang, Hong Jie. A novel clustering algorithm for grouping and cascade utilization of retired Li-ion batteries. *J Energy Storage* 2020,29:101303. https://doi.org/10.1016/j.est.2020.101303.

[6] Adnan Nuhic, Tarik Terzimehic, Thomas Soczka-Guth, Michael Buchholz, Klaus Dietmayer. Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods. *J Power Sources* 2013, 239:680–688. https://doi.org/10.1016/j.jpowsour.2012.11.146.

[7] Xiaosong Hu, Fei Feng, Kailong Liu, Lei Zhang, Jiale Xie, Bo Liu. State estimation for advanced battery management: Key challenges and future trends. *Renew Sustain Energy Rev* 2019, 114:109334. https://doi.org/10.1016/j.rser.2019.109334.

[8] Qian Lin, Jun Wang, Rui Xiong, Weixiang Shen, Hongwen He. Towards a smarter

battery management system: A critical review on optimal charging methods of lithium ion batteries. *Energy* 2019, 183:220–234. https://doi.org/10.1016/j.energy.2019.06.128.

[9] Changfu Zou, Chris Manzie, Dragan Nesic. A Framework for Simplification of PDE-Based Lithium-Ion Battery Models. *IEEE Trans Control Syst Technol* 2016, 24:1594–1609. https://doi.org/10.1109/TCST.2015.2502899.

[10] Lu, J., Wu, T. and Amine, K. State-of-the-art characterization techniques for advanced lithium-ion batteries. *Nat Energy* 2017**,** 2:17011. https://doi.org/10.1038/nenergy.2017.11.

[11] S. Panchal, J. McGrory, J. Kong, I. Dincer, M. Agelin-Chaab, R. Fraser, M. Fowler. Cycling degradation testing and analysis of a LiFePO4 battery at actual conditions. *International Journal of Energy Research* 2017, 41(15):2565-2575. https://doi.org/10.1002/er.3837.

[12] Samadani, E., Farhad, S., Panchal, S., Fraser, R., Fowler, M. Modeling and Evaluation of Li-Ion Battery Performance Based on the Electric Vehicle Field Tests. SAE Technical Paper, 2014-01-1848, 2014. https://doi.org/10.4271/2014-01-1848.

[13] Wladislaw Waag, Dirk Uwe Sauer. Adaptive estimation of the electromotive force of the lithium-ion battery after current interruption for an accurate state-of-charge and capacity determination. *Appl Energy* 2013, 111:416–427. https://doi.org/10.1016/j.apenergy.2013.05.001.

[14] Lin Chen, Huimin Wang, Bohao Liu, Yijue Wang, Yunhui Ding, Haihong Pan. Battery state-of-health estimation based on a metabolic extreme learning machine combining degradation state model and error compensation. *Energy* 2021, 215:119078. https://doi.org/10.1016/j.energy.2020.119078.

[15] Jinhao Meng, Lei Cai, Daniel-Ioan-Stroe, Junpeng Ma, Guangzhao Luo, Remus Teodorescu. An optimized ensemble learning framework for lithium-ion Battery State of Health estimation in energy storage system. *Energy* 2020, 206:118140. https://doi.org/10.1016/j.energy.2020.118140.

[16] Yuanwang Deng, Hejie Ying, Jiaqiang E, Hao Zhu, Kexiang Wei, Jingwei Chen, Feng Zhang, Gaoliang Liao. Feature parameter extraction and intelligent

estimation of the State-of-Health of lithium-ion batteries, *Energy* 2019, 176:91-102. https://doi.org/10.1016/j.energy.2019.03.177.

[17] Sheng Shen, Mohammadkazem Sadoughi, Xiangyi Chen, Mingyi Hong, Chao Hu A deep learning method for online capacity estimation of lithium-ion batteries. *J Energy Storage* 2019, 25:100817. https://doi.org/10.1016/j.est.2019.100817.

[18] Xing Shu, Guang Li, Jiangwei Shen, Zhenzhen Lei, Zheng Chen, Yonggang Liu. A uniform estimation framework for state of health of lithium-ion batteries considering feature extraction and parameters optimization. *Energy* 2020, 204:117957. https://doi.org/10.1016/j.energy.2020.117957.

[19] Duo Yang, Xu Zhang, Rui Pan, Yujie Wang, Zonghai Chen. A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve. *J Power Sources* 2018,384:387–395. https://doi.org/10.1016/j.jpowsour.2018.03.015.

[20] Robert R Richardson, Michael A Osborne, David A Howey. Gaussian process regression for forecasting battery state of health. *J Power Sources* 2017,357:209–219. https://doi.org/10.1016/j.jpowsour.2017.05.004.

[21] Piyush Tagade, Krishnan S Hariharan, Sanoop Ramachandran, Ashish Khandelwal, Arunava Naha, Subramanya Mayya Kolake, Seong Ho Han. Deep Gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis. *J Power Sources* 2020, 445:227281. https://doi.org/10.1016/j.jpowsour.2019.227281.

[22] Jinhao Meng, Lei Cai, Guangzhao Luo, Daniel Ioan Stroe, Remus Teodorescu. Lithium-ion battery state of health estimation with short-term current pulse test and support vector machine. *Microelectron Reliab* 2018,88–90:1216–1220. https://doi.org/10.1016/j.microrel.2018.07.025.

[23] Matthieu Dubarry, Bor Yann Liaw, Mao-Sung Chen, Sain-Syan Chyan, Kuo-Chang Han, Wun-Tong Sie, She-Huang Wu. Identifying battery aging mechanisms in large format Li ion cells. *J Power Sources* 2011, 196:3420–3425. https://doi.org/10.1016/j.jpowsour.2010.07.029.

[24] Caihao Weng, Yujia Cui, Jing Sun, Huei Peng. On-board state of health monitoring

of lithium-ion batteries using incremental capacity analysis with support vector regression. *J Power Sources* 2013, 235:36–44. https://doi.org/10.1016/j.jpowsour.2013.02.012.

[25] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, P. Van den Bossche. Critical review of state of health estimation methods of Li-ion batteries for real applications. *Renew Sustain Energy Rev* 2016, 56:572–587. https://doi.org/10.1016/j.rser.2015.11.042.

[26] Xiaoyu Li, Changgui Yuan, Zhenpo Wang. State of health estimation for Li-ion battery via partial incremental capacity analysis based on support vector regression. *Energy* 2020, 203:117852. https://doi.org/10.1016/j.energy.2020.117852.

[27] Bo Jiang, Haifeng Dai, Xuezhe Wei. Incremental capacity analysis based adaptive capacity estimation for lithium-ion battery considering charging condition. *Appl Energy* 2020, 269:115074. https://doi.org/10.1016/j.apenergy.2020.115074.

[28] National Big Data Alliance of New Energy Vehicles. Available online: http://www.ndanev.com (accessed on 20 Dec 2020).

[29] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung and Henry H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. Royal Soc. A* 1998, 454: 903-995. https://doi.org/10.1098/rspa.1998.0193.

[30] E.S. Gardner Jr, Exponential smoothing: The state of the art—Part II, *Int J Forecast* 2006, 22: 637-666. https://doi.org/10.1016/j.ijforecast.2006.03.005.