eTransportation 3 (2020) 100051

Contents lists available at ScienceDirect

eTransportation

journal homepage: www.journals.elsevier.com/etransportation

A reliable approach of differentiating discrete sampled-data for battery diagnosis

Xuning Feng [a, b], Yu Merla [c], Caihao Weng [d], Minggao Ouyang [b, *], Xiangming He [a, **], Bor Yann Liaw [e], Shriram Santhanagopalan [f], Xuemin Li [f], Ping Liu [g], Languang Lu [b], Xuebing Han [b], Dongsheng Ren [b], Yu Wang [b], Ruihe Li [b], Changyong Jin [b, h], Peng Huang [b, h], Mengchao Yi [b, i], Li Wang [a], Yan Zhao [c], Yatish Patel [c], Gregory Offer [c, ***]

[a] Institute of Nuclear and New Energy Technology, Tsinghua University, Beijing, 100084, China
[b] State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing, 100084, China
[c] Department of Mechanical Engineering, Imperial College London, Exhibition Road, South Kensington Campus, London, SW7 2AZ, UK
[d] Department of Naval Architecture and Marine Engineering, University of Michigan, Ann Arbor, MI, 48109, USA
[e] Idaho National Laboratory, P.O. Box 1625, Idaho Falls, ID, 83415, USA
[f] Transportation and Hydrogen System Center, National Renewable Energy Laboratories, Golden, CO, 80401, USA
[g] Department of NanoEngineering, University of California, San Diego, CA, 92093-0448, USA
[h] College of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, 200093, China
[i] College of Engineering, China Agricultural University, Beijing, 100083, China

ARTICLE INFO

Article history:
Received 14 January 2020
Received in revised form
13 February 2020
Accepted 16 February 2020
Available online 22 February 2020

Keywords:
Battery aging
Lithium-ion batteries
Energy storage
Differential analysis
Incremental capacity analysis
Differential thermal voltammetry

ABSTRACT

Over the past decade, major progress in diagnosis of battery degradation has had a substantial effect on the development of electric vehicles. However, despite recent advances, most studies suffer from fatal flaws in how the data are processed caused by discrete sampling levels and associated noise, requiring smoothing algorithms that are not reliable or reproducible. We report the realization of an accurate and reproducible approach, as "Level Evaluation ANalysis" or LEAN method, to diagnose the battery degradation based on counting the number of points at each sampling level, of which the accuracy and reproducibility is proven by mathematical arguments. Its reliability is verified to be consistent with previously published data from four laboratories around the world. The simple code, exact fitting, consistent outcome, computational availability and reliability make the LEAN method promising for vehicular application in both the big data analysis on the cloud and the online battery monitoring, supporting the intelligent management of power sources for autonomous vehicles.

1. Introduction

The world is becoming increasingly reliant on batteries for multiple facets, such as portable electronic devices [1,2], electric vehicles [3–5], and stationary electrical storage [6]. Long cycle life is one of the most significant features for the batteries in application [7,8]. The diagnosis of battery degradation is of great significance in the intelligent energy management of all kinds of electrochemical power sources [9–11]. Yet, degradation is highly unpredictable creating huge uncertainties over lifetime and total cost of ownership, setting a critical barrier for their widespread economic adoption. Researchers have developed a wide range of diagnostic techniques that can be used to understand degradation [12,13]. These techniques range from destructive ex situ experiments to in situ experiments. A number of the most promising in situ techniques, particularly those with the potential to be used during application, can be classified as 'differential techniques', which compare the rate of change in current to the rate of change in voltage, or vice versa, such as the incremental capacity analysis (ICA) or the differential voltage analysis (DVA), or the rate of change in temperature to the rate of change in voltage, such as the differential thermal voltammetry (DTV).

In 1967, Balewski and Brenet [14] published the first paper using

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: ouymg@tsinghua.edu.cn (M. Ouyang), hexm@tsinghua.edu.cn (X. He), gregory.offer@imperial.ac.uk (G. Offer).

https://doi.org/10.1016/j.etran.2020.100051
2590-1168/

ICA to characterize the electrochemical reactivity of manganese dioxide. Dahn et al. [15] and Tillement and Quarton [16] endorsed the use of ICA by applying it to the characterization of the electrochemical insertion of cell materials. Since 2005, Dubarry et al. [17,18] have pioneered the application of ICA for full-cell degradation analysis. The ICA technique is similar to cyclic voltammetry (CV) [19] in characterizing battery behaviour [20,21] but is typically faster and can also be applied to characterize the performance of large-format batteries [22]. Using ICA to characterize battery performance has received increasing attention, and the number of research papers has continuously increased over the past five years. DVA is complementary to ICA because their results are reciprocals [19]. Bloom et al. [23] first used DVA to characterize capacity degradation, followed by Honkura et al. [24] and others [25,26]. In a significant recent development, Wu et al. [27] proposed a new approach termed differential thermal voltammetry (DTV) to track battery degradation in a similar fashion to ICA but using temperature as the alternative signal to current, which makes it possible to analyse cells in parallel and does not require constant temperature conditions.

However, there is a challenge that prevents ICA, DVA and DTV from becoming standardized tools for degradation analysis, either in the lab or during application. Taking derivatives of discrete signals with sampling levels and significant noise introduces multiple problems [28], as shown in Fig. 1, requiring data to be manipulated in various ways before it can be used. There is currently no standardized approach to manipulate data, and therefore, it is often not possible to compare data. The challenge of calculating the differentiation between discrete sampling signals that contain levels due to digital sampling with specific resolution and noise has thwarted scientists for years. An intuitive solution is to transfer signals with discrete levels and random noise into smoothed and differentiable functions, so-called “curve fitting”, as the fitted curves have continuous functions that are more easily differentiable.

In 1964, Savitzky and Golay [29] proposed an approach to smooth and differentiate sampled data using piecewise polynomial curve fitting with optimality obtained by least squares fitting. The method has been widely accepted as the famous SG method and has been cited more than 15,000 times. Craven and Wahba [30] improved upon the SG method, which can give large fluctuations in the results of high-order polynomial coefficients, by regulating the derivatives using generalized cross-validation techniques. Samad et al. [31] successfully applied the SG method to calculate the ICA curves for battery degradation analysis. In recent years, techniques for optimal curve fitting have been greatly improved upon through

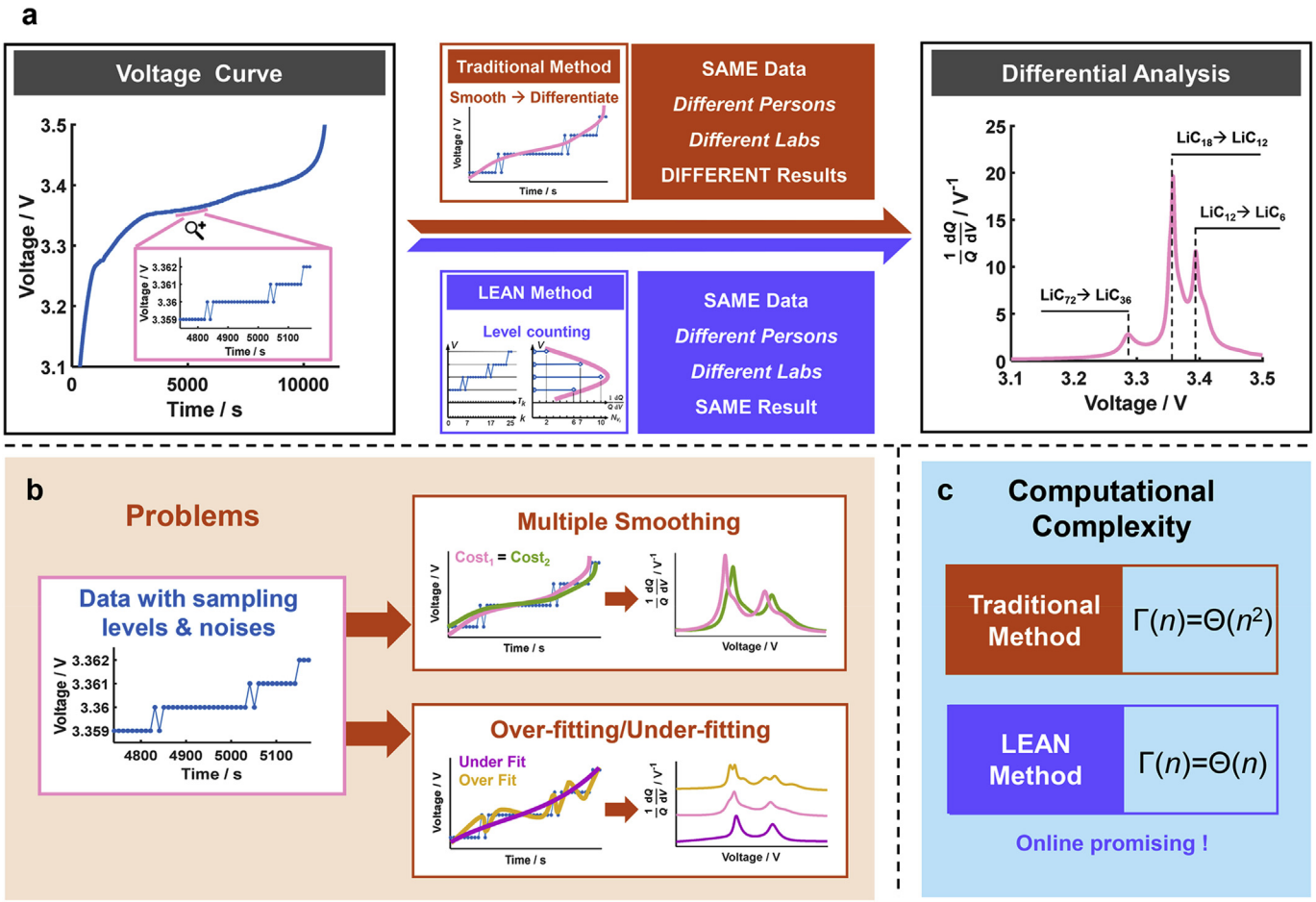

Fig. 1. Overview of the method and motivation. a (*Left*) The voltage curve under 1/3C rate charging for a 60Ah commercial lithium-ion cell with $LiFePO_4$/graphite electrodes. The magnified figure shows the actual raw data, including sampling levels and noise. **a** (*Right*) The ideal IC curve that contains three characteristic peaks, which reflect the phase changes in the graphite anode. **a** (*Middle-Top*) The traditional method that uses smoothing techniques first and then calculates the differentiation of the smoothed function. **a** (*Middle-Bottom*) The proposed method, termed Level Evaluation Analysis (LEAN), which directly counts the numbers at each voltage level to acquire the IC curve. **b** The problem of the traditional method. First, multiple optimal results of curve fitting are possible during smoothing. Second, over-fitting/under-fitting misrepresents the shape and location of the IC peaks. **c** Comparison of the computational complexity $\Gamma(n)$. $\Gamma(n)$ is usually $\Theta(n^2)$ for the curve-fitting method, whereas $\Gamma(n) = \Theta(n)$ for the proposed LEAN method.

the use of artificial intelligence [32]. Weng et al. [33] and Wang et al. [34] used support vector machines, which is a basic technique in pattern recognition for kernel-based curve fitting, to smooth the sampling data and calculate the ICA curves by linearized support vector regression (SVR). Weng et al. [35] further extended the usage of SVR from processing the IC curves for cells to that for a battery pack.

However, smoothing approaches, including both SG and SVR, share inherent deficiencies: ① The adjustable parameters in the smoothing algorithms influence the result of curve fitting, especially the peak position and height of the IC curve, which are critical indicators of battery thermodynamics; ② Under-fitting or over-fitting is possible, thereby distorting the information in the raw data; ③ Due to the deficiencies in ① and ②, different researchers in different labs will have diverse results in ICA/DVA/DTV, thereby undermining the confidence of the results in publications and reports; ④ The computational complexity of curve-fitting algorithms is always $\Theta(n^2)$, where n is the total length of the data to be smoothed, making it impractical to apply curve-fitting methods for the purpose of online performance characterization.

These implicit facts ①-④ are rarely discussed but significantly limit the usefulness of such techniques in real-world implementations. In this paper, we propose an approach named "Level Evaluation ANalysis" (LEAN) to calculate the differentiations by counting the numbers at the sampling levels. The LEAN method was proved to be equivalent to ICA and DVA after careful consideration of battery degradation data and curve-fitting codes from Tsinghua University [36–38], Imperial College London [39], and University of Michigan [40]. Further discussions convinced us that the LEAN method can overcome all four deficiencies of traditional methods of differential analysis and demonstrated that the LEAN method has the following advantages: (i) very simple code with only a few parameters to adjust; (ii) no under-fitting or over-fitting; (iii) different researchers at different labs will derive the same results for ICA/DVA/DTV for the same set of data; and (iv) the computational complexity is $\Theta(n)$, making the LEAN method promising for further vehicular application in both the big data analysis on the cloud and the online battery monitoring, supporting the intelligent management of all kinds of electrochemical energy storage systems for electric vehicles, portable electronics, and power station.

2. Experimental

Battery degradation data from Tsinghua University (THU) [36–38], Imperial College London (ICL) [39], and University of Michigan, Ann Arbor (UoM) [40] were collected for cross validation of the LEAN method, as in Table 1. Cells A, C, E and F were tested by THU, Cell B was tested by UoM, and Cell D was tested by ICL. To ensure the universality of the LEAN method, the six selected cells cover various chemistries under commercial use, with a wide range of capacities from 5 Ah to 60 Ah and galvanostatic current rates (both charging and discharging) from 1/3C–2C, and also at different aging status.

3. Methods

3.1. The sampling data of battery tests

The sampling data of battery tests usually displays the stair-like feature as shown in Fig. 2(a), when we magnify the voltage curves to see the details. The stair-like feature is naturally formed when the monitoring system is using digital sampling chips. Direct differentiation by subtracting adjacent data points seems impossible, because the results contain many zeros (0) or infinities (∞). Intuitively, textbook tells us that curve-fitting is essential to smooth the sampling data y with an explicit function $f(x)$, then the differentiation of the sampling data can be calculated by $f'(x)$.

3.2. The curve fitting method

3.2.1. The canonical form

Assume x and y represent the input and output data to be fitted, respectively. For the battery sample-data for differentiation purpose, x can be the time τ, the charged/discharged capacity Q, or the state-of-charge SOC, whereas y is the sampled voltage V. Let vector $\boldsymbol{x} = \{x_1, x_2, x_3 \dots x_n\}$ and $\boldsymbol{y} = \{y_1, y_2, y_3 \dots y_n\}$ be the data set of x and y, respectively. n is the total number of data for curve fitting. For the Canonical Form, the output data y is fitted using a function $f(x)$:

$$\widehat{y} = f(x) \tag{1}$$

where $\widehat{y}$ is approximation of the fitted curve. The estimation error $\boldsymbol{e}$ is defined by:

$$\boldsymbol{e} = \boldsymbol{y} - \widehat{\boldsymbol{y}} \tag{2}$$

All kinds of curve fitting method are trying to find an optimized $f(x)$ that can minimize some forms of $\boldsymbol{e}$. For instances, to find $f(x)$ that minimizes $J = \boldsymbol{e}^T\boldsymbol{e}$ is common in quadratic optimization. The differentiations $\frac{dy}{dx}$ or $\frac{dx}{dy}$ can be calculated by the differentiation of $f(x)$, which has explicit expression:

$$\begin{cases} \dfrac{dy}{dx} = f'(x) \\ \dfrac{dx}{dy} = \dfrac{1}{f'(x)} \end{cases} \tag{3}$$

3.2.2. The SG method

The SG method was created by Savitzky and Golay[29] in 1964 for smoothing and differentiating sampling data using piecewise polynomial functions with optimality obtained by least squares fitting. To acquire the differentiations of $V(y)$ at time $\tau(x)$, the

Table 1
Degradation test for validation of the LEAN method. The abbreviations for the cathode are as follows: 1) LFP = $LiFePO_4$, 2) NCM=$Li(Ni_xCo_yMn_{1-x-y})O_2$, and 3) LMO = $LiMn_2O_4$. The abbreviations for the anode are as follows: 1) C = carbon- or graphite-based anode and 2) LTO = $Li_4Ti_5O_{12}$. The voltage ranges comply with the instructions provided by the manufacturers.

Cell	Cathode	Anode	Capacity (Ah)	Voltage (V)	Current	Institute	Data Ref.
A	LFP	C	60	2.75–3.6	1/3C CHA	THU	[36]
B	LFP	C	1.1	3.0–3.55	1/2C CHA	UoM	[40]
C	NCM	C	48	3.0–4.2	1/3C CHA	THU	[36]
D	NCM	C	5	2.7–4.2	2C DIS	ICL	[39]
E	NCM + LMO	C	20	2.5–4.2	1/2C DIS	THU	[38]
F	NCM	LTO	15	1.5–2.7	1/3 CHA	THU	[37]

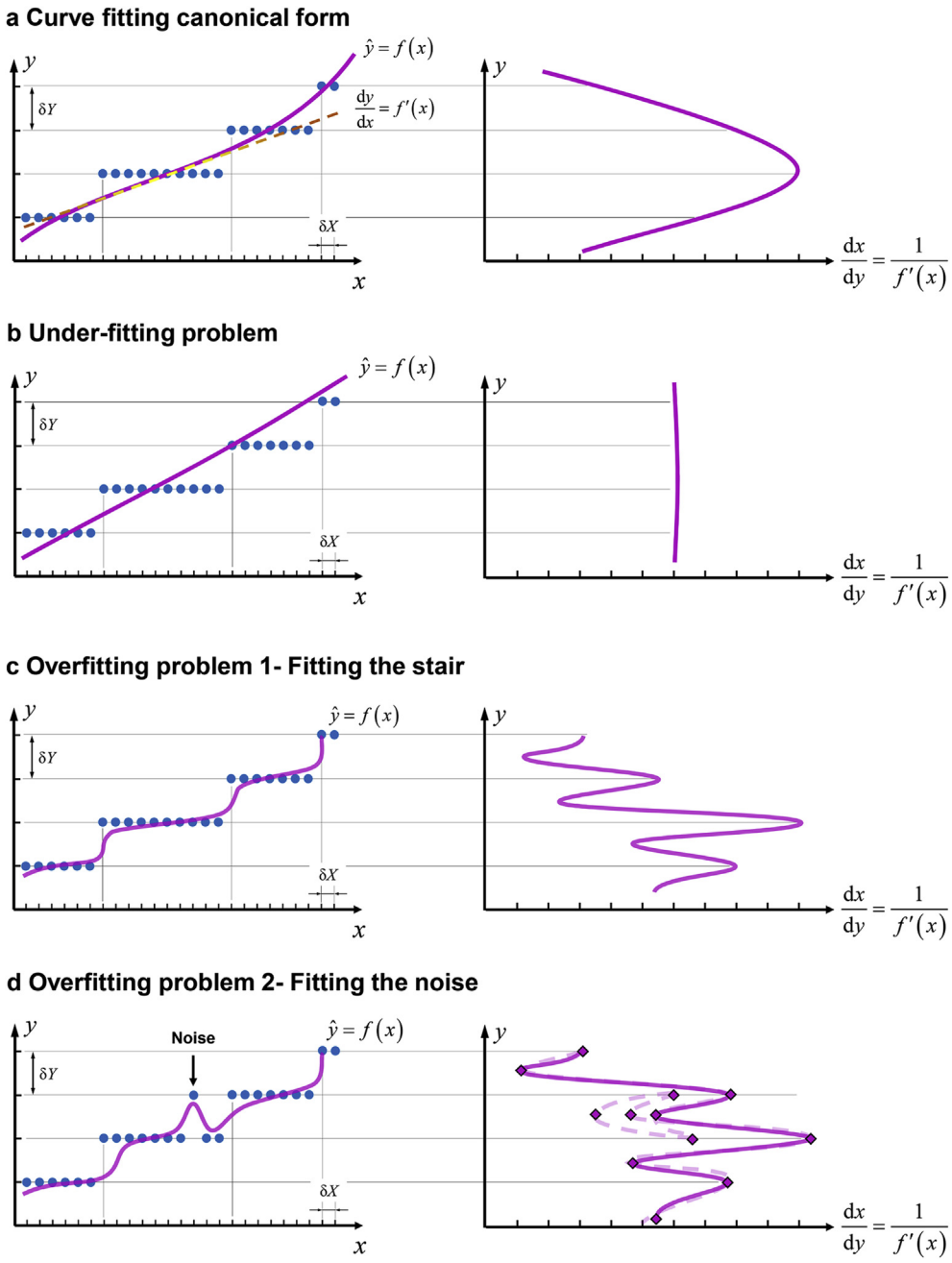

Fig. 2. The curve fitting methods. a The canonical form of curve fitting, and the differentiation is calculated from the fitted function. **b** The under-fitting problem that is common for curve fitting methods. **c** The overfitting problem caused by fitting the sampling stairs. **d** The overfitting problem caused by fitting the noise.

piecewise sampling data $\boldsymbol{y}$ are fitted as:

$$\boldsymbol{y} = [y_{k-w}, y_{k-w+1}, \ldots y_{k-1}, y_k, y_{k+1}, \ldots y_{k+w-1}, y_{k+w}]^T \tag{4}$$

where the window length is $2w+1$. The SG method tries to fit y by $\hat{y}$ using polynomial functions f with an order of p:

$$\hat{y} = f(x) = \boldsymbol{c}^T \cdot \boldsymbol{\kappa}(x) = \sum_{i=0}^{p} c_i \cdot x^i \tag{5}$$

where $\boldsymbol{c} = [c_0, c_1, \ldots c_p]^T \in R^{(p+1)\times 1}$ and $\boldsymbol{\kappa}(x) = [1, x, x^2, \ldots x^p]^T$. To compare the fitting errors calculated from the piecewise data, a vector $\hat{\boldsymbol{y}}$ with a length of $2w+1$ is generated as follows:

$$\hat{\boldsymbol{y}} = [f(-w), f(-w+1), \ldots f(-1), f(0), f(1), \ldots f(w-1), f(w)]^T = G \cdot \boldsymbol{c} \tag{6}$$

where the matrix $G = [\boldsymbol{g_0}, \boldsymbol{g_1}, \boldsymbol{g_2}, \ldots \boldsymbol{g_p}]^T \in R^{(2w+1)\times(p+1)}$ and the vector $\boldsymbol{g_k} = [-w^k, (-w+1)^k, \ldots, (-1)^k, 0^k, 1^k, \ldots (w-1)^k, w^k]^T$. Define $\boldsymbol{e}$ as the vector for the errors calculated from the piecewise data such that:

$$\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - G \cdot \boldsymbol{c} \tag{7}$$

and the SG method performs the following optimization:

$$\min_{\boldsymbol{c}} \boldsymbol{e}^T\boldsymbol{e} = \min_{\boldsymbol{c}} (\boldsymbol{y} - G\cdot\boldsymbol{c})^T(\boldsymbol{y} - G\cdot\boldsymbol{c}) \tag{8}$$

$\boldsymbol{c}$ is the optimal solution for the quadratic optimization, and the incremental capacity (IC) at time τ_k is the reciprocal of c_1.

$$\boldsymbol{c} = \left(G^TG\right)\cdot G^T\boldsymbol{y} \tag{9}$$

$$\left.\frac{\mathrm{d}Q}{\mathrm{d}V}\right|_{\tau=\tau_k} = I\cdot\left.\frac{\mathrm{d}\tau}{\mathrm{d}V}\right|_{\tau=\tau_k} = I\cdot\frac{\mathrm{d}x}{\mathrm{d}y} = I\cdot\frac{1}{f'(x)} = I\cdot\frac{1}{c_1} \tag{10}$$

3.2.3. The SVR method

The SVR method is the application of a support vector algorithm in artificial intelligence, which was initially developed by Drucker et al. at&T Bell Laboratories for data regression [41]. The data are fitted using a linearized parametric model:

$$\widehat{y} = f(x) = \boldsymbol{c}^T\cdot\boldsymbol{\kappa}(x) + \mu = \sum_{k=0}^{n} c_k\cdot\kappa(x_k, x) + \mu \tag{11}$$

where $\boldsymbol{c} = [c_0, c_1, \ldots c_n]^T \in R^{n\times 1}$, n is the total sampling number in S^V, μ is the offset constant, $\boldsymbol{\kappa}(x) = [\kappa(x_1, x),\ \kappa(x_2, x), \ldots\kappa(x_n, x)]^T$ is the vector of kernels, and x can be the time τ or the state of charge *SOC*. The arbitrary kernel function $\kappa(x_k, x)$ is defined as:

$$\kappa(x_k, x) = \exp\left(\frac{-\|x_k - x\|^2}{2\sigma^2}\right) \tag{12}$$

where σ is a pre-set parameter to control the shape of $\kappa(x_k, x)$. Define ε as the precision parameter that set some tolerance for the fitting error and slack variables ξ_k^+ and ξ_k^- to cope with infeasible constraints such that:

$$\xi_k = \begin{cases} \xi_k^+ = y - \widehat{y} - \varepsilon, & (y > \widehat{y} + \varepsilon) \\ \xi_k^- = \widehat{y} - \varepsilon - y, & (y < \widehat{y} - \varepsilon) \\ 0, \quad \text{otherwise} \end{cases} \tag{13}$$

Then, SVR using l_1 regularization formulates the optimization problem as follows:

$$\min_{\boldsymbol{c},\mu,\xi^+,\xi^-} \|\boldsymbol{c}\|_1 + \lambda\sum_{k=1}^{n}(\xi_k^- + \xi_k^+)$$
$$\text{subject to} \begin{cases} y_k - \widehat{y}_k \le \varepsilon + \xi_k^+ \\ \widehat{y}_k - y_k \le \varepsilon + \xi_k^- \\ \xi_k^+ \ge 0 \\ \xi_k^- \ge 0 \end{cases} \tag{14}$$

where λ is the weighting factor and $\|\cdot\|_1$ denotes the l_1 norm in the coefficient space. The optimization problem is solved using the function *linprog* in MATLAB®. As the computational complexity of the problem is $\Theta(n^2)$, the data are always re-sampled with a distance of D ($D = 200$ in this paper); therefore, the optimization can be downsized to n/D. The optimal result usually gives a near-zero value for most of the c_i, and those c_i that are much larger than zero ($>10^{-4}$) are regarded as significant. The corresponding x_i is called the support vector sv_i, of which the total number is N_{sv}. Therefore, the curve-fitting model is built as:

$$\widehat{y} = f(x) = \boldsymbol{c}^T\cdot\boldsymbol{\kappa}(x) + \mu = \sum_{i=0}^{N_{sv}} c_i\cdot\kappa(sv_i, x) + \mu \tag{15}$$

Furthermore, the incremental capacity can be calculated by:

$$\left.\frac{\mathrm{d}Q}{\mathrm{d}V}\right|_{\tau=\tau_k} = I\cdot\left.\frac{\mathrm{d}\tau}{\mathrm{d}V}\right|_{\tau=\tau_k} = I\cdot\frac{\mathrm{d}x}{\mathrm{d}y} = I\cdot\frac{1}{f'(x)} \tag{16}$$

3.2.4. The under-fitting/overfitting problem

Most of the curve fitting methods suffer from under-fitting or overfitting problem. Fig. 2(b)–(d) illustrates the possible under-fitting and overfitting problems that undermines the reliability of curve fitting in calculating differentiations. Fig. 2(b) shows a case for under-fitting, which means that although there are changes in the curvature, the fitting algorithm neglects the details and reports a smoothed constant. Under-fitting is common when the constraints for optimization are too loose. Fig. 2(c) illustrates an overfitting case caused by fitting the stair-like sampled voltage, which should be smoothed out during optimization. Extra peaks can be observed in the dx/dy curves at each level of the voltage stairs, as long as the curve fitting method tries to track the shape of stairs. Moreover, overfitting may occur if there are noises in the sampling sequence, as shown in Fig. 2(d). The curve fitting methods may try to fit the noise when the constraints for optimization are too tight. However, how to define "loose" and "tight" in setting the optimization problem for curve fitting relies highly on the experience of the operator, thereby causing that different researchers in different labs will have diverse results in fitting a same sampled-data.

3.3. The LEAN method

3.3.1. The Theorem of the LEAN method

Here we propose a new method called the "Level Evaluation ANalysis" (LEAN) to solve the problem of data differentiation, especially for processing the sampled-data from battery tests. The LEAN method is proficient to cope with the differentiation problem for the data that has features as shown in the left side of Figs. 2 and 3. Fig. 3 illustrates the Theorem of the LEAN method. Consider the sampled-data sequence $\boldsymbol{y} = \{y_1, y_2, y_3 \ldots y_n\}$, and $\boldsymbol{x} = \{x_1, x_2, x_3 \ldots x_n\}$ are both quasi-monotonic. For $\boldsymbol{x}$, the interval between x_i and x_{i+1} is constant for arbitrary i, and define $\delta X = x_{i+1} - x_i$. For $\boldsymbol{y}$, there is $y_1 \le y_2 \le \ldots y_k \le \ldots y_{n-1} \le y_n$. The sampling sequence $\boldsymbol{y}$ falls into a discretized set of $Y = \{\psi_1, \psi_2, \psi_3, \ldots \psi_m\}$, where the digital sampling interval between ψ_i and ψ_{i+1} is also constant for arbitrary i, and define $\delta Y = \psi_{i+1} - \psi_i$. Assume N_{ψ_i} refers to that ψ_i appears N_{ψ_i} times in $\boldsymbol{y}$, then there holds the **LEAN Theorem**:

$$\begin{cases} \left.\dfrac{\mathrm{d}x}{\mathrm{d}y}\right|_{y=\psi_i} = N_{\psi_i}\cdot\dfrac{\delta X}{\delta Y} \\ \left.\dfrac{\mathrm{d}y}{\mathrm{d}x}\right|_{y=\psi_i} = \dfrac{1}{N_{\psi_i}}\cdot\dfrac{\delta Y}{\delta X} \end{cases} \tag{17}$$

Proof of the theorem

Fig. 3 helps illustrate the proof of the LEAN Theorem. Look at a specific ψ_i, the target is to calculate the local slope of the solid red line in Fig. 3. The slope of the red line $\tan\gamma = \left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{Y=\psi_i}$, whereas the inverse of the slope $\cot\gamma = \left.\frac{\mathrm{d}x}{\mathrm{d}y}\right|_{Y=\psi_i}$. Then the proof comes out into Eqn. (18):

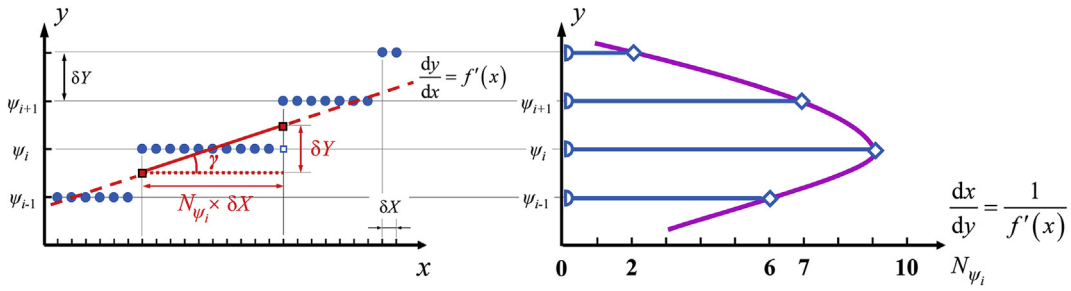

Fig. 3. The Theorem of the LEAN method.

$$\cot\gamma = \left.\frac{dx}{dy}\right|_{Y=\psi_i} = \frac{N_{\psi_i}\cdot\delta X}{\delta Y} = N_{\psi_i}\cdot\frac{\delta X}{\delta Y} \tag{18}$$

where N_{ψ_i} denotes that ψ_i appears N_{ψ_i} times in the sampled-data set $\boldsymbol{y}$. The second formula is the reciprocal of the first formula. ☒

3.3.2. The Corollary of the LEAN theorem

Corollary 1 of the theorem:

Suppose $\Pi = \left\{ \pi_i | \pi_i = \left.\frac{dx}{dy}\right|_{Y=\psi_i}, \ i = 1, 2, \ldots m \right\}$ is the result of the LEAN method, and $\boldsymbol{y} = \{y_1, y_2, y_3 \ldots y_n\}$ is quasi-monotonic, then $\boldsymbol{y}$ can be reformulated by Π. In other words, the LEAN result Π and the sampled-signal $\boldsymbol{y}$ are equivalent, as shown in Fig. 4(a).

Proof of corollary 1:

Corollary 1 of the Theorem is obviously true, because for each segment of $\boldsymbol{y}$, $\boldsymbol{y}_{\text{seg},i} = \{y_k | y_k = \psi_i\} \in \boldsymbol{y}$, its length $\boldsymbol{y}_{\text{seg},i} = N_{\psi_i}$ can be inferred by Π, and the segments are arranged in a monotonic order, both in the x and y axis. ☒

Corollary 2 of the theorem:

The LEAN Theorem still holds for data set $\boldsymbol{y}$ that is "approximately" quasi-monotonic. The word "approximately" means that some sampled data may jump to adjacent values in real sampling data set, as the red/black arrows shown in Fig. 4(b). This kind of jumping is usually regarded as the sampling noise.

Proof of corollary 2:

In real data sampling, the jump is common at the boundary where $y_k = x_i$, $y_{k+1} = x_{i+1}$, caused by noise or the sampling drift. Let $\Pr(y_{k+1}\downarrow)$ denotes the probability of one sampled-data jumping down δY, whereas $\Pr(y_k\uparrow)$ denotes the probability of one sampling data jumping up δY. Considering the independency and randomness of data sampling, we have $\Pr(y_{k+1}\downarrow) = \Pr(y_k\uparrow)$, therefore there will be no change in the overall LEAN results, as shown in Fig. 4(b). ☒

3.3.3. Using the LEAN method to calculate the incremental capacity

For the incremental capacity analysis, we are calculating:

$$\left.\frac{dQ}{dV}\right|_{V=v_i} = I\cdot\left.\frac{d\tau}{dV}\right|_{V=v_i} \tag{19a}$$

where Q is the charge/discharge capacity during the battery test, V is the voltage of the battery cell, I=*Const* is the current used during IC testing, and τ is the time. According to the LEAN Theorem, substitute y with V, and x with τ, the ICA can thus be calculated by:

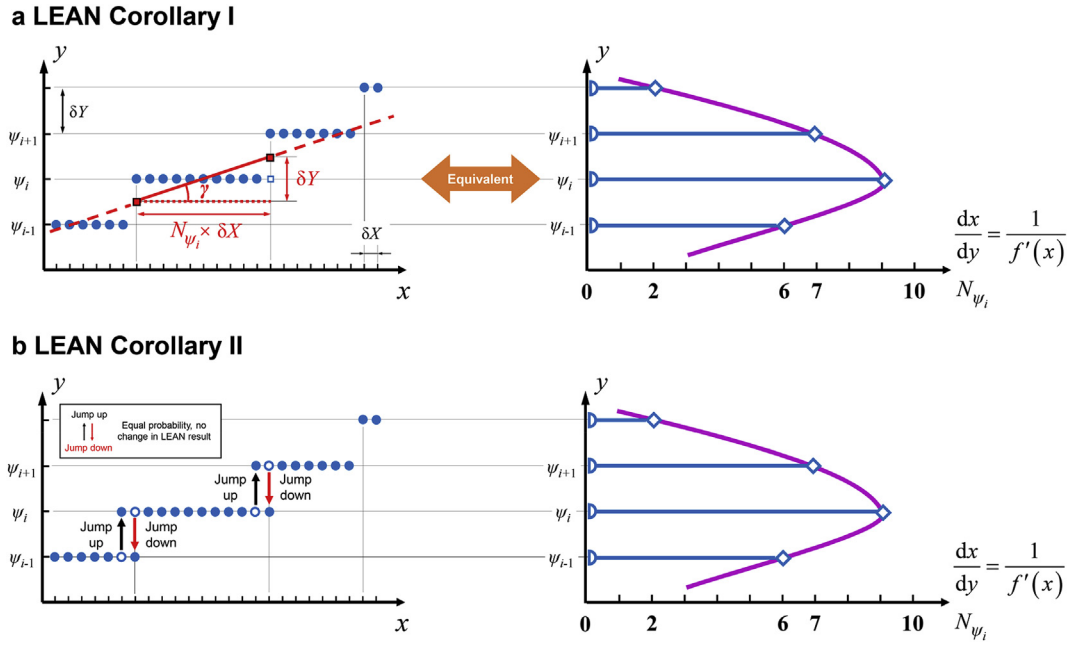

Fig. 4. The Corollaries of the LEAN method. a LEAN Corollary I, equivalency between raw data and differentiation result. **b** LEAN Corollary II, robustness to noise.

$$\left.\frac{dQ}{dV}\right|_{V=v_i} = I \cdot \left.\frac{d\tau}{dV}\right|_{V=v_i} = I \cdot \cot\gamma = I \cdot N_{v_i} \cdot \frac{\delta\tau}{\delta V} \tag{19b}$$

$$\cot\gamma = \left.\frac{d\tau}{dV}\right|_{V=v_i} = N_{v_i} \cdot \frac{\delta\tau}{\delta V} \tag{20}$$

where $\delta\tau = Const$ is the sampling interval of time, and δV is the sampling interval of voltage. N_{v_i} is the counted frequency of $V = v_i$. Fig. 5(a) illustrates the correctness of Eqn. (19) because the slope $\cot\gamma = \left.\frac{d\tau}{dV}\right|_{V=v_i}$ is in exact proportion to the counted frequency N_{v_i}, as in Eqn. (20). For the convenience of further discussion, the sampled-data set for voltage is defined as $S^V = \{V_1, V_2, V_3 \ldots V_n\}$, corresponding to $\boldsymbol{y} = \{y_1, y_2, y_3 \ldots y_n\}$ in the definitions of curve fitting.

4. Results and discussions

4.1. The equivalence of the LEAN method and ICA/DVA

The equivalence of the LEAN method and ICA/DVA has been represented mathematically in Eqn. (19). The thermodynamics of a cell can be calculated by the derivative of the state-of-charge over voltage (dz/dV), which is a normalized value of ICA, as calculated in Eqn. (21), where $Q = n \cdot I \cdot \delta\tau$, and n is the total number of samples in one test. Therefore, polynomial fitting is no longer required to conduct differential analysis of the battery voltage, and the LEAN method can be used to calculate the derivatives of voltage.

$$\text{Incremental Capacity} = \left.\frac{dz}{dV}\right|_{V=v_i} = \frac{1}{Q}\left.\frac{dQ}{dV}\right|_{V=v_i} = \frac{N_{v_i}}{n} \cdot \frac{1}{\delta V} \tag{21}$$

Fig. 5(a) demonstrates an example data: the sampling data set S^V are marked by blue dots, with a total length of $n = 25$. Fig. 5(b) is the measure voltage of full state-of-charge. The discretized data range is $V = \{v_1 = 3.359\text{ V}, v_2 = 3.360\text{ V}, v_3 = 3.361\text{ V}, v_4 = 3.362\text{ V}\}$ with $\delta V = 1$ mV, and the sampling interval is $\delta\tau = 0.1$s. The set of counting numbers is $N_V = \{N_{v_1} = 6, N_{v_2} = 10, N_{v_3} = 7, N_{v_4} = 2\}$. Therefore, dz/d$V$ can be further calculated using Eqn. (21).

Reference ICA curves are derived using published smoothing methods, SG[29] and SVR[33] (see Sec. III-Methods) to validate the outcome of the LEAN method. Fig. 5(c) compares the IC curves derived by the LEAN method, the SG method, and the SVR method, with the characteristic peaks and their heights marked with coloured triangles. The shapes of the ICA curves for the LEAN, SG and SVR methods appear similar, with similar peak locations, indicating

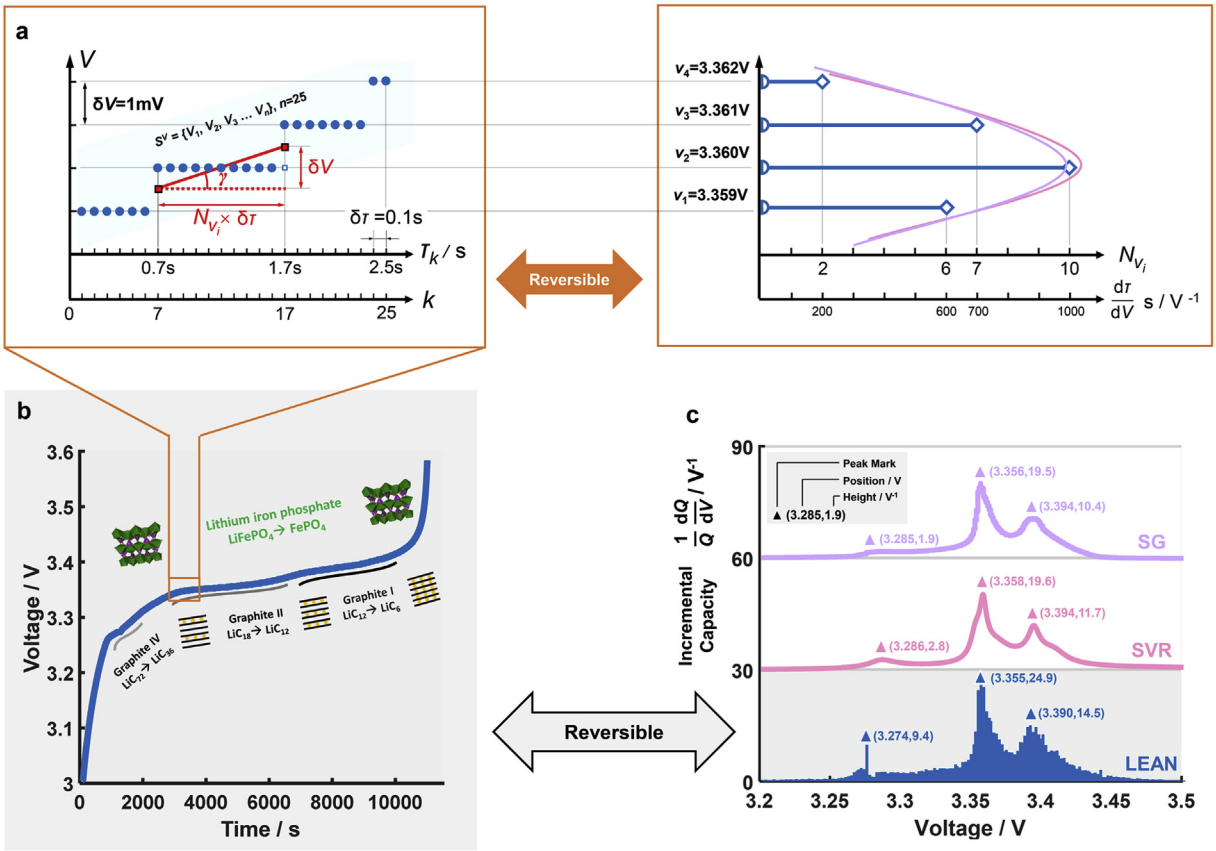

Fig. 5. Proof of the equivalence of the LEAN method and ICA. a The equivalence proof using sampling segments. τ is the time, Q is the charge/discharge capacity, and V is the cell voltage. The prefix "d" denotes differentiation, and "δ" denotes the sampling interval; here $\delta\tau = 0.1$ s and $\delta V = 1$ mV. The subscript k denotes the kth sample measured during the test, e.g., τ_k is the kth sampled time. A voltage sequence of $S^V = \{V_1, V_2, \ldots V_k, \ldots V_n\}$ is measured, $n = 25$ for Fig. 5(a). Under discrete sampling conditions, each V_k falls into the discretized voltage set $V = \{v_i \,|\, i = 1, 2, \ldots m, v_{i+1}\text{-}v_i = \delta V\}$, e.g., for the LFP sampled-data, the set V can be {2.750 V, 2.751V, 2.752 V, …3.199 V, 3.200 V, 3.201 V, …, 3.599 V, 3.600 V}; $m = 4$ and the voltage set $V = \{v_1 = 3.359\text{ V}, v_2 = 3.360\text{ V}, v_3 = 3.361\text{ V}, v_4 = 3.362\text{ V}\}$. Counting the frequency N_{v_i} such that $V_k = v_i$ according to S^V, we know that v_i appears N_{v_i} times in the sampled-data set S^V. **b** The voltage curve of a commercial lithium-ion cell with $LiFePO_4$/graphite electrodes. **c** Comparison between the results of ICA using traditional curve-fitting methods (SG and SVR) and that of the LEAN method. The characteristic peaks are marked by solid triangles, with their locations and heights given in brackets.

that the LEAN method is a reliable substitute for traditional data-processing methods. The peak heights reported by the LEAN method are always higher than those of the SG and SVR methods, because the curve-fitting method always averages adjacent data with an optimal solution, which weakens the concentration of data points at voltage plateaus.

The correctness of the LEAN method in calculating the IC curve is further validated using the battery data from THU, ICL, UoM and NREL, as shown in Fig. 6 and Fig. 7. Fig. 6 compares the ICA results derived by both the LEAN method and the reference method using the traditional curve-fitting method (SVR or SG) considering capacity degradation for Cells A-F (see Table 1). The results strongly demonstrate that the LEAN method and ICA are equivalent because the ICA results derived by the LEAN method appear very similar to those derived by the SVR and SG methods for all types of cell chemistries at different state-of-healths. The differences in the location of the ICA peaks (collected in Table 2) are usually small, confirming the equivalency of the LEAN method and ICA. The accuracy is further confirmed because the LEAN method can well capture the voltage platforms of the graphite, as shown in Fig. 7(a). And Fig. 7 further demonstrates the validity of LEAN method in analysing cells with different chemistries.

4.2. Setting benchmarks for the curve-fitting methods

Setting the LEAN method as a benchmark in the differential analysis of battery performance is here to be justified. As shown in Figs. 3 and 5(a), the LEAN method simply counts the frequency of a specific voltage; therefore, the LEAN method will never "under-fit" or "over-fit" the data. Furthermore, the raw voltage data (although certain types of noise are permanently lost, but this is not a bad thing) can be rebuilt given the LEAN result if the counting bucket δV equals the sampling resolution δR and the data set $S^V = \{V_1, V_2, V_3 \dots V_n\}$ is quasi-monotonic. In contrast, for methods using curve fitting to smooth the data first, the fitness can be controlled by tuning the cost function during optimization; therefore, a researcher may show a desired peak location and height by unfairly adjusting the parameters in the curve-fitting algorithm. This explains why different researchers obtain different results when performing differential analysis using the same data, which continuously undermines the confidence in the published ICA/DVA/DTV results.

Curve fitting is still useful, and many researchers will prefer to interpret the smooth continuous functions that it provides; however, the LEAN method provides a benchmark to judge the curve-fitting results. Under-fitting may be much easier to identify through the fitting results of the voltage curve, as shown in Fig. 8(a). However, distinguishing "over-fitting" and "good fitting" is more difficult because the fitted curves in these two cases are quite similar. According to the benchmark set by the LEAN method in the IC curve, it is easy to distinguish "over-fitting", as shown in Fig. 8(b). If over-fitting or under-fitting are identified according to the benchmark set by the LEAN method, then an analyst knows that they must use another set of parameters in their curve-fitting algorithm.

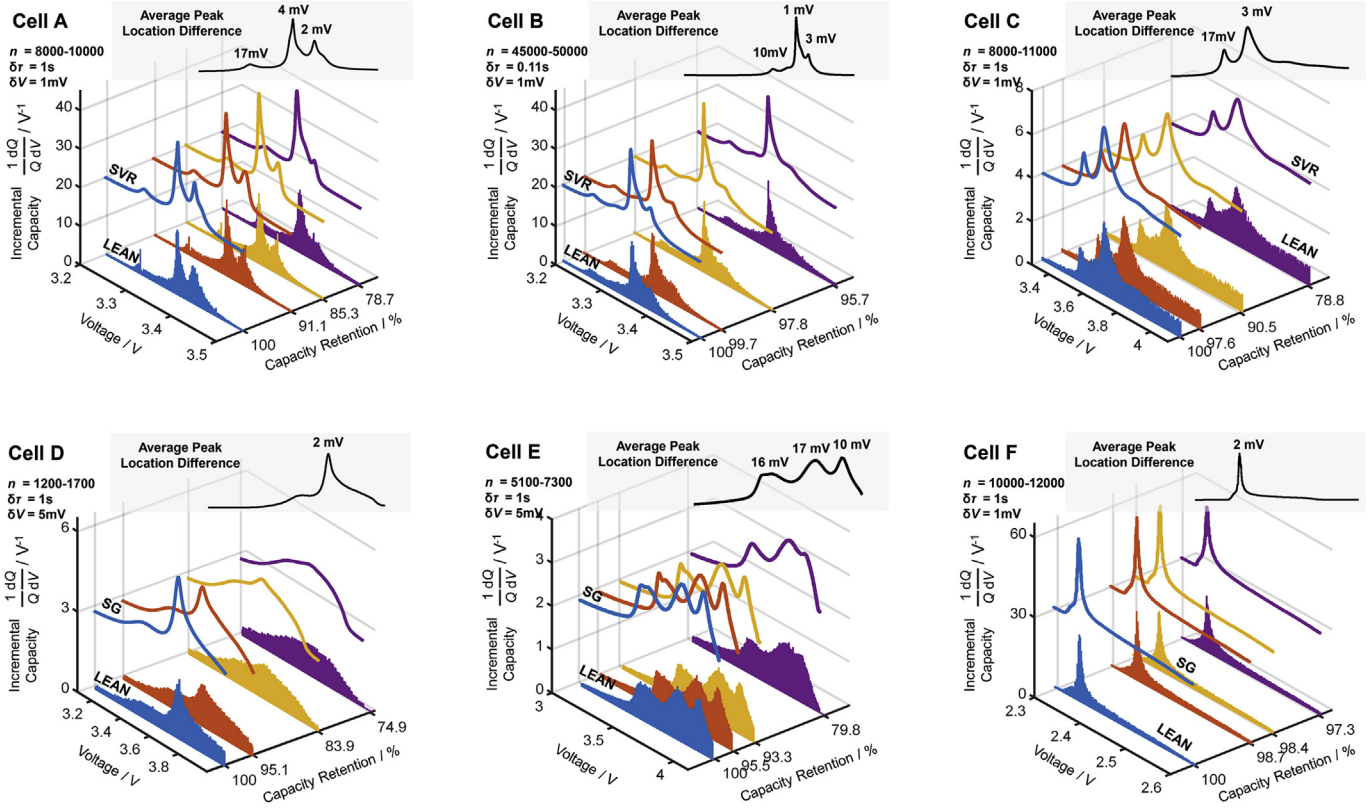

Fig. 6. Validation of the equivalence of the LEAN method and ICA. Comparison between the results of differential analysis using traditional methods and that of the LEAN method for 6 cells with different chemistries (see Table 1) from Tsinghua University, Imperial College London, and University of Michigan, Ann Arbor. Degradation data are used to confirm the validity of the LEAN method throughout the full cycle life. The parameters $\{n, \delta\tau, \delta V\}$, used to calculate the ICA curves using Eqn. (21), are noted in the upper left corner under the cell label. The average difference in the location of the characteristic peaks reported by the traditional curve-fitting method and that reported by the LEAN method is also marked.

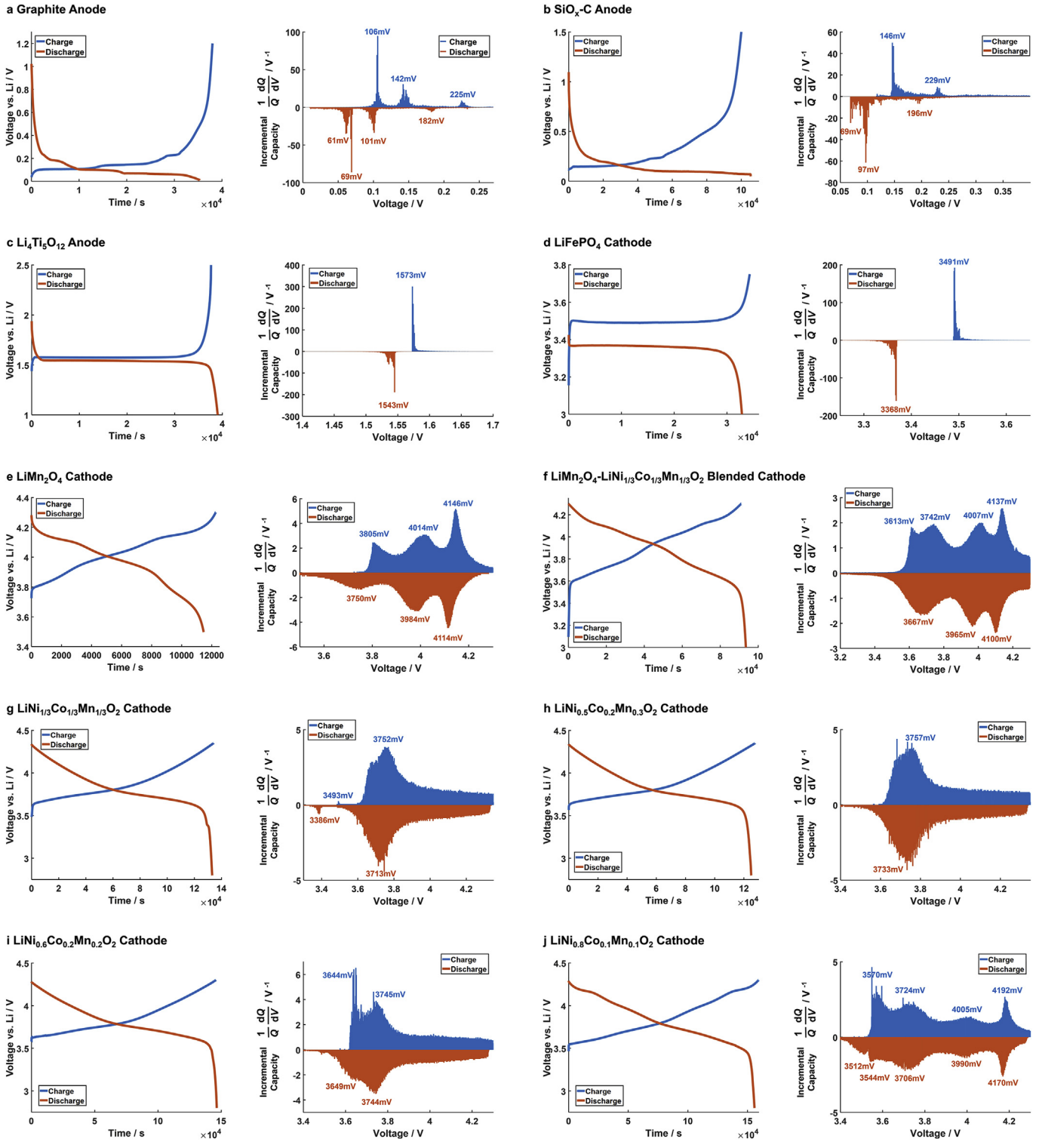

Fig. 7. Validation of the LEAN method in calculating incremental capacity curves for cells with different chemistries. a Graphite vs. Li half-cell, data from National Renewable Energy Laboratory, U.S., including the charge and discharge curve with 1/10C current, and the incremental capacity calculated by the LEAN method, the locations of the peaks conform to those in "Ohzuku, T., Iwakoshi, Y., Sawai, K. *J. Electrochem. Soc.* 140(9), 2490−2498." **b** Silicon Oxide-Graphite vs. Li half-cell, data from Tsinghua University, including charge and discharge curve with 1/30C current, and the incremental capacity calculated by the LEAN method. **c** $Li_4Ti_5O_{12}$ vs Li half-cell, with 1/10C current. **d** $LiFePO_4$ vs Li half-cell, with 1/10C current. **e** $LiMn_2O_4$ vs Li half-cell, with 1/3C current. **f** $LiMn_2O_4$−$LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ vs Li half-cell, with 1/30C current. **g** $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ vs Li half-cell, with 1/30C current. **h** $LiNi_{0.5}Co_{0.2}Mn_{0.3}O_2$ vs Li half-cell, with 1/30C current. **i** $LiNi_{0.6}Co_{0.2}Mn_{0.2}O_2$ vs Li half-cell, with 1/40C current. **j** $LiNi_{0.8}Co_{0.1}Mn_{0.1}O_2$ vs Li half-cell, with 1/40C current. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Comparison of the peak position and height calculated by the LEAN method and other methods.

Cell	Capacity Retention	LEAN Peak Locations/V	SVR Method Diff./mV	Cell	Capacity Retention	LEAN Peak Locations/V	SG Method Diff./mV
A	100%	3.274, 3.355, 3.390	+12, −4, −5	**D**	100%	3.495, 3.680	−24, −3
	91.1%	3.275, 3.358, 3.399	+16, 0, 0		95.1%	3.480, 3.650	−55, −2
	85.3%	3.275, 3.358, 3.399	+17, +2, +2		83.9%	3.615	−16
	78.7%	3.275, 3.369, 3.399	+21, −7, 0		74.9%	3.570	−4
B	100%	3.257, 3.348, 3.388	+11, 0, +2	**E**	100%	3.500, 3.815, 3.990	−4, +3, +1
	99.7%	3.257, 3.348, 3.388	+11, +1,+3		95.5%	3.475, 3.815, 3.990	+6, +2, +2
	97.8%	3.257, 3.351, 3.388	+12, 0, +5		93.3%	3.485, 3.815, 3.990	−1, −9, +6
	95.7%	3.266, 3.351, 3.400	+4, +2, 0		79.8%	3.460, 3.815, 3.93	+32, −32, −19
C	100%	3.522, 3.659	+18, −3	**F**	100%	2.352	+1
	97.6%	3.509, 3.659	+17, +1		98.7%	2.354	+3
	90.5%	3.509, 3.668	+19, +2		98.4%	2.356	+1
	78.8%	3.526, 3.677	+15, +5		97.3%	2.354	+1

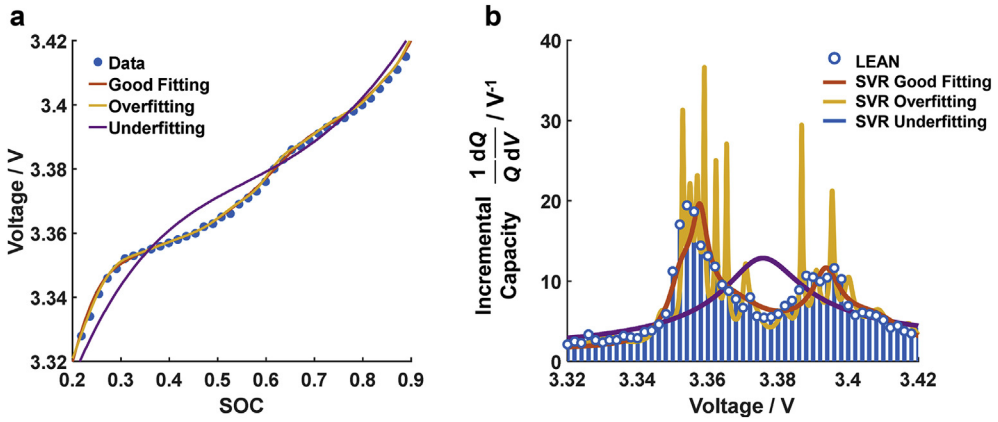

Fig. 8. LEAN method regarded as a benchmark for calculating differentiations of battery voltage curves. a The curve fitting for Cell A using both the LEAN method and the SVR method. Note that with different settings of curve-fitting parameters, the SVR may return result of good-fitting, overfitting, or under-fitting. **b** The voltage differentiation for Cell A using the LEAN method is set as a benchmark for the SVR method.

4.3. The equivalence of the LEAN method and DTV

The LEAN method can calculate the differentiations in temperature curves, which includes changing the voltage "*V*" into "*T*" to yield Eqn. (21):

$$\left.\frac{\mathrm{d}z}{\mathrm{d}T}\right|_{T=\theta_j} = \frac{1}{Q}\left.\frac{\mathrm{d}Q}{\mathrm{d}T}\right|_{T=\theta_j} = \frac{N_{\theta_j}}{n}\cdot\frac{1}{\delta T} \tag{22}$$

For the data set $S^V = \{V_1, V_2, V_3 \ldots V_n\}$, discretized sampling set $V = \{v_i | i = 1, 2, \ldots m,\ v_{i+1} - v_i = \delta V\}$, and counting set $N_V = \{N_{v_i} | i = 1, 2, \ldots m\}$ for the voltage signal, we get $S^T = \{T_1, T_2, T_3 \ldots T_n\}$, $T = \{\theta_j | j = 1, 2, \ldots l,\ \theta_{j+1} - \theta_j = \delta T\}$ and $N_T = \{N_{\theta_j} | i = 1, 2, \ldots l\}$ for the temperature signal. The LEAN method can then calculate the DTV, d*T*/d*V* as in Ref. [39]. First, the counting sets N_T and N_V are generated separately using the LEAN method. Second, N_T and N_V at a specific time τ_k are divided, and the DTV curve is derived:

$$\left.\frac{\mathrm{d}T}{\mathrm{d}V}\right|_{\tau=\tau_k} = \frac{\mathrm{d}T}{\mathrm{d}z}\cdot\left.\frac{\mathrm{d}z}{\mathrm{d}V}\right|_{\tau=\tau_k} = \frac{N_{v_i}}{N_{\theta_j}}\cdot\frac{\delta T}{\delta V},\ (V_k = v_i, T_k = \theta_j) \tag{23}$$

The DTV curves over different degradation stages derived by the LEAN method, as shown in Fig. 9, look similar to those in Ref. [39], which used a mixed curve-fitting method considering both the moving average and polynomial functions. The DTV results further validate that the LEAN method is correct.

4.4. Bucket selection for frequency counting

The selection of proper δV when applying the LEAN method is critical. The peak location and height are the two most important characteristics of the IC curve, and most research papers studying battery degradation base their conclusions on the variation of these two characteristics. Intuitively, δV not only influences the peak height according to Eqn. (21) but also determines the resolution of the peak location. δV should be a multiple of the original sampling resolution δR ($\delta V = K \cdot \delta R$, where K is an integer) to guarantee fair counting in the LEAN method, as shown in Fig. 10(a). Unfair counting ($\delta V \neq K \cdot \delta R$) leads to periodic fluctuations in the curves, whereas fair counting ($\delta V = 2 \cdot \delta R$) leads to good curves, as shown in Fig. 10(b). However, sometimes fair counting also leads to fluctuations in the curves when $\delta V = \delta R$, as shown in Fig. 10(b). A sampling error or noise in the raw data with a magnitude of σ_V causes instability in the counting process, as shown in Fig. 10(c). The instability caused by σ_V can be eliminated by expanding the sampling bucket to $\delta V = K \cdot \delta R > \sigma_V$. In summary, in the trade-off between high-resolution peak location and the elimination of fluctuations, the optimal sampling bucket should be selected as $\delta V = K \cdot \delta R > \sigma_V$, with $(K\text{-}1) \cdot \delta R \leq \sigma_V$.

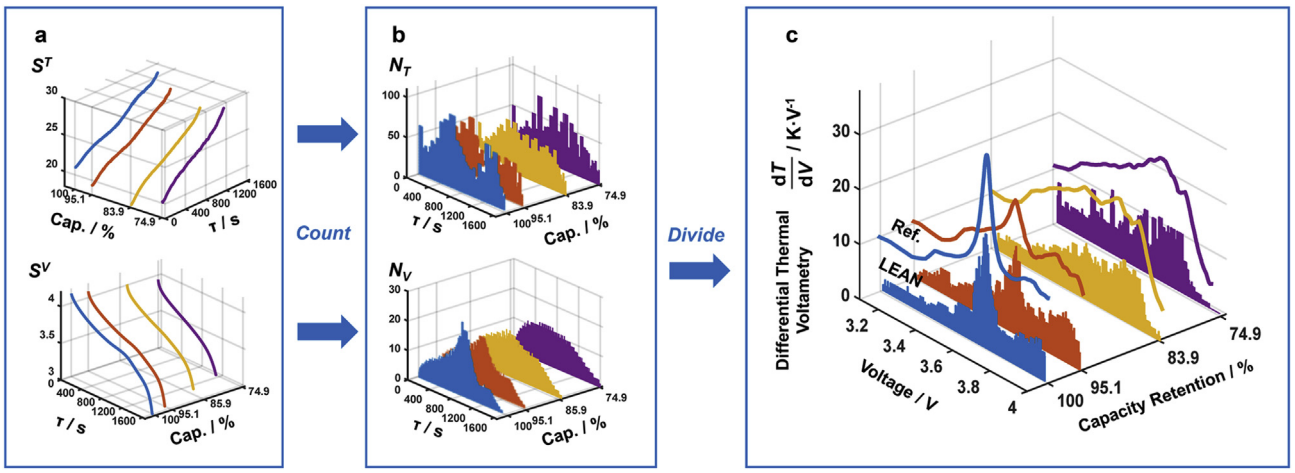

Fig. 9. Using the LEAN method to calculate differentiations in temperature and DTV. **a** The raw temperature and voltage data. **b** The counting numbers at sampled levels. **c** The calculated DTV using the LEAN method compared with that in Ref. [39].

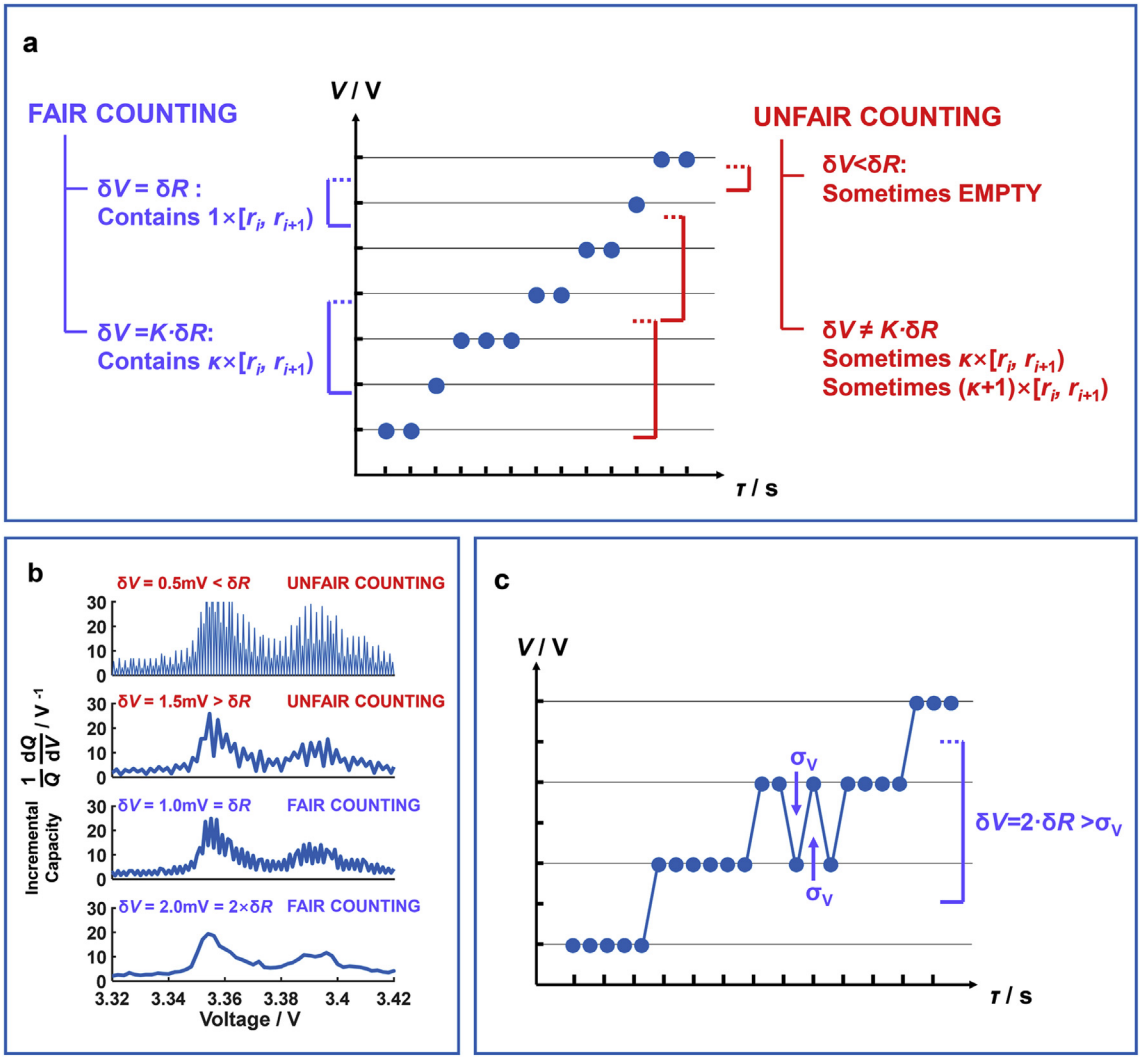

Fig. 10. The principle of selecting the proper bucket length for counting in the LEAN method. **a** Fair counting and unfair counting. Fair counting means that the counting bucket δV is K (K is an integer) times the sampling resolution δR, where each counting bucket $[v_i, v_{i+1})$ contains κ times sampling intervals $[r_i, r_{i+1})$. κ is the rounded integer of K. Unfair counting means that δV is not integer times δR. Therefore, $[v_i, v_{i+1})$ sometimes contains κ times $[r_i, r_{i+1})$ and sometimes $\kappa+1$ times, resulting in fluctuations in the LEAN results. **b** LEAN results using different counting buckets. For unfair counting when $\delta V \neq K \cdot \delta R$, large fluctuations can be seen in the LEAN results. For fair counting if δV=1 mV is not larger than $\sigma_V = 1$ mV, there are still some fluctuations. When δV=2 mV>$\sigma_V = 1$ mV, the LEAN results appear better. **c** The counting bucket can be expanded to contain fluctuations or noise in the raw data to reduce the fluctuations in the LEAN results. Here, the voltage fluctuation σ_V is 1 mV; therefore, ideally, $\delta V = 2 \cdot \delta R = 2$ mV>σ_V.

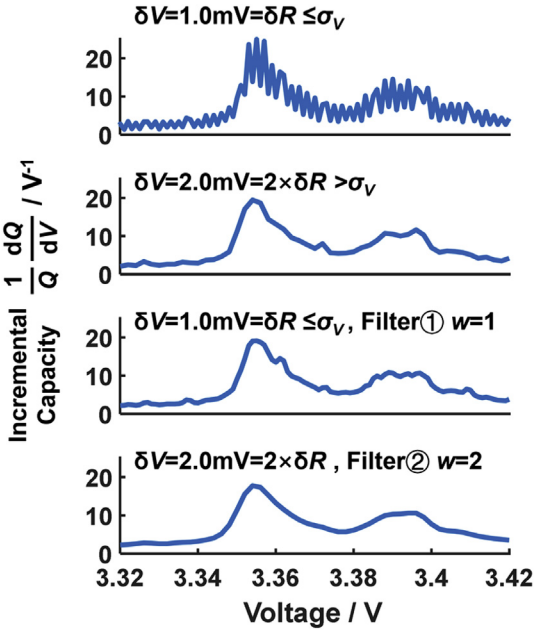

Fig. 11. Reversible smoothing after level evaluation, ensuring the universality of the LEAN method. When $\delta V = 1$ mV $\leq \sigma_V$, the LEAN results show fluctuations. The fluctuations are attenuated once we choose δV=2 mV>σ_V. Filter ① in Table 3, $\alpha = [0.25, 0.5, 0.25]^T$ with $w = 1$, can smooth the LEAN results with $\delta V = 1$ mV, whereas Filter ② in Table 3, $\alpha = [0.0668, 0.2417, 0.3830, 0.2417, 0.0668]^T$ with $w = 2$, can further smooth the LEAN results with $\delta V = 2$ mV.

4.5. Reversible smoothing ensures universality of the LEAN method

Smoothing is possible after calculating the differentiations using the LEAN method. The after-smoothing can remove fluctuations in the ICA curve for $\delta V \leq \sigma_V$, with the same effect brought by expanding the bucket width, as shown in Fig. 11. If the smoothing is reversible, the universality of the LEAN method still holds because we can still rebuild the raw data from the result. To simplify the discussion, let vector $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots \pi_m]^T$, $\pi_i = \frac{dz}{dV}\Big|_{V=v_i}$ represents the result calculated by the LEAN method, whereas vector $\boldsymbol{\pi}^* = [\pi_1^*, \pi_2^*, \ldots \pi_m^*]^T$ represents the result after smoothing. $\boldsymbol{\alpha}$=[α_{-w}, α_{-w+1}, …α_{-1}, α_0, α_1, …, α_{w-1}, α_w]^T is a filtering vector with a length of $2w+1$. The sum of the elements in $\boldsymbol{\alpha}$ should be 1 ($\sum_{j=-w}^{j=w} \alpha_j = 1$) to avoid deformation of the original result. Take $\boldsymbol{\pi}_k^{\boldsymbol{w}} = [\pi_{k-w} \ldots \pi_{k-1},\ \pi_k, \pi_{k+1} \ldots \pi_{k+w}]^T$ as a piecewise vector extracted from $\boldsymbol{\pi}$ centred at π_k, then:

$$\pi_k^* = \boldsymbol{\alpha}^T \cdot \boldsymbol{\pi}_k^w \tag{24}$$

Define a filtering matrix A_m that is expanded by the filtering vector $\boldsymbol{\alpha}$ such that:

$$\boldsymbol{\pi}^* = A_m \cdot \boldsymbol{\pi} \tag{25}$$

The filtering matrix $A_m \in R^{m \times m}$ is presented in Eqn. (20), where m is the total number of the sampled discrete data V, $V = \{v_i | i = 1, 2, \ldots m, v_{i+1} - v_i = \delta V\}$. If A_m is reversible, we can infer $\boldsymbol{\pi}$ from $\boldsymbol{\pi}^*$ without information loss. Therefore, different researchers in different labs will have identical IC results derived from the raw data using the LEAN method if they share the same A_m.

$$A_m = \begin{bmatrix}
\alpha_0 & \alpha_1 & \cdots & \alpha_{-w+2} & \alpha_{-w+1} & \alpha_{-w} & 0 & \cdots & 0 & 0 & 0 \\
\alpha_{-1} & \alpha_0 & \cdots & \alpha_{-w+3} & \alpha_{-w+2} & \alpha_{-w+1} & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
\alpha_{-w+1} & \alpha_{-w+2} & \cdots & \alpha_0 & \alpha_{-1} & \alpha_{-2} & 0 & \cdots & 0 & 0 & 0 \\
\alpha_{-w} & \alpha_{-w+1} & \cdots & \alpha_1 & \alpha_0 & \alpha_{-1} & 0 & \cdots & 0 & 0 & 0 \\
0 & \alpha_{-w} & \cdots & \alpha_2 & \alpha_1 & \alpha_0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & \cdots & \alpha_3 & \alpha_2 & \alpha_1 & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & \alpha_{w-1} & \alpha_{w-2} & \alpha_{w-3} & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & \cdots & \alpha_w & \alpha_{w-1} & \alpha_{w-2} & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & \cdots & 0 & \alpha_w & \alpha_{w-1} & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \alpha_{w-1} & \cdots & \alpha_2 & \alpha_1 & \alpha_0 \\
0 & 0 & \cdots & 0 & 0 & 0 & \alpha_w & \cdots & \alpha_3 & \alpha_2 & \alpha_1 \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & \alpha_4 & \alpha_3 & \alpha_2 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & \alpha_{w-1} & \alpha_{w-2} & \alpha_{w-3} \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & \alpha_w & \alpha_{w-1} & \alpha_{w-2} \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & \alpha_w & \alpha_{w-1} \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0
\end{bmatrix}_{m} \tag{26}$$

Table 3
Recommended reversible smoothing filter α for the LEAN method.

No.	α	w	Purpose
①	$[0.25, 0.50, 0.25]^T$	1	Average of the adjacent value
②	$[0.0668, 0.2417, 0.3830, 0.2417, 0.0668]^T$	2	For noise ε~$N(0,\sigma)$, $\sigma = \delta V$
③	$[0.1059, 0.121, 0.1745, 0.1972, 0.1745, 0.121, 0.1059]^T$	3	For noise ε~$N(0,\sigma)$, $\sigma = 2\cdot\delta V$
④	$[(2w+1)^{-1}, \ldots, (2w+1)^{-1}, \ldots, (2w+1)^{-1}]^T$	Arbitrary value	Moving average

We can provide some simple $\boldsymbol{\alpha}$ that can guarantee that A_m is reversible, as listed in Table 3. The construction of $\boldsymbol{\alpha}$ may consider the probability of noisy fluctuations, such as the standardized normal distribution $N(0,\sigma)$. The filters have good effects on smoothing the ICA curves, as shown in Fig. 11. Researchers can also deduct their customized filter $\{\boldsymbol{\alpha}, A_m\}$ but must remember to state $\{\boldsymbol{\alpha}, A_m\}$ when sharing their results.

4.6. The computational complexity $\Theta(n)$ for online implementation

The computational complexity of the LEAN method is $\Theta(n)$, making it promising for processing real-time data online. The pseudocode for the LEAN method is given as in Table 4. The total cost during one LEAN calculation is expected to be $\Gamma(n) = n+2(m-1)+O(1) = O(n)$, given that the number of the sampling bucket m is far less than the total length n in the battery degradation data. The low computational load ($\Gamma(n) = \Theta(n)$) indicates that LEAN is a promising approach for online state-of-health evaluation.

Table 5 compares the computational complexity of the LEAN method with common smoothing algorithms (SVR and SG), considering their accuracies. The first row collects the critical parameters in the algorithms that can influence the computational complexity $\Gamma(n)$. The SVR method always has $\Gamma(n) = \Theta(n^2)$, indicating a long-time computation when processing big data with large n. Therefore, in current applications[33], the trick is to resample and downsize the data from n to n/D. However, resampling the data leads to information loss. Although both the $\Gamma(n)$ of the SG method and that of the LEAN method are $\Theta(n)$, the $\Gamma(n)$ for the SG method is a multivariate function that can be expressed by $\Theta(w)$ or $\Theta(p^2)$, where p is the highest order for the polynomial used in fitting, and $p \leq 3$ is commonly used. However, w for the SG method should be large for the least squares algorithm to find a meaningful optimum. In Ref. [31], w is set as 475, and in this paper, w is always larger than 100 to obtain good smoothing. Hence, the computational load for the LEAN method is commonly less than 1/100 compared with that for the SG method. Fig. 12 compares the computation time for the LEAN method, the SG method and the SVR method, using a personal computer. The computational time of the LEAN method is obviously lower than that of the SG method and the SVR method, especially for big data with large length n, indicating that the LEAN method is promising to process large scale data.

As mentioned, the accuracy of the location and height of the peaks in ICA/DVA/DTV are critical for interpreting the battery degradation data. However, previous methods always risk over-fitting or under-fitting by improper parameter settings, and there is no benchmark to evaluate the accuracy of the ICA/DVA/DTV peak. In this paper, the benchmark is set by the LEAN method, which retains the full information of the original data. Nevertheless, the LEAN method may require curve smoothing after counting. The $\Gamma(n)$ will be influenced by both $\Theta(n)$ and $\Theta(w)$. Fortunately, Fig. 11 shows that w can be very small (generally $w \leq 3$) because most of the filtering load has already been done by the LEAN counting.

Table 4
Pseudocode for the LEAN method.

Step	Pseudocode
(1)	**Initialization:** The counting bucket $B_1 = [v_1, v_2)$, such that $V_1 \in B_1$, set the index of bucket $i = 1$
(2)	**for: *k* from 1 to *n*** (Go through each sampled data in S^V)
(3)	**while** $V_k \notin B_i = [v_i, v_{i+1})$ (find the proper bucket)
(4)	**If** $V_k \in B_i$, counter $N_{v_i} = N_{v_i} + 1$, break the while loop and jump to (2). **Otherwise** go to (5).
(5)	**If** $V_k < v_i$, set $i = i$-1; **Otherwise** $V_k \geq v_{i+1}$, set $i = i$+1. **Until** $V_k \in B_i$
(6)	**while end**
(7)	**for end**

4.7. Applications of the LEAN method

Based on the above discussions, the invention of LEAN method will significantly help the utilization of differential analysis (ICA, DVA, DTV) in the field of battery aging diagnosis. The application of differential analysis for battery aging diagnosis can be tracked back to Refs. [36,42], which are both classical in aging mechanism identification for lithium-ion batteries. The invention of LEAN method guarantees that different researchers at different laboratories with a same set of battery aging data will get same result of differential analysis, significantly improving the universality of ICA/DVA/DTV in field applications. In this paper, the central message is that how the LEAN method help processing the discrete sampled-data to generate ICA/DVA/DTV curves. We do not go into the details on how to use differential analysis to diagnose the battery

Table 5
Comparison of the computational complexity and accuracy of the LEAN method and other methods.

Differentiation method	SG in Ref. [29]	SVR in Ref. [33]	LEAN in this paper
Critical parameters	Window length: w Polynomial order: p	Resampling divider: D	Bucket broad: δV Window length: w (for smoothing)
Computational complexity $\Gamma(n)$	$\Gamma(n) = 2wp^2n+O(n) = \Theta(n)$, w can be very large, $w = 475$ in Ref. [31]	$\Gamma(n) = \frac{n^2}{D^2} + O(n^2) = \Theta(n^2)$ given the code in Ref. [35]	$\Gamma(n) = n+O(1) = \Theta(n)$
Accuracy	Possible over-fitting or under-fitting	Possible over-fitting or under-fitting; information loss when downsize the sample set by dividing by D	No over-fitting or under-fitting; depends on δV
Smoothing after calculation	No need	No need	$\Gamma(n) = 2wn+O(1) = \Theta(n)$, $w \leq 3$ commonly
Same data, same differentiation result	No, every lab varies	No, every lab varies	Yes, given δV and α

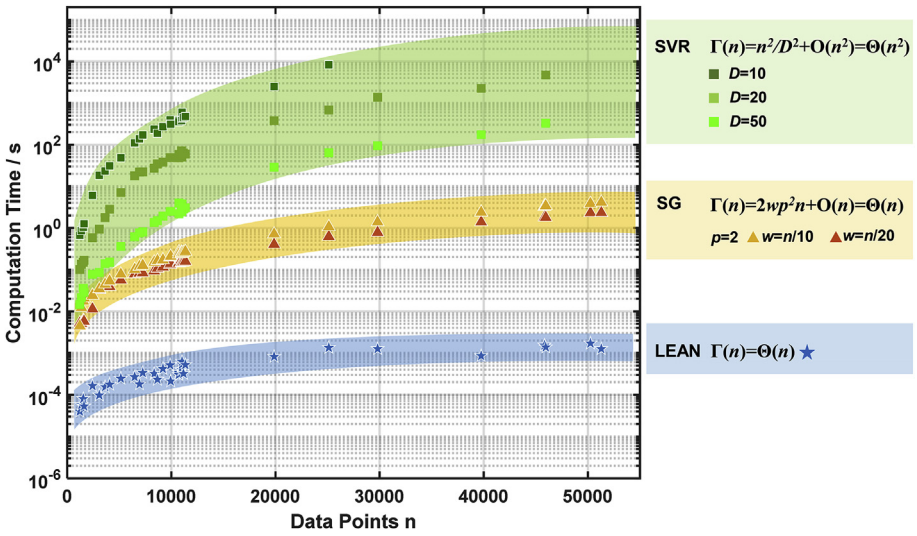

Fig. 12. Comparison of the computation time for different curve fitting approaches. The computation is conducted by a laptop with Intel® Core™ i7-6820HK CPU @2.70 GHz, with 16 GB RAM. The code is run in MATLAB R2018a®, using m code attached in the Appendix.

aging behaviours. Interested readers can find more recent references that are studying the battery aging mechanisms using ICA/DVA. There are emerging researches [35,43,44] discussing the applications of ICA/DVA on associations of cells (or we call "battery pack"). The LEAN method sets a benchmark for data processing, therefore the application in both cell and pack level will be convincible to reflect the aging characteristics reported by the sampled-data. The classical ICA/DVA is using low current (C/25, C/20, etc.) to guarantee thermodynamic equilibrium in generating battery voltage curves [18,42]. Of course, the LEAN method is competent to process the differentiation of voltage under low current charge/discharge. We can also see that the LEAN method can process voltage data that is generated by relatively high current charge/discharge, as listed in Table 1. Large currents, although undermines the assumption of "equilibrium" thermodynamic state of battery cell, may provide "quasi-equilibrium" state for aging characterization. Successful practices can be seen in Refs. [27,31,33,36,45], in which the applied currents range from C/5 to 2C. This paper already demonstrates that the sampled-data in those papers can use the LEAN method to do differential analysis. Moreover, partial charging/discharging is quite common in the field application of electrochemical energy storage systems, especially for electric vehicles. Correlated demonstrations of the usage of ICA/DVA in online battery diagnosis using partial charging curves can be seen in Refs. [46,47]. In summary, although this paper did not go into the detailed applications of the differential analysis (ICA/DVA/DTV) in field applications, the properties of the LEAN method are promising for further vehicular application in both the big data analysis on the cloud and the online battery monitoring, supporting the intelligent management of all kinds of electrochemical energy storage systems for electric vehicles, portable electronics, and power stations.

We should also remind that the LEAN method can process the differentiation of discrete sampled-data not just for battery tests, for voltage, or for temperature curves. According to the universality of the Theorem of the LEAN method, it can do the differentiation work for all kinds of discrete sampled-data. As we only have data of battery aging, we add "for battery diagnosis" in the title to be strict with the content of this paper. The deeper meaning of the LEAN method for differentiating discrete sampled-data might be still waiting for mathematicians to discover in the future.

5. Conclusion

A universal approach named the "Level Evaluation ANalysis" (LEAN) method is proposed for calculating differentiations in data for battery aging analysis. The LEAN method has been shown to work effectively for the derivation of different types of differential curves (ICA/DVA/DTV), with minimal loss of information and no risk of creating information or over-fitting/under-fitting. The LEAN method will allow results to be standardized and compared because every lab will obtain identical result using the LEAN method when analysing the same data. Therefore, the LEAN method is proposed as a benchmark for all types of curve-fitting methods that are used to derive ICA/DVA/DTV curves. Furthermore, the computational complexity of the LEAN method is only $\Gamma(n) = \Theta(n)$, making it promising for online diagnosis algorithms in the near future. The conclusions above are based on strict mathematical deductions. Moreover, every conclusion is supported by experimental validation, the data of which come from laboratories in China, the U.K., and the U.S., representing a snapshot of data from around the world. Future work will focus on applying the LEAN method to specific cases, such as the vehicular applications in both the big data analysis on the cloud and the online battery monitoring, the intelligent management of all kinds of electrochemical energy storage systems for electric vehicles, portable electronics, and power stations.

The good properties of the LEAN method in processing the differentiation of discrete sampled-data are also interesting in the view of math. As we have not seen that relative literature in math is talking about similar issues, we sincerely welcome mathematicians and scientists in signal processing to collaborate to explore the truth of the LEAN method.

Additional information

Correspondence and requests for materials/data should be addressed to X.F.

Declaration of competing interest

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

CRediT authorship contribution statement

Xuning Feng: Conceptualization, Methodology, Data curation, Writing - original draft, Software. **Yu Merla:** Writing - review & editing. **Caihao Weng:** Writing - review & editing. **Minggao Ouyang:** Supervision, Conceptualization, Resources. **Xiangming He:** Writing - review & editing. **Bor Yann Liaw:** Writing - original draft, Formal analysis. **Shriram Santhanagopalan:** Formal analysis. **Xuemin Li:** Formal analysis. **Ping Liu:** Writing - review & editing, Formal analysis. **Languang Lu:** Writing - review & editing. **Xuebing Han:** Writing - review & editing. **Dongsheng Ren:** Writing - original draft. **Yu Wang:** Formal analysis. **Ruihe Li:** Formal analysis. **Changyong Jin:** Formal analysis. **Peng Huang:** Formal analysis. **Mengchao Yi:** Formal analysis. **Li Wang:** Writing - review & editing. **Yan Zhao:** Writing - review & editing. **Yatish Patel:** Validation, Resources, Data curation. **Gregory Offer:** Writing - original draft.

Acknowledgements

This work is supported by the Ministry of Science and Technology of China under Grant No. 2016YFE0102200, the National Natural Science Foundation of China of China under Grant No. 51706117 and U1564205, and the China Postdoctoral Science Foundation funded project 2019T120087 and 2017M610086. The first author recognizes the support of the “Young Elite Scientist Sponsorship Program” from the China Association for Science and Technology under Grant No. 2018QNRC001.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.etran.2020.100051.

References

[1] Van Noorden R. The rechargeable revolution: a better battery. Nat News 2014;507(7490):26.
[2] Tomaszewska A, Chu Z, Feng X, O'Kane S, Liu X, Chen J, Ji C, Endler E, Li R, Liu L, Li Y, Zheng S, Vetterlein S, Gao M, Du J, Parkes M, Ouyang M, Marinescu M, Offer G, Wu B. Lithium-ion battery fast charging: a review. eTransportation 2019;1:100011.
[3] Zhang ZJ, Fang W, Ma R. Brief review of batteries for XEV applications. eTransportation 2019;2:100032.
[4] Jiang L, Cheng XB, Peng HJ, Huang JQ, Zhang Q. Carbon materials for rraffic power battery. eTransportation 2019:100033.
[5] Zhang Y, Wang H, Li W, Li C. Quantitative identification of emissions from abused prismatic Ni-rich lithium-ion batteries. eTransportation 2019;2:100031.
[6] Stephan A, Battke B, Beuse MD, Clausdeinken JH, Schmidt TS. Limiting the public cost of stationary battery deployment by combining applications. Nat Energy 2016;1(7):16079.
[7] 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.
[8] Ren D, Hsu H, Li R, Feng X, Guo D, Han X, Lu L, He X, Gao S, Hou J, Li Y, Wang Y, Ouyang M. A comparative investigation of aging effects on thermal runaway behavior of lithium-ion batteries. eTransportation 2019;2:100034.
[9] Tian J, Xiong R, Shen W. A review on state of health estimation for lithium ion batteries in photovoltaic systems. eTransportation 2019:100028.
[10] Rodríguez A, Plett GL, Trimboli MS. Comparing four model-order reduction techniques, applied to lithium-ion battery-cell internal electrochemical transfer functions. eTransportation 2019;1:100009.
[11] Zhang YZ, Xiong R, He HW, Qu X, Pecht M. Aging characteristics-based health diagnosis and remaining useful life prognostics for lithium-ion batteries. eTransportation 2019;1:100004.
[12] Waldmann T, Iturrondobeitia A, Kasper M, Ghanbari N, Aguesse F, Bekaert E, Daniel L, Genies S, Gordon IJ, Löble MW, Vito ED, Wohlfahrt-Mehrens M. Post-mortem analysis of aged lithium-ion batteries: disassembly methodology and physico-chemical analysis techniques. J Electrochem Soc 2016;163(10):A2149–64.
[13] Dubarry M, Truchot C, Liaw BY. Synthesize battery degradation modes via a diagnostic and prognostic model. J Power Sources 2012;219:204–16.
[14] Balewski L, Brenet JP. A new method for the study of the electrochemical reactivity of manganese dioxide. Electr Technol 1967;5:527–31.
[15] Dahn JR, McKinnon WR. Thermodynamics of lithium intercalation from high resolution electrochemical measurements. J Electrochem Soc 1984;131:1823–8.
[16] Tillement O, Quarton M. Theoretical study of ordering effects during electrochemical insertion. J Electrochem Soc 1993;140(7):1870–6.
[17] Dubarry M, Svoboda V, Hwu R, Liaw BY. Incremental capacity analysis and close-to-equilibrium OCV measurements to quantify capacity fade in commercial rechargeable lithium batteries. Electrochem Solid-State Lett 2006;9(10):A454–7.
[18] Dubarry M, Svoboda V, Hwu R, Liaw BY. Capacity and power fading mechanism identification from a commercial cell evaluation. J Power Sources 2007;165(2):566–72.
[19] Feng X, Li J, Ouyang M, Lu L, Li J, He X. Using probability density function to evaluate the state of health of lithium-ion batteries. J Power Sources 2013;232:209–18.
[20] Hunt IA, Zhao Y, Patel Y, Offer GJ. Surface cooling causes accelerated degradation compared to tab cooling for lithium-ion pouch cells. J Electrochem Soc 2016;163(9):A1846–52.
[21] Torai S, Nakagomi M, Yoshitake S, Yamaguchi S, Oyama N. State-of-health estimation of $LiFePO_4$/graphite batteries based on a model using differential capacity. J Power Sources 2016;306:62–9.
[22] Ouyang M, Feng X, Han X, Lu L, Li Z, He X. A dynamic capacity degradation model and its applications considering varying load for a large format Li-ion battery. Appl Energy 2016;165:48–59.
[23] Bloom I, Jansen AN, Abraham DP, Knuth J, Jones SA, Battaglia VS, Henriksen GL. Differential voltage analyses of high-power, lithium-ion cells: 1. Technique and application. J Power Sources 2005;139(1–2):295–303.
[24] Honkura K, Takahashi K, Horiba T. Capacity-fading prediction of lithium-ion batteries based on discharge curves analysis. J Power Sources 2011;196(23):10141–7.
[25] Schindler S, Danzer MA. A novel mechanistic modeling framework for analysis of electrode balancing and degradation modes in commercial lithium-ion cells. J Power Sources 2017;343:226–36.
[26] Ecker M, Nieto N, Käbitz S, Schmalstieg J, Blanke H, Warnecke A, Sauer DU. Calendar and cycle life study of $Li(NiMnCo)O_2$-based 18650 lithium-ion batteries. J Power Sources 2014;248:839–51.
[27] Wu B, Yufit V, Merla Y, Martinez-Botas RF, Barndon NP, Offer GJ. Differential thermal voltammetry for tracking of degradation in lithium-ion batteries. J Power Sources 2015;273:495–501.
[28] Christophersen JP, Shaw SR. Using radial basis functions to approximate battery differential capacity and differential voltage. J Power Sources 2010;195(4):1225–34.
[29] Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 1964;36(8):1627–39.
[30] Craven P, Wahba G. Smoothing noisy data with spline functions. Numer Math 1978;31(4):377–403.
[31] Samad NA, Kim Y, Siegel JB, Stefanopoulou AG. Battery capacity fading estimation using a force-based incremental capacity analysis. J Electrochem Soc 2016;163(8):A1584–94.
[32] Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.
[33] Weng C, Cui Y, Sun J, Peng H. 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.
[34] Wang D, Miao Q, Pecht M. Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model. J Power Sources 2013;239:253–64.
[35] Weng C, Feng X, Sun J, Peng H. State-of-health monitoring of lithium-ion battery modules and packs via incremental capacity peak tracking. Appl Energy 2016;180:360–8.
[36] 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.
[37] Han X, Ouyang M, Lu L, Li J. Cycle life of commercial lithium-ion batteries with lithium titanium oxide anodes in electric vehicles. Energies 2014;7(8):4895–909.
[38] Ren D, Lu L, Ouyang M, Feng X, Li J, Han X. Degradation identification of individual components in the $Li_yNi_{1/3}Co_{1/3}Mn_{1/3}O_2$-$Li_yMn_2O_4$ blended cathode for large format lithium ion battery. Energy Procedia 2017;105:2698–704.
[39] Merla Y, Wu B, Yufit V, Brandon NP, Martinez-Botas RF, Offer GJ. Novel application of differential thermal voltammetry as an in-depth state-of-health diagnosis method for lithium-ion batteries. J Power Sources 2016;307:308–19.
[40] Weng C, Sun J, Peng H. Model parametrization and adaptation based on the invariance of support vectors with applications to battery state-of-health monitoring. IEEE Trans Veh Technol 2014;64(9):3908–17.
[41] Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V. Support vector regression machines. Advances in neural information processing systems. MIT Press; 1997. p. 155–61.
[42] Dubarry M, Liaw BY. Identify capacity fading mechanism in a commercial $LiFePO_4$ cell. J Power Sources 2009;194(1):541–9.

[43] Tanim TR, Shirk MG, Bewley RL, Dufek EJ, Liaw BY. Fast charge implications: pack and cell analysis and comparison. J Power Sources 2018;381:56–65.
[44] Tanim TR, Dufek EJ, Walker LK, Ho CD, Hendricks CE, Christophersen JP. Advanced diagnostics to evaluate heterogeneity in lithium-ion battery modules. eTransportation 2020:100045.
[45] Feng X, Sun J, Ouyang M, He X, Lu L, Han X, Fang M, Peng H. Characterization of large format lithium ion battery exposed to extremely high temperature. J Power Sources 2014;272:457–67.
[46] Cai Y, Lu L, Shen P, Feng X, Wang H, Ouyang M. Online weld breakage diagnosis for the battery of electric vehicle: a data-driven approach. In: 2016 IEEE vehicle power and propulsion conference (VPPC). IEEE; 2016. p. 1–5.
[47] Feng X, Weng C, He X, Han X, Lu L, Ren D, Ouyang M. Online state-of-health estimation for Li-ion battery using partial charging segment based on support vector machine. IEEE Trans Veh Technol 2019;68(9):8583–92.