nature portfolio

https://doi.org/10.1038/s42256-021-00312-3

Supplementary information

Machine learning pipeline for battery state-of-health estimation

In the format provided by the authors and unedited

Supplementary material: Machine learning pipeline for battery state of health estimation

Darius Roman*, Saurabh Saxena, Valentin Robu, Michael Pecht, David Flynn
*e-mail: dvr1@hw.ac.uk

Supplementary material

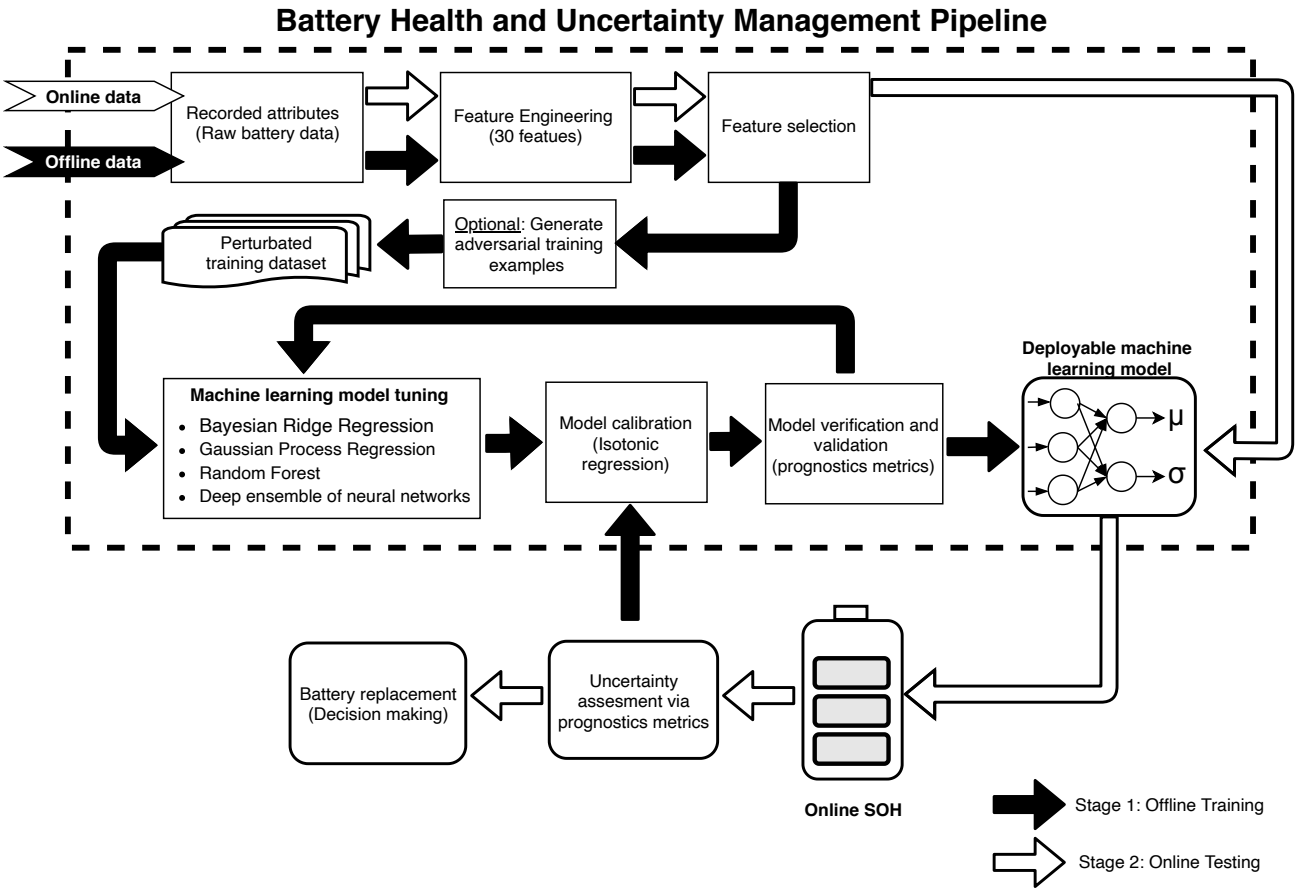

Figure 1: BHUMP flowchart

Supplementary Note 1. Domain explanation of features

Features are generated by mathematical manipulation, involving pattern recognition and information theory principles, of voltage and current charge curves. Any charging protocol finishes with both electrodes materials at their most extreme potential (and most reactive states) [1], namely the highest for the positive electrode and the lowest for the negative electrode. The diffusion of lithium ions inside an electrode is a complex process involving both microscopic and macroscopic processes that can potentially be partially captured by charge curves. During charging two crucial processes occur at the anode side (graphite-based batteries considered here), namely the intercalation of lithium ions into the active material and lithium plating [2], [3], [4]. Due to intercalation kinetics at the anode, cathode deintercalates faster than the anode can intercalate, and thus during charging, the current is the main limiting factor in a graphite-based lithium-ion battery. [2], [5], [4] Consequently, any charging protocol suffers from such limitations. The charging protocols typically go through a constant-current (CC) mode, followed by a constant-voltage (CV) mode, see Supplementary Figure 3 for a typical CC-CV charge protocol.

Zhang et al. [6] investigated the effects of charging protocols in $LiCoO_2$ based batteries by creating a bespoke three-electrode cell. The authors emphasise that lithium-ion plating coexists with the intercalation process in the anode and it occurs in the late period of the CC despite the graphite not being fully lithiated. Similarly, Zhou et al. [7] also mention that kinetically, under high-current charging conditions, the negative electrode can be polarised to such an extent that it's potential drops below 0 V, facilitating lithium metal deposition onto the surface of the electrode particles. It is known that the duration of the CC captures the polarisation phenomenon.[7] Therefore, as the battery ages, the constant current charge time (CCCT) decreases. Upon the start of the CV charging, as the current decreases, the negative electrode slowly recovers to a nominal potential value. The CV mode duration is thus crucial to eliminate the polarisation effect caused during the CC mode allowing for the anode to recover and thus fully charge the battery. With aging, the constant voltage charge time (CVCT) increases as demonstrated in [6] and [8].

A feasibility study of CCCT and constant voltage charge time (CVCT) as proxies for battery state of health was carried out in [8]. CVCT has already been considered as input to SOH methods in the additional studies [9], [10]. To reduce diagnostics time, we only use sections of the charge curves as input to the algorithm. The availability of the entire charge curve in real-life applications is limited. Hence it is advantageous to design features that could be extracted from segments of such curves. The benefits of the approach are a lower diagnostics time (as little as 15 min) and the possibility of battery SOH estimation even in partial discharge conditions.

During discharge, the process of lithium extraction/insertion happens in reverse from anode to cathode. Since discharge currents vary with usage, we only extract one feature from the discharge curve, namely the pseudo linear resistance as introduced by Saxena et al. [11]. This is due to the instant drop in voltage associated with internal battery impedance on the application of load current. We estimate this resistance as the ratio of the observed voltage drop and the applied load current. It is understood that as the battery degrades the internal resistance of the battery increases, and hence an estimate of this internal resistance can be used as a proxy for battery SOH.[11] We used a lagged version of this feature, i.e. pseudo linear resistance from the previous cycle to estimate the SOH at the end of a charge cycle. For a mathematical explanation of all engineered features in Supplementary Material Table 2 refer to Supplementary Note 3 Feature engineering.

Supplementary Note 2. Voltage threshold values

We first define V_h to be equal to charge cut-off voltage, $V_{cut-off}$, while V_l is defined using the below formula:

$$V_l = V_h - \Delta V \tag{1}$$

where ΔV is a predefined voltage range. The recorded curve between V_l and V_h with each charge as illustrated in Figures 2a, 2a is then normalised on the interval $[0, 1]$ by subtracting the minimum value and dividing by the resulted maximum value. Following the normalisation procedure, we proceed on mathematically deriving the features. This allows for training different batteries types and designs on the same training dataset provided they underwent the same charging protocol. To overcome issues resulting from battery terminal voltage increase after previous discharge cycle and to capture the late period of the CC charging phase (when lithium plating occurs) we make use of a ΔV equal to 0.3V. A high V_l value accommodates for the increase in battery terminal voltage upon removal of load current after each discharge cycle. A behaviour commonly observed with battery ageing as referenced in Supplementary Figure 9. Furthermore, a high V_l threshold reduces the time necessary to record the CC charge curve while accommodating for partial discharge of the battery. Note, ΔV value and corresponding V_l and V_h threshold values could be adjusted based on battery type, application and user behaviour, end of life threshold, data storage capacity and processing power.

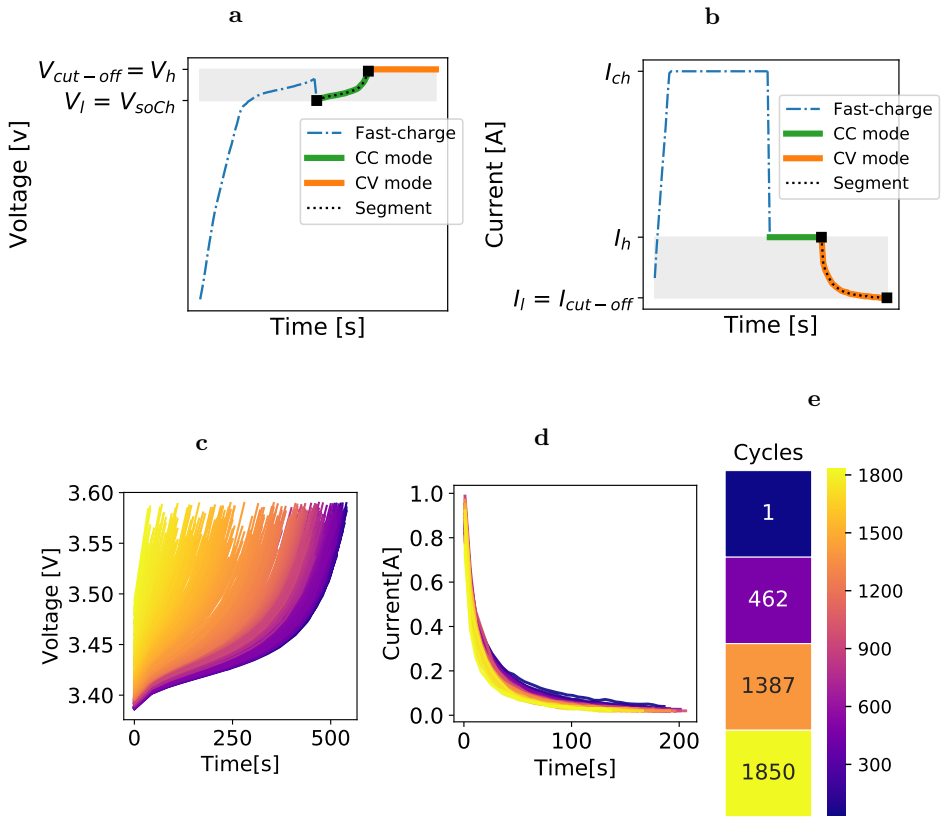

Figure 2: The 2 step fast-charge protocol and extracted ageing segment of the curves for a Li-ion pouch cell. a Voltage during charge protocol. **b** Current during charge protocol, **c** Extracted ageing voltage curve segments corresponding to marked grey area, **d** Extracted ageing current curve segments corresponding to marked grey area, **e** Heatmap of ageing with cycle number.

Supplementary Note 3. Feature engineering.

Capacity (Q) is calculated based the charge/discharge current (I) and it is given by:

$$Q = \int_{t_0}^{t_{end}} I dt \tag{2}$$

Energy (E) is calculated based on capacity (Q) and voltage (V) given by:

$$E = \int_{t_0}^{t_{end}} V(t) \cdot I dt \tag{3}$$

Attribute	Target variable
Cycle time	Discharge Capacity
Discharge C-rate	
Charge C-rate	
Operational time	
Voltage vs. Time	
Current vs. Time	
Charge times	

Table 1: Parameters recorded during cycling tests.

From pattern recognition domain, three features are derived, signal mean, kurtosis coefficient and skewness coefficient. Skewness coefficient and kurtosis coefficient are calculated based on the following formulas:

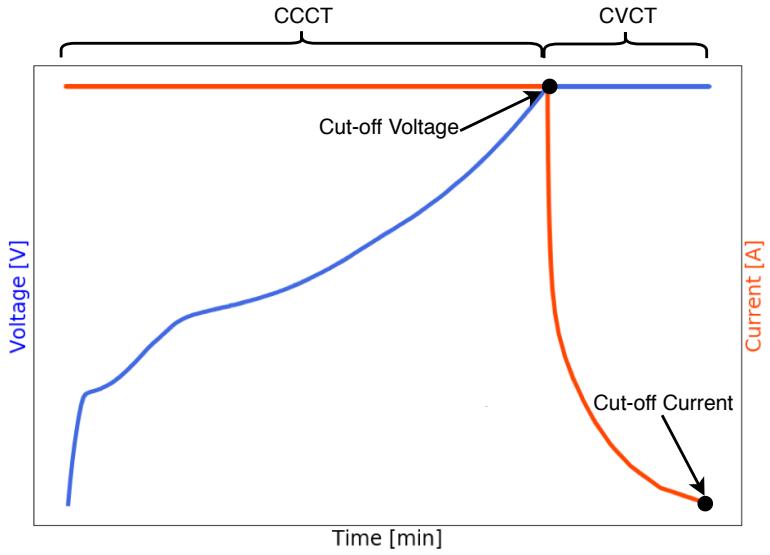

Figure 3: Typical constant current - constant voltage charge protocol. Note: CCCT=constant voltage charge time, CVCT=constant voltage charge time

	Feature	**Target variable**
Battery specific data	Nominal Capacity [Ah]	Discharge Capacity [Ah]
	Charge Current [A]	
	Discharge Current [A]	
Cumulative (historical) data	Cumulated Discharge Capacity [Ah]	
	Cumulated Discharge Energy [Wh]	
1 Cycle Lagged Data	Lagged Cycle Time [s]	
	Lagged Pseudo Resistance [Ω]	
Instantaneous charge data*	Terminal Voltage @ Start of charge [V]	
	Charge time of CC segment of charge curve [s]	
	Charge time of CV segment of charge curve [s]	
	Mean current during CC segment of the curve [A]	
	Mean voltage during CV segment of the curve [A]	
	Slope of CCCV-CCCT segment of the curve	
	Slope of CVCC-CVCT segment of the curve	
	Energy during CCCV-CCCT segment of the curve [Wh]	
	Energy during CVCC-CVCT segment of the curve [Wh]	
	Energy ratio CCCV-CCCT / CVCC-CVCT segment of the curve	
	Energy Difference between the curve segments (CCCV-CCCT) - (CVCC-CVCT)	
	Entropy of CCCV-CCCT segment of the curve eq 8	
	Entropy of CCCV-CCCT segment of the curve eq 8	
	Shannon entropy of CCCV segment of the curve	
	Shannon entropy of CVCC segment of the curve	
	Skewness coefficient of CCCV-CCCT segment of the curve eq 4	
	Skewness coefficient of CVCC-CVCT segment of the curve eq 4	
	Kurtosis coefficient of CCCV-CCCT segment of the curve eq 5	
	Kurtosis coefficient of CVCC-CVCT segment of the curve eq 5	
	Frechet Distance of CCCV-CCCT segment of the curve eq 7	
	Frechet Distance of CVCC-CVCT segment of the curve eq 7	
	Hausdorff Distance of CCCV-CCCT segment of the curve eq 6	
	Hausdorff Distance of CVCC-CVCT segment of the curve eq 6	

Table 2: Engineered features based on recorded parameters in Table 1. Note: CC = consatnt current, CV = constant voltage, CCCV = constant current charge voltage, CVCC = contant voltage charge current, CCCT = constant current charge time, CVCT = constant voltage charge time

$$skewness = \frac{\sum_{i=1}^{n}(x(i) - \bar{x})^3}{(n-1)\sigma_x^3} \tag{4}$$

$$kurtosis = \frac{\sum_{i=1}^{n}(x(i) - \bar{x})^4}{(n-1)\sigma_x^4} \tag{5}$$

where $\bar{x}$ and σ_x represent the mean and standard deviation of feature x.

In addition to pattern recognition based features, distance measurements from a predetermined reference curve to CVCC - CVCT curve and CCCV - CCCT have also been considered. We choose here as reference a simple line defined by the equation $y = mx + c$ where y represents current or voltage depending on the curve under scrutiny, and x represents time. An illustration of the two curves and their reference lines are shown in figures 4 and 5. Instead of simple Euclidean distance, we employ here two different measurements, namely Directed Hausdorff (DH) and Frechet (FD) distance. Both methods are well established in various domains and thoroughly explained in [12], [13] and [14]. We only consider here Directed Hausdorff distance from charge curve to reference line and not vice-versa. DH distance between two point sets $A(a_1, a_2)$ and $B(b1, b2)$, where a_1, a_2, b_1, b_2 are 2D coordinates, is calculated as maximum distance between each point $x \epsilon A$ to its nearest neighbour $y \epsilon B$ and is given by:

$$H(A, B) = max_{x \epsilon A}\{min_{y \epsilon B}\{||x, y||\}\} \tag{6}$$

where $||x, y||$ can be any norm, including the Euclidean distance. Note that $H(A, B) \neq H(B, A)$, in other words, DH is not symmetric.

The point set A is represented by one of the two charge curves namely, CCCV-CCCT or CVCC-CVCT, whereas B is represented by a line of 30-40 points as shown in 4 and 5.

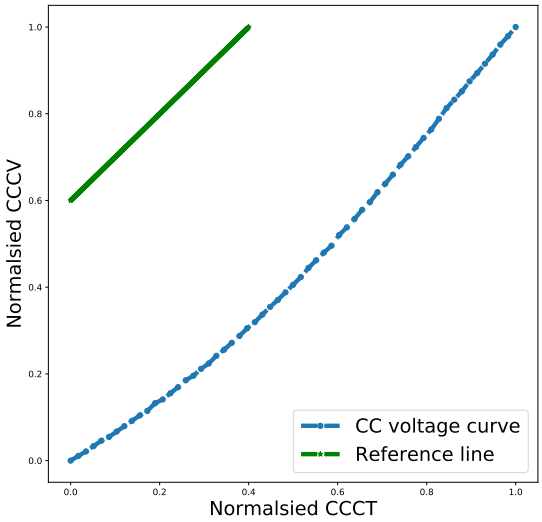

Figure 4: Typical constant current (CC) charge curve with associated reference line of equation $y = mx + c$

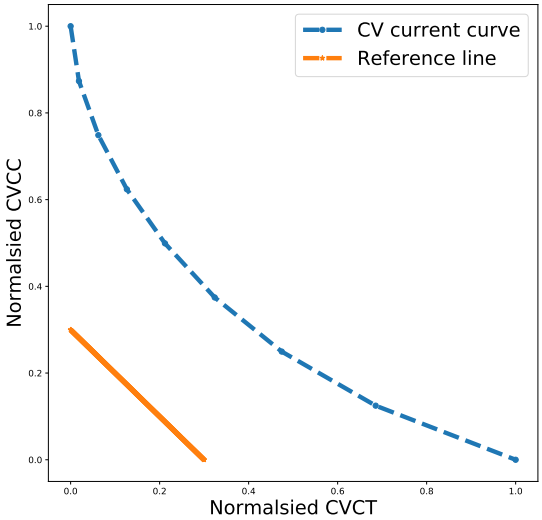

Figure 5: Typical constant voltage (CV) charge curve with associated reference line of equation $y = mx + c$

Frechet distance of two curves A , B has been generally described as the minimal length of a leash required to connect a dog to its owner, as they walk along A or B, respectively, without backtracking. In contrast to distance notions such

as the Hausdorff distance, it takes into account the order of the points along the curve, and thus better captures the similarity as perceived by human observers.[14] In mathematical terms, however, the Frechet distance between two curves is defined as:

$$FD(A, B) = min\{max||A(\alpha(t)), B(\beta(t))||\} \tag{7}$$

where $\alpha(t)$ and $\beta(t)$, range over continuous and increasing functions with $\alpha, \beta, t\ \epsilon[0, 1]$. Again, $||...||$ can be any norm, including Euclidian distance. A more elaborate mathematical explanation is beyond the scope of the present material, however, a thorough mathematical explanation can be found in [15]

The entropy of CVCC-CVCT and CCCV-CCCT curves is also considered as a feature. In information theory, entropy is the average rate at which information is produced by a stochastic source of data [16], whereas in statistical mechanics, entropy is an extensive property of a thermodynamic system. Thermodynamic property of curves has been thoroughly analysed in [17], [18], [19]. Authors in [19] provide an algorithmic procedure to compute curve entropy, and it has been adopted here with slight modification as follows. Curve entropy (EC) is defined by:

$$EC = \frac{\log_2 (\frac{2L}{D})}{\log_2 (N - 1)} \tag{8}$$

where L is the length of the plane curve, D is the diameter of the smallest hypersphere covering the curve, and $N - 1$ is the number of segments approximating the line. For a thorough mathematical explanation on how all variables have been calculated refer to reference [19].

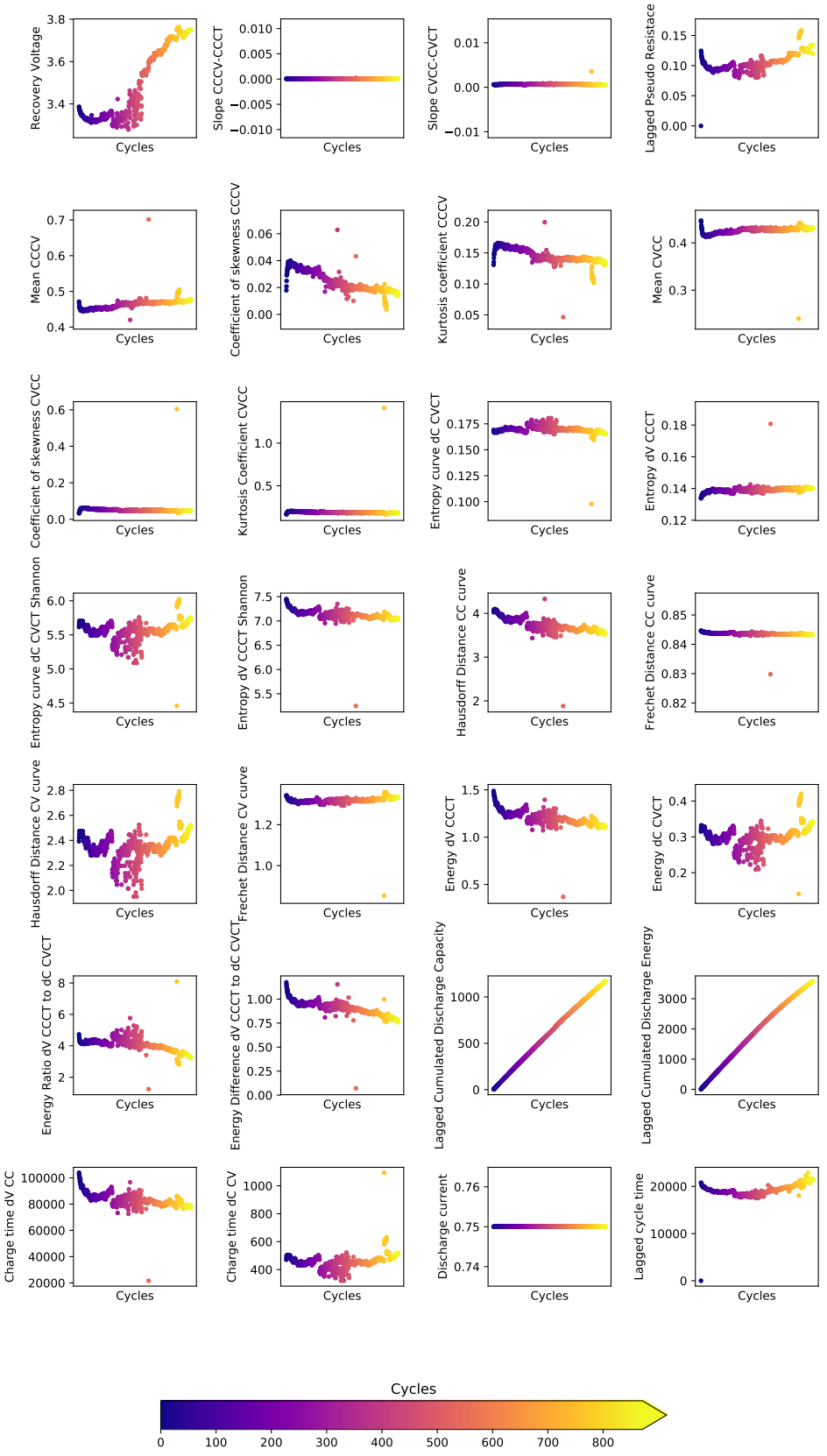

Figure 6: Example visualisation of derived features for Group I datasets cell no. 11.

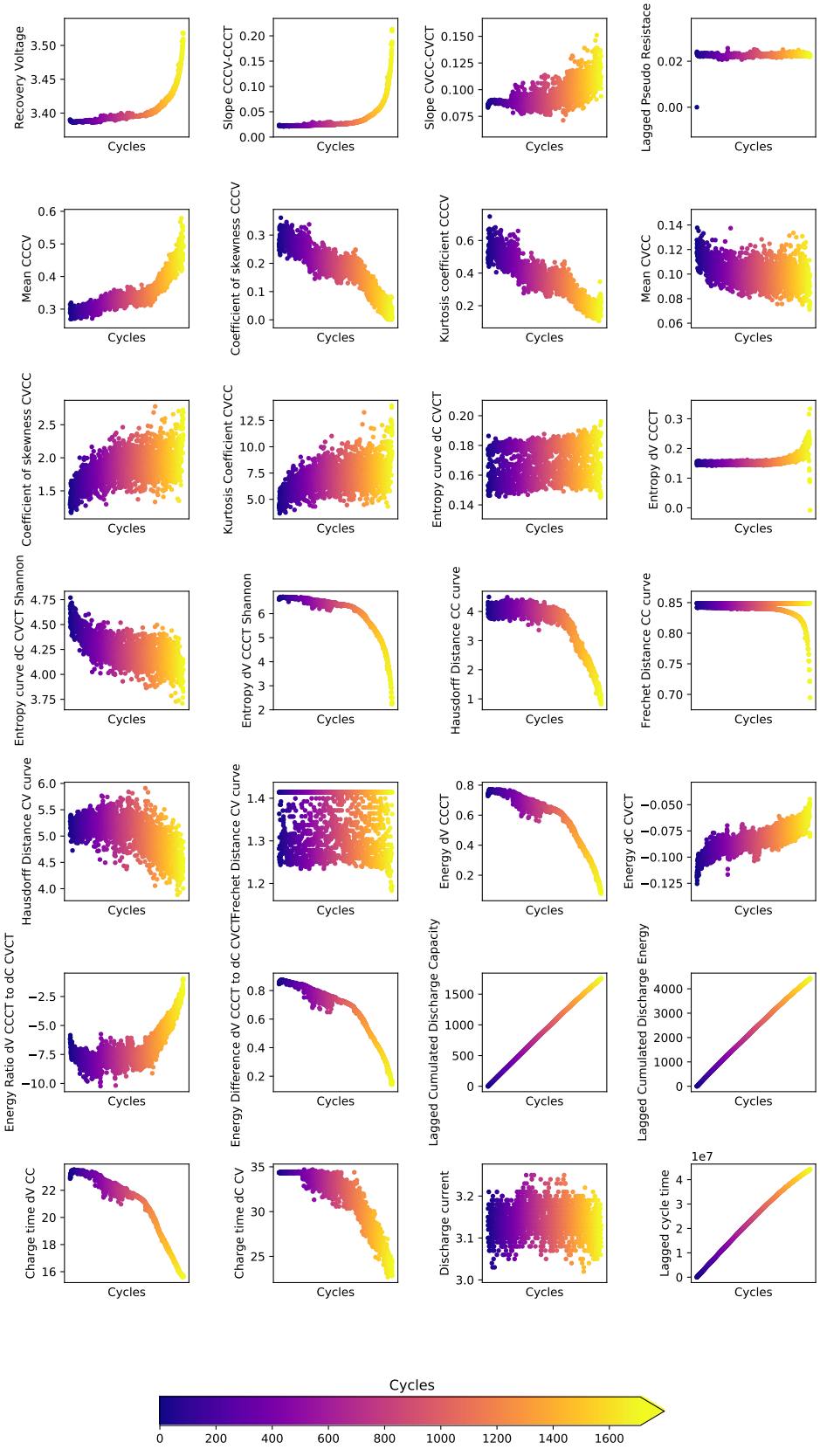

Figure 7: Example visualisation of derived features for Group 2 datasets cell no. 1.

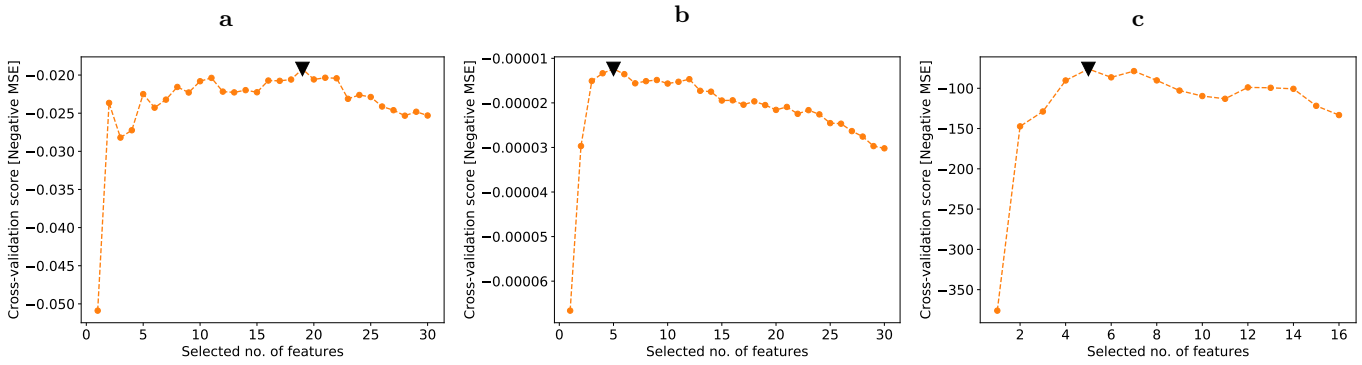

Figure 8: Automatic feature selection with RF-RFE-CV - Note: black triangle indicates selected no. of features. **a** 18 features selected for Group I. **b** 5 features selected for Group 2, **c** 5 features selected for Group 3.

Data type	Feature	Feature no.
Battery specific data	Nominal Capacity [Ah]	1
	Charge Current [A]	2
Cumulative (historical) data	Cumulated Discharge Capacity [Ah]	3
	Cumulated Discharge Energy [Wh]	4
1 Cycle Lagged Data	Lagged Cycle Time [s]	5
	Terminal Voltage @ Start of charge [V]	6
Instantaneous Charge Data	Charge time of CC segment of charge curve [s]	7
	Charge time of CV segment of charge curve [s]	8
	Mean current during CC segment of the curve [A]	9
	Slope of CCCV-CCCT segment of the curve	10
	Slope of CVCC-CVCT segment of the curve	11
	Energy during CCCV-CCCT segment of the curve [Wh]	12
	Energy during CVCC-CVCT segment of the curve [Wh]	13
	Energy ratio CCCV-CCCT / CVCC-CVCT segment of the curve	14
	Energy Difference between curve segments (CCCV-CCCT) - (CVCC-CVCT)	15
	Entropy of CCCV-CCCT segment of the curve based on \ref{}	16
	Shannon entropy of CCCV segment of the curve	17
	Frechet Distance of CCCV-CCCT segment of the curve	18

Table 3: Selected features using RF-RFE-CV for Group I. Note: CC = consatnt current, CV = constant voltage, CCCV = constant current charge voltage, CVCC = contant voltage charge current, CCCT = constant current charge time, CVCT = constant voltage charge time

Data type	Feature	Feature no.
Instantaneous charge data	Energy during CCCV-CCCT segment of the curve [Wh]	1
	Energy Difference between curve segments (CCCV-CCCT) - (CVCC-CVCT)	2
	Hausdorff Distance of CCCV-CCCT segment of the curve	4
	Shannon entropy of CCCV segment of the curve	3
	Frechet Distance of CCCV-CCCT segment of the curve	5

Table 4: Selected features using RF-RFE-CV for Group 2. Note: CC = consatnt current, CV = constant voltage, CCCV = constant current charge voltage, CVCC = contant voltage charge current, CCCT = constant current charge time, CVCT = constant voltage charge time

Supplementary Note 4. Data overview

Irrespective of dataset, input data consistency is ensured by removing outliers in the training data, possibly introduced due to inherent cell variability and measurement errors. The data preprocessing step involves filtering of the raw data based on erroneous capacity measurements by utilizing Random Sample Consensus (RANSAC) algorithm [20]. Training data that contains a significant percentage of gross errors in capacity from one cycle to another is removed as illustrated in the examples of Supplementary Figure 10. Note, test data has not been processed for outliers to simulate a realistic deployment scenario.

Data type	Feature	Feature no.
Cumulative (historical) data	Cumulated Discharge Capacity [Ah]	1
	Cumulated Discharge Energy [Wh]	2
1 Cycle Lagged Data	Lagged Cycle Time [s]	3
Instantaneous Charge Data	Capacity during CCCV-CCCT segment of the curve [Ah]	4
	Energy during CCCV-CCCT segment of the curve [Wh]	5

Table 5: Selected features using RF-RFE-CV for Group 3. Note: CC = consatnt current, CV = constant voltage, CCCV = constant current charge voltage, CVCC = contant voltage charge current, CCCT = constant current charge time, CVCT = constant voltage charge time

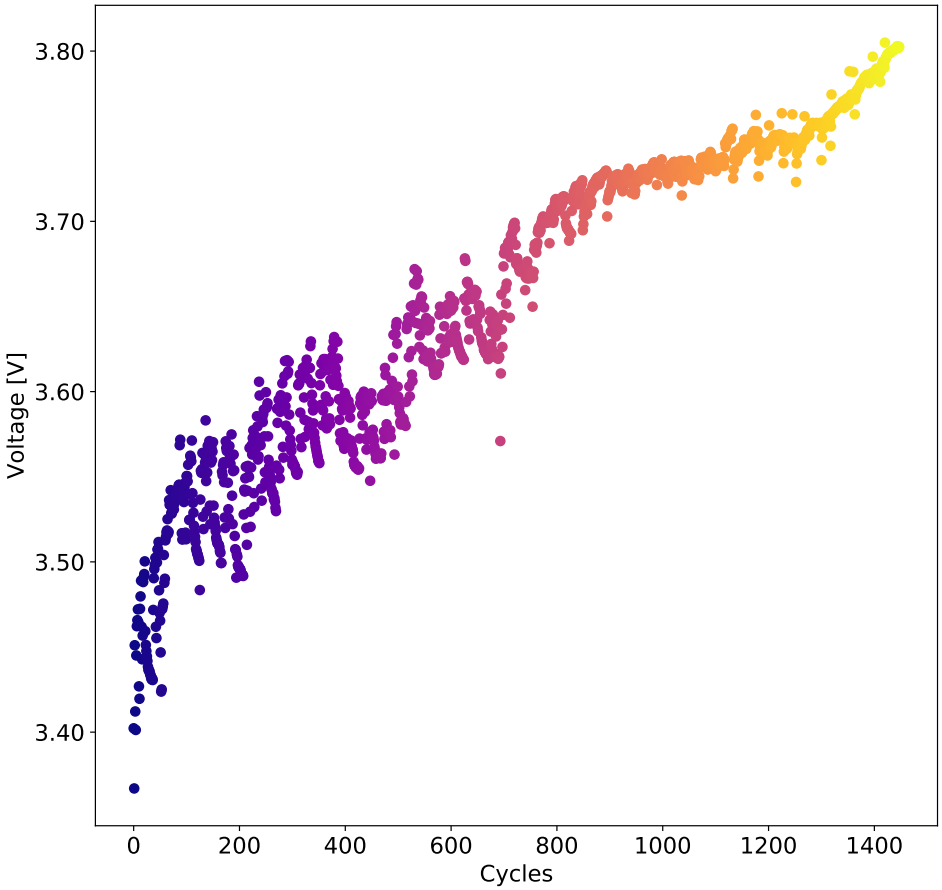

Figure 9: Increase in start of charge voltage between test cycles for a $LiCoO_2$ prismatic battery. The cell underwent a full depth of discharge at a current value of 1 C-rate with constant current - constant voltage charging.

CALCE dataset

Data sourced from CALCE battery group consists of three batteries. For ease of reference, we preserve the original dataset names as per their website `https://web.calce.umd.edu/batteries/data.htm`. All cells in the dataset underwent the same charging profile, the standard CC-CV. The CC phase of charging profile includes a 0.5 C-rate charging current until the voltage reached the cut-off threshold value of 4.2V. The CV top-up phase sustained the previously reached 4.2V until the current dropped to a value of 0.05 C-rate, at which point the charging is complete. Except for batteries in CALCE PL dataset, which were discharged at 1 C-rate, the other two datasets have been discharged at both 0.5 C-rate and 1 C-rate until the battery voltage reached the pre-defined discharge cut-off voltage of 2.7V. A schematic of the charge profile together with a detailed summary of discharge conditions for each battery can be found in Figure 3 and in Supplementary Table 6, respectively.

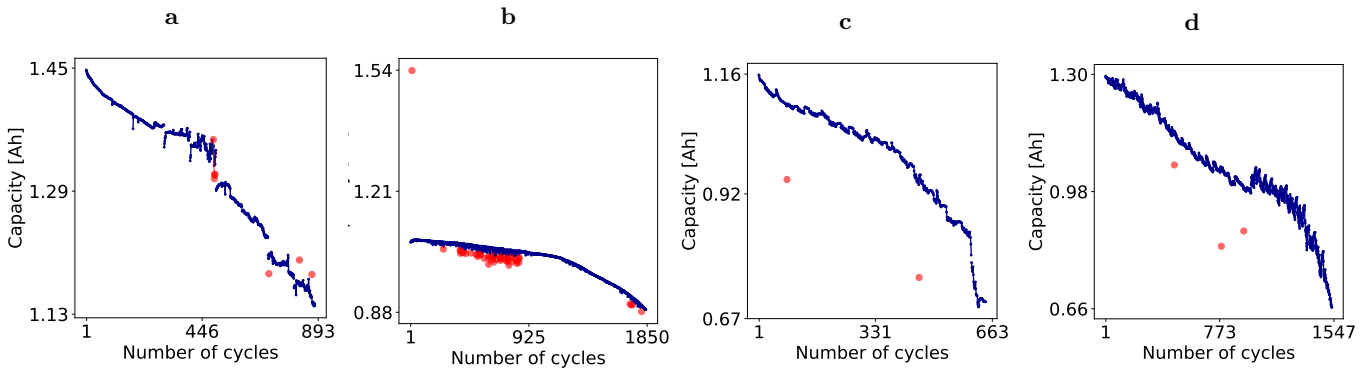

Figure 10: Training data outlier removal with RANSAC (red denotes outliers, blue denotes inlier). **a** Cylindrical A123 LFP/graphite training cell 11 Group II. **b** Pouch LCO cell 2 Group I, **c** Prismatic CS2 LCO training cell 34 Group I, **d** Prismatic CX2 LCO training cell 34 Group I.

NASA dataset

NASA data can be retrieved from the public NASA Ames Prognostics Centre of Excellence website `https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/` and includes two datasets. The first repository, the battery dataset denoted here by NASA5, includes a mixture of constant discharge current and squared wave-based discharge current experiments at different temperatures. The second repository, the randomised battery usage dataset, denoted here by NASA11, includes batteries that are continuously cycled with randomly generated current profiles. The randomised nature of the load profiles is an ideal representation of practical battery usage. Both NASA5 and NASA11 dataset follow the traditional CC-CV charge protocol. CC charging mode was carried at 1.5A until the battery voltage reached 4.2V and then continued in a CV fashion until the charge current dropped to 0.02A, at which point the battery was deemed fully charged. In terms of discharge, NASA5 discharge was carried out at a constant current level of 2A or square wave loading profile of 4A until the battery voltage fell to 2.7V, 2.5V or 2.2V. Whereas, NASA11 undergone a randomised discharge profile of varying duration ranging from 5 minutes to 3 hours as well as varying discharge current values ranging from 0A to 5A. All cells underwent a periodic characterisation test whereby a 2A CC and 0.02A CV current cut-off charge protocol and a 2A constant current discharge was applied. The characterisation test data was used in BHUMP to evaluate battery health as a function of capacity, as opposed to cyclic data. Details of charging and discharging profiles per battery are found in Supplementary Table 7.

TRI dataset

The work supported by Toyota Research Institute in partnership with MIT and Stanford generated a lifecycle battery dataset consisting of 124 cells, available at `https://data.matr.io/1/projects/5c48dd2bc625d700019f3204`. The dataset was used in [21] where more details on battery type, manufacturer and testing equipment can be found. All cells in the dataset were cycled with a total of 72 different fast charging-polices but identically discharged with a current 4 C-rate between 3.6V and 2.0V. The charging protocol included a two-step fast charge between 0% to 80% SOC. The fast charge section is followed by a CC protocol, i.e. a uniform charge current value of 1 C-rate to 3.6V until the voltage reaches the cut-off value of 3.6V, immediately followed by a CV top-up phase until current dropped to 0.02 C-rate. The raw data from each cycle is used as input to BHUMP pipeline. Details regarding the charge profile as well as the cycling regimes for each battery can be found in [21], whilst Figure 2 illustrates the charging protocol and Supplementary Table 9 indicates which cells have been used for training and testing of the algorithms.

Oxford dataset

The Oxford Battery Degradation Dataset and can be accessed at `https://ora.ox.ac.uk/objects/uuid:03ba4b01-cfed-46d3-9b1a-7d4a7bdf6fac`. A comprehensive explanation of the testing method, equipment and battery specific characteristic is found in [22]. The data consists of ageing experiments by repeatedly cycling the cells via a CC charge profile coupled with the ARTEMIS urban drive cycle discharge profile. The CC charge protocol uses a 2 C-rate current to a voltage of 4.2V. The discharge profile voltage range is 4.2V to 2.7V. After every 100 cycles of repeated charge-discharge using the protocol mentioned above, a characterisation test (incorporating a full constant current charge-discharge at C/18.5 (40 mA), repeated every 100 drive cycles.) is carried out. The characterisation test data

is used in this work for battery health degradation estimation purposes. Supplementary Table 10 indicates which cells have been used for training/testing the algorithms.

Supplementary Note 5. Data partitioning

Group I

Out of the 47 cells in Group I, we use 23 cells for training (out of this 10 are used for feature selection), 5 cells for calibration and remaining 19 for evaluating the algorithm performance (the cells used during training-testing can be found in Supplementary Tables 6, 7, 8). Note that the calibration dataset is neither used in training nor testing to prevent overfitting.

Group II

Group II dataset is randomly split into 63 cells for training (out of which 37 cells are used for feature selection), 10 for calibration and the remainder 51 cells for testing, refer to Supplementary Table 9 for cell partition in each dataset.

Group III

Group III dataset is split into 3 cells for training (cells 1 to 3), one cell for calibration (cell no. 4), and the remainder of 4 cells for testing (see Supplementary Table 10 for details).

Cell name	Discharge condition	Dataset
CS2 - 33	0.5 C-rate	Test
CS2 - 34	0.5 C-rate	Train
CS2 - 35	1 C-rate	Train & Feature Selection
CS2 - 36	1 C-rate	Train & Feature Selection
CS2 - 37	1 C-rate	Calibration
CS2 - 38	1 C-rate	Test
CX2 - 33	0.5 C-rate	Test
CX2 - 34	0.5 C-rate	Train
CX2 - 35	0.5 C-rate	Train & Feature Selection
CX2 - 36	0.5 C-rate	Calibration
CX2 - 37	0.5 C-rate	Train & Feature Selection
CX2 - 38	0.5 C-rate	Test
PL - 11	0.5 C-rate	Train
PL - 13	0.5 C-rate	Test

Table 6: Group I: CALCE battery data discharge conditions and train, calibration and test split. For complete details on test conditions access `https://web.calce.umd.edu/batteries/data.htm`.

Cell name	Discharge condition	Dataset
B0005	2A	Train & Feature Selection
B0006	2A	Test
B0007	2A	Train
B0018	2A	Test
B0025	Square wave @ 4A	Test
B0026	Square wave @ 4A	Train & Feature Selection
B0027	Square wave @ 4A	Train
B0028	Square wave @ 4A	Calibration

Table 7: Group I: NASA 5 battery data discharge conditions and train, calibration and test split. For complete details on test conditions access `https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/`.

Cell name	Discharge condition	Dataset
RW1	Random Sequence	Train & Feature Selection
RW2	Random Sequence	Train
RW3	Random Sequence	Train
RW4	Random Sequence	Train
RW5	Random Sequence	Test
RW6	Random Sequence	Test
RW7	Random Sequence	Test
RW8	Random Sequence	Test
RW9	Random Sequence	Train & Feature Selection
RW10	Random Sequence	Train
RW11	Random Sequence	Calibration
RW12	Random Sequence	Test
RW13	Random Sequence	Train
RW14	Random Sequence	Train
RW15	Random Sequence	Test
RW16	Random Sequence	Test
RW20	Random Sequence	Train & Feature Selection
RW21	Random Sequence	Train & Feature Selection
RW22	Random Sequence	Train
RW23	Random Sequence	Test
RW24	Random Sequence	Test
RW25	Random Sequence	Train & Feature Selection
RW26	Random Sequence	Train
RW27	Random Sequence	Test
RW28	Random Sequence	Calibration

Table 8: Group I: NASA 11 battery data discharge conditions and train, calibration and test split. Note: batteries are discharged to 3.2V using a randomized sequence of discharging loads between 0.5A and 4A. For complete details on test conditions access `https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/`.

Dataset	Cell number	Number of cells
Feature Selection	2, 6, 8, 14, 18, 19, 26, 28, 32, 35, 37, 45, 51, 53, 55, 58, 60, 61, 65, 69, 72, 76, 79, 83, 90, 91, 92, 103, 107, 109, 110, 113, 115, 116, 119, 120, 124	37
Training	2, 3, 6, 8, 9, 13, 14, 16, 18, 19, 20, 21, 23, 25, 26, 28, 32, 35, 37, 42, 45, 46, 50, 51, 53, 55, 56, 58, 60, 61, 63, 64, 65, 66, 69, 72, 73, 76, 79, 83, 84, 86, 88, 90, 91, 92, 94, 95, 98, 100, 103, 105, 106, 107, 109, 110, 113, 115, 116, 118, 119, 120, 124	63
Calibration	7, 12, 22, 48, 54, 59, 68, 77, 82, 108	10
Testing	1, 4, 5, 10, 11, 15, 17, 24, 27, 29, 30, 31, 33, 34, 36, 38, 39, 40, 41, 43, 44, 47, 49, 52, 57, 62, 67, 70, 71, 74, 75, 78, 80, 81, 85, 87, 89, 93, 96, 97, 99, 101, 102, 104, 111, 112, 114, 117, 121, 122, 123	51

Table 9: Group II: TRI dataset splitting for: feature selection, training, calibration and testing. For complete details on test conditions access `https://data.matr.io/1/projects/5c48dd2bc625d700019f3204`.

Dataset	Cell number	Total number of cells
Feature Selection	1, 3	2
Training	1, 2, 3	3
Calibration	4	1
Testing	5, 6, 7, 8	4

Table 10: Group III: Oxford dataset splitting for: feature selection, training, calibration and testing. For complete details on test conditions access `https://ora.ox.ac.uk/objects/uuid:03ba4b01-cfed-46d3-9b1a-7d4a7bdf6fac`.

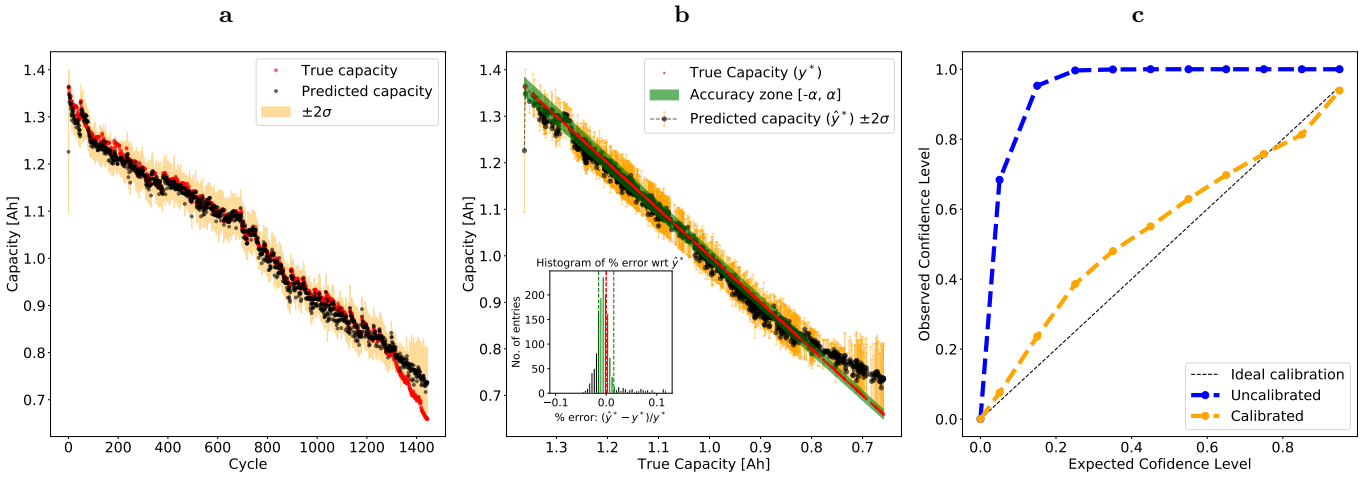

Figure 11: Prediction results with BRR Group I cell no. 38. **a** Prediction as a function of cycle numbers, **b** Actual vs. predicted capacity, **c** Calibration results.

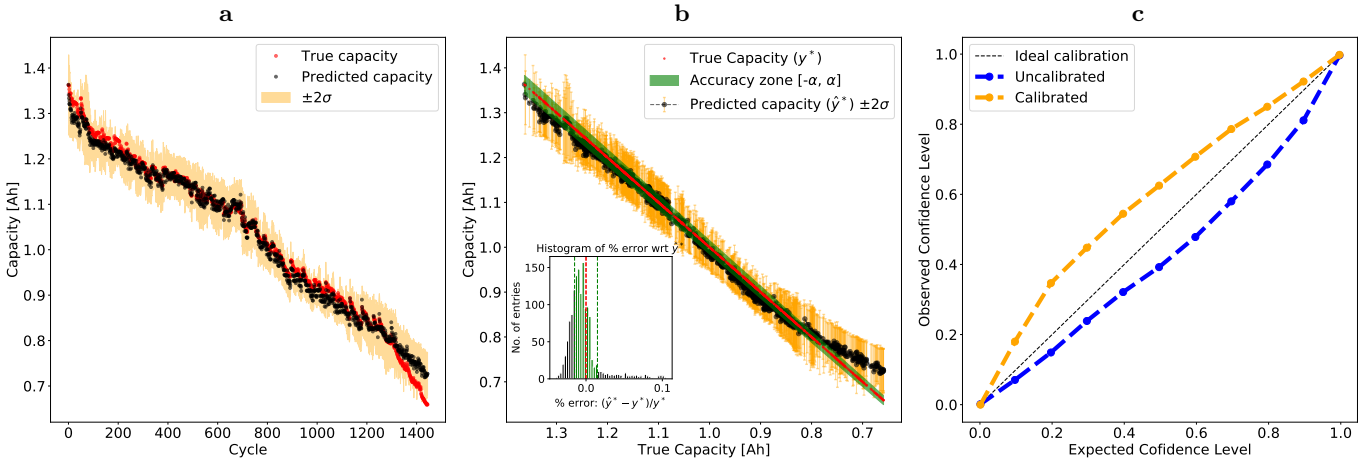

Figure 12: Prediction results with GPR Group I cell no. 38. **a** GPR prediction as a function of cycle numbers, **b** GPR actual vs. predicted capacity, **c** GPR calibration results.

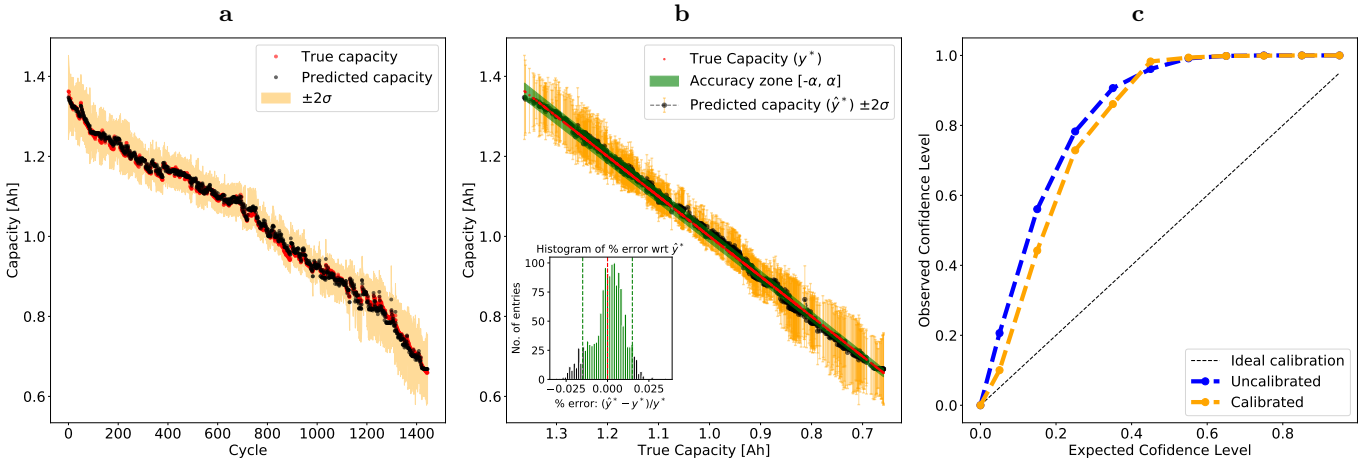

Figure 13: Prediction results with RF Group I cell no. 38. **a** RF prediction as a function of cycle numbers, **b** RF actual vs. predicted capacity, **c** RF calibration results.

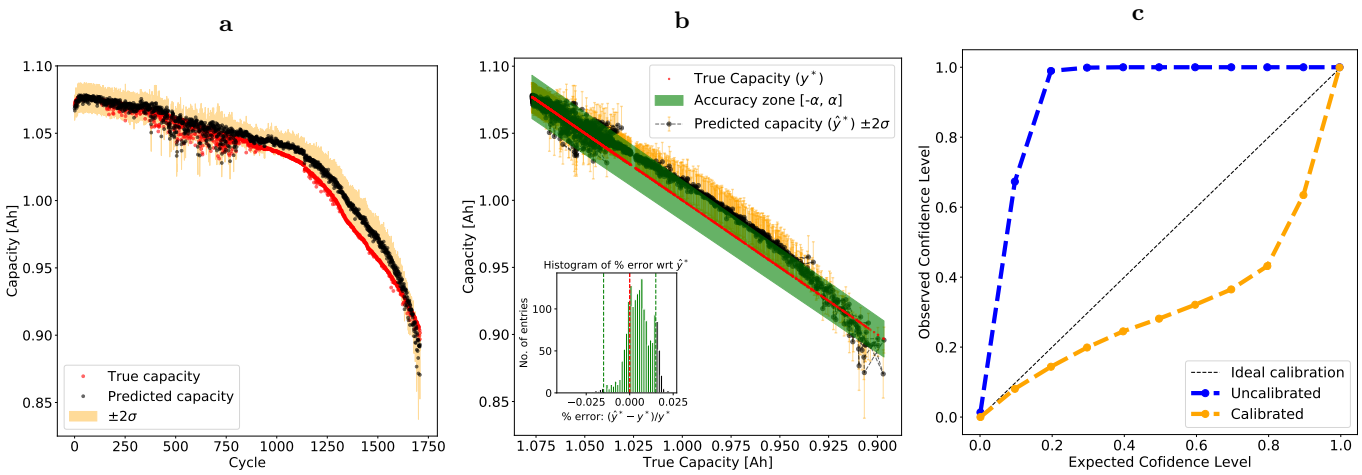

Figure 14: Prediction results with BRR Group II cell no. 1. **a** BRR prediction as a function of cycle numbers, **b** BRR actual vs. predicted capacity, **c** BRR calibration results.

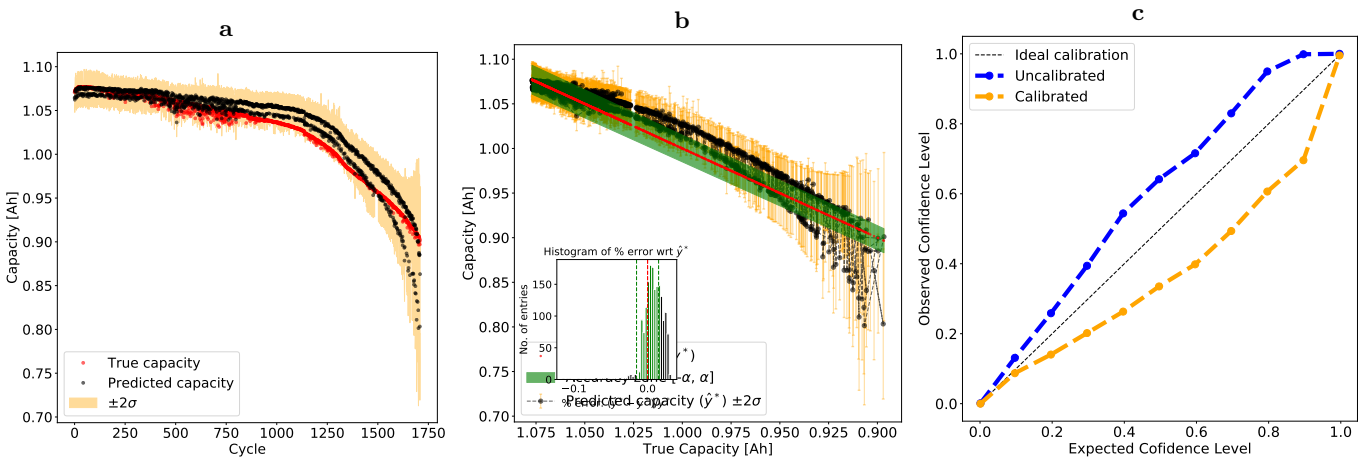

Figure 15: Prediction results with GPR Group II cell no. 1. **a** GPR prediction as a function of cycle numbers, **b** GPR actual vs. predicted capacity, **c** GPR calibration results.

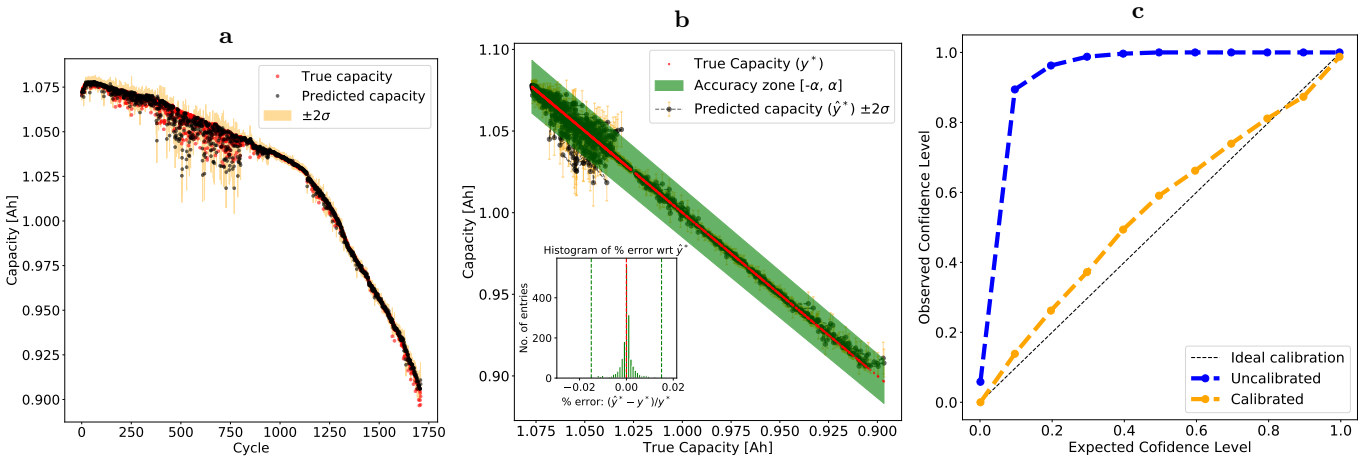

Figure 16: Prediction results with RF Group II cell no. 1. **a** RF prediction as a function of cycle numbers, **b** RF actual vs. predicted capacity, **c** RF calibration results.

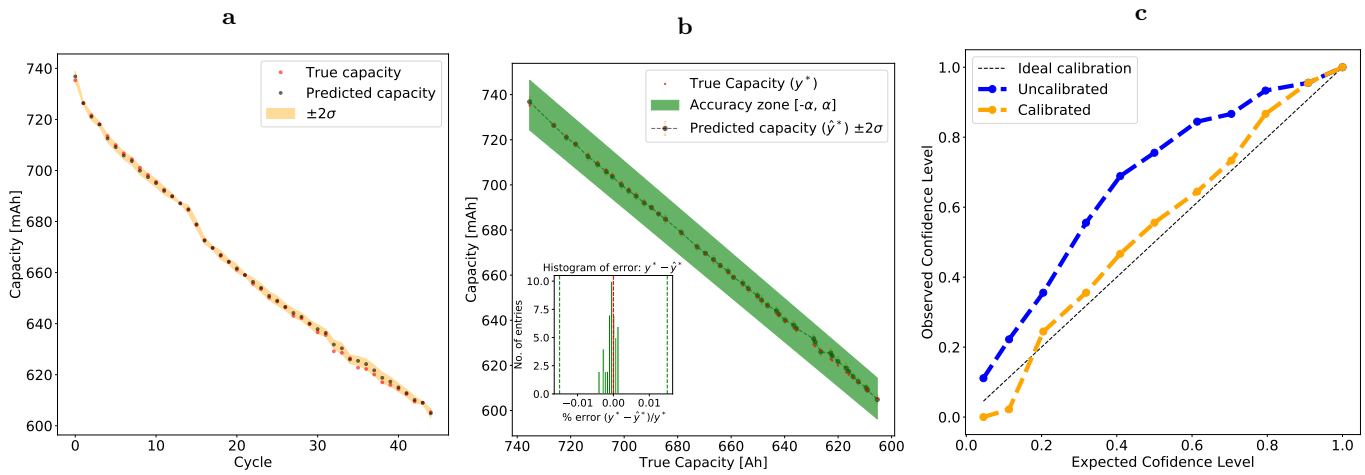

Figure 17: Prediction results with BRR Group III cell no. 5. a BRR prediction as a function of cycle numbers, **b** BRR actual vs. predicted capacity, **c** BRR calibration results.

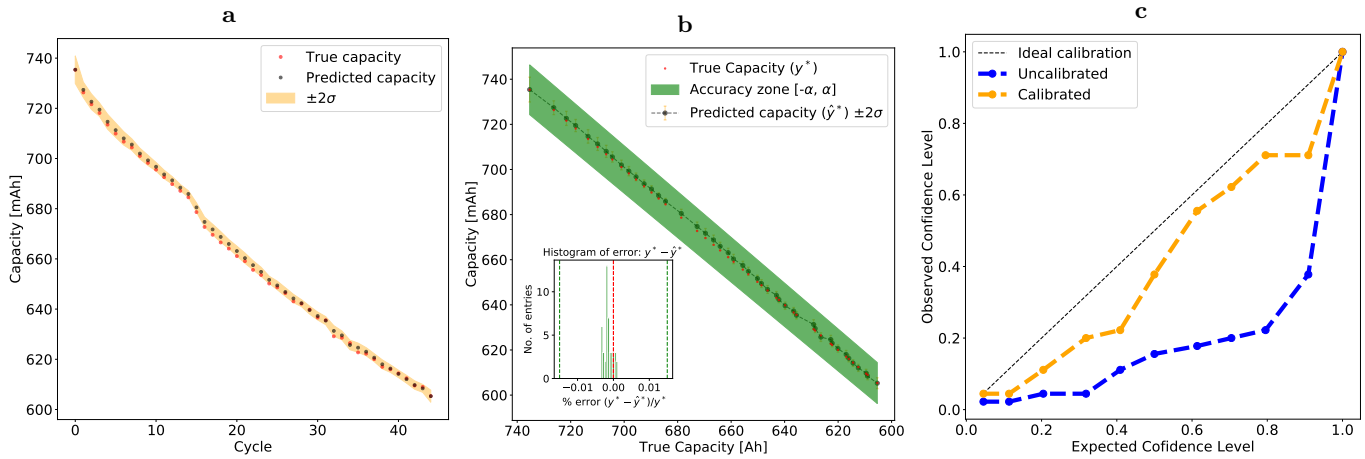

Figure 18: Prediction results with GPR Group III cell no. 5. a GPR prediction as a function of cycle numbers, **b** GPR actual vs. predicted capacity, **c** GPR calibration results.

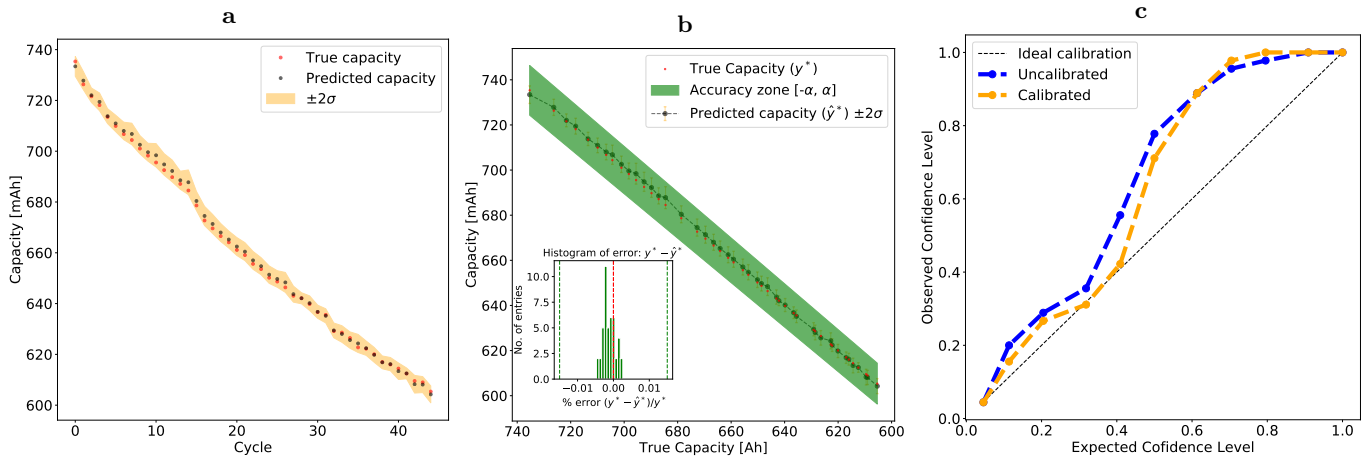

Figure 19: Prediction results with RF Group III cell no. 5. a RF prediction as a function of cycle numbers, **b** RF actual vs. predicted capacity, **c** RF calibration results.

Supplementary References

1. Palacın, M. R. Understanding ageing in Li-ion batteries: a chemical issue. *Chemical Society Reviews* **47,** 4924–4933 (2018).
2. Liu, Q. *et al.* Understanding undesirable anode lithium plating issues in lithium-ion batteries. *RSC Advances* **6,** 88683–88700 (2016).
3. Liu, C., Neale, Z. G. & Cao, G. Understanding electrochemical potentials of cathode materials in rechargeable batteries. *Materials Today* **19,** 109–123 (2016).
4. Aurbach, D., Zinigrad, E., Teller, H. & Dan, P. Factors which limit the cycle life of rechargeable lithium (metal) batteries. *Journal of The Electrochemical Society* **147,** 1274–1279 (2000).
5. Keil, P. & Jossen, A. Charging protocols for lithium-ion batteries and their impact on cycle life—An experimental study with different 18650 high-power cells. *Journal of Energy Storage* **6,** 125–141 (2016).
6. Zhang, S. S., Xu, K. & Jow, T. Study of the charging process of a LiCoO2-based Li-ion battery. *Journal of Power Sources* **160,** 1349–1354 (2006).
7. Zhou, J. & Notten, P. Studies on the degradation of Li-ion batteries by the use of microreference electrodes. *Journal of power Sources* **177,** 553–560 (2008).
8. Williard, N., He, W., Osterman, M. & Pecht, M. Comparative Analysis of Features for Determining State of Health in Lithium-Ion Batteries. *International Journal of Prognostics and Health Management, ISSN 2153-2648* (2013).
9. Eddahech, A., Briat, O. & Vinassa, J.-M. Determination of lithium-ion battery state-of-health based on constant-voltage charge phase. *Journal of Power Sources* **258,** 218–227 (2014).
10. Wang, Z., Zeng, S., Guo, J. & Qin, T. State of health estimation of lithium-ion batteries based on the constant voltage charging curve. *Energy* **167,** 661–669 (2019).
11. Saxena, A. *et al. Designing data-driven battery prognostic approaches for variable loading profiles: Some lessons learned* in *European conference of prognostics and health management society* (2012), 72–732.
12. Taha, A. A. & Hanbury, A. An efficient algorithm for calculating the exact Hausdorff distance. *IEEE transactions on pattern analysis and machine intelligence* **37,** 2153–2163 (2015).
13. Efrat, A., Fan, Q. & Venkatasubramanian, S. Curve matching, time warping, and light fields: New algorithms for computing similarity between curves. *Journal of Mathematical Imaging and Vision* **27,** 203–216 (2007).
14. Bringmann, K. *Why walking the dog takes time: Frechet distance has no strongly subquadratic algorithms unless SETH fails* in *2014 IEEE 55th Annual Symposium on Foundations of Computer Science* (2014), 661–670.
15. Eiter, T. & Mannila, H. Computing Discrete Frechet Distance (1994).
16. Shannon, C. E. A mathematical theory of communication. *Bell system technical journal* **27,** 379–423 (1948).
17. Dupain, Y., Kamae, T. & Mendés, M. Can one measure the temperature of a curve? *Archive for Rational Mechanics and Analysis* **94,** 155–163 (1986).
18. Moore, R., Van Der Poorten, A., *et al. On the thermodynamics of curves and other curlicues* in *Conference on Geometry and Physics, Canberra* (1989), 82–109.
19. Balestrino, A., Caiti, A. & Crisostomi, E. Generalised entropy of curves for the analysis and classification of dynamical systems. *Entropy* **11,** 249–270 (2009).
20. Fischler, M. A. & Bolles, R. C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **24,** 381–395 (1981).
21. Severson, K. A. *et al.* Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy* **4,** 383 (2019).
22. Birkl, C. R. Diagnosis and Prognosis of Degradation in Lithium-Ion Batteries (2017).