Engineering Testing Protocols for Machine Learning-Based SoH Estimation in Lithium Metal Batteries

Journal of The Electrochemical Society, 2025 **171** 120531
1945-7111/2025/171(12)/120531/15/$40.00 © 2025 The Electrochemical Society ("ECS"). Published on behalf of ECS by IOP Publishing Limited.

Engineering Testing Protocols for Machine Learning-Based SoH Estimation in Lithium Metal Batteries

Wenting Ma,[1] Sara Ha,[2] Gabriele Pozzato,[1] Alireza Saatchi,[3] Nojan Aliahmad,[3] Karl Littau,[3] and Simona Onori[1,z]

[1]*Energy Science and Engineering, Stanford University, Stanford, 94305 California, United States of America*
[2]*Mechanical Engineering, Stanford University, Stanford, 94305 California, United States of America*
[3]*Sakuu Corporation, San Jose, California 95138, United States of America*

Lithium metal batteries (LMB) represent one of the most promising battery technologies due to their high energy density. However, the commercial adoption of LMBs has been hindered by their limited lifespan and safety concerns resulting from lithium metal's high electrochemical reactivity. Effective health estimation could enable optimized LMB operation to improve utilization efficiency and reduce failure risks, as well as accelerate LMB development iteration. In this work, we first propose novel diagnostic tests and aging cycling protocols that enable the tracking of LMB degradation from voltage-current measurements through extracting internal resistance, relaxation voltage statistics, and charging impedance descriptors. We present an integrated data-driven health estimation framework for LMBs, using Bagged Gaussian Process Regression (Bagged-GPR). Validation over test cells demonstrates accurate health tracking performance throughout LMB lifetime using one cycle of voltage-current data. The proposed model provides degradation insights for LMBs without the need for additional internal sensors.
 [DOI: 10.1149/1945-7111/ad9cc8]

Manuscript submitted October 6, 2024; revised manuscript received November 16, 2024. Published January 6, 2025.

Background.—The future of electric mobility and clean energy storage hinges on the development of next-generation high-energy battery technologies. Lithium metal batteries (LMBs), with their high anode theoretical specific capacity (~ 10 times higher than that of graphite anode) and low anode redox potential (−3.04 V versus standard hydrogen electrode), represent one of the most promising energy storage technologies.[1] However, the commercial adoption of LMBs with liquid electrolyte has been hindered by distinct failure mechanisms emanating from lithium metal's high electrochemical reactivity and significant volume change.[2] The confluence of these mechanisms often leads to premature sudden death caused by soft shortening or high cell impedance and poses safety hazards.[3] In LMBs, Li deposition and dissolution occur at the Li metal-electrolyte interface during charge/discharge cycles.[4] Under certain conditions, lithium deposits form dendritic structures that can penetrate the separator, causing internal short circuits—a major safety risk.[5] Uneven lithium deposition and large volume expansions in the lithium anode during charge and discharge destabilizes the solid electrolyte interface (SEI) layer, leading to cracks in the SEI and exposes fresh lithium to electrolyte, promoting further dendrite growth and electrolyte depletion.[6] During cell discharge, lithium is stripped from the lithium metal surface, causing dendritic structures to detach and become isolated from the lithium metal surface, forming "dead Li" layers on the anode surface. With cycling, the accumulation of "dead Li" and increasing porosity in the anode hinder ion transport, leading to capacity loss.[7] Recent efforts to extend LMB lifespan have focused on stabilizing the anode and preventing dendrite growth through improved cell design and manufacturing, including advancements in anode structures, electrolyte composition, interface engineering, SEI layers, and separators.[1,2,6,8] Rapid, scalable LMB development requires clear traceability of cell health throughout its life cycle, making the construction of health estimation models a critical design step.

Existing methods in the literature to conduct LMB health estimation fall into two categories. One category is by simulating LMB failure mechanisms using physics-based models. Hong and Viswanathan[9] developed a phase-field model to simulate lithium metal electrodeposition using MOOSE software. Liu and Lu[5] created a coupled model for lithium dendrite growth, SEI formation, and SEI regrowth. Xu et al.[10] modeled dead Li layer growth, capturing voltage and capacity behavior as the layer thickens. Uppaluri et al.[11] modeled SEI growth as a moving boundary impacting ion transport. These models can extrapolate to certain operating conditions unseen during their construction and calibration, linking capacity fade to mechanistic insights and suggesting aging-mitigation strategies. However, they assume a single dominant failure mechanism, while real-world conditions likely involve multiple, coupled mechanisms, many of which are not fully understood. Additionally, the computational cost and parameter identifiability limit the deployment of these high-fidelity models.[12] The other category is by adopting data-driven approaches. They have proven effective in designing LMB electrolyte and interphase layers to enhance fresh-cell performance, including Coulombic Efficiency (CE) and component stability.[13–17] Some studies highlight the potential of these methods for LMB health monitoring. Iurilli et al.[18] identified six non-invasive health indicators for detecting capacity knee in Li/NMC LMBs. However, their detection thresholds, selected empirically, achieve a maximum success rate of 78%, limiting generalizability to new conditions. Liu et al.[19] developed an online lifetime prediction model using linear regression and time history decomposition for LMBs with NMC and LCO cathodes, attaining a 2.8% error in trajectory estimation when trained with historical usage data up to 90% state-of-health (SoH).

This paper aims to develop an integrated and general pipeline for inferring LMB health status from aging cycle data (Fig. 1). The pipeline utilizes data from purposely designed cycling protocols to extract internal resistance, relaxation voltage, and charging impedance. Bagged Gaussian Process Regression (Bagged-GPR) models are constructed from these features to estimate battery health based on discharge capacity decrease. The model's accuracy is validated across various charge/discharge rates and temperatures.

The paper is organized as follows: LMB Cell Cycling Dataset describes the LMB cycling experiments and dataset. Battery Performance Metrics defines the metrics for assessing LMB performance and aging, used in the health estimation model. Feature Engineering covers data pre-processing, feature extraction, and selection. Model Training details model training and hyperparameter selection. Results and Analysis validates the health estimation method, and Conclusion offers concluding remarks.

LMB Cell Cycling Dataset

The LMB cells used in this work were produced and cycled by Sakuu Corp. in San Jose, California, USA from February to August 2023. The dataset includes 36 nominally identical LMB cells, with specifications summarized in Table I. As the battery ages, the C-rates

[z]E-mail: sonori@stanford.edu

Figure 1. Health estimation from cycling data.

in this study continue to be calculated based on the nominal capacity, Q_0.

In this dataset, most cells went through three sequential testing phases after assembly: formation cycle, RPT cycle, and repeated aging cycling until reaching their end of life. 22 cells are tested at room temperature around 21 °C with possible temperature fluctuations around ±3 °C (referred to as T21 Tests), while a group of 14 cells are cycled in a thermal chamber at controlled temperature of 40 °C (referred to as T40 Tests). Cell voltage and current are continuously measured throughout testing. In all testing phases, charging and discharging terminate when the voltage reaches 4.3 V and 3.0 V, respectively, or when the charging or discharging capacity reaches 3.5 mAh, whichever occurs first. Each cell is labeled as Config*i*-Cell*j*-T*k*, where $i \in \{1, \ldots, 4\}$ denotes the aging cycling protocol, $j \in \{1, \ldots, 7\}$ indicates the test repetition, and $k \in \{21, 40\}$ represents the testing temperature. All cells, except Config1-Cell6-T21, Config2-Cell6-T21, and Config2-Cell7-T21, underwent a formation cycle consisting of a C/15 constant-current (CC) charge, followed by a C/15 CC discharge, and then two C/10 CC charge-discharge cycles, with 10-minute rests between each cycle (Fig. 2(a)). The sampling interval is 60 seconds for T21 tests and 30 seconds for T40 tests. Data from the formation cycle for Config1-Cell3-T21 and Config1-Cell5-T21 were not properly recorded and are excluded from this study.

The RPT cycle was designed to evaluate the fresh cells' discharge capacity Q_{dchg} and internal resistance R0 with a single test (Fig. 2(b)). It includes one C/5 CC charge - C/10 CC discharge cycle (RPT1) and one C/10 CC charge - C/5 CC discharge cycle (RPT2). Special segments are inserted at 20%, 50%, and 80% state-of-charge (SoC) to measure the internal resistances R0. In RPT1, each segment includes an 80-second C/2 discharge pulse, a 100-second rest, and an 80-second C/2 charge pulse. In RPT2, each segment consists of an 80-second C/2 charge pulse, a 100-second rest, and an 80-second C/2 discharge pulse. There are no rest periods before or after these segments. Aging cycling is conducted at various charge/discharge C-rates as detailed in Table II. The experimental design aims to explore cell degradation patterns across different operating temperatures and C-rates. Multiple repeats (*j*) are performed for each condition (*i*) to account for cell-to-cell variability, with only successful experiments analyzed in this work. Repeats with premature terminations are discarded, leading to some missing indexes in Table II. Figure 2c illustrates the 'Config1' aging cycling protocol, which includes a C/5 CC charge followed by a 1C CC discharge. Similarly, small pulses are superimposed at 20%, 50%, and 80% SoC during discharge to non-invasively track R0 and relaxation voltage throughout battery lifetime. During data collection for the RPT and aging cycles, the sampling time is 60 seconds when the current magnitude is constant and 1 second during current changes. Specifically, data are sampled every second for 5 seconds before and 10 seconds after each current change to accurately capture voltage dynamics and calculate R0.

Battery Performance Metrics

SoC is defined as the amount of charge currently available in the cell relative to its nominal capacity, with the charge calculated through coulomb counting. Specifically, we have

Table I. LMB cell specifications.

Cell Attribute	Description
Cathode	NCA
Anode	Li metal
Form Factor	Coin cells
Nominal Cell Voltage	3.8 V
Min/Max Voltage	3/4.3 V
Nominal capacity, Q_0	3.3 mAh

Table II. Charge and discharge C-rates for each aging cycling protocol are detailed as follows: *i* represents the aging cycling protocol index, and *j* denotes the repeated test index under the same protocol.

Config # (*i*)	Charge C-rate	Discharge C-rate	T21 Test # (*j*)	T40 Test # (*j*)
1	C/5	1C	1,2,3,4,5,6	1,2,3,4
2	C/5	3C	1,2,4,5,6,7	1,2,3,4
3	C/2	1C	1,2,3,4,5	1,2,3,4
4	C/2	3C	1,2,3,4,5	2,5

$$\mathrm{SoC}(t) = \begin{cases} \mathrm{SoC}(n,t) = 100\% + \dfrac{\int_{t_{0,dchg}^{(n)}}^{t} I_{batt}(\tau)d\tau}{3600 \cdot Q_{dchg}^{(n)}} \times 100\%, & \text{discharge} \\ \mathrm{SoC}(n,t) = \dfrac{\int_{t_{0,chg}^{(n)}}^{t} I_{batt}(\tau)d\tau}{3600 \cdot Q_{chg}^{(n)}} \times 100\%, & \text{charge} \end{cases} \quad [1]$$

Here, the applied current I_{batt} is positive for charging and negative for discharging. The coefficient 3600 in the denominator converts the integrated capacity from mA-seconds to mAh. The index n denotes the n-th raw aging cycle. $Q_{dchg}^{(n)}$ and $Q_{chg}^{(n)}$ represent the discharge and charge capacity values at the n-th cycle, measured in mAh:

$$Q_{dchg}^{(n)} = -\frac{1}{3600} \int_{t_{0,dchg}^{(n)}}^{t_{f,dchg}^{(n)}} I_{batt}(\tau)d\tau \quad [2]$$

$$Q_{chg}^{(n)} = \frac{1}{3600} \int_{t_{0,chg}^{(n)}}^{t_{f,chg}^{(n)}} I_{batt}(\tau)d\tau \quad [3]$$

where $t_{0,dchg}^{(n)}$ and $t_{f,dchg}^{(n)}$ are the beginning and ending time of the discharge phase at the n-th raw aging cycle, and $t_{0,chg}^{(n)}$ and $t_{f,chg}^{(n)}$ are

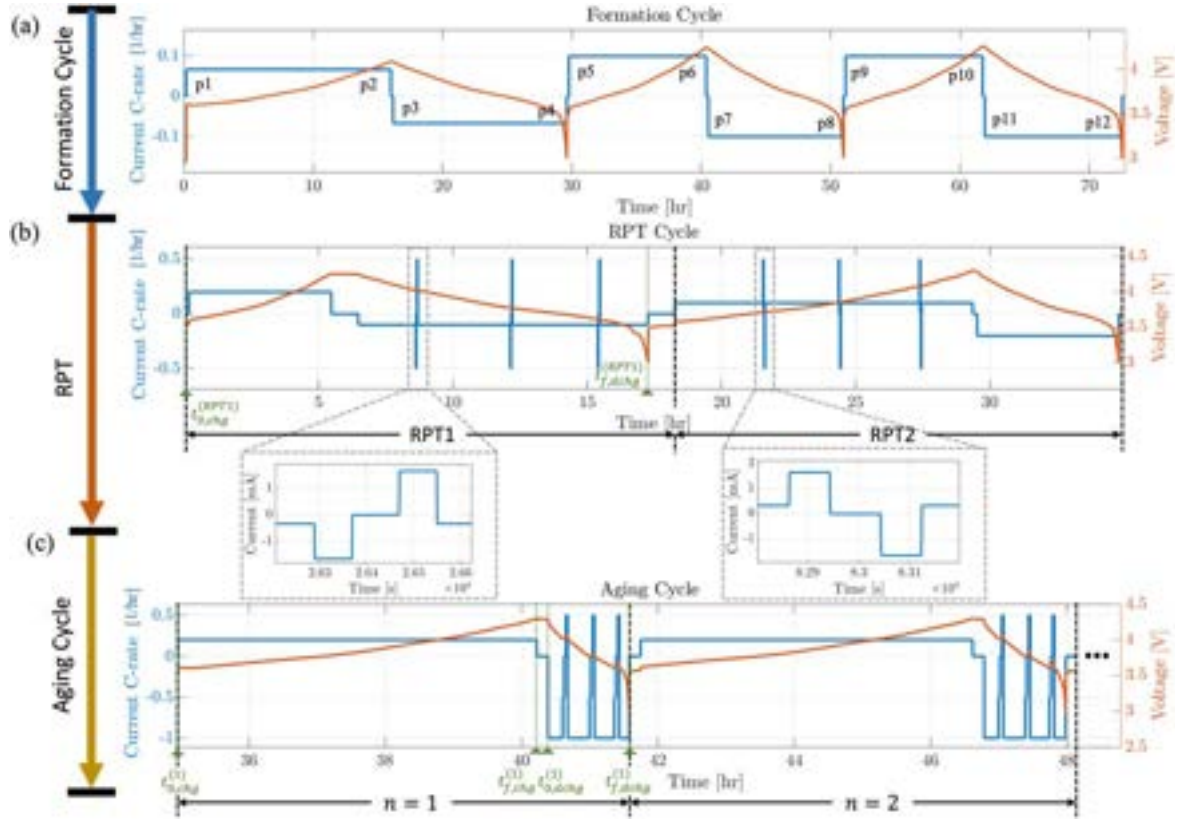

Figure 2. Testing protocol description. Current values are positive for charging and negative for discharging. (a) Formation cycle protocol. (b) RPT cycle protocol executed at the beginning of life. (c) Example of an aging cycling protocol, with 0.2C charging and 1C discharging.

the beginning and ending time of the charge phase in the n-th raw aging cycle, respectively.

The SoH of a cell at aging cycle n is defined as the discharge capacity with respect to discharge capacity of aging cycle 2[20]:

$$\text{SoH}^{(n)} = \frac{Q_{dchg}^{(n)}}{Q_{dchg}^{(2)}} \quad [4]$$

In this work, we consider the battery to reach its end-of-life when SoH drops below 0.8:

$$n_{end} = \min\{n|\text{SoH}^{(n)} \leqslant 0.8\} \quad [5]$$

We select the discharge capacity of the second aging cycle $Q_{dchg}^{(2)}$ as a normalizer in Eq (4), rather than that of the first aging cycle $Q_{dchg}^{(1)}$, because $Q_{dchg}^{(1)}$ defined in Eq (2) tends to be an outlier, which is not related to degradation but rather the cell stabilization process. The proposed definition of SoH assumes $\text{SoH}^{(2)} = 100\%$ for all cells.

For any cells at a given raw aging cycle n, the battery ampere-hour throughput is defined as the cumulative charge in and out of the cell from the beginning of the RPT cycle ($t_{0,chg}^{(RPT1)}$) until the end of cycle n:

$$\text{Ah} - \text{Throughput}^{(n)} = \frac{1}{3600}\int_{t_{0,chg}^{(RPT1)}}^{t_{f,dchg}^{(n)}} |I_{batt}(\tau)|d\tau \quad [6]$$

The equivalent full cycle (EFC) at raw aging cycle n is defined as the cell's ampere-hour throughput at raw aging cycle n normalized by $2Q_0$, where Q_0 is the nominal capacity:

$$\text{EFC}^{(n)} = \frac{\text{Ah} - \text{Throughput}^{(n)}}{2Q_0} \quad [7]$$

Figure 3 shows the SoH trajectories of all 36 cells in this dataset. The SoH trajectories of Sakuu cells reported in this paper are not representative of cycle life performance in the latest cell batches from Sakuu Corp.

Cell initial capacity is calculated from the C/10 discharge cycle in RPT1 using Eq (2). Cell internal resistance R0 is defined as the ratio of voltage change upon a step change of applied current over the magnitude of current step change,[21] as shown in Fig. 4:

$$R0(t) = \frac{V(t+\Delta t) - V(t)}{I_{batt}(t+\Delta t) - I_{batt}(t)} \quad [8]$$

where $V(t)$ is the measured cell terminal voltage at time t, $I_{batt}(t) \neq 0$ is the current at time t, and Δt is the time step over which resistance is calculated. Figure 5 provides the values of cell initial capacity $Q_{dchg}^{(RPT1)}$ and initial internal resistance $R_{dchg}^{(RPT1)}$ in the dataset. Cell initial internal resistance is calculated using Eq. (8) upon the removal of C/10 discharge current in RPT1 at 0% SoC:

$$R_{dchg}^{(RPT1)} = R_0(t_{f,dchg}^{(RPT1)}) \quad [9]$$

with $\Delta t = 5$ s.

All cells exhibit consistent initial capacity ($Q_{dchg}^{(RPT1)}$) except for 3 outliers which correspond to the 3 cells that did not go through the formation cycle. In contrast to $Q_{dchg}^{(RPT1)}$, a large spread of initial resistance is observed, which could be attributed to variations in the manufacturing process.

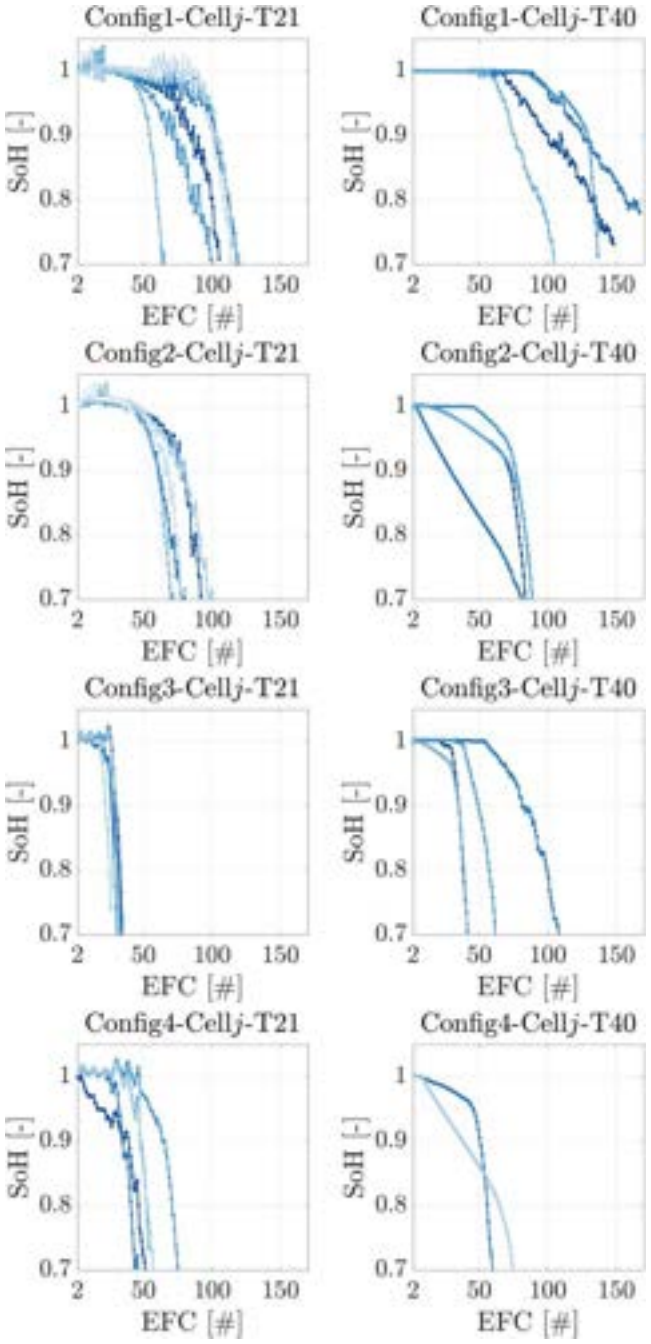

Figure 3. The SoH trajectories throughout the EFC progression are presented for each test condition, with each curve representing a repeated test. The SoH for raw aging cycle 1 is excluded from the figure because it is an outlier compared to the other data points.

Feature Engineering

In this section, we first explain the data pre-processing procedure. Thereafter, we construct a comprehensive inventory of features based on LMB voltage and current measurements. The features defined here are, however, universally applicable across batteries but could indicate different degradation mechanisms for cells beyond the scope of this work. Subsequently, we propose two feature selection algorithms that pick a subset of most aging-sensitive features from the feature inventory, specific to the training cells.

Data pre-processing.—The formation cycle data for the T21 Tests and T40 Tests were originally collected using different sampling intervals—30 seconds and 60 seconds, respectively. To maintain consistency in the extracted features, we resampled the formation cycle voltage and current signals to a uniform 60 second interval using interpolation.

We divided the dataset into two sets for model training and testing. Due to significant cell-to-cell variability, the selection of test cells plays a crucial role in feature selection, model development, and validation. For the remainder of this work, we selected 8 test cells: Config1–Cell6–T21, Config2–Cell6–T21, Config3–Cell4–T21, Config4–Cell3–T21, Config1–Cell2–T40, Config2–Cell3–T40, Config3–Cell3–T40, and Config4–Cell2–T40. These cells represent all the temperature and charge/discharge C-rate combinations in the dataset, and their SoH trajectories, as shown in Fig. 3, are similar to most other cells operating under the same aging cycling protocol. The remaining 28 cells form the training set for health estimation.

Health estimation features.—We extract three groups of features from aging cycle data for health estimation: internal resistance (R0) features, relaxation voltage (RV) features, and charging impedance (Z) features. These features are particularly valuable due to their combined benefits of interpretability and accessibility. For instance, tracking aging through R0 only requires a pulse, while tracking aging through capacity requires constant-current discharge, which might take several hours. RV features provide insights into complex, nonlinear side reactions throughout the cycle, which are not observable from fixed currents alone. Z features are extracted using a short moving window lasting a few seconds over standard charging protocols. All features require only brief operational bursts around targeted state-of-charge, significantly reducing the need for extended diagnostic tests like constant-current discharge characterization. Feature values vary across aging cycles in each cell and may reflect the cell degradation process. Additionally, three experimental-condition-specific features—temperature, charge C-rate, and discharge C-rate—are considered as cycle-invariant health estimation features.

Internal resistance features.—Within each aging cycle, R0 features are calculated from 16 current step changes following Eq. (8), as shown in Fig. 6. These features are denoted following the notation R0_cyc_pi_{R0}_j_{R0}s, where $i_{R0} \in \{1, \cdots, 16\}$ represents the step change index in the cycle, and $j_{R0} \in \{1, 3, 5\}$ indicates the time step in seconds for resistance calculation. This results in 48 R0 features per aging cycle per cell, allowing for the comparison of R0 feature importance at discrete SoC points and various time resolutions. The extraction of R0 features is usually conducted from the hybrid pulse power characterization (HPPC) test.[21] While HPPC tests typically use current pulses after a rest period to reduce polarization effects and isolate ohmic losses, our approach simplifies the process by inserting pulses directly into the aging cycle discharge profile without rest periods. As a result, the extracted R0 includes multiple polarization mechanisms along with ohmic loss.[22] Changes in R0 over aging are considered indicative of diffusion processes across the dead Li and SEI layer.[23] Figure 7 shows the variation trend for the R0_cyc_p14 feature. After around EFC 15, R0 increases with aging, as commonly observed in LIBs. However, before EFC 15, R0 first increases and then decreases, a trend not associated with battery degradation. This behavior may be linked to LMB anode surface stabilization through cycling.[24] To better isolate the correlation between R0 and capacity fade, all feature values before EFC 15 are excluded from health estimation model training.

Relaxation voltage features.—Relaxation voltage (RV) is defined as the time series of voltage measurements when the applied current $I_{batt} = 0$. For fresh LMBs, the open circuit voltage is primarily determined by the lithium concentration in the cathode, with the redox potential of the Li metal anode assumed to be zero.[10] As LMB cells age, factors such as SEI formation, dead lithium, and dendrite growth[25] alter the open circuit potential of the Li metal anode. In this

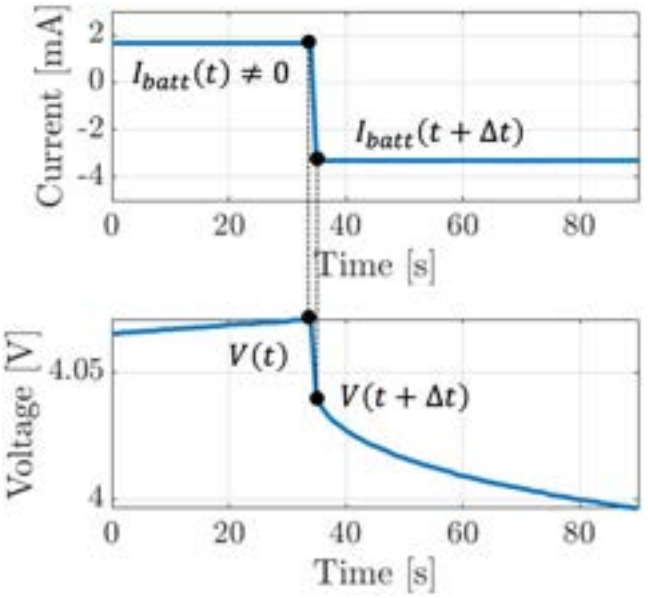

Figure 4. Calculating internal resistance R0 using Eq. (8) from voltage change given a current magnitude step change.

study, we assume that changes in the Li metal anode's open circuit potential with aging are negligible. Under this assumption, the transient dynamics of RV during charge or discharge reflect the disappearance of concentration gradients in the cathode and electrolyte.[23] In the LMB dataset, each aging cycle includes 5 rest periods, as shown in Fig. 6. During each rest period, the transient voltage response is characterized by 11 features, detailed in Table III, which quantify statistical attributes of the relaxation voltage curve. Thus, a total of 55 RV features can be extracted per aging cycle per cell, covering the entire SoC range from 0% to 100%. Figure 8 shows the relaxation voltage curves for cell Config2-Cell2-T40 throughout aging after charging to 100% SoC (RV1 in Fig. 6). The curves display a downward shift and an increasing relaxation time constant over successive cycles, indicating impeded diffusion.[25] While a decrease in steady-state relaxation voltage after charging is typically attributed to electrolyte depletion,[25] the absence of CV charging following CC charging in this LMB dataset suggests that increasing charging impedance might also contribute. Thus, the observed decrease in steady-state relaxation voltage likely reflects both lithium loss and internal resistance growth.

Charging impedance features.—For a given aging cycle n, the battery charging impedance is defined as the ratio of battery voltage increase over a moving time window of length Δt to the constant charge current:[26]

$$\begin{aligned} Z^{(n)}(t) &= \frac{V(t+\Delta t) - V(t)}{I_{batt}^{(n)}} \\ &= \frac{\Delta V}{\Delta Q}\Delta t \\ &= DV\,\Delta t \end{aligned} \qquad [10]$$

$Z(t) > 0$ since the current I_{batt} is positive and $V(t + \Delta t) > V(t)$ during charging. Eq. (10) links charging impedance with the differential voltage (DV) curve using the scaling time coefficient Δt. DV analysis is a battery characterization technique that examines degradation by analyzing the differential of terminal voltage with respect to charged capacity. Voltage plateaus on the $V - Q$ (voltage-capacity) plane are transformed into valleys in the $\Delta V/\Delta Q$ - Q plane, which are in turn correlated with cell degradation.[27]

Assuming a constant anode potential, variations in the $\Delta V/\Delta Q$ curve are attributed to cathode and electrolyte degradation. Since $\Delta V/\Delta Q$ analysis requires equilibrium, a low charging C-rate is necessary. In this work, LMB cells are charged at C/5 or C/2, which introduces polarization effects in terminal voltage.[28] The charging impedance Z aggregates these electrochemical effects into a single indicator of battery health. The time window length Δt influences the smoothness of charging impedance curves. A longer window reduces noise from current measurements but may lessen the detail

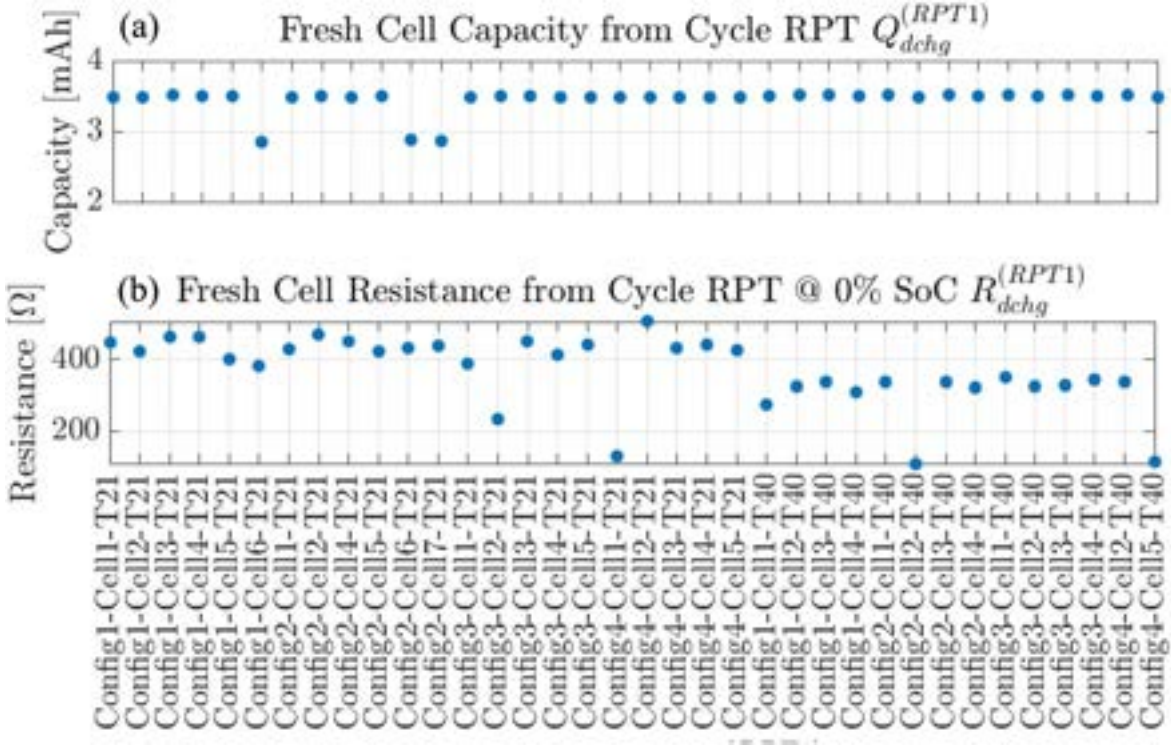

Figure 5. Distribution of (a) fresh cell capacity $Q_{dchg}^{(RPT1)}$ and (b) fresh cell resistance $R_{dchg}^{(RPT1)}$.

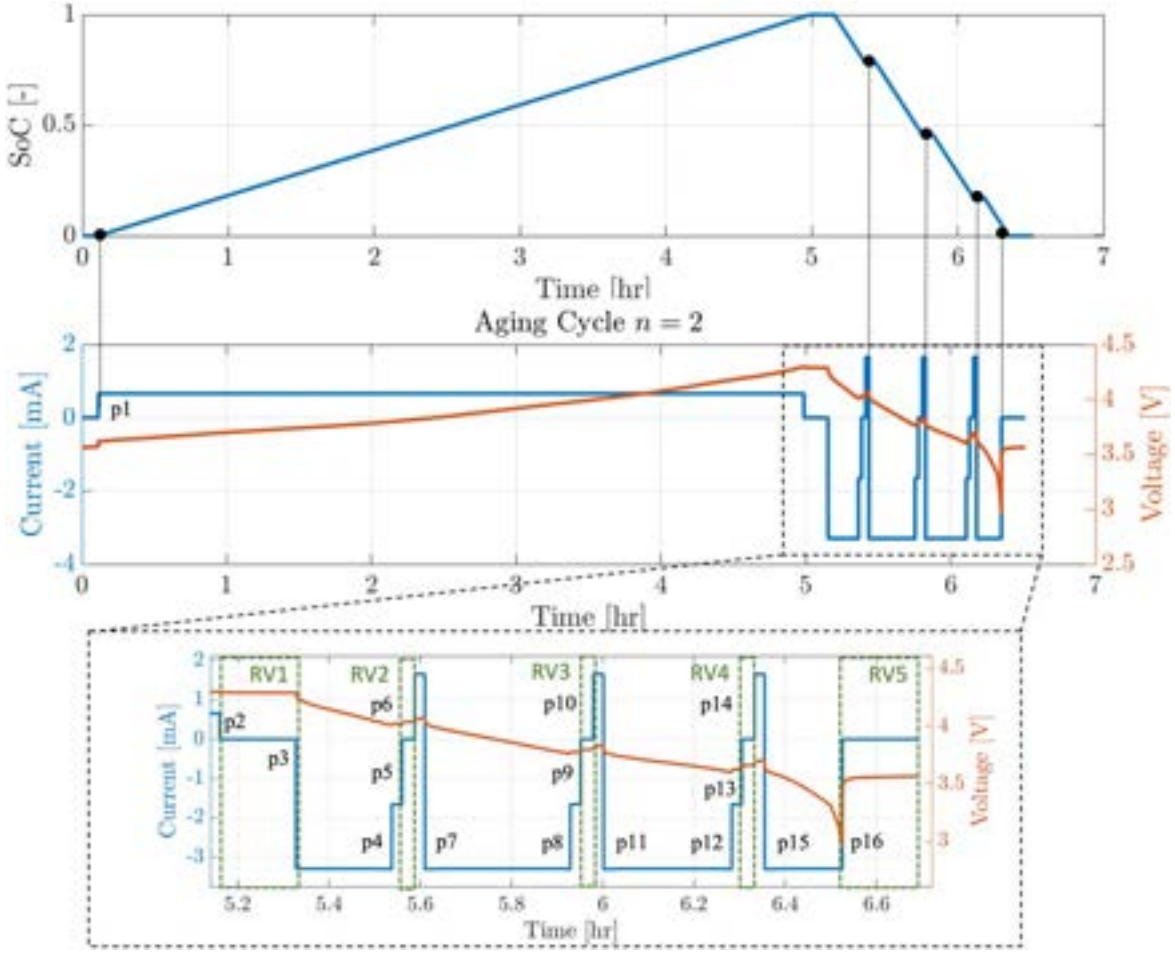

Figure 6. The SoC, current, and voltage signals of the aging cycle are analyzed. R0 features are extracted at the current step changes during the discharge phase, while the signals during the five rest periods (RV1...RV5) are used to extract RV features within each aging cycle.

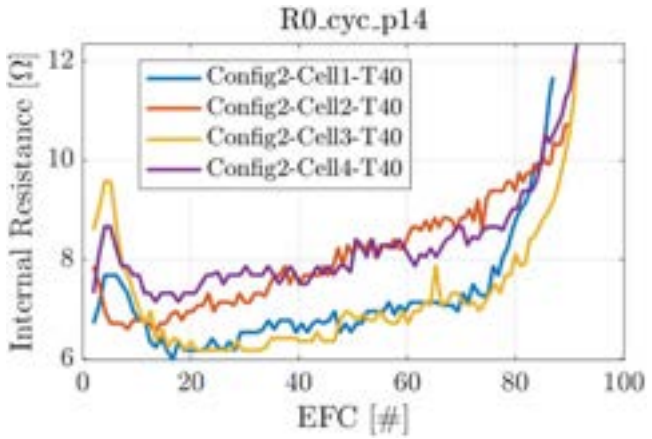

Figure 7. Example of trends of variation for R0 features extracted at current step change p14. Test cell: Config2-Cell$j$-T40, $j \in \{1, \cdots, 4\}$.

in the impedance profile. We use $\Delta t = 120$ s for Config1 and Config2 (C/5 charge) and $\Delta t = 60$ s for Config3 and Config4 (C/2 charge), ensuring that four valleys, i.e. local minima in Z, are consistently observed as in previous studies on DV curves for NCA cathodes.[29] Figure 9(a) shows an example of Z vs charge capacity Q_{chg}. Each valley on the curve is described by six distinct features, as detailed in Table IV.

Specifically, a valley's prominence reflects its distinctiveness from the surroundings, and is defined as the least increase in height necessary to get from the valley's lowest point to any lower points on the curve.[30] A detailed descritpion of features extracted on the charging impedance profile is provided in the caption of Fig. 9. Analyzing Z vs SoC, rather than Z vs Q_{chg}, avoids the issue of horizontal shifts in valleys caused by charge capacity fade. The plots in Fig. 9(b) show Z vs Q_{chg} and Z vs SoC over aging. As cycling progresses, Valley 1 rises, while Valleys 2, 3, and 4 shift toward the upper-left corner.

Feature normalization using fresh cell properties.—Feature normalization is crucial in data-driven model construction to reduce the effects of cell-to-cell manufacturing variations and enhance model performance. It also aids in training convergence by standardizing feature value ranges. We normalize single-cycle features for any raw aging cycle n based on the cells' starting conditions, specifically using the feature values from raw aging cycle 2:

$$\tilde{x}_i^{(n)} = \frac{x_i^{(n)} - x_i^{(2)}}{x_i^{(2)}} \qquad [11]$$

where x_i is the i-th feature, $i \in \{1, \cdots, 127\}$.

Feature selection.—Following the feature extraction methodology, we obtain a total of 127 single-cycle features for health estimation. Given the high-dimensional input space, only a subset of features is expected to show consistent associations with cell degradation. Therefore, an automatic feature selection scheme is necessary to reduce model overfitting, eliminate misleading features, and decrease the computational burden for subsequent models. The goal of feature

Table III. RV features are defined for the five rest periods (RV1, ...RV5) shown in green windows in Fig. 6. The timestamp t corresponds to a given point on the relaxation voltage curve, with t_0 indicating the start time of the rest period. The derivative of the relaxation voltage is given by $dV(t) = \frac{V(t+\Delta t) - V(t)}{\Delta t}$, where Δt is the sampling time resolution, and $\overline{V}$ and σ_V represent the mean and standard deviation of the voltage time series during the rest period, respectively: $\overline{V} = \frac{1}{T}\sum_{t=1}^{T} V(t)$, $\sigma_V = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(V(t) - \overline{V})^2}$.

Feature Name	Definition (over rest period RVi)	Mathematical Expression
min_RVi	Minimum voltage	$\min_t\{V(t)\}$
max_RVi	Maximum voltage	$\max_t\{V(t)\}$
range_RVi	Range of voltage	$\max_t\{V(t)\} - \min_t\{V(t)\}$
skew_RVi	Skewness	$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{V(t)-\overline{V}}{\sigma_V}\right)^3$
kurt_RVi	Kurtosis	$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{V(t)-\overline{V}}{\sigma_V}\right)^4$
tau_RVi	Relaxation time constant	$t_d - t_0$, where $V(t_d) = V(t_0) + 63\% \times [V(t) - V(t_0)]$
min_dRVi	Minimum voltage derivative	$\min_t\{dV(t)\}$
max_dRVi	Maximum voltage derivative	$\max_t\{dV(t)\}$
raneg_dRVi	Voltage derivative range	$\max_t\{dV(t)\} - \min_t\{dV(t)\}$
skew_dRVi	Skewness of voltage derivative	$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{dV(t)-\overline{dV}}{\sigma_{dV}}\right)^3$
kurt_dRVi	Kurtosis of voltage derivative	$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{dV(t)-\overline{dV}}{\sigma_{dV}}\right)^4$

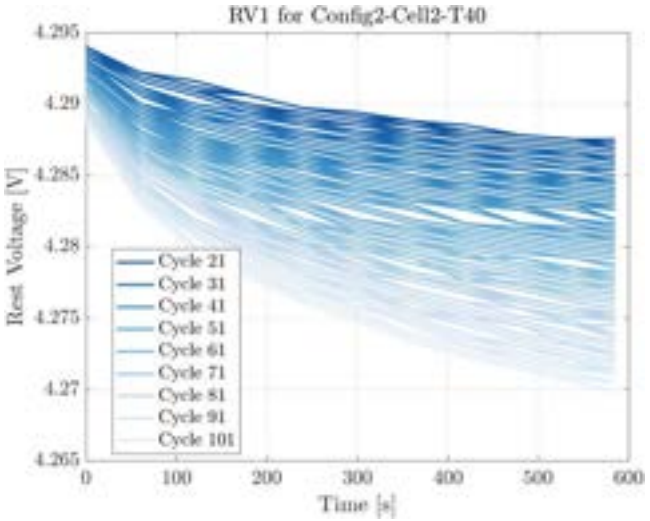

Figure 8. Example of relaxation voltage curve variation after EFC 15. Test cell: Config2-Cell2-T40.

selection is to identify an optimal set of features that best represent the target output. To assess the similarity among features and between features and target outputs, we use the Spearman correlation coefficient:[32]

$$\rho(x, y) = 1 - \frac{6\sum_{j=1}^{m} d_j(x, y)^2}{m(m^2 - 1)} \quad [12]$$

where $d_j(x, y)$ denotes the difference between the rank of variable x and y for the j-th sample in the dataset, and m is the number of samples. This metric measures how well the relationship between two variables x and y can be described using a monotonic function, which is not limited to a linear function. The Spearman correlation coefficient ranges from $[-1,1]$, where $+1$ indicates a perfect monotonically increasing relationship, -1 represents a perfect monotonically decreasing relationship, and 0 implies no apparent correlation between the variables. Specifically, consider selecting features from an initial feature set $\mathcal{X}_0 = \{x_1, \cdots, x_d\}$ with d features for the purpose of estimating variable y, according a training dataset $\mathcal{D} = \{(\mathbf{x}^{(1)}, y^{(1)}), \cdots, (\mathbf{x}^{(m)}, y^{(m)})\}$ with m samples. The general feature selection algorithm $\mathcal{X} = f_{select}(\mathcal{X}_0, \mathcal{D}, \alpha, \beta)$ is explained in Algorithm 1.

Algorithm 1. Feature selection algorithm.

Input $\mathcal{X}_0 \neq \varnothing$, $\mathcal{D} \neq \varnothing$, $0 \leqslant \alpha, \beta \leqslant 1$
Output $\mathcal{X}$

Phase 1 - Select features highly correlative with y into set $\mathcal{U}$:
$\mathcal{U} = f^{\alpha}_{select}(\mathcal{X}_0, \mathcal{D})$
Initialize $\mathcal{U} \leftarrow \varnothing$
for $x_i \in \mathcal{X}_0$
  if $|\rho(x_i, y)|_{\mathcal{D}}| > \alpha$
    Add x_i to $\mathcal{U}$
  end if
end for

Phase 2 - Group features highly correlative with other features in set $\mathcal{U}$ into set $\mathcal{C}$: $\mathcal{C} = f^{\beta}_{select}(\mathcal{U}, \mathcal{D})$
Initialize the set of selected features $\mathcal{X} \leftarrow \varnothing$
while $\mathcal{U} \neq \varnothing$
  $x_j \leftarrow$ pop($\mathcal{U}$)
  Initialize the set of correlated features $\mathcal{C} \leftarrow \{x_j\}$
  for $x_i \in \mathcal{U}$
    if $|\rho(x_i, x_j)|_{\mathcal{D}}| > \beta$
      Add x_i to $\mathcal{C}$
    end if
  end for
  Initialize *highest_correlation* $\leftarrow 0$
  for $x_i \in \mathcal{C}$
    if $|\rho(x_i, y)|_{\mathcal{D}} >$ *highest_correlation* **then**
      highest_correlation $\leftarrow |\rho(x_i, y)|_{\mathcal{D}}|$
      $x_{best} \leftarrow x_i$
    end if
    Remove x_i from $\mathcal{U}$ ▷ Remove visited features from set $\mathcal{U}$
  end for
  Add x_{best} to $\mathcal{X}$ ▷Pick the most informative features for y from set $\mathcal{C}$ and move to set $\mathcal{X}$
end while

A key consideration in feature selection is addressing inter-cell variability in feature correlations. Therefore, this work explores two schemes for merging cell-specific feature spaces, as detailed below.

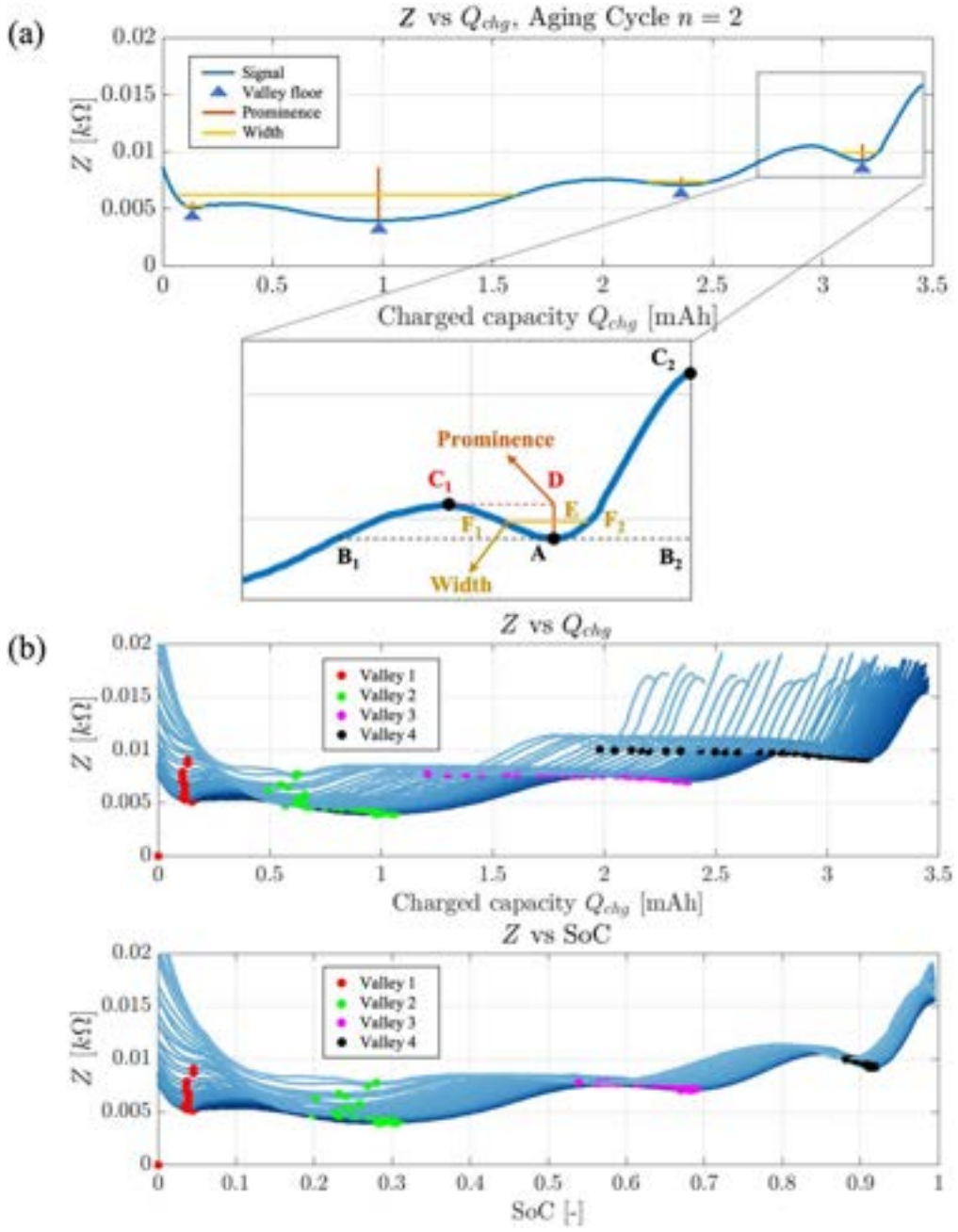

Figure 9. a. Extracted Z features from each valley of charging impedance. b. Example of charge impedance curves starting from EFC 15. Test cell: Config2-Cell1-T40. To calculate the prominence of Valley 4, first of all, one has to locate the lowest point of that valley, i.e. point A defined as the valley floor. Next, we draw a horizontal black dashed line B_1B_2 from the valley floor A toward both sides until it either crosses the curve because there is a lower valley (the case for point B_1), or it reaches the left or right end of the curve (the case for point B_2). Two search intervals AB_1 and AB_2 are defined, one on each side of the valley. The maximum of the curve in each of the two intervals is denoted as C_1 and C_2, respectively. Now, we extend a horizontal red dashed line from the lower of the two interval maxima, in this case, point C_1. Extend an orange vertical line from the valley floor A upward until it crosses the red dashed line at point D. The prominence of Valley 4 is the length of line AD, i.e. the distance along the y axis from the black dashed line B_1B_2 to the red dashed line C_1D. Locate the middle point of line AD as E, i.e. length of AE equals length of DE. Extend a yellow horizontal line F_1F_2 from point E toward both sides until it crosses the curve at F_1 and F_2. The width of Valley 4 is the length of line F_1F_2. The prominence and width of valleys are extracted using MATLAB function `findpeaks`[31].

- **Intersection.** Execute Phase 1 in Algorithm 1 for each training cell, and intersect the obtained z feature sets first to obtain an intermediate feature set $\mathcal{U}$:

$$\begin{aligned} \mathcal{X}^{(k)} &= f^{\alpha}_{select}(\mathcal{X}^{(k)}_0, \mathcal{D}^{(k)}), \quad k = 1, \cdots, z \\ \mathcal{U} &= \mathcal{X}^{(1)} \cap \cdots \cap \mathcal{X}^{(N)} \end{aligned}$$

Then execute Phase 2 in Algorithm 1 over the intersected feature set:

$$\mathcal{X} = f^{\beta}_{select}(\mathcal{U}, \mathcal{D})$$

This scheme includes features that exhibit consistent correlations across all cells, thereby reducing inter-cell variability in the ensemble.

Table IV. Definition of charging impedance features for each valley i_Z, where $i_Z \in \{1, 2, 3, 4\}$.

Feature Name	Definition	Unit
pks_i_Z_Z	Value of Z at the valley floor i_Z	$V \cdot \text{mAh}^{-1}$
loc_i_Z_Q	Location of the valley floor i_Z represented by charged capacity	mAh
loc_i_Z_SoC	Location of the valley floor i_Z represented by SoC	—
prom_i_Z_Z	Prominence of valley i_Z	$V \cdot \text{mAh}^{-1}$
width_i_Z_Q	Width of valley i_Z represented by charged capacity	mAh
width_i_Z_SoC	Width of valley i_Z represented by SoC	—

• **Concatenation.** Concatenate each feature across all cells first to create an aggregate dataset, and then perform feature selection:

$$\mathcal{D} = \{\mathcal{D}^{(1)}, \cdots, \mathcal{D}^{(z)}\}$$
$$\mathcal{X} = f_{select}(\mathcal{X}_0, \mathcal{D}, \alpha, \beta)$$

This scheme treats all cells identically and is straightforward to implement. The three experimental-condition-specific features—temperature, charge C-rate, and discharge C-rate—are deemed significant and independent of the other features. These features are not included in the feature selection process and are directly used as inputs in model training and validation.

Model Training

This section outlines the model training procedures for LMB health estimation. The Bagged-GPR model used in this work provides unique advantages by efficiently extracting aging patterns from a modest-sized battery dataset while quantifying uncertainty. Hyperparameter optimization, combined with k-fold cross-validation, is employed to tune both the feature selection algorithm and the Bagged-GPR model for optimal performance.

Bagged gaussian process regression.—This work adopts the Bagged-GPR model in the same spirit as in Takahashi et al.[33] GPR is a non-parametric Bayesian probabilistic algorithm to learn input-output mapping using a Gaussian process model:[34]

$$P(y|f(\mathbf{x}, \mathbf{x}))\sim N(h(\mathbf{x}) + f(\mathbf{x}), \sigma^2) \quad [13]$$

where $\mathbf{x}$ is the input vector and y is the output. h: $\mathbb{R}^d \rightarrow \mathbb{R}$ is a deterministic basis function, where d is the dimension of $\mathbf{x}$, σ is the noise variance. In this work, we select the basis function to be constant: $h(\mathbf{x}) = 1$. $f(\mathbf{x})\sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ is a latent random variable of Gaussian process, with $m(\mathbf{x})$ being the mean function, and $k(\mathbf{x}, \mathbf{x}')$ being the covariance function:

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \quad [14]$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x} - m(\mathbf{x}))][(f(\mathbf{x}' - m(\mathbf{x}'))] \quad [15]$$

GPR assumes that the latent variable ($\mathbf{x}$) for all observed samples in the training set form a joint Gaussian distribution. When given the inputs of a new sample $\mathbf{x}$, the model can therefore estimate the distribution of the output y based on previously observed samples.

In this work, we select a zero mean function $m(\mathbf{x}) = 0$ and the Matern 5/2 covariance function k_f[35] to capture typical smoothness of physical processes due to its finite differentiability:

$$k_f(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\mathbf{x} - \mathbf{x}'}{\sigma_l} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{\mathbf{x} - \mathbf{x}'}{\sigma_l} \right) \quad [16]$$

where Γ is the gamma function, K_ν is the modified Bessel function, σ_f is the signal variance of the input variables, σ_l is the length scale, and $\nu = 5/2$ is a smoothness parameter. When fitting to data, the hyper-parameters to be optimized in GPR include basis function coefficients β, the noise variance σ^2, the signal variance σ_f^2, and the length scale σ_l. We utilize `fitrgp` function in MATLAB[36] for optimizing these hyper-parameters by maximizing the log marginal likelihood function.

GPR models are widely used in prognostics[37] due to their non-parametric and Bayesian nature. These models do not assume a specific functional form for the underlying dynamics, allowing them to learn complex patterns from limited data. This flexibility is particularly advantageous for LMB health estimation, where degradation behaviors are complex and nonlinear, and existing datasets typically include hundreds to thousands of samples due to limitations in test infrastructure and testing timespan. Simple machine learning models, such as elastic net regression (ENR) and support vector regression (SVR), struggle to capture LMB degradation behaviors effectively because of their limited representational capacity, while neural networks often suffer from overfitting due to their large number of parameters. A key advantage of GPR in battery health estimation is its ability to quantify uncertainty in predictions, which provides insights into prediction quality and addresses cell-to-cell variation. The Bayesian framework of GPR helps prevent overfitting and enhances interpolation performance by incorporating domain knowledge about the smoothness of LMB aging processes through the choice of covariance function. A major drawback of GPR models is their lack of scalability, with computational and storage requirements scaling as $O(m^3)$ and $O(m^2)$ respectively, where m is the number of training samples. In this work, with a training set of $m =$ 1893 samples for health estimation, the computational cost for model fitting becomes prohibitive. To address this issue, the bagging technique is employed with GPR models. Specifically, p GPR models are used as meta-estimators, each trained on a subset of q samples, known as bags, which are randomly sampled with replacement from the original training set. For prediction, the final output is obtained as the weighted average of the predictions from each GPR model:

$$\hat{y} = \frac{\sum_{a=1}^{p} w_a \hat{y}_a}{\sum_{a=1}^{p} w_a} \quad [17]$$

$$\hat{\sigma_y}^2 = \frac{b \sum_{a=1}^{p} w_a (\hat{y}_a - \hat{y})^2}{(b-1) \sum_{a=1}^{p} w_a} \quad [18]$$

where $\hat{y}$ is the aggregated prediction from Bagged GPR, y_a is the prediction made by the a-th model out of total p GPR models, $\hat{\sigma_y}^2$ is the predicted variation of $\hat{y}$, and b is the number of nonzero weights. w_a is the weight of the a-th GPR model defined by:

$$w_a = \frac{1}{\sigma_a} \quad [19]$$

where σ_a is the predicted standard deviation of the a-th model. The weighted prediction thus favors models with higher certainty within the ensemble. Bagging aggregates predictions from multiple models, which helps mitigate overfitting and enhances model stability with unseen inputs. Additionally, bagging enables parallel computing due to the independent nature of each bagged model. For comparison, GPR without bagging is used as a baseline model to highlight the performance improvements achieved with the Bagged-GPR approach.

Hyper-parameter optimization.—Hyper-parameter optimization is essential for enhancing machine learning model performance on unseen samples, assuming similar distributions to the training data. In the feature selection algorithm, α sets the Spearman coefficient threshold for selecting features correlated with the target variable, with higher values requiring stronger correlation. β sets the threshold

Table V. Range and grid size of the hyper-parameter search space.

Hyper-parameter	Range	Grid size
α	[0.5,0.9]	0.1
β	[0.5,0.9]	0.1
p	[10,100]	10
q	[25,400]	25

for feature independence, with lower values enforcing stricter separation. Higher α and lower β generally lead to a smaller feature set $\mathcal{X}$. For each α and β combination, we optimize the Bagged-GPR model's p (number of bags) and q (samples per bag). Increasing p improves prediction accuracy over outlier cells, while increasing q enhances representativeness over average cells, though both increase computational time. The hyper-parameter ranges and grids are detailed in Table V, based on the training set of $m = 1893$ samples.

We use stratified 5-fold cross-validation based on different aging cycling conditions to avoid overfitting hyper-parameters to the training set. The objective for optimization is the average mean percentage error (MPE) across all validation cells:

$$\overline{MPE}^{val} = \frac{1}{z^{val}} \sum_{k=1}^{z^{val}} MPE(\mathbf{y}^{(k)}, \hat{\mathbf{y}}^{(k)}) \quad [20]$$

$$z^{val} = \sum_{h=1}^{5} z_h^{val}$$

$$MPE(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{m} \sum_{j=1}^{m} \frac{|\, y_j - \hat{y}_j \,|}{y_j} \times 100\% \quad [21]$$

where $\mathbf{y}^{(k)}$ and $\hat{\mathbf{y}}^{(k)}$ are ground truth and model predictions of output variable y for the k-th validation cell containing m_k samples. $m_k = n_{end} - n_{EFC15}$ where $n_{EFC15} = \min\{n|\mathrm{EFC}^{(n)} \geqslant 15\}$.$z^{val}$ is the total number of validation cells from 5 validation sets, with the h-th validation set containing z_h^{val} validation cells.

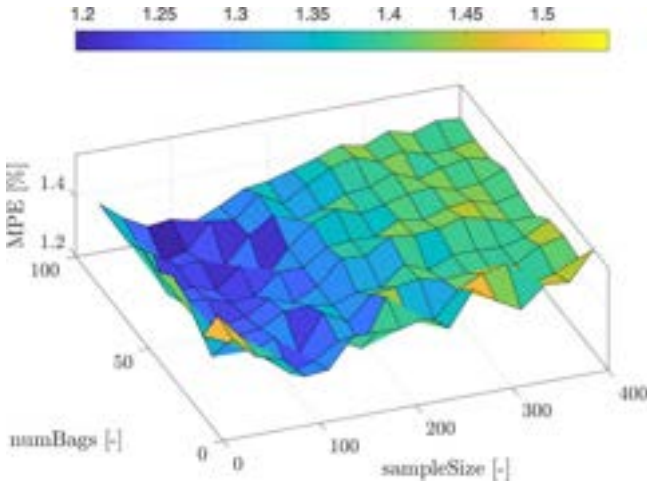

Figure 10. SoH estimation errors as a function of number of bags p (numBags) and samples per bag q (sampleSize), for feature set 0.5-0.8-Conc. The plotted error represents the average MPE across validation cells from 5 iterations.

Results and Analysis

Selected features.—Table VI presents five representative sets of selected features for health estimation, using the two proposed inter-cell feature merging schemes and various feature selection hyper-parameter values.

These feature sets, representing abundant, moderate, and minimal input features, are highlighted from all α and β combinations. Table VI shows that Scheme 2 (Intersection) is stricter than Scheme 1 (Concatenation) for the same α and β, leading to different features.

Table VI. Selected features for various choices of feature selection hyper-parameters α and β.

Scheme 1: Concatenation

α	β	R0	RV	Z	Total # of Features
0.5	0.8	—	range_RV2, skew_RV2, max_RV3, range_RV3, kurt_RV3, tau_RV3, max_RV4, max_dRV4, min_RV5, max_RV5, skew_RV5	pks_1_Z, loc_1_Q, pks_2_Z, pks_3_Z, loc_3_SOC, width_3_SOC, prom_3_Z, pks_4_Z, loc_4_Q, loc_4_SOC	21
0.6	0.8	—	range_RV2, range_RV3, kurt_RV3, max_RV4, max_RV5	pks_2_Z , pks_3_Z, loc_3_SOC, loc_4_Q	9
0.9	0.9	—	—	loc_4_Q,	1

Scheme 2: Intersection

α	β	R0	RV	Z	Total No. of Features
0.5	0.8	R0_cyc_p13_5s	—	loc_3_Q, pks_2_Z	3
0.6	0.8	—	range_RV4	pks_2_Z	2

Table VII. SoH estimation performance for each feature set with optimized p and q compared to the baseline.

Feature Set	Model	$RMSPE_{median}$	$RMSPE_{mean}$	MPE_{median}	MPE_{mean}
0.5-0.8-Inter	Bagged-GPR	2.54	2.39	1.37	1.42
	GPR	3.53	3.97	2.31	2.97
0.6-0.8-Inter	Bagged-GPR	2.31	2.89	1.60	1.96
	GPR	4.45	4.89	3.33	3.44
0.5-0.8-Conc	Bagged-GPR	1.40	1.49	0.75	0.78
	GPR	1.62	1.82	1.12	1.11
0.6-0.8-Conc	Bagged-GPR	1.85	1.79	0.87	1.04
	GPR	1.57	2.83	1.17	1.82
0.9-0.9-Conc	Bagged-GPR	1.41	1.86	1.12	1.30
	GPR	1.58	1.90	1.16	1.28

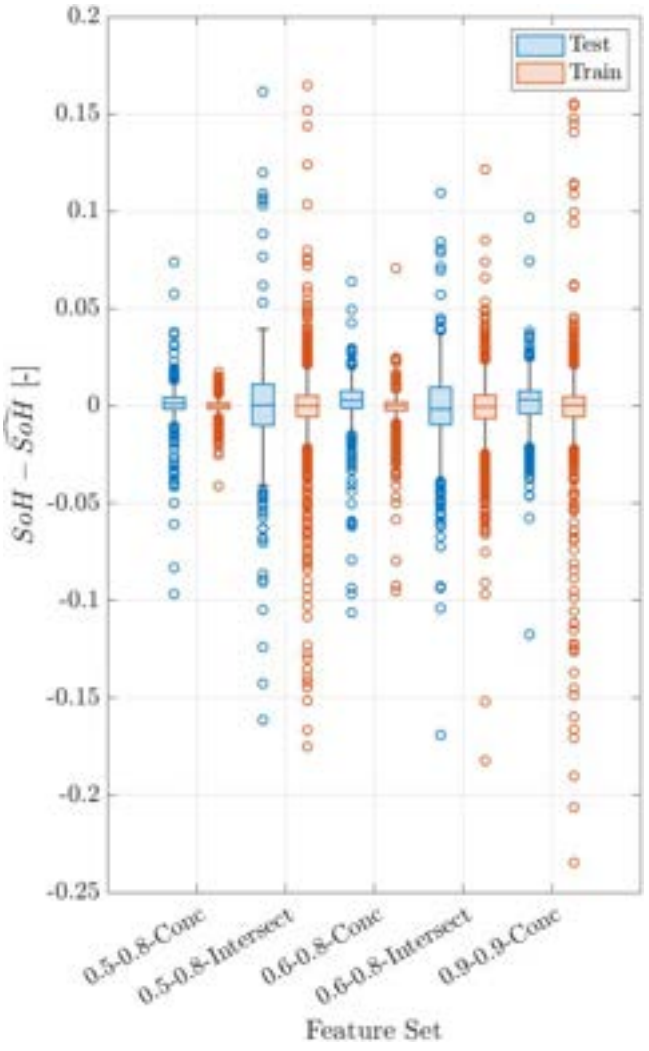

Figure 11. Box plots displaying the median, 25th and 75th percentiles, and outliers (points beyond the interquartile range) of SoH estimation error ($SoH - \widehat{SoH}$) for each feature set with optimized p and q.

Scheme 1 excludes R0 features but includes many RV features and mid- to high-SoC Z features. Scheme 2 selects one R0 feature at low-SoC and one RV feature at mid-SoC. Feature sets are referred to as α-β-scheme, where where $(\alpha, \beta) \in \{(0.5, 0.8), (0.6, 0.8), (0.9, 0.9)\}$ and scheme $\in \{\text{Intersect}, \text{Conc}\}$.

Effect of hyper-parameters.—Figure 10 shows the optimziation results from the number of bags p and samples per bag q in the Bagged-GPR model using feature set 0.5-0.8-Conc, with accuracy evaluated by $\overline{MPE}^{val}$ as defined in Eq (20). Accuracy improves slightly with more bags p but drops with more samples q per bag, likely due to overfitting with larger sample sizes. Using fewer samples per bag and a moderate number of bags mitigates overfitting. However, feature set 0.9-0.9-Conc, which has only one feature, shows better accuracy with more samples per bag. Overall, a moderate p value and a small q value offer significant computational savings, with training times under 30 seconds on a Quad-Core i5 CPU @ 2 GHz, reducing to less than 1 second with fewer features, a 20-fold improvement over the baseline.

Table VIII. Optimized Bagged-GPR hyper-parameters p and q and computation time for each feature set. t_{train} and t_{test} denote the training and testing time, respectively.

Feature Set	p	q	t_{train}^{bagged}	t_{test}^{bagged}	$t_{train}^{baseline}$	$t_{test}^{baseline}$
0.5-0.8-Intersect	10	75	0.7 s	0.6 s	35.1 s	0.5 s
0.6-0.8-Intersect	50	75	2.6 s	0.8 s	60.2 s	0.6 s
0.5-0.8-Conc	90	75	25.8 s	1.1 s	276.2 s	0.7 s
0.6-0.8-Conc	60	50	5.3 s	1.7 s	126.1 s	1.8 s
0.9-0.9-Conc	10	250	2.1 s	1.6 s	20.8 s	0.6 s

Health estimation performance with optimized hyper-parameters.—Table VII lists error metrics for health estimation models with optimized Bagged-GPR hyper-parameters p and q. It includes the median and mean of RMSPE and MPE across 8 test cells. RMSPE is defined as:

$$RMSPE(\mathbf{y}, \hat{\mathbf{y}}) = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(\frac{y_j - \hat{y}_j}{y_j}\right)^2} \times 100\% \quad [22]$$

where $\mathbf{y}$ and $\hat{\mathbf{y}}$ are ground truth and model predictions of output variable y. MPE is defined as in Eq (21). The Bagged-GPR model consistently outperforms the baseline across all feature sets, except for 0.9-0.9-Conc, where performance is comparable. This demonstrates the bagging technique's effectiveness in enhancing prediction accuracy and computational efficiency. Table VII shows that increasing the number of input features can improve SoH estimation accuracy to under 1% MPE and 1.5% RMSPE, though it increases the training time. The proposed framework achieves satisfactory performance with less than 1.5% MPE and 2% RMSPE using just one feature. Box plots in Fig. 11 reveal low median errors (under 1% SoH error) for all models, with feature sets from scheme 1 slightly underestimating and those from scheme 2 slightly overestimating SoH. Generally, scheme 2 (Concatenation) provides better estimation performance with a narrower error spread, as it maintains inter-cell correlations during model training.

Figure 12 shows the estimated aging trajectories with a $\pm 2\sigma$ uncertainty range compared to the observed SoH for 8 test cells using feature set 0.5-0.8-Conc. The confidence interval narrows

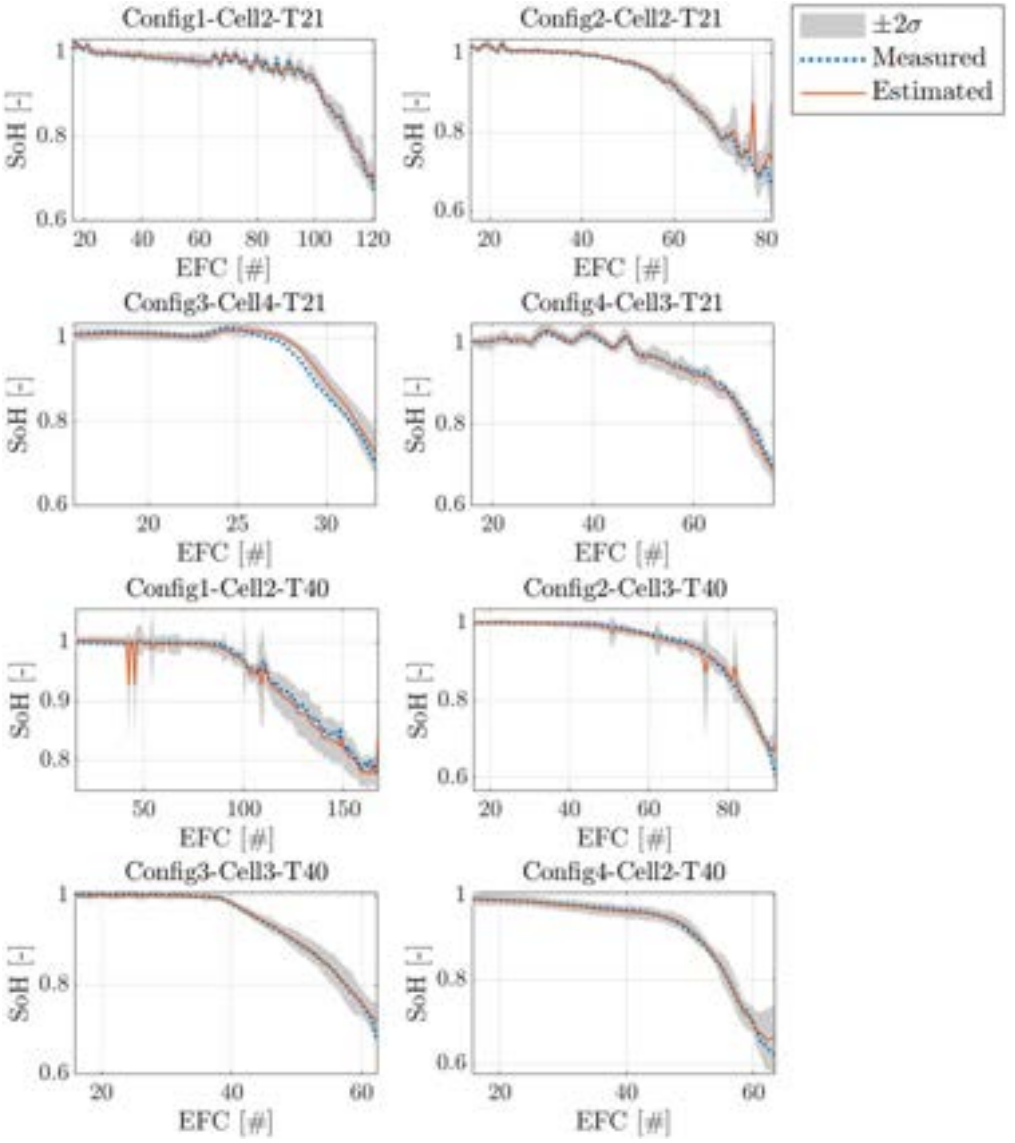

Figure 12. Estimated and actual aging trajectories for 8 test cells using feature set 0.5-0.8-Conc.

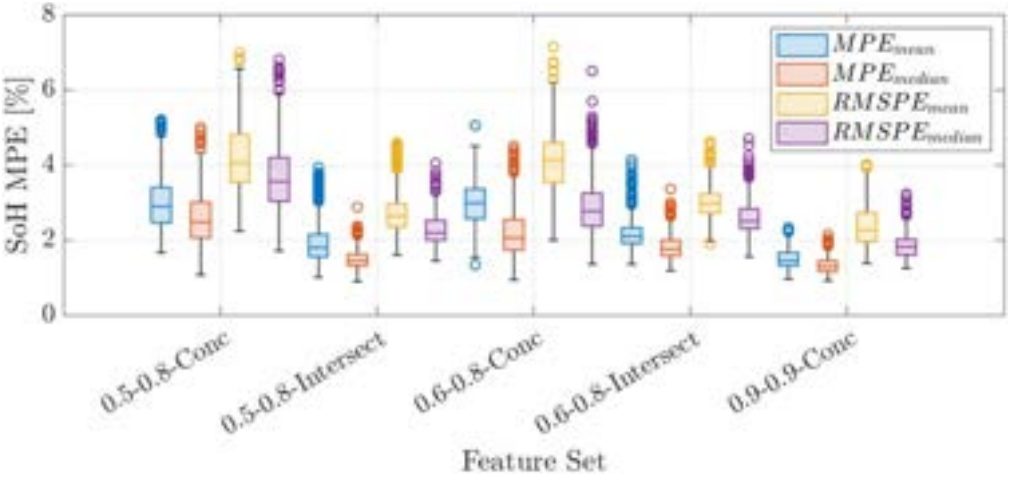

Figure 13. Distribution of SoH estimation errors under 1000 different train-test splits.

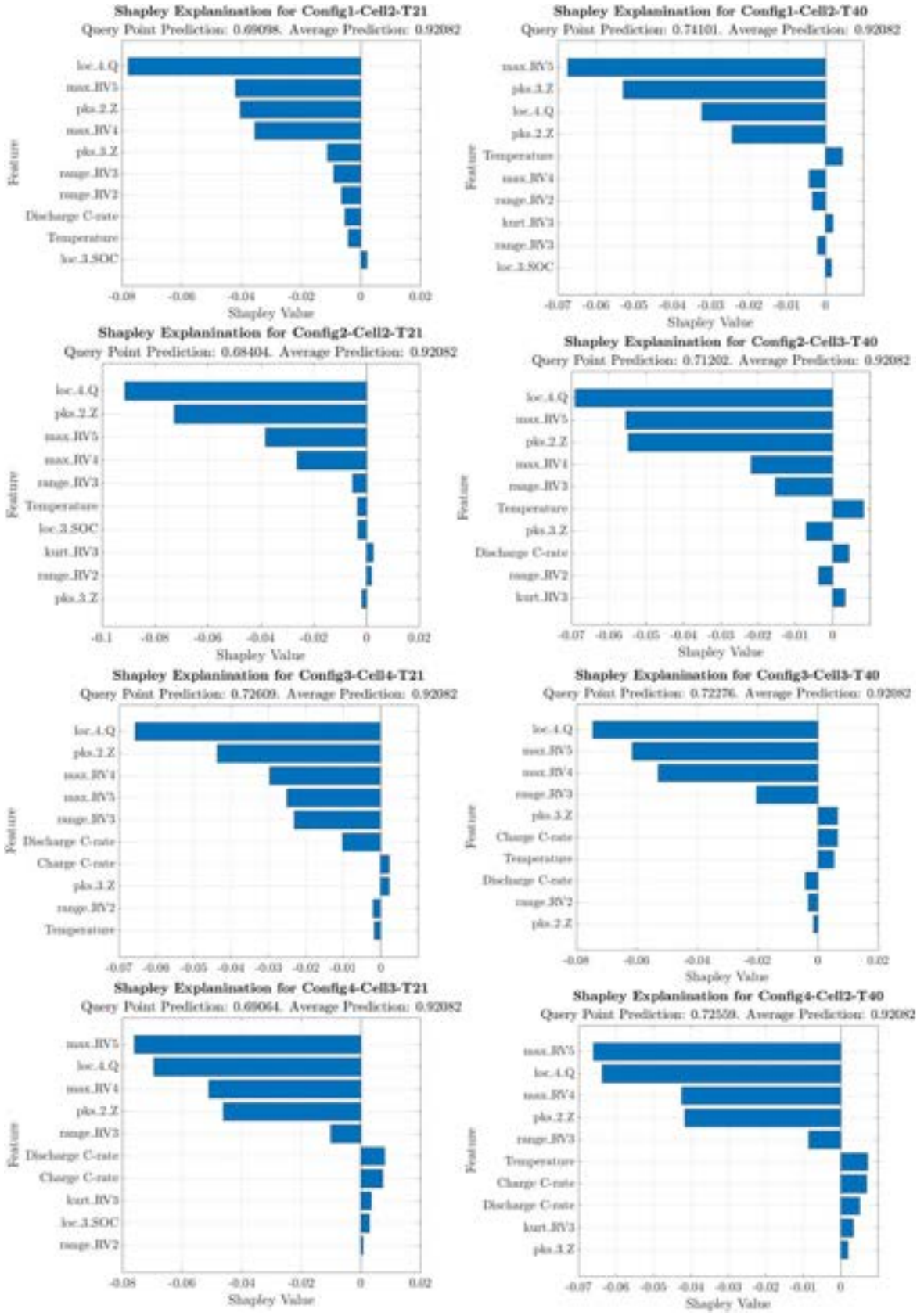

Figure 14. SHAP analysis conducted at cells' end of life for health estimation features in feature set 0.6-0.8-Conc.

when estimates are close to the ground truth and widens with less accurate predictions, reflecting variability. The model shows higher confidence during the cells' early life, where aging is minimal. As cells degrade, the uncertainty range broadens, but the mean estimates remain consistent with measurements across various operating conditions.

Table IX. Comparison of SoH estimation performance with limited feature sets. 'Full' refers to the baseline scenario with all features available.

Scenario #	$RMSPE_{median}$	$RMSPE_{mean}$	MPE_{median}	MPE_{mean}
1	1.36	1.57	0.79	0.86
2	2.16	2.34	1.03	1.14
3	1.34	1.70	0.85	0.92
Full	1.40	1.49	0.75	0.78

Sensitivity analysis on the train-test split.—It is essential to assess health estimation performance by varying the choice of test cells in the dataset, rather than relying on a single arbitrary split. This approach ensures a fair and generalizable evaluation of model accuracy, accounting for cell-to-cell variability. Maintaining a 70–30 train-test split ratio, we vary the combination of 8 test cells, ensuring each test cell represents a different test condition to cover all aging cycling protocols in the dataset. With 115,200 possible train-test splits, iterating through all options is impractical. Instead, we randomly sampled 1,000 train-test splits from this set, performed feature selection and model training with fixed Bagged-GPR hyperparameters (p and q from Table VIII), and evaluated SoH estimation performance. The results are depicted in Fig. 13.

We use the same SoH performance metrics as in Table VII to evaluate model performance. Notably, for feature sets with $(\alpha, \beta) = (0.5, 0.8)$ and $(\alpha, \beta) = (0.6, 0.8)$ in scheme 2, the SoH models exhibit a wider error spread, despite lower average errors in single train-test splits (Table VII). In contrast, models trained using scheme 1 and $(\alpha, \beta) = (0.9, 0.9)$ in scheme 2 demonstrate better performance to train-test split variations. This aligns with scheme 2's design to minimize cell-to-cell variability by intersecting feature sets. Scheme 1, with more features, may improve accuracy for certain splits but risks overfitting and poor performance on outlier test cells. Overall, the worst-case MPE is below 8% across all feature sets, with 0.9-0.9-Conc achieving 4% worst-case MPE.

Feature importance evaluation: SHAP analysis.—We obtained informative feature sets for SoH estimation, raising questions about each feature's contribution to LMB health estimation. While correlation alone does not imply causation, SHAP analysis offers a theoretically grounded method for interpreting machine learning models, based on coalitional game theory.[38] The Shapley value, a concept from game theory, fairly allocates gains or costs among participants in a coalition. In this context, the Bagged-GPR model represents the game, and the prediction is the payout. After training the model, we calculate the Shapley value for each feature in a given data sample to quantify its contribution to the final prediction.

Figure 14 illustrates the Shapley values for feature set 0.6-0.8-Conc at the end of life for each test cell. Notably, the feature loc_4_Q emerges as a strong health indicator, consistently showing significant impact across all test cells. This feature's left-shift movement in correlation with charge capacity suggests a loss of active material (LAM) in the cathode, as supported by the shift observed in the DV curve.[39] Additionally, the disappearance of peaks in the incremental capacity (IC) is typically associated with loss of lithium inventory (LLI).[23] However, since all DV valleys persist throughout aging, LLI does not seem to be a major factor in this dataset. Furthermore, features like pks_2_Z and pks_3_Z, which are indicative of cathode LAM, also contribute significantly to the SoH estimation,[40,41] The maximum voltage during rest periods at mid- and low-SoC (max_RV4 and max_RV5) is a key indicator of capacity fade, reflecting the effective amount of active material in the cathode. This sensitivity might be due to reduced reaction surface from electrolyte depletion and intensified side reactions. Additionally, Z valley depth at medium SoC (pks_2_Z and pks_3_Z) also influences SoH estimation.[23] Increased RV after discharge may indicate SEI growth and dead lithium in the anode, raising impedance. The RV sensitivity at low SoC likely results from reduced cathode lithiation capacity due to LAM.

Sensitivity analysis on available features.—The proposed health estimation framework has been evaluated using test data with fixed charge and discharge protocols. However, real-world applications of LMBs present additional complexities, such as discharge profiles that may not follow constant current, SoC operation windows that differ from the 0%–100% range in this dataset, and insufficient rest periods between usage cycles. These scenarios limit feature applicability. Despite the absence of realistic operation data, we evaluate feature sensitivity to gauge the framework's adaptability. Based on Health Estimation Performance with Optimized Hyper-parameters, the feature merging scheme 2 (Concatenation) with $\alpha = 0.6$, $\beta = 0.8$, and Bagged-GPR parameters $p = 90$, $q = 75$ shows the best performance. We use this setting for the following analysis. We evaluated health estimation performance under scenarios with limited feature sets. The first scenario involved missing test data for specific discharge segments, leading to absent R0 features (current step changes p4-p15) and RV features (rest periods RV2-RV4). The second scenario involved missing data at high SoC ($\geqslant$85%) and low SoC ($\leqslant$15%), which resulted in missing R0 features (current step changes p1, p2, p3, p16), RV features (rest periods RV1, RV5), and Z features (Valley 1, Valley 4). The third scenario involved missing data for all rest periods, which resulted in absent RV features.

Table IX summarizes the health estimation performance for each scenario. Accuracy loss is most pronounced in Scenario 2, due to the key features loc_4_Q and max_RV5 being unavailable. Scenarios 1 and 3, which involve missing discharge segments or rest periods, show minimal accuracy loss. Overall, all scenarios maintain $<$1.5% MPE, demonstrating satisfactory performance of the proposed framework to limited feature sets.

Conclusion

Despite the high theoretical energy density of lithium metal batteries, their limited lifespan and variability in aging due to multiple failure mechanisms have been significant challenges. This study shows that well-designed diagnostic tests and aging protocols can enhance the detectability of LMB aging behaviors through external voltage-current measurements, by extracting a comprehensive set of aging-informative features. Using Bagged-GPR models, we achieved SoH estimation with $<$1% MPE from a single aging cycle and $<$8% worst-case MPE through sensitivity analysis on train-test splits.

We quantified the contribution of each feature, offering insights into the mechanistic origins of LMB aging based on existing literature. Furthermore, our approach highlights the potential to integrate data-driven models, such as our Bagged-GPR-based SoH estimation, with more rigorous physics-based models or their reduced-order versions[42,43] to develop hybrid frameworks. Such hybrid models could combine the predictive capability of machine learning with the mechanistic insights of physics-based approaches, enhancing the interpretability and effectiveness of the SoH estimation framework.

Overall, our data-driven health estimation framework shows potential for LMB cell design improvements and on-board battery management systems. The framework is built on a novel feature-prompting cycling protocol that captures key LMB behaviors from which one can extract health indicators, such as internal resistance, relaxation voltage, and charging impedance. This protocol integrates seamlessly with existing lab and industry practices. With 127 features and an automated machine learning pipeline, the framework adapts to new LMB cells, addressing cell-to-cell variability and capacity fade. The identified feature correlations lay the groundwork for deeper first-principles modeling and experimental research into LMB aging mechanisms. However, the study's limitations include

the modest dataset size and cell-to-cell variability. Further validation with larger datasets, diverse test conditions, and different electrolyte compositions is needed to confirm the generalizability of the findings.

Acknowledgments

This work was financially supported by Sakuu Corp., San Jose, CA, USA, through the StorageX initiative at Stanford University. The authors thank the Sakuu Corp. team for their valuable contributions to data collection and their insightful comments.

ORCID

Wenting Ma https://orcid.org/0009-0009-7358-1401
Sara Ha https://orcid.org/0009-0005-9878-3537
Gabriele Pozzato https://orcid.org/0000-0001-5253-0271
Simona Onori https://orcid.org/0000-0002-6556-2608

References

1. X.-B. Cheng, R. Zhang, C.-Z. Zhao, and Q. Zhang, "Toward safe lithium metal anode in rechargeable batteries: a review." *Chemical Reviews*, **117**, 10403 (2017).
2. D. Lin, Y. Liu, and Y. Cui, "Reviving the lithium metal anode for high-energy batteries." *Nature Nanotechnology*, **12**, 194 (2017).
3. C. Niu et al., "Balancing interfacial reactions to achieve long cycle life in high-energy lithium metal batteries." *Nature Energy*, **6**, 723 (2021).
4. W. Xu, J. Wang, F. Ding, X. Chen, E. Nasybulin, Y. Zhang, and J.-G. Zhang, "Lithium metal anodes for rechargeable batteries." *Energy Environ. Sci.*, **7**, 513 (2014).
5. G. Liu and W. Lu, "A Model of Concurrent Lithium Dendrite Growth, SEI Growth, SEI Penetration and Regrowth." *J. Electrochem. Soc.*, **164**, 1826 (2017).
6. Q. Wang, B. Liu, Y. Shen, J. Wu, Z. Zhao, C. Zhong, and W. Hu, "Confronting the Challenges in Lithium Anodes for Lithium Metal Batteries." *Advanced Science*, **8**, 2101111 (2021).
7. D. Lu et al., "Failure mechanism for fast-charged lithium metal batteries with liquid electrolytes." *Advanced Energy Materials*, **5**, 1400993 (2015).
8. S. Liu, K. Jiao, and J. Yan, "Prospective strategies for extending long-term cycling performance of anode-free lithium metal batteries." *Energy Storage Materials*, **54**, 689 (2023).
9. Z. Hong and V. Viswanathan, "Phase-Field simulations of lithium dendrite growth with open-source software." *ACS Energy Letters*, **3**, 1737 (2018).
10. S. Xu, K.-H. Chen, N. P. Dasgupta, J. B. Siegel, and A. G. Stefanopoulou, "Evolution of dead lithium growth in lithium metal batteries: experimentally validated model of the apparent capacity loss." *J. Electrochem. Soc.*, **166**, 3456 (2019).
11. M. Uppaluri, K. Shah, V. Viswanathan, and V. R. Subramanian, "The importance of a moving boundary approach for modeling the sei layer growth to predict capacity Fade." *J. Electrochem. Soc.*, **169**, 040548 (2022).
12. G. Pozzato and S. Onori, "Combining physics-based and machine learning methods to accelerate innovation in sustainable transportation and beyond: a control perspective." *2023 American Control Conference (ACC)*, San Diego, CA, USAMay 31 - June 2, 2023(IEEE)640 (2023).
13. S. C. Kim et al., "Data-driven electrolyte design for lithium metal anodes." *Proceedings of the National Academy of Sciences*, **120**, 2214357120 (2023).
14. X. L. Wang, "High-performance liquid electrolytes for lithium metal batteries discovered by machine learning and high-throughput experimentation." *2022 ECS Meeting*, Atlanta, GA, USAOctober 9-13, 2022121 (2022).
15. Q. Zhang, C. Zhou, D. Zhang, D. Kramer, C. Peng, and D. Xue, "Data-driven discovery and intelligent design of artificial hybrid interphase layer for stabilizing lithium-metal anode." *Matter*, **6**, 2950 (2023).
16. Y.-C. Gao, N. Yao, X. Chen, L. Yu, R. Zhang, and Q. Zhang, "Data-driven insight into the reductive stability of ion–solvent complexes in lithium battery Electrolytes." *Journal of the American Chemical Society*, 3 (2023).
17. K. A. Severson et al., "Data-driven prediction of battery cycle life before capacity degradation." *Nature Energy*, **4**, 383 (2019).
18. P. Iurilli, L. Luppi, and C. Brivio, "Non-invasive detection of lithium-metal battery degradation." *Energies*, **15**, 6904 (2022).
19. X. Liu, X.-Q. Zhang, X. Chen, G.-L. Zhu, C. Yan, J.-Q. Huang, and H.-J. Peng, "A generalizable, data-driven online approach to forecast capacity degradation trajectory of lithium batteries." *Journal of Energy Chemistry*, **68**, 548 (2022).
20. A. Cordoba-Arenas, S. Onori, and G. Rizzoni, "A control-oriented lithium-ion battery pack model for plug-in hybrid electric vehicle cycle-life studies and system design with consideration of health management." *Journal of Power Sources*, **279**, 791 (2015).
21. S. Ha, G. Pozzato, and S. Onori, "Electrochemical characterization tools for lithium-ion batteries." *Journal of Solid State Electrochemistry*, **28**, 1131–1157 (2024).
22. A. Barai, K. Uddin, W. D. Widanage, A. McGordon, and P. Jennings, "A study of the influence of measurement timescale on internal resistance characterisation methodologies for lithium-ion cells." *Sci. Rep.*, **8**, 21 (2018).
23. N. Gao et al., "Fast Diagnosis of Failure Mechanisms and Lifetime Prediction of Li Metal Batteries." *Small Methods*, **5**, 2000807 (2021).
24. B. D. Adams, J. Zheng, X. Ren, W. Xu, and J. Zhang, "Accurate Determination of Coulombic Efficiency for Lithium Metal Anodes and Lithium Metal Batteries." *Advanced Energy Materials*, **8**, 1702097 (2018).
25. K.-H. Chen, K. N. Wood, E. Kazyak, W. S. LePage, A. L. Davis, A. J. Sanchez, and N. P. Dasgupta, "Dead lithium: mass transport effects on voltage, capacity, and failure of lithium metal anodes." *Journal of Materials Chemistry A*, **5**, 11671 (2017).
26. G. Pozzato, A. Allam, L. Pulvirenti, G. A. Negoita, W. A. Paxton, and S. Onori, "Analysis and key findings from real-world electric vehicle field data." *Joule*, **7**, 2035 (2023).
27. C. Julien, A. Mauger, A. Vijh, and K. Zaghib, *Lithium Batteries: Science and Technology* (Springer) (2016), https://link.Springer.com/10.1007/978-3-319-19108-9.
28. A. Fly and R. Chen, "Rate dependency of incremental capacity analysis (dQ/dV) as a diagnostic tool for lithium-ion batteries." *Journal of Energy Storage*, **29**, 101329 (2020).
29. P. Keil and A. Jossen, "Calendar Aging of NCA Lithium-Ion Batteries Investigated by Differential Voltage Analysis and Coulomb Tracking." *J. Electrochem. Soc.*, **164**, 6066 (2017).
30. A. Kirmse and J. De Ferranti, "Calculating the prominence and isolation of every mountain in the world." *Progress in Physical Geography: Earth and Environment*, **41**, 788 (2017).
31. "MATLAB: findpeaks - Find local maxima." https://www.mathworks.com/help/signal/ref/findpeaks.html (2024).
32. C. Wissler, "The Spearman Correlation Formula." *Science*, **22**, 309 (1905).
33. A. Takahashi, A. Allam, and S. Onori, "Evaluating the feasibility of batteries for second-life applications using machine learning." *iScience*, **26**, 106547 (2023).
34. E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on gaussian process regression: modelling, exploring, and exploiting functions." *Journal of Mathematical Psychology*, **85**, 1 (2018).
35. R. R. Richardson, M. A. Osborne, and D. A. Howey, "Gaussian process regression for forecasting battery state of health." *Journal of Power Sources*, **357**, 209 (2017).
36. "MATLAB: fitrgp - Fit a Gaussian process regression (GPR) model." https://www.mathworks.com/help/stats/fitrgp.html (2024).
37. Y. Li, K. Liu, A. M. Foley, A. Zülke, M. Berecibar, E. Nanini-Maury, J. Van Mierlo, and H. E. Hoster, "Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review." *Renewable and Sustainable Energy Reviews*, **113**, 109254 (2019).
38. S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions." *NIPS'17*, Long Beach, CA, USADecember 4 - 9, 2017 (2017).
39. A. Barai, K. Uddin, M. Dubarry, L. Somerville, A. McGordon, P. Jennings, and I. Bloom, "A comparison of methodologies for the non-invasive characterisation of commercial Li-ion cells." *Progress in Energy and Combustion Science*, **72**, 1 (2019).
40. I. Zilberman, S. Ludwig, M. Schiller, and A. Jossen, "Online aging determination in lithium-ion battery module with forced temperature gradient." *Journal of Energy Storage*, **28**, 101170 (2020).
41. M. Schindler, J. Sturm, S. Ludwig, A. Durdel, and A. Jossen, "Comprehensive analysis of the aging behavior of nickel-rich, silicon-graphite lithium-ion cells subject to varying temperature and charging profiles." *J. Electrochem. Soc.*, **168**, 060522 (2021).
42. A. K. De Souza, W. Hileman, M. S. Trimboli, and G. L. Plett, "A closed-form impedance model for rechargeable lithium-metal batteries." *IFAC-PapersOnLine*, **56**, 7114 (2023).
43. A. Kawakita De Souza, W. Hileman, M. S. Trimboli, and G. L. Plett, "A control-oriented reduced-order model for lithium-metal batteries." *IEEE Control Systems Letters*, **7**, 1165 (2023).